The Austronesian languages

Revised Edition

Robert Blust

A-PL 008

This is the first single-authored book that attempts to describe the Austronesian language family in its entirety. Topics covered include: the physical and cultural background, official and national languages, largest and smallest languages in all major geographical regions, language contact, sound systems, linguistic palaeontology, morphology, syntax, the history of scholarship on Austronesian languages, and a critical assessment of the reconstruction of Proto Austronesian phonology.

Robert Blust is a professor in the Department of Linguistics at the University of Hawaii. He has authored over 200 publications, mostly in the field of Austronesian comparative linguistics, but with forays into linguistically-guided ethnology.

First published: 2013

Contents

x

Maps, Figures, Tables

Preface

This book is long overdue. In January 1978, at the Second International Conference on Austronesian Linguistics (SICAL), hosted by the Department of Linguistics at the Research School of Pacific Studies, Australian National University, R.M.W. Dixon, who had written the lead volume in the Cambridge Language Survey series, approached me about doing a book on Austronesian. I hesitated, due to the pressure of other commitments, and it was about two years before I began work on the book. Progress went well for five or six months, when I became distracted by competing research agendas, in particular the need to move toward a new comparative dictionary of the Austronesian languages to replace the pioneering, but by then already outdated work of the German scholar Otto Dempwolff (1938). This required systematic searching and comparison of the available lexical data for more than 150 languages. After nearly a decade of groundwork which resulted in a number of lengthy publications funding for the dictionary project was obtained in 1990, and by 1995 a substantial, but still incomplete annotated comparative dictionary with over 5,000 reconstructed bases and many more affixed forms together with full supporting evidence, was available by request on the internet. Just as the funding for this project was ending I became involved in salvage lexicography with Thao, one of the most endangered Austronesian languages of Taiwan, a task that took up large parts of the next five years. As a result, it has been nearly three decades since the first portions of this book were written, and considerable revision has been necessary in bringing it up to date between 2004 and 2007.

Several general sources on Austronesian languages are already available. Dahl (1976) offers a technical treatment of issues in comparative phonology, but is narrowly focused and assumes a good deal of previous background. Tryon (1995) provides a broad typological coverage for some 80 languages with highly condensed data samples. These provide some serviceable introductions to structural properties, but much of this five-volume work is taken up with a comparative vocabulary that has less value than it might, because it is modeled, at least in part, on Indo-European semantic categories. Moreover, the Introduction contains a number of unfortunate errors (Blust 1997e). More useful are the recent broad areal summaries of Lynch, Ross, and Crowley (2002) for the Oceanic branch of Austronesian, and Adelaar and Himmelmann (2005) for the Austronesian languages of Asia and Madagascar. The present book differs from these in being the product of a single author, with all that entails: greater internal integration of ideas on the one hand, but lacking the insights that others might provide on the other.

The form that a book takes depends in important ways on its perceived readership. There are at least four types of readers that this book might serve: 1. specialists in Austronesian linguistics, 2. speakers of Austronesian languages who want to known more about their own 'roots', 3. the general linguistic public, and 4. scholars in collateral disciplines such as archaeology, history, ethnology, or biological phylogenetics. Any attempt to serve all types of reader equally would be too ambitious, and probably would result in an unwieldy and overlong text that would be of only limited use to any given

person. This book is therefore directed primarily toward the general linguist, although it pays relatively little attention to the burgeoning formalist literature as represented by e.g. the Austronesian Formal Linguistics Association, since much of this literature is arguably concerned primarily with testing general theories, and the use of Austronesian language material is secondary to that aim. However, the decision to write a survey of Austronesian linguistics for the non-specialist does not mean that no thought has been given to other types of readers. Whereas most chapters aim at the general reader, some chapters (especially those on historical linguistics) are highly technical. Moreover, although this book is intended to serve as an introduction to a field that will be new to many, it also aims to make some original contributions. The challenge in writing it, then, has been in how to strike a balance between providing information that is broad and generally useful, while allowing some of the detail and density that is the passion of the specialist.

Even with a reasonably clear idea of readership, it can be difficult to know what to include and what to leave out of a book. The Austronesian language family appears to be second only to Niger-Congo in size, and even though very limited descriptive material is available for many languages, the Austronesian literature is enormous, and growing at an ever-increasing pace. Various ways were considered to restrict the scope of this volume and thereby simplify the task of both writer and reader. One might, for example, discuss only publications that have appeared during the past twenty years. But this would fail to provide historical perspective on a scholarly literature that already possessed real scientific merit by the time of Wilhelm von Humboldt in the 1830s, and had reached a comparatively high level of sophistication in the work of the Dutch Indonesianist H.N. van der Tuuk by the early 1860s. Alternatively, we might consider restricting ourselves to the literature in English, which has been the dominant language of scientific discourse for at least the past half century, but this would be equally pointless. Much of the most important literature on Austronesian languages prior to the Second World War was in Dutch or German, and French, Spanish and Japanese are equally important descriptive languages for particular areas (Madagascar, mainland Southeast Asia and the various French possessions in the Pacific for the first, the Philippines and some early publications on Chamorro for the second, and Taiwan for the last). Or, then, we might decide to provide an overview of the typology of the languages for the general reader but omit any discussion of historical reconstruction, which tends to be of more specialised interest. This, too, would be a serious mistake. With well over 1,000 members and a history of roughly six millennia, the Austronesian language family is not only large and internally complex, but it was the most widespread of all language families prior to the European colonial expansions of the past four centuries.

Despite this scope the fact of relationship can readily be established for many languages (but not all) by a comparison of the numerals 1-10, the personal or possessive pronouns, or such relatively stable body-part terms as the words for 'eye', 'breast', or 'liver'. But historical linguistics has gone far beyond establishing an Austronesian language family – something that was already done in 1708 by the Dutch scholar Hadrian Reland. The Austronesian language family today probably is second only to Indo-European in the progress of comparative research. Not only has it proved to be a rich laboratory for testing the universality of the comparative method as developed in Indo-European studies, but it has been at the forefront in establishing correlations between linguistics and archaeology (Blust 1976b, 1995c, Green and Pawley 1999, Kirch and Green 2001) and the reconstructed vocabulary for early proto languages is arguably equal in quantity and

quality to that of any other language family, including Indo-European (Blust and Trussel ongoing, Ross, Pawley and Osmond 1998, 2003, 2008, 2011, 2013).

Although many Indo-Europeanists specialise in one branch of the family, it is a fairly common practice for comparativists to deal with Indo-European as a whole. By contrast, in Austronesian the areal fragmentation of scholarship is much more marked. While the scholar specializing in problems of Indo-European reconstruction and change can achieve significant results with attention to five or six written languages representing older stages of development, this option is not open to Austronesianists. Consequently, most comparative work has been carried out on lower levels: within Polynesian or the slightly larger Central Pacific group (Polynesian-Fijian-Rotuman), within the Nuclear Micronesian languages, within the Philippines, etc. Only a handful of scholars over the past 150 years have taken the entire Austronesian language family as their chosen domain. As a result, it is not unusual for researchers who are familiar with one part of the Austronesian world to be poorly informed about others. In writing a survey of this language family it is thus difficult to distinguish between the general linguistic public and the community of Austronesian specialists, since a specialist in, say, Philippine linguistics may know little about Polynesian linguistics, or vice-versa. One of the aims of this book, therefore, is to provide a comprehensive overview of Austronesian languages which integrates areal interests into a broader perspective.

There is an old philosophical conundrum about a glass containing water up to the mid-point: we can look at the glass as half-full, or as half-empty. In describing the state of research on the Austronesian languages we are faced with just such a problem. In a language family with over 1,000 members, many with fewer than 1,000 speakers, it goes without saying that hundreds of languages – particularly the smaller ones – are represented only by short comparative vocabularies, if at all. But to conclude from this that the level of scholarship in Austronesian linguistics is undeveloped would be unjust, since a great deal of valuable descriptive and comparative work has been done over the past century and a half. Although the Austronesian languages cannot boast of a record of textual documentation for the earlier periods comparable to that in Indo-European, many modern languages are documented by substantial dictionaries or grammars (or both), and these are representative of most major geographical areas, most areas of typological variation, and several major subgroups of the family. Fine-grained dialect geography has been done or is underway in Java, Fiji, and the central Philippines, and as already noted, lexical reconstruction is perhaps as advanced as that of any other language family. Even with many languages still poorly described, then, there is an enormous literature that a book such as this must cover to be truly representative.

As already stated, the long delay in producing this book was partly due to the need to attend to other research which in some ways seemed more pressing. In addition, however, from the outset I was faced with a quandary. The manuscript was originally intended for the Cambridge Language Surveys, but it was apparent from early on that it would be very large, and as it grew toward its final form I had to face the fact that the length limits of this series could not be satisfied, even with the most dedicated effort. That publication agreement was therefore terminated and the manuscript submitted to Pacific Linguistics. I am grateful to Andrew Pawley and Malcolm Ross for their assistance in making this publication transition, and to many friends who read and commented on parts of the manuscript and offered suggestions for improvement, or who came to my assistance with answers to particular questions. A complete list of those who helped is probably impossible to reconstruct at this time, but would include Juliette Blevins, John Bowden, Abigail Cohn,

Jim Collins, Nick Evans, James J. Fox, Stefan Georg, Paul Geraghty, Ives Goddard, George Grace, Chuck Grimes, Marian Klamer, Harold Koch, Uli Kozok, John Kupcik, Paul Jen-kuei Li, Tsai-hsiu (Dorinda) Liu, John McLaughlin, John Lynch, Miriam Meyerhoff, Yuko Otsuka, Bill Palmer, Andrew Pawley, Kenneth Rehg, Lawrence Reid, Laura Robinson, Malcolm Ross, Laurent Sagart, Hiroko Sato, Thilo Schadeberg, Albert Schütz, Graham Thurgood, Brent Vine, Alexander Vovin, and Elizabeth Zeitoun. I owe special thanks to Hsiu-chuan Liao for valuable comments on the syntax of Philippine languages, and editorial suggestions on the form of the manuscript, to Peter Lincoln for correcting errors of fact, and to Jason Lobel for reading most of the manuscript, offering information on all things Philippine, and for his assistance with the maps. Byron Bender, my former teacher and colleague of many years, was helpful in more ways than I can hope to repay.

In connection with the second edition I am particularly grateful for help received from four graduate students at the University of Hawai'i: Katie Butler who provided invaluable assistance in redoing the maps, James Hafford and Emerson Odango, who read and commented on portions of the book in areas that they know particularly well, and especially Tobias Bloyd, who spent many hours of his own time helping me turn a mediocre first-try at an index into something genuinely useful, and providing assistance in other ways as well. K. Alexander Adelaar and Alexandre François, longstanding colleagues and friends, also sent me many helpful suggestions for improvement that I appreciate immensely. Nothing can be more rewarding to an author than to have dedicated readers who do not let the smallest detail escape their attention, and to all of those mentioned here, and any others who I may have inadvertently omitted, I express my deepest appreciation.

1   Abbreviations and conventions

To a very large extent the use of grammatical terminology follows that of the sources cited. This shows wide variation, so that the same phenomenon may be represented in the literature under more than one name. The following abbreviations reflect that fact. In addition, grammatical elements are bolded in the glosses, a hyphen marks a morpheme boundary, and a dot between elements of a gloss indicates fusion of meanings or functions in a single form, as with Tetun/Tetum *nia n-aklelek* 's/he abuses verbally' (= **3sg 3sg-**speak.abuse), or Kambera *ku* '**1sg.nom**'. Despite the presence of a morpheme boundary, reduplications are glossed as single morphemes with fused meaning, as in Malay/Indonesian *Meréka ber-lihat-lihat* 'They looked at each other' (= **3pl intr-**see.**recip**). Elements that could not be glossed are repeated in the glosses, as with the element *de* in Taba *n-yol calana de n-ha-totas* (> natotas) 'She took the trousers and washed them' (= **3sg-**take trousers **de 3sg-caus-**wash).

abl	ablative
abs	absolutive
ac	accomplished
acc	accusative
act	active
act.part	active participle
af	actor focus
aff	affirmative

ag	animate goal
all	allative
an.sg	singular anaphor
aor	aorist
ap	adversative passive
appl	applicative
apt	aptative
art	article
as	animate subject; away from speaker
asp	aspect
att	attributive
aux	auxiliary
av	actor voice/active voice
ben	benefactive
bn	bound nominative pronoun
bv	benefactive voice
card	cardinal
caus	causative
cf	counterfactual; causative focus
cl	clitic
clas	classifier
coe	co-enunciation
comp	completive
conj	conjunction
conn	connective (ligature)
cont	continuative aspect
cop	copula
cv	circumstantial voice
dat	dative
def	definite
deic	deictic
dem	demonstrative
dep	dependent
dir	directional
dis	distributive
dist	distal
dl	dual
dp	direct passive; dominant possession
dtr	determinative
dur	durative
dx.1	first-order deictic
emph	emphasis; emphatic
es	echo subject
excl	exclusive
foc	focused NP
fut	future
gen	genitive
gol	goal

hab	habitual
hort	hortative
hum	human
hyp	hypothetical
ia	involuntary action
if	instrument focus
imm	imminent
imp	imperative
impf	imperfective
inc	inceptive
incl	inclusive
indef	indefinite
inst	instrumental
int	intensifier
intr	intransitive
intv	intermediary voice
io	indirect object
lp	instrumental passive
irr	irrealis
itr	iteration
iv	instrumental voice
lig	ligature
loc	locative, location
lp	local passive
ltr	local transitive
lv	locative voice
mod	mood marker
mot	motion
mult	multiple
neg	negative
nf	non-focused
nom	nominative
obj	object
obl	oblique
of	object focus
om	object marker
p	past
pass	passive voice
pc	possessive classifier
perf	perfective
pf	patient focus
pl	plural
pm	person marker
pol	polite
poss	possessive
poss.dr	drinkable possession
poss.ed	edible/alimentary possession
poss.gnr	general/neutral possession

pot	potential
pr	person (without respect to number)
pr.st	static presentative
pres	present
prox	proximal
prp	proper
psm	possessive marker
purp	purpose
pv	patient voice/passive voice
q	question/interrogative
qm	question marker
real	realis
recip	reciprocal
red	reduplication
rel	relativizer
res	resultative
resp	respect form
res:pro	resumptive proform
rf	referent(ial) focus/voice
rltr	relator
sg	singular
seq	sequential
sf	subject focus
so	specific object
sp	subordinate possession
stat	stative
sub	subordinator
top	topic
tr	transitive
ts	toward speaker
vbl	verbalizer
ven	venitive
vet	vetative
vp	verbal particle; verb phrase
vtf	ventif/centripetal directional
wh	question word as relativizer
1	first person
2	second person
3	third person

2   Orthography

In writing a book which uses data from scores or even hundreds of languages the problem of uniform vs. individualised orthographies becomes a thorny issue. On the one hand, it is more helpful for the general reader if the same sound is represented by a single symbol across many languages. Adhering to this principle would favor using the IPA, at least for attested languages. On the other hand, applying the IPA to reconstructed languages, where the phonetic character of proto phonemes is often uncertain, and

longstanding conventions are well-embedded in a voluminous literature, would introduce needless chaos. Likewise, changes to well-established orthographic traditions can lead to native speaker rejection of the presentation, or can complicate the matter of locating forms in an original source.

To facilitate retrieval of information from dictionaries I have usually adhered to the orthography of the sources, except when giving phonetic forms within square brackets, where I use IPA. Thus, for Paiwan I retain *tj*, *ts* and *dj* rather than replacing them with *č*, *c* and *ǰ*, since doing so would require the reader to make major adjustments in searching for forms. However, I have replaced *l* with the phonetically more revealing *lʸ*, which does not involve a displacement in alphabetical order.

I have made major departures from the orthography of the sources in four cases (except in quoted material, where the original spelling is retained). First, I represent the velar nasal as /ŋ/. For most languages this means only replacing the digraph *ng*, and so should cause no serious problems in searching the dictionaries or grammars from which the material has been taken. However, for a few languages the representation of the velar nasal as /ŋ/ may be jarring. This is particularly true for Fijian and Samoan, which have a longstanding missionary-based orthographic tradition in which the velar nasal is represented as /g/. Fijian and Samoan specialists are therefore forewarned that the orthographic tradition they are familiar with has been modified in this way.

Parallel to this alteration, but affecting a far smaller number of languages, is the use of /ñ/ to represent the palatal nasal. This mainly affects languages in western Indonesia, where the palatal nasal generally is represented as *ny* or *nj* (the latter primarily in earlier Dutch sources). For languages such as Chamorro, where the standard sources already represent the palatal nasal as *ñ* there is no change.

Third, I represent the glottal stop as [ʔ]. This modification affects the standard orthography of a number of languages. In major Philippine languages such as Tagalog it is conventional to represent the glottal stop and stress placement together with diacritics: *batà* = [bátaʔ] 'child', *upà* = [úpaʔ] 'pit, excavation', *punò* = [púnoʔ] 'chief', but *mulî* = [mulíʔ] 'again, once more', *tukô* = [tukóʔ] 'gecko', *walâ* = [waláʔ] 'none, nothing'. For these languages I write the glottal stop and the stress separately, as is already done for most Philippine minor languages. In many other languages the glottal stop is represented in published sources by a generic apostrophe. In Polynesian languages this is often a raised and inverted comma, as in Tongan *'one* 'sand' or Hawaiian *'ewa* 'crooked'. In other languages, as Arosi of the southeast Solomons, it is a vertical apostrophe, and in Chamorro it is an acute accent to the right of a vowel. This change will most seriously affect the orthography of Palauan where, following a early German-based system the glottal stop (then heard as [x]) is represented as *ch*. Scholars working with Palauan and users of McManus and Josephs (1977) are therefore forewarned that words such as *chull* 'rain' or *rasech* 'blood' are written in this book as *ʔull* and *rasəʔ* respectively.

Fourth, as in the preceding Palauan example, I write the schwa as /ə/ wherever I am sure that a traditional orthographic 'e' represents a mid-central rather than a mid-front vowel. For languages that have little written tradition, as with many of the languages of Borneo, this will create no difficulty, but in such relatively well-described languages as Malay and Javanese, where the schwa is conventionally written 'e' and mid-front vowels as 'é' (Malay) or 'é' and 'è' (Javanese) it affects many forms, through changing 'e' to /ə/ 'é' to /e/, and 'è' to /ɛ/. In most Philippine languages the reflex of PAN *e is a high-central or high back unrounded vowel rather than a mid-central vowel. For simplicity I write all of these as schwa. The schwa is also represented as *e in higher-level reconstructions, and

here I have left it unchanged. With only minor exceptions, all other orthographic features of the sources have been retained.

In addition, some language names have older and newer orthographic forms, as with Isneg or Isnag, Ifugao or Ifugaw, Ponapean or Pohnpeian, and Trukese or Chuukese. It is important to note that since both versions are found in the literature I have tolerated a certain amount of variation in such usage. Related to this, citations of Chinese forms in Roman script appear in the Wade-Giles system for older publications and in pinyin for most newer publications. I have given preference to pinyin, but have included the corresponding Wade-Giles transcription in parentheses where it is the form that is used in certain frequently cited publications.

3   Austronesian languages cited in the text

The following Austronesian languages are cited in the text. (SHWNG = South Halmahera-West New Guinea; * = extinct).

Language	Subgroup	Location
Abaknon	Sama-Bajaw	Capul Island
Acehnese	Malayo-Chamic	Sumatra
Adzera	North New Guinea	New Guinea
Agta, Dupaningan	Northern Luzon	Luzon
Agta, Isarog	Greater Central Philippines	Luzon
*Agta, Mt. Iraya	Greater Central Philippines	Luzon
Agutaynen	Greater Central Philippines	Palawan
Ajië	South New Caledonian	New Caledonia
Aklanon	Greater Central Philippines	Bisayas
‘Ala‘ala	Papuan Tip	New Guinea
Alta, Northern	Northern Luzon	Luzon
Alta, Southern	Northern Luzon	Luzon
Alune	Three Rivers	Seram
Amahai	East Piru Bay	Seram
Amara	North New Guinea	New Britain
Ambae, Northeast	East Vanuatu	Vanuatu
Ambae, West/Duidui	East Vanuatu	Vanuatu
Ambai	SHWNG	New Guinea
Ambelau	West Central Maluku	Ambelau Island
Ambrym, North	East Vanuatu	Vanuatu
Ambrym, Southeast	East Vanuatu	Vanuatu
Amis	East Formosan	Taiwan
Anakalangu	Sumba-Hawu	Sumba
Anejom	Southern Melanesian	Vanuatu
Anuki	Papuan Tip	New Guinea
Anus	SHWNG	New Guinea
Anuta	Polynesian	Anuta
*Aore	West Santo	Vanuatu
Apma	Pentecost	Vanuatu
Araki	West Santo	Vanuatu
‘Āre’āre	Southeast Solomonic	Malaita

Arhâ	South New Caledonian	New Caledonia
Arhö	South New Caledonian	New Caledonia
Arop-Lokep	North New Guinea	New Guinea
Arosi	Southeast Solomonic	San Cristobal/Makira
*Arta	Northern Luzon	Luzon
As	SHWNG?	New Guinea
Asi/Bantoanon	Greater Central Philippines	Banton, etc.
Asilulu	Piru Bay	Ambon Island
Asumboa	Utupua-Vanikoro	Santa Cruz islands
Ata	Greater Central Philippines	Negros
Atayal	Atayalic	Taiwan
Ati/Inati	Inati	Panay
Atoni/Dawan	West Timor	Timor
Atta	Northern Luzon	Luzon
Atta, Faire	Northern Luzon	Luzon
Atta, Pudtol	Northern Luzon	Luzon
Avasö	Northwest Solomonic	Choiseul
Aveteian	Malakula Interior?	Vanuatu
Ayta, Bataan	Central Luzon?	Luzon
*Ayta, Sorsogon	Greater Central Philippines	Luzon
Babatana	Northwest Solomonic	Choiseul
Babar, North	Central Malayo-Polynesian	Babar Islands
*Babuza/Favorlang	Western Plains	Taiwan
Bada/Besoa	Kaili-Pamona	Sulawesi
Baelelea	Southeast Solomonic	Malaita
Baetora	East Vanuatu	Vanuatu
Bahasa Indonesia	Malayo-Chamic	Indonesia
Bahonsuai	Bungku-Tolaki	Sulawesi
Balaesang	Tomini-Tolitili	Sulawesi
Balangaw	Northern Luzon	Luzon
Balantak	Saluan	Sulawesi
Bali/Uneapa	Meso-Melanesian	French Islands
Balinese	Bali-Sasak	Bali
Baluan	Admiralties	Baluan Island
Bam	North New Guinea	New Guinea
Banggai	Saluan	Banggai Archipelago
Bangsa	Malakula Interior?	Vanuatu
Banjarese	Malayo-Chamic	Kalimantan
Banoni	Northwest Solomonic	Bougainville
Bantik	Sangiric	Sulawesi
Barang-Barang	South Sulawesi	Sulawesi
Baras	Kaili-Pamona	Sulawesi
Barok	Meso-Melanesian	New Ireland
*Basay	East Formosan	Taiwan
Batak, Angkola	Barrier Island-Batak	Sumatra
Batak, Karo	Barrier Island-Batak	Sumatra
Batak, Mandailing	Barrier Island-Batak	Sumatra
Batak, Palawan	Greater Central Philippines	Palawan
Batak, Simalungun	Barrier Island-Batak	Sumatra

Batak, Toba	Barrier Island-Batak	Sumatra
Batin	Malayo-Chamic	Sumatra
Bauro	Southeast Solomonic	San Cristobal/Makira
Bekatan	?	Sarawak
Belait	North Sarawak	Brunei
Berawan	North Sarawak	Sarawak
Besemah	Malayo-Chamic	Sumatra
Betawi/Jakarta Malay	Malayo-Chamic	Java
Bidayuh/Land Dayak	Land Dayak	Sarawak
Bieria	Epi	Vanuatu
Big Nambas	Malakula Interior	Vanuatu
Bikol	Greater Central Philippines	Luzon
Bilaan/Blaan	Bilic	Mindanao
Bimanese	?	Sumbawa
Bina	Papuan Tip	New Guinea
Bintulu	North Sarawak	Sarawak
Binukid	Greater Central Philippines	Mindanao
Bipi	Admiralties	Manus
Bisaya, Limbang	Greater Dusunic	Brunei
Bisayan, Cebuano	Greater Central Philippines	Bisayas
Bisayan, Samar-Leyte/ Waray(-Waray)	Greater Central Philippines	Bisayas
Boano	Tomini-Tolitoli	Sulawesi
Bola	Meso-Melanesian	New Britain
Bolinao	Central Luzon	Luzon
Bonfia/Bobot	East Seram	Seram
Bonggi	Sabahan	Banggi Island
Bonkovia	Epi	Vanuatu
Bontok	Northern Luzon	Luzon
Budong-Budong	South Sulawesi	Sulawesi
Buginese	South Sulawesi	Sulawesi
Bugotu	Guadalcanal-Nggelic	Solomon Islands
Buhid	Greater Central Philippines	Mindoro
Bukat	?	Sarawak
Bukawa/Bugawac	North New Guinea	New Guinea
Buli	SHWNG	Halmahera
Bulu	Meso-Melanesian	New Britain
Bunun	Bunun	Taiwan
Buruese	West Central Maluku	Buru
Burumba/Baki	Epi	Vanuatu
Carolinian	Micronesian	Saipan
Cemuhî	South New Caledonian	New Caledonia
Cham, Eastern	Malayo-Chamic	Vietnam
Cham, Western	Malayo-Chamic	Cambodia
Chamorro	?	Mariana Islands
Cheke Holo	Meso-Melanesian	Santa Isabel
Chru	Malayo-Chamic	Vietnam
Chuukese/Trukese	Micronesian	Caroline Islands
Dai	Central Malayo-Polynesian	Babar Islands

*Dali'	North Sarawak	Sarawak
Damar, West	Central Malayo-Polynesian	Damar Island
Dampal	Tomini-Tolitoli	Sulawesi
Dampelas	Tomini-Tolitoli	Sulawesi
Dangal	North New Guinea	New Guinea
Dawawa	Papuan Tip	New Guinea
Dawera-Daweloor	Central Malayo-Polynesian	Babar Islands
Dehu	Loyalty Islands	Lifu, Loyalty Islands
Dhao/Ndao	Sumba-Hawu	Dhao
Dixon Reef	Malakula Interior	Vanuatu
Dobel	Aru	Aru Islands
Dobuan	Papuan Tip	New Guinea
Dohoi	Greater Barito	Kalimantan
Doura	Papuan Tip	New Guinea
Duano'	Malayo-Chamic	Malay peninsula
Dumagat, Casiguran	Northern Luzon	Luzon
Duri	South Sulawesi	Sulawesi
Dusner	SHWNG	New Guinea
Dusun, Central	Dusunic	Sabah
Dusun Deyah	Barito	Kalimantan
Dusun, Kadazan	Dusunic	Sabah
Dusun, Kimaragang	Dusunic	Sabah
Dusun Malang	Barito	Kalimantan
Dusun, Rungus	Dusunic	Sabah
Dusun,Tindal	Dusunic	Sabah
Dusun Witu	Barito	Kalimantan
Efate, North/Nakanamanga	Central Vanuatu	Vanuatu
Efate, South	Central Vanuatu	Vanuatu
Elat	Central Malayo-Polynesian	Banda Islands
Elu	Admiralties	Manus
Ende	East Flores	Flores
Ere	Admiralties	Manus
Emplawas	Central Malayo-Polynesian	Babar Islands
Enggano	?	Barrier Islands, Sumatra
Erai	Central Malayo-Polynesian	Wetar
Erromangan	Southern Melanesian	Erromango
Fagauvea/West Uvean	Polynesian	Loyalty Islands
Fijian, Eastern	Central Pacific	Fiji
Fijian, Western/Wayan	Central Pacific	Fiji
Fordata	Kei-Fordata	Tanimbar Archipelago
Fortsenal	West Santo	Vanuatu
Futuna-Aniwa	Polynesian	Vanuatu
Futunan, East	Polynesian	Futuna, New Caledonia
Fwâi	North New Caledonian	New Caledonia
Gabadi/Abadi	Papuan Tip	New Guinea
Gaddang	Northern Luzon	Luzon
Galeya	Papuan Tip	New Guinea
Gane/Gimán	SHWNG	Halmahera
Gapapaiwa	Papuan Tip	New Guinea

Gasmata	North New Guinea	New Britain
Gayō	Barrier Island-Batak	Sumatra
Gedaged	North New Guinea	New Guinea
Geser-Goram	East Central Maluku	Seram Laut Islands
Getmata	North New Guinea	New Britain
Ghari	Southeast Solomonic	Guadalcanal
Giangan Bagobo	Bilic	Mindanao
Gitua	North New Guinea	New Guinea
Gomen	South New Caledonian	New Caledonia
Goro	South New Caledonian	New Caledonia
Gorontalo	Gorontalic	Sulawesi
Guramalum	Meso-Melanesian	New Ireland
Gweda/Garuwahi	Papuan Tip	New Guinea
Haeke	North New Caledonian	New Caledonia
Halia	Meso-Melanesian	Buka
Hanunóo	Greater Central Philippines	Mindoro
Haroi	Malayo-Chamic	Vietnam
Haruku	Piru Bay	Haruku Island
Hatue	East Seram	Seram
Hatusua	Piru Bay	Seram
Hawaiian	Polynesian	Hawai'i
Hawu	Sumba-Hawu	Savu Island
Helong	West Timor?	Timor
Hiligaynon/Ilonggo	Greater Central Philippines	Negros, etc.
Hitu	East Piru Bay	Ambon Island
Hitulama	East Piru Bay	Ambon Island
Hila	Piru Bay	Ambon Island
Hiw	Torres	Vanuatu
*Hoanya	Western Plains	Taiwan
Hoava	Meso-Melanesian	New Georgia Archipelago
Hoti	East Seram	Seram
*Hukumina	?	Buru
Hulung	Three Rivers	Seram
Iaai	Loyalty Islands?	Loyalty Islands
Iakanaga	Epi?	Vanuatu
Ianigi	Epi?	Vanuatu
Ibaloy	Northern Luzon	Luzon
Iban	Malayo-Chamic	Sarawak
Ibanag	Northern Luzon	Luzon
Ibatan	Bashiic	Babuyan Islands
Ida'an Begak	Ida'an	Sabah
Ifugao	Northern Luzon	Luzon
Iliun	Central Malayo-Polynesian	Wetar
Ilokano	Northern Luzon	Luzon
Ilongot	Northern Luzon	Luzon
Imroing	Central Malayo-Polynesian	Babar Islands
Inagta, Alabat Island	Northern Luzon	Luzon
Indonesian	Malayo-Chamic	Indonesia
Irarutu	SHWNG?	New Guinea

Iresim	SHWNG	New Guinea
Isinay	Northern Luzon	Luzon
Isneg	Northern Luzon	Luzon
Itbayaten	Bashiic	Batanes Islands
Itneg	Northern Luzon	Luzon
Ivatan	Bashiic	Batanes Islands
I-wak	Northern Luzon	Luzon
Jakun	Malayo-Chamic	Malay peninsula
Jarai	Malayo-Chamic	Vietnam
Javanese	?	Java
Javanese, New Caledonian	?	New Caledonia
Jawe	North New Caledonian	New Caledonia
Kaagan	Greater Central Philippines	Mindanao
Kadazan, Coastal	Dusunic	Sabah
Kadazan, Labuk	Dusunic	Sabah
Kagayanen	Greater Central Philippines	Cagayancillo Island
Kaibobo	Piru Bay	Seram
Kairiru	North New Guinea	New Guinea
Kalagan	Greater Central Philippines	Mindanao
Kalao	Muna-Buton	Sulawesi
Kaliai-Kove	North New Guinea	New Britain
Kalinga	Northern Luzon	Luzon
Kallahan	Northern Luzon	Luzon
Kamarian	East Piru Bay	Seram
Kamayo	Greater Central Philippines	Mindanao
Kambera	Sumba-Hawu	Sumba
Kanakanabu	Kanakanabu-Saaroa	Taiwan
*Kaniet	Admiralties	Kaniet Islands
Kankanaey	Northern Luzon	Luzon
Kanowit	Melanau-Kajang	Sarawak
Kapampangan	Central Luzon	Luzon
Kapingamarangi	Polynesian	Caroline Islands
Kara	Meso-Melanesian	New Ireland
Karao	Northern Luzon	Luzon
Kaulong	North New Guinea?	New Britain
Kavalan	East Formosan	Taiwan
Kawi/Old Javanese	?	Java
Kayan	Kayan-Murik-Modang	Sarawak/Kalimantan
Kayeli	Nunusaku	Buru
Kayupulau	North New Guinea	New Guinea
Keapara	Papuan Tip	New Guinea
Kédang	Central Malayo-Polynesian	Lomblen Island
Kedayan	Malayo-Chamic	Brunei
Kei	Kei-Fordata	Kei Islands
Kejaman	Melanau-Kajang	Sarawak
Kelabit	North Sarawak	Sarawak/Kalimantan
Kemak/Ema	Central Timor	East Timor
Kendayan Dayak	Malayo-Chamic	Kalimantan
Keninjal	Malayo-Chamic	Kalimantan

Kenyah	North Sarawak	Sarawak/Kalimantan
Keo	West Flores	Flores
Kerebuto	Southeast Solomonic	Guadalcanal
Kesui	East Seram	Kesui Island
Kiandarat	East Seram	Seram
Kilenge	North New Guinea	New Britain
Kilivila	Papuan Tip	Trobriand Islands
Kinamigin	Greater Central Philippines	Camiguin Island
Kinaray-a	Greater Central Philippines	Panay
Kiput	North Sarawak	Sarawak
Kiribati/Gilbertese	Micronesian	Kiribati
Kis	North New Guinea	New Guinea
Kisar	Luangic-Kisaric	Kisar
Kodi	Sumba-Hawu	Sumba
Kokota	Meso-Melanesian	Santa Isabel
Komodo	?	Komodo
Konjo	South Sulawesi	Sulawesi
Koroni	Bungku-Tolaki	Sulawesi
Kosraean/Kusaiean	Micronesian	Caroline Islands
Kove	North New Guinea	New Britain
Kowiai/Koiwai	?	New Guinea
Kroe/Krui	Lampungic	Sumatra
Kuap	Land Dayak	Sarawak
Kubu	Malayo-Chamic	Sumatra
Kulisusu	Muna-Buton	Sulawesi
*Kulon	NW Formosan?	Taiwan
Kumbewaha	Bungku-Tolaki	Sulawesi
Kuni	Papuan Tip	New Guinea
Kunye/Kwenyii	South New Caledonian	Isle of Pines
Kurudu	SHWNG	Halmahera
Kuruti	Admiralties	Manus
Kwaio	Southeast Solomonic	Malaita
Kwamera	Southern Melanesian	Vanuatu
Kwara'ae	Southeast Solomonic	Malaita
Label	Meso-Melanesian	New Ireland
Labuʔ	North New Guinea	New Guinea
Laghu	Meso-Melanesian	Santa Isabel
Lahanan	Melanau-Kajang	Sarawak
Lakalai/Nakanai	Meso-Melanesian	New Britain
Lala	Papuan Tip	New Guinea
Lamaholot	East Flores	Solor Archipelago
Lamaholot, Southwest	East Flores	Solor Archipelago
*Lamay	?	Lamay Island, Taiwan
Lamboya	Sumba-Hawu	Sumba
Lamenu	Epi	Vanuatu
Lampung	Lampungic	Sumatra
Langalanga	Southeast Solomonic	Malaita
Lara	Land Dayak	Sarawak
Larike	West Piru Bay	Ambon Island

Lau	Southeast Solomonic	Malaita
Lauje	Tomini-Tolitoli	Sulawesi
Laukanu	Papuan Tip?	New Guinea
Laura	Sumba-Hawu	Sumba
Lavongai/Tungag	Meso-Melanesian	New Hanover Island
Lawangan	Greater Barito	Kalimantan
Ledo Kaili	Kaili-Pamona	Sulawesi
Lehalurup	Torres-Banks	Vanuatu
Leipon	Admiralties	Pityilu Island
*Lelak	North Sarawak	Sarawak
Lele	Admiralties	Manus
Lemerig	Torres-Banks	Vanuatu
Lemolang	South Sulawesi	Sulawesi
Lenakel	Southern Melanesian	Vanuatu
Lengilu	North Sarawak	Kalimantan
Lengo	Southeast Solomonic	Guadalcanal
Lenkau	Admiralties	Rambutyo Island
Leti/Letinese	Central Malayo-Polynesian	Leti-Moa Archipelago
Levei	Admiralties	Manus
Leviamp	Malakula Interior?	Vanuatu
Lewo	Epi	Vanuatu
Liabuku	Muna-Buton	Sulawesi
Lihir	Meso-Melanesian	Lihir Island
Liki	North New Guinea	New Guinea
Likum	Admiralties	Manus
Lindrou	Admiralties	Manus
Lio	West Flores	Flores
Litzlitz	Malakula Interior	Vanuatu
Lödai	?	Santa Cruz Islands
Lolsiwoi	Aoba	Vanuatu
Lom/Bangka Malay	Malayo-Chamic	Bangka
Lömaumbi	Northwest Solomonic	Choiseul
Loncong	Malayo-Chamic	Sumatra
Longgu	Southeast Solomonic	Guadalcanal
*Longkiau	Paiwan?	Taiwan
Loniu	Admiralties	Manus
Lonwolwol	East Vanuatu	Vanuatu
Lorediakarkar	East Santo	Vanuatu
Lou	Admiralties	Lou Island
*Loun	Three Rivers	Seram
Luang	Luangic-Kisaric	Leti-Moa Archipelago
Luangiua/Ontong Java	Polynesian	Solomon Islands
Lubu	Barrier Island-Batak	Sumatra
*Luilang	?	Taiwan
Lun Dayeh	North Sarawak	Sarawak/Sabah
Lundu	Land Dayak	Sarawak
Lungga	Meso-Melanesian	Solomon Islands
Ma'anyan	Greater Barito	Kalimantan
Madak/Mendak	Meso-Melanesian	New Ireland

Madurese	?	Madura
Mafea	East Santo	Vanuatu
Magindanao	Greater Central Philippines	Mindanao
Magori	Papuan Tip	New Guinea
Maisin	Papuan Tip	New Guinea
Makasarese	South Sulawesi	Sulawesi
Makian Dalam	SHWNG	Makian
Makura/Namakir	Central Vanuatu	Vanuatu
Malagasy	Greater Barito	Madagascar
Malagasy, Antaimoro	Greater Barito	Madagascar
Malagasy, Antambahoaka	Greater Barito	Madagascar
Malagasy, Antandroy	Greater Barito	Madagascar
Malagasy, Antankarana	Greater Barito	Madagascar
Malagasy, Betsileo	Greater Barito	Madagascar
Malagasy, Betsimisaraka	Greater Barito	Madagascar
Malagasy, Merina	Greater Barito	Madagascar
Malagasy, Sakalava	Greater Barito	Madagascar
Malagasy, Tañala	Greater Barito	Madagascar
Malagasy, Tsimihety	Greater Barito	Madagascar
Malakula, Northeast	Malakula Coastal	Vanuatu
Malay, Ambon	Malayo-Chamic	Ambon Island, etc.
Malay, Baba	Creole?	Malay peninsula
Malay, Brunei	Malayo-Chamic	Brunei
Malay, Kedah	Malayo-Chamic	Malay peninsula
Malay, Kupang	Malayo-Chamic	Timor
Malay, Malaccan Creole	Malayo-Chamic	Malay peninsula
Malay, Negri Sembilan	Malayo-Chamic	Malay peninsula
Malay, Papuan	Malayo-Chamic	New Guinea
Malay, Pattani	Malayo-Chamic	Thailand
Malay, Sarawak	Malayo-Chamic	Sarawak
Malay, Sri Lankan	Malayo-Chamic	Sri Lanka
Malay, Standard	Malayo-Chamic	Malaysia
Malay, Ternate	Malayo-Chamic	Ternate
Malay, Trengganu	Malayo-Chamic	Malay peninsula
Malmariv	Central Santo	Vanuatu
Maloh	Greater South Sulawesi?	Kalimantan
Mamanwa	Greater Central Philippines	Mindanao
Mamuju	South Sulawesi	Sulawesi
Manam	North New Guinea	New Guinea
Mandar	South Sulawesi	Sulawesi
Mandaya	Greater Central Philippines	Mindanao
Mangarevan	Polynesian	Gambier Islands
Manggarai	West Flores	Flores
Manobo, Cotabato	Greater Central Philippines	Mindanao
Manobo, Ilianen	Greater Central Philippines	Mindanao
Manobo, Sarangani	Greater Central Philippines	Mindanao
Manobo, Tigwa	Greater Central Philippines	Mindanao
Manobo, Western Bukidnon	Greater Central Philippines	Mindanao
Mandaya	Greater Central Philippines	Mindanao

Mansaka	Greater Central Philippines	Mindanao
Manusela	Patakai-Manusela	Seram
Maori	Polynesian	New Zealand
*Mapia	Nuclear Micronesian	Mapia Island
Mapos Buang	North New Guinea	New Guinea
Mapun	Greater Barito	Cagayan de Sulu Island
Maragus/Tape	Malakula Interior	Vanuatu
Maranao	Greater Central Philippines	Mindanao
Marquesan, Northwest	Polynesian	Marquesas
Marquesan, Southeast	Polynesian	Marquesas
Marshallese	Micronesian	Marshall Islands
Masbatenyo	Greater Central Philippines	Bisayas
Masela, Central	Central Malayo-Polynesian	Babar Islands
Masimasi	North New Guinea	New Guinea
Masiwang	East Seram	Seram
Massenrempulu	South Sulawesi	Sulawesi
Matae/Navut	West Santo	Vanuatu
Matanvat	Malakula Coastal?	Vanuatu
Matbat	SHWNG	Raja Ampat Islands
Ma'ya	SHWNG	Raja Ampat Islands
Mbwenelang	Malakula Interior?	Vanuatu
Medebur	North New Guinea	New Guinea
Mekeo	Papuan Tip	New Guinea
Melanau, Dalat	Melanau-Kajang	Sarawak
Melanau, Mukah	Melanau-Kajang	Sarawak
Mele-Fila/Ifira-Mere	Polynesian	Vanuatu
Mengen/Poeng	North New Guinea	New Britain
Mentawai	Barrier Islands-Batak?	Barrier Islands, Sumatra
Merei	Central Santo	Vanuatu
Middle Malay	Malayo-Chamic	Sumatra
Minangkabau	Malayo-Chamic	Sumatra
Mindiri	North New Guinea	New Guinea
Minyaifuin	SHWNG?	New Guinea
Misima	Papuan Tip	New Guinea
Moa	Central Malayo-Polynesian	Leti-Moa Archipelago
Modang	Kayan-Murik-Modang	Kalimantan
Moken/Selung	?	Thailand
Mokerang	Admiralties	Manus
Mokilese	Micronesian	Caroline Islands
Moklen/Chau Pok	?	Thailand
Molbog	Palawanic?	Balabac Island
Molima	Papuan Tip	New Guinea
Mondropolon	Admiralties	Manus
Mongondow	Gorontalic	Sulawesi
Moor	SHWNG	New Guinea
Moriori	Polynesian	Chatham Islands
Moronene	Bungku-Tolaki	Sulawesi
Mortlockese	Micronesian	Caroline Islands
Mota	Torres-Banks	Vanuatu

Motu	Papuan Tip	New Guinea
Motu, Hiri	Papuan Tip	New Guinea
Mpotovoro	Malakula Coastal	Vanuatu
Mukawa/Are	Papuan Tip	New Guinea
Muko-Muko	Malayo-Chamic	Sumatra
Mumeng	North New Guinea	New Guinea
Muna	Muna-Buton	Sulawesi
Munggui	SHWNG	New Guinea
Murik	Kayan-Murik-Modang	Sarawak
Murung	Greater Barito	Kalimantan
Murut, Okolod	Murutic	Sabah
Murut, Selungai	Murutic	Sabah
Murut, Timugon	Murutic	Sabah
Mwesen	Torres-Banks	Vanuatu
Mwotlap	Torres-Banks	Vanuatu
Nahati/Nāti	Malakula Interior	Vanuatu
Naka'ela	Three Rivers	Seram
Nalik	Meso-Melanesian	New Ireland
Naman	Malakula Interior?	Vanuatu
Nanggu/Nagu	Reefs-Santa Cruz	Santa Cruz Island
Narum	North Sarawak	Sarawak
Nasal	?	Sumatra
Nasarian	Malakula Interior	Vanuatu
Nāti/Nahati	Malakula Interior?	Vanuatu
Nauna	Admiralties	Nauna Island
Nauruan	Micronesian	Nauru and Banaba Islands
Navenevene	Ambae-Maewo	Vanuatu
Navwien	Malakula Interior?	Vanuatu
Ndrehet/Drehet	Admiralties	Manus
Neku	South New Caledonian	New Caledonia
Nêlêmwa	North New Caledonian	New Caledonia
Nemboi	Reefs-Santa Cruz	Santa Cruz Island
Nemi	North New Caledonian	New Caledonia
Nengone	Loyalty Islands	Maré, Loyalty Islands
Neve'ei/Vinmavis	Malakula Interior	Vanuatu
Ngadha	West Flores	Flores
Ngaibor	Central Malayo-Polynesian	Aru Islands
Ngaju Dayak	Greater Barito	Kalimantan
Ngatikese	Nuclear Micronesian	Caroline Islands
Nggela/Gela	Southeast Solomonic	Florida Island
Nguna/North Efate	Shepherds-North Efate	Vanuatu
Nias	Barrier Island-Batak	Barrier Islands, Sumatra
Niuean	Polynesian	Niue
Nivat	Malakula Interior?	Vanuatu
Niviar	Malakula Interior?	Vanuatu
Nómwonweité/Namonuito	Micronesian	Caroline Islands
Nakanamanga	Shepherds-North Efate	Vanuatu
Notsi	Meso-Melanesian	New Ireland
Nuaulu	East Seram	Seram

Nuguria	Polynesian	Solomon Islands
Nukumanu	Polynesian	Solomon Islands
Nukuoro	Polynesian	Caroline Islands
Numbami	North New Guinea	New Guinea
Numfor/Biak	SHWNG	New Guinea
Nusa Laut	East Piru Bay	Nusa Laut Island
Nyelâyu	North New Caledonian	New Caledonia
Old Javanese	?	Java
Olrat	Torres-Banks	Vanuatu
Onin	Yamdena-Sekar	New Guinea
Orang Kanaq	Malayo-Chamic	Malay peninsula
Orang Seletar	Malayo-Chamic	Malay peninsula
Orap	Malakula Coastal?	Vanuatu
Orkon	Ambrym-Paama	Vanuatu
Oroha	Southeast Solomonic	Malaita
Osing	?	Java
Ot Danum	Greater Barito	Kalimantan
Pááfang/Hall Islands	Micronesian	Caroline Islands
Paamese	Ambrym-Paama	Vanuatu
Paicî	South New Caledonian	New Caledonia
Paiwan	Paiwan	Taiwan
Pak	Admiralties	Pak Island
Paku	Greater Barito	Kalimantan
Palauan	?	Palau
Palawano	Greater Central Philippines	Palawan
Palu'e	West Flores?	Flores
Pamona/Bare'e	Kaili-Pamona	Sulawesi
Paneati	Papuan Tip	New Guinea
Pangasinan	Northern Luzon	Luzon
*Pangsoia-Dolatak	?	Taiwan
Papitalai	Admiralties	Manus
*Papora	Western Plains	Taiwan
Patpatar	Meso-Melanesian	New Ireland
*Paulohi	Piru Bay	Seram
Pazeh	NW Formosan?	Taiwan
Pekal	Malayo-Chamic	Sumatra
Pelipowai	Admiralties	Manus
Penchal	Admiralties	Rambutyo Island
Penesak	?	Sumatra
Penrhyn/Tongareva	Polynesian	Cook Islands
Peterara	Ambae-Maewo	Vanuatu
Pije	North New Caledonian	New Caledonia
Pileni	Polynesian	Santa Cruz Islands
Pingilapese	Micronesian	Caroline Islands
Piru	West Piru Bay	Seram
Pitu Ulunna Salo	South Sulawesi	Sulawesi
Pohnpeian/Ponapean	Micronesian	Caroline Islands
Ponam	Admiralties	Manus
Port Sandwich	Malakula Coastal	Vanuatu

Pukapukan	Polynesian	Pukapuka
Pulo Annian	Micronesian	Caroline Islands
Puluwat	Micronesian	Caroline Islands
Punan Aput	Kayan-Murik-Modang	Kalimantan
Punan Batu	?	Sarawak
Punan Merah	Kayan-Murik-Modang?	Kalimantan
Punan Merap	?	Kalimantan
Puyuma	Puyuma	Taiwan
Pwaamei	North New Caledonian	New Caledonia
Pwapwa	North New Caledonian	New Caledonia
Qae	Southeast Solomonic	Guadalcanal
*Qauqaut	?	Taiwan
Raga	Pentecost	Vanuatu
Rajong	West Flores?	Flores
Rakahanga-Manihiki	Polynesian	Cook Islands
Ramoaaina/Duke of York	Meso-Melanesian	New Ireland
Rapa	Polynesian	Austral Islands
Rapanui/Easter Island	Polynesian	Easter Island
Rarotongan	Polynesian	Cook Islands
Ratagnon	Greater Central Philippines	Mindoro
Ratahan/Toratán	Sangiric	Sulawesi
Rejang	?	Sumatra
Rembong	West Flores	Flores
Rennell-Bellona	Polynesian	Solomon Islands
Rhade/Rade	Malayo-Chamic	Vietnam
Ririo	Meso-Melanesian	Choiseul
Riung	West Flores	Flores
Roglai, Cacgia	Malayo-Chamic	Vietnam
Roglai, Northern	Malayo-Chamic	Vietnam
Roglai, Southern/Rai	Malayo-Chamic	Vietnam
Roma	Central Malayo-Polynesian	Roma Island
Rongga	West Flores	Flores
Roria	West Santo	Vanuatu
Roro	Papuan Tip	New Guinea
Rotinese	West Timor	Roti
Rotuman	Central Pacific	Rotuma
Roviana	Meso-Melanesian	New Georgia Archipelago
Rowa	Torres-Banks	Vanuatu
Rukai, Budai	Rukai	Taiwan
Rukai, Maga	Rukai	Taiwan
Rukai, Mantauran	Rukai	Taiwan
Rukai, Tanan	Rukai	Taiwan
Rukai, Tona	Rukai	Taiwan
Sa'a	Southeast Solomonic	Malaita
Saaroa	Kanakanabu-Saaroa	Taiwan
Sa'ban	North Sarawak	Sarawak
Sa'dan Toraja	South Sulawesi	Sulawesi
Saisiyat	NW Formosan?	Taiwan
Sakao	East Santo	Vanuatu

Salas	?	Seram
Saliba	Papuan Tip	New Guinea
Samal, Central	Greater Barito	Sulu Archipelago
Sambal	Central Luzon	Luzon
Samoan	Polynesian	Samoa
Sangil	Sangiric	Mindanao
Sangir	Sangiric	Sangir Islands
Saparua	Piru Bay	Saparus Island
Sasak	Bali-Sasak	Lombok
Satawalese	Micronesian	Caroline Islands
Sawai	SHWNG	Halmahera
Sebop	North Sarawak	Sarawak
Seediq	Atayalic	Taiwan
Seimat	Admiralties	Ninigo Lagoon
Sekar	Yamdena-Sekar	New Guinea
Seke	East Vanuatu	Vanuatu
Seko	South Sulawesi	Sulawesi
Selako	Malayo-Chamic	Kalimantan
Selaru	Yamdena-Sekar?	Tanimbar Archipelago
Selau	Northwest Solomonic	Bougainville
Selayarese	South Sulawesi	Sulawesi
Sengga	Northwest Solomonic	Choiseul
Sepa	Piru Bay	Seram
Sera	North New Guinea	New Guinea
Seraway	Malayo-Chamic	Sumatra
*Seru	?	Sarawak
Serui-Laut	SHWNG	New Guinea
Shark Bay	East Santo	Vanuatu
Sian	?	Sarawak
Siang	Greater Barito	Kalimantan
Siar	Meso-Melanesian	New Ireland
Sichule	Barrier Island-Batak	Barrier Islands, Sumatra
Sika	East Flores	Flores
Sikaiana	Polynesian	Solomon Islands
Simbo	Northwest Solomonic	Simbo Island
Simeulue/Simalur	Barrier Island-Batak	Barrier Islands, Sumatra
Sinaugoro	Papuan Tip	New Guinea
Singhi	Land Dayak	Sarawak
*Siraya	East Formosan	Taiwan
Sissano	North New Guinea	New Guinea
So'a	West Flores?	Flores
Soboyo	West Central Maluku	Sula Archipelago
Solorese	East Flores	Solor Archipelago
Solos	Northwest Somomonic	Bougainville
Sonsorol	Micronesian	Caroline Islands
Sori	Admiralties	Manus
Sörsörian	Malakula Interior?	Vanuatu
South Gaua	Banks	Vanuatu
Suau	Papuan Tip	New Guinea

Sula	West Central Maluku	Sula Archipelago
Sumbawanese	Bali-Sasak	Sumbawa
Sundanese	?	Java
Sursurunga	Meso-Melanesian	New Ireland
Surua Hole	Malakula Interior?	Vanuatu
Sye	Southern Melanesian	Vanuatu
Taba/Makian Dalam	SHWNG	Halmahera
Taboyan	Greater Barito	Kalimantan
Tae'	Kaili-Pamona	Sulawesi
Tagakaulu	Greater Central Philippines	Mindanao
Tagalog	Greater Central Philippines	Luzon
Tagbanwa, Aborlan	Greater Central Philippines	Palawan
Tagbanwa, Central	Greater Central Philippines	Palawan
Tagbanwa, Kalamian	Greater Central Philippines	Kalamian Islands
Tahitian	Polynesian	Society Islands
Taiof	Northwest Somomonic	Bougainville
*Taivuan	East Formosan?	Taiwan
Taje	Tomini-Tolitoli	Sulawesi
*Takaraian/Makatau	?	Taiwan
Takia	North New Guinea	New Guinea
Takuu	Polynesian	Solomon Islands
Talaud	Sangiric	Talaud Islands
Talise	Southeast Solomonic	Guadalcanal
Taloki	Bungku-Tolaki	Sulawesi
Talondo'	South Sulawesi	Sulawesi
Tambotalo	East Santo	Vanuatu
Tandia	SHWNG	New Guinea
Tanema	Utupua-Vanikoro	Santa Cruz Islands
Tanga	Meso-Melanesian	New Ireland
Tangoa	East Santo?	Vanuatu
Tanimbili	Utupua-Vanikoro	Santa Cruz Islands
Tanjong	?	Sarawak
*Taokas	Western Plains	Taiwan
Tarangan, West	Aru	Aru Islands
Tasaday	Greater Central Philippines	Mindanao
Tasmate	West Santo	Vanuatu
Tausug	Greater Central Philippines	Sulu Archipelago, etc.
Tawala	Papuan Tip	New Guinea
Tboli/Tagabili	Bilic	Mindanao
Teanu/Buma	Utupua-Vanikoro	Santa Cruz Islands
Tela-Masbuar	Central Malayo-Polynesian	Babar Islands
Temuan	Malayo-Chamic	Malay peninsula
Tenis	St. Mathias	Tenis Island
Teop	Northwest Solomonic	Bougainville
Terebu	North New Guinea	New Guinea
Tetun	Central Timor	Timor, East Timor
Thao	Western Plains	Taiwan
Tigak	Meso-Melanesian	New Ireland
Tihulale	Piru Bay	Ambon Island

Tikopia	Polynesian	Solomon Islands
Tingguian	North Luzon	Luzon
Tinrin/Tĩrĩ	South New Caledonian	New Caledonia
Tiruray	Bilic	Mindanao
Titan	Admiralties	Manus
Tobati/Yotafa	Sarmi Coast	New Guinea
Tobi	Micronesian	Caroline Islands
Tokelauan	Polynesian	Tokelau Archipelago
Tolai/Kuanua	Meso-Melanesian	New Britain
Tolaki	Bungku-Tolaki	Sulawesi
Tolomako	East Santo	Vanuatu
Tombonuwo	Paitanic	Sabah
Tondano	Minahasan	Sulawesi
Tongan	Polynesian	Tonga
Tonsawang	Minahasan	Sulawesi
Tontemboan	Minahasan	Sulawesi
Toqabaqita	Southeast Solomonic	Malaita
Torau-Uruava	Northwest Solomonic	Bougainville
Tring	North Sarawak	Sarawak
*Trobiawan	East Formosan	Taiwan
Tsat	Malayo-Chamic	Hainan Island, China
Tsou	Tsou	Taiwan
Tuamotuan	Polynesian	Tuamotu Archipelago
Tubetube	Papuan Tip	New Guinea
Tubuai-Rurutu	Polynesian	Austral Islands
Tukang Besi	Tukang Besi	Tukang Besi Islands
Tungag	Meso-Melanesian	New Ireland
Tunjung	?	Kalimantan
Tuvaluan	Polynesian	Tuvalu
Ubir	Papuan Tip	New Guinea
Ukit	?	Sarawak
Ulawa	Northwest Solomonic	Contrariété Island
Ulithian	Micronesian	Caroline Islands
Umbrul	Malakula Interior?	Vanuatu
Unmet	Malakula Interior	Vanuatu
Unya	South New Caledonian	New Caledonia
Ura	Southern Melanesian	Vanuatu
Urak Lawoi'	Malayo-Chamic	Malay peninsula
Uruangnirin	Yamdena-Sekar	New Guinea
Uvol	North New Guinea	New Britain
Valpei	West Santo?	Vanuatu
Vamale	North New Caledonian	New Caledonia
Vanikoro	Utupua-Vanikoro	Santa Cruz Islands
Vano	Utupua-Vanikoro	Santa Cruz Islands
Vao	Malakula Coastal?	Vanuatu
Varisi	Northwest Solomonic	Choiseul
Vaturanga	Southeast Solomonic	Guadalcanal
Vehes	North New Guinea	New Guinea
Vitu/Muduapa	Meso-Melanesian	French Islands

Vowa	Epi	Vanuatu
Wab	North New Guinea	New Guinea
Wae Rana	West Flores	Flores
Wailengi	East Vanuatu	Vanuatu
Waima'a/Waimaha	East Timor?	East Timor
Wallisian/East Uvean	Polynesian	East Uvea, New Caledonia
Wampar	North New Guinea	New Guinea
Wanukaka	Sumba-Hawu	Sumba
Warloy	Central Malayo-Polynesian	Aru Islands
Waropen	SHWNG	New Guinea
Waru	Bungku-Tolaki	Sulawesi
Watubela	East Seram?	Watubela Islands
Wayan/Western Fijian	Central Pacific	Fiji
Weda	SHWNG	Halmahera
Wedau	Papuan Tip	New Guinea
Wemale	Three Rivers	Seram
Wetan	Central Malayo-Polynesian	Babar Islands
Weyewa	Sumba-Hawu	Sumba
Whitesands	Southern Melanesian	Vanuatu
Windesi	SHWNG	New Guinea
Wogeo	North New Guinea	New Guinea
Woleaian	Micronesia	Caroline Islands
Wolio	Wotu-Wolio	Buton Island
Wotu	Wotu-Wolio	Sulawesi
Wuvulu	Admiralties	Wuvulu and Aua Islands
Xârâcùù/Canala	South New Caledonian	New Caledonia
Yabem	North New Guinea	New Guinea
Yakan	Greater Barito	Basilan Island
Yamdena	Yamdena-Sekar	Tanimbar Archipelago
Yami	Bashiic/Batanic	Orchid Island, Taiwan
Yapese	?	Caroline Islands
Yoba	Papuan Tip	New Guinea
Yogad	Northern Luzon	Luzon
Zazao/Kilokaka	Meso-Melanesian	Santa Isabel
Zenag	North New Guinea	New Guinea
Zire	South New Caledonian	New Caledonia

1   *The Austronesian world*

1.0   Introduction

Many aspects of language, especially in historical linguistics, require reference to the physical environment in which speakers live, or the culture in which their use of language is embedded. This chapter sketches out some of the physical and cultural background of the Austronesian language family before proceeding to a discussion of the languages themselves. The major topics covered include 1. location, 2. physical environment, 3. flora and fauna, 4. physical anthropology, 5. social and cultural background, 6. external contacts, and 7. prehistory.

1.1   Location

As its name ('southern islands') implies, the AN language family has a predominantly insular distribution in the southern hemisphere. Many of the more westerly islands, however, lie partly or wholly north of the equator. The major western island groups include the great Indonesian, or Malay Archipelago, to its north the smaller and more compact Philippine Archipelago, and still further north at 22 to 25 degrees north latitude and some 150 kilometres from the coast of China, the island of Taiwan (Formosa). Together these island groups constitute insular (or island) Southeast Asia. Traditionally, the major eastern divisions, each of which includes several distinct island groups, are Melanesia (coastal New Guinea and adjacent islands, the Admiralty Islands, New Ireland, New Britain, the Solomons, Santa Cruz, Vanuatu, New Caledonia and the Loyalty Islands), Micronesia (the Marianas, Palau, the Caroline Islands, the Marshalls, Nauru and Kiribati), and Polynesia (Tonga, Niue, Wallis and Futuna, Samoa, Tuvalu, Tokelau, Pukapuka, the Cook Islands, the Society Islands, the Marquesas, Hawai'i, Rapanui or Easter Island, New Zealand, and others). Because a number of Polynesian 'Outlier' languages are also spoken in Melanesia and Micronesia, the Polynesian heartland is often distinguished as 'Triangle Polynesia', defined by a northern apex in Hawai'i, and a southern base connecting New Zealand to Easter Island. Three of these regions thus take their names from characteristics of the land forms within them ('Indian islands', 'small islands', 'many islands'), while the fourth ('black islands') takes its name from a physical characteristic of its inhabitants. A few cases, as the Fijian Islands and the tiny island cluster of Rotuma, resist easy categorisation. Together these large geographical regions constitute Oceania. In more recent treatments the terms 'Near Oceania', describing the larger and generally intervisible islands of the western Pacific, and 'Remote Oceania', describing the smaller and more widely scattered islands of the central and eastern Pacific, have taken precedence over the terms 'Melanesia', 'Micronesia' and 'Polynesia', particularly among Pacific archaeologists (Pawley and Green 1973, Green 1991). Rather surprisingly, at the western edge of the Indian Ocean is a lonely outpost of the Austronesian world—the large, and geologically long-isolated island of Madagascar. In addition, a few AN languages are spoken on the Asian mainland, including Malay in the southern third of the Malay peninsula, Moken off

the western coast of peninsular Burma and Thailand, and members of the Chamic group, numbering some seven or eight languages in Vietnam and Cambodia, and a single language (Tsat) on Hainan Island in southern China.

The boundaries of the AN world, proceeding clockwise, are as follows. In the west the Great Channel separates the 1,600 kilometres long and entirely AN-speaking island of Sumatra (together with Sabang and other near offshore islands) from the small Austroasiatic-speaking Nicobar Islands which stretch in a north-south chain some 160 kilometres to the north between the Bay of Bengal and the Andaman sea. In contrast to the distinct break made possible by a sea interval, language families on the Asian mainland interlock in a bewildering ethnolinguistic puzzle. In the southern third of the Malay Peninsula the typically coastal Malays yield ground in the upper courses of the major rivers to the Austroasiatic-speaking 'Orang Asli' (Malay for 'original people') of the interior rainforest. On the Malaysia-Thailand border, and continuing northward for some distance, speakers of phonologically aberrant Malay dialects commingle with speakers of Thai. North of the Malays on the west coast of peninsular Thailand and in Myanmar (Burma), the AN-speaking Moken, sometimes called 'sea gypsies' from their migratory life in houseboats, wander over the numerous islands of the Mergui Archipelago and parts of the adjacent mainland, where they come into contact with speakers of Thai (Tai-Kadai), Burmese, and Karen (Sino-Tibetan). Across the Andaman sea, some 500 kilometres to the west, lie the Andaman Islands, once home to speakers of languages that belong to two widely divergent groups (North Andamanese, South Andamanese), long thought to have no external linguistic relatives, but assigned by Joseph Greenberg in 1971 to a highly speculative and generally rejected superfamily called 'Indo-Pacific' (Blust 1978c, Pawley 2009a). Only a small population of South Andamanese speakers survives today.

The northern boundary of the AN language family in Asia is relatively sharp. All of the 15 surviving aboriginal languages of Taiwan and the dozen or so that are extinct are AN, whereas the Ryukyu Islands north of Taiwan are home to various forms of Ryukyuan, regarded either as divergent dialects of Japanese, or as a distinct language or languages closely related to Japanese.

Since the Polynesian languages extend to the easternmost inhabited islands of the Pacific, it might be said that the eastern boundary of the AN language family falls between these and the west coast of the Americas. But a large island of alien speech lies between Indonesia and the Pacific. With an estimated 750 languages belonging to a number of distinct families, the mountainous island of New Guinea (about one and one half times the area of France) is perhaps the nearest real-life equivalent of the biblical Tower of Babel. Although the languages and the population of the greater part of the island are often called 'Papuan', in its linguistic sense this term has never meant anything more than 'non-Austronesian.' Over the past three decades evidence has accumulated that roughly two-thirds of the Papuan languages of New Guinea probably belong to a single large, diffuse genetic grouping that the pioneering Papuanist S.A. Wurm in the 1970s christened the 'Trans-New Guinea phylum' (Pawley, Attenborough, Golson, and Hide 2005). The remaining non-AN languages of the region are partitioned between ten other 'phyla.' Foley (1986:3) adopts a more conservative position, recognizing "…upwards of sixty Papuan language families plus a number of Papuan languages, probably a couple of dozen, which are isolates."

In eastern Indonesia non-AN languages are found on Timor, Alor, Pantar, and Kisar in the Lesser Sunda Islands, and on Halmahera in the northern Moluccas. A language that appears to have been non-AN was also spoken near the western tip of Sumbawa in the

Lesser Sundas until the first decade of the nineteenth century. This language, known only from a vocabulary of 40 words collected during the Raffles governorship of Java, disappeared following the catastrophic eruption of Mount Tambora in 1815. Donohue (2007:520) argues, largely on the basis of typological traits in the inferred phonology, that this was "a Papuan language spoken by a trading population of southern Indonesia."

Other non-AN languages are spoken on New Ireland and New Britain in the Bismarck Archipelago, on Rossel Island in the Louisiade Archipelago southeast of New Guinea, on Bougainville and the smaller islands of Vella Lavella, Rendova, New Georgia, the Russell Islands, and Savo in the western and central Solomons. Greenberg (1971:807) has maintained that "the bulk of non-Austronesian languages of Oceania from the Andaman Islands in the west on the Bay of Bengal to Tasmania in the southeast forms a single group of genetically related languages for which the name Indo-Pacific is proposed. The major exception to this generalisation is constituted by the indigenous languages of Australia, nearly all of which are generally accepted as related to each other." Since the Australian family shows no evidence of relationship to AN, the southern boundary of the AN language family in insular Southeast Asia falls between the island world to the north and the continent of Australia.

Finally, although distant genetic relationship has been suggested between AN and various language families of mainland Asia or Japan, the classification of particular languages as AN is rarely problematic. As will be seen, the distribution of genetically problematic languages nonetheless shows a distinct geographical bias: whereas the western boundary has been seriously disputed only once (vis-à-vis the position of the Chamic languages), and then through an error that was later widely recognised as such, the boundary between AN and Papuan sometimes still presents difficulties in the classification of the languages of Melanesia.

1.2   Physical environment

Most of the AN world lies within ten degrees of the equator, making it almost exclusively tropical or sub-tropical. Many of the islands are volcanic in origin, and several areas, including the island of Hawai'i (from which the Hawaiian chain is named), parts of Vanuatu and western Melanesia, and an extensive zone skirting the southern and eastern boundaries of Indonesia and extending northward through the Philippines, are centers of active volcanism and seismic activity. The violent and destructive eruptions of Mount Tambora in 1815, of the islet of Krakatau in the Sunda strait between Java and Sumatra in 1883, of Gunung Agung on the island of Bali in 1962, and of Mount Pinatubo in the Zambales Mountains of western Luzon in 1991 are only among the more recent and spectacular instances of volcanic activity which has been a continuing feature of the environment of many AN-speaking peoples for millennia. Reflexes of *linuR or *luniR 'earthquake' are widespread in Taiwan, the Philippines, and western Indonesia, but no widely distributed cognate set for 'volcano' is known, although the structural collocation 'fire mountain' appears in a number of languages.

The islands of Indonesia are commonly divided into Greater Sunda and Lesser Sunda groups, a distinction based in part on size and in part on geological origin. Among the former are Borneo, Sumatra (third and sixth largest in the world), Java, and Bali. The Lesser Sunda chain includes the smaller islands from Lombok east to Timor and beyond, where the eastern Lesser Sundas and southern Moluccas merge across a vaguely defined boundary. Although not generally enumerated among the Greater Sunda Islands, the

smaller islands flanking Sumatra, Borneo and Java, including the Barrier Islands west of Sumatra, Bangka and Belitung (Billiton) between Sumatra and Borneo, Madura off the north coast of Java, Bali just east of Java, and Palawan in the southwestern Philippines, like their larger neighbors, rest on the submerged Sunda Shelf, a submarine extension of the Asian mainland that was exposed during the last glacial maximum.

The Aru Islands in the southern Moluccas, like the great island of New Guinea of which they are geologically a part, lie on the Sahul Shelf, a submarine extension of Australia. All other islands in Indonesia and the Philippines, including the Moluccas (once famous for their cloves, mace, and nutmeg), and the relatively large island of Sulawesi in central Indonesia, formerly called 'the Celebes' or 'the orchid of the equator' from its curious shape, occupy Wallacea, a zone of geological instability between these shelves named after the nineteenth-century British naturalist Alfred Russel Wallace. During glacial maxima the area that now includes insular Southeast Asia and Australia-New Guinea thus consisted of three large divisions: 1. Sundaland, an extension of the Asian mainland, 2. Sahulland, a single landmass which during glacial minima split into New Guinea, Australia and Tasmania, and 3. Wallacea, a shifting island world between these larger, more stable continental blocks.

An important geological boundary in the Pacific is the Andesite Line. Islands that lie to the west of this line rest on the continental shelf of Australia (e.g. New Caledonia, Fiji), while those lying to the east are true Oceanic islands (e.g. the Societies or Hawai'i). The latter, being of volcanic origin and never having been connected to any continental land mass, suffer from varying degrees of biological impoverishment.

On some of the larger islands, as Borneo, where dense vegetation, high rainfall and dangerous or noxious animals can greatly impede progress by foot, the rivers form natural avenues to the interior. The subgrouping of the languages of northern Sarawak and Sabah suggests that the settlement of northern and western Borneo by AN speakers proceeded along the coast and then up the major river systems. It is also along such a relatively open route (the Markham valley) that AN-speakers made their only significant penetration of the hinterland of New Guinea. In some areas heavy rainfall produces considerable loss of topsoil, which is carried downriver to the sea. The resulting alluvial deposits around the mouths of major rivers have created large sections of eastern Sumatra and southern Borneo within the relatively recent past. Deforestation resulting first from swidden agriculture and more recently from international logging undoubtedly have accelerated this process.

A number of the inhabited islands of Micronesia and some elsewhere in the Pacific, as the Tuamotus of French Polynesia, are coral atolls rising no more than a few meters above sea level. Micronesian atolls are a particularly precarious habitat, as many lie in the typhoon belt that runs from the region of Chuuk (Truk) in the eastern Carolines, west and northwest to the Philippines, Taiwan and southern Japan. Typhoon damage to vegetation may require six or seven years for recovery, and in a fragile atoll environment that in any case offers limited opportunities for food production, this can be disastrous (Alkire 1977).

There is considerable seasonal rainfall over much of the AN world, although in general the region can be characterised as wet. In the monsoon regime of island Southeast Asia and western Melanesia sailing conditions and other facets of economic life are greatly affected by the seasonal variation in dominant rain-bearing winds. That these conditions have been important to AN-speaking peoples for millennia seems likely from such linguistic expressions as Malay *mata aŋin*, Fijian *mata ni caŋi* (lit. 'eye of the wind') as the general term for 'direction, point of the compass', and by such specific reconstructed directional terms as *habaRat 'west monsoon' and *timuR 'east monsoon.'

The more low-lying areas of many of the larger islands of the AN world are hot and humid, and malaria is a serious problem in much of Melanesia. Surrounded as they are by cooling waters and gentle sea breezes, however, the smaller and more remote islands of Polynesia have been regarded by European romantics with some justification as earthly paradises. Much the same could be said for the generally even smaller islands of Micronesia, although the majority of these are atolls and have failed to capture the European imagination to the same extent as the more striking high islands of Samoa, Tahiti or Hawai'i. The more elevated inhabited areas of the larger islands, such as the Imerina plateau of central Madagascar, or the Kerayan-Kelabit or Usun Apau Highlands of central Borneo, are often quite cool at night and are subject to occasional hailstorms. In only a few areas of extreme altitude (the 4,200 meter Mauna Kea and Mauna Loa volcanoes on the of Hawai'i), high latitude (the south island of New Zealand), or a combination of these in more moderate degree (various peaks rising from 3,000 to over 4,000 meters in central Taiwan) is snow seen.

1.3   Flora and fauna

Most islands in the AN world present a similar array of shore trees. Prominent among these is the ubiquitous coconut (*Cocos nucifera*). Other trees that are frequently encountered just back of the beach are the pine-like casuarina (*Casuarina equisetifolia*), the shade trees *Calophyllum inophyllum*, *Barringtonia asiatica*, and *Terminalia catappa*, some of which produce valued fruits or nuts, and such economically useful shrubs or low trees as the pandanus, or screw-pine (*Pandanus tectorius* and *Pandanus odoratissimus*), and the brightly flowering hibiscus (*Hibiscus tiliaceus*). In swampy coastal areas extensive mangrove forests are sometimes found, sending down their long prop roots into the salty or brackish water where they provide a haven for small fish or crustaceans, and a place for the attachment of oysters.

Important non-food plants include the nipa palm (*Nipa fruticans*), the leaves of which—like the leaves of the sago palm—are widely used in island Southeast Asia as material for walls and roofing, a littoral pandanus (*Pandanus odoratissimus*) from which mats are woven for floor coverings and (in the Pacific) as material for canoe sails, the *Hibiscus tiliaceus*, the bark of which is used for cordage, various types of bamboo of which the larger species are used in island Southeast Asia as vessels for carrying water or cooking food, rattan and various vines used for tying, *Derris elliptica*, the pulverised root of which is mixed with river water to immobilise fish, and a great variety of trees which yield timber for the construction of houses, canoes, etc.

Among the more important food plants common to much of the AN world are the coconut, banana (*Musa* sp.), breadfruit (*Artocarpus* sp.), sago palm (*Metroxylon sagu*), yam (*Dioscorea alata*), and taro (principally *Colocasia esculenta*, although the giant swamp taro *Cyrtosperma chamissonis* is important in some parts of the Pacific). Some plants were traditionally prized both for their food value and for other kinds of practical uses, as the *Artocarpus*, which yields the large edible breadfruit as well as a sticky sap used as birdlime. Rice is important virtually everywhere in island Southeast Asia, although its centrality in the economy diminishes in moving eastward through Indonesia, where sago assumes increasingly greater importance as a staple. East of the Moluccas grain crops are entirely absent, except in the Mariana Islands, where rice evidently was introduced by the ancestral Chamorros some 3,500 years ago. Millet is also important in parts of eastern Indonesia, as well as in Taiwan.

The interior of most islands is covered by tropical rainforest. Exceptions are the southern side of Timor and neighboring islands which lie in the path of the seasonal hot, dry winds sweeping north from the desert of central Australia, and islands at some distance from the equator (Taiwan, New Zealand). In some parts of island Southeast Asia and New Guinea extensive tracts of abandoned agricultural land have been taken over by sawgrass (*Imperata cylindrica*), and so transformed into permanent savanna.

It is impossible to discuss the animal life of the AN world meaningfully without reference to geological history. In 1869 the English naturalist Alfred Russel Wallace published his observations concerning the natural history of what he called the 'Malay Archipelago.' The most important of these observations concerned a curious division of the terrestrial fauna and of certain groups of birds between two very distinct faunal zones—a western zone which shows close affinities with mainland Southeast Asia and India, and an eastern zone which shows much stronger affinities with Australia. The break between these two zones is in some areas surprisingly abrupt. Wallace (1962:11) noted, for example, that the neighboring islands of Bali and Lombok quite unexpectedly contain radically different faunal assemblages: "In Bali we have barbets, fruit-thrushes, and woodpeckers; on passing over to Lombock these are seen no more, but we have abundance of cockatoos, honeysuckers, and brush-turkeys, which are equally unknown in Bali, or any island further west. The strait is here fifteen miles wide, so that we may pass in two hours from one great division of the earth to another, differing as essentially in their animal life as Europe does from America."

Among terrestrial mammals characteristic of one or more of the western islands are the elephant, tapir, rhinoceros, wild ox, sambhur deer (*Cervus equinus*), muntjac or barking deer (*Muntiacus muntjac*), and the mousedeer (*Tragulus kanchil*), the Malayan sun bear (*Ursus malayanus*), the tiger and clouded leopard, the pangolin or scaly anteater (*Manis javanica* in western Indonesia, but *Manis pentadactyla* in Taiwan), porcupine, wild pig, civet cat, orangutan and gibbon, various monkeys, the tupai (a tree shrew), slow loris, tarsier, otter, badger and the rat. Terrestrial mammals characteristic of the eastern islands include various species of cuscus, bandicoot (marsupial rat), tree kangaroos (Aru Islands and New Guinea), the echidna, or spiny anteater, and the rat. Wallace showed that this faunal distribution could be explained most simply if the Greater Sunda Islands exclusive of Sulawesi once formed an extension of continental Asia. Similarly, the eastern islands were once connected with or in closer proximity with Australia, but the western and eastern biotic zones have long been separated by a water barrier that is impassible by most terrestrial mammals and land birds. Wallace's inferences from faunal distribution were later found to correspond closely to measured sea depths, and in his honor this major zoogeographical boundary was named the 'Wallace Line.'

It is now known that the boundary between the Indian and Australian biotic zones is somewhat less clear that might be imagined from a 'line' drawn between them. The large island of Sulawesi partakes to some extent of both faunal regions, having a lemur (*Tarsius spectrum*), and apparently indigenous species of monkeys and wild pig characteristic of the western islands, as well two species of cuscus, and a megapode characteristic of the eastern islands. In other respects Sulawesi is zoologically unique, with two species of dwarf buffalo (genus *Bubalus*) found nowhere else, and the strange long-tusked *babirusa* (Malay for 'pig deer'), a member of the pig family found only on Sulawesi and a few smaller adjacent islands.

The northern continuation of the Wallace Line has been a matter of some controversy, but it seems clear that if the Wallace Line is taken to mark the western limit of marsupials

the Philippine Islands lie in the Indian biotic zone, although only the Island of Palawan and the smaller Calamian and Cuyo Islands near it rest on the Sunda shelf. Various types of monkeys, the sambhur deer, wild pig, and civet cat are found on most of the Philippine Islands, and within historical times a pangolin was found on Palawan and the neighboring Calamian Islands to the north. Related forms of all of these are found on Taiwan, along with species of buffalo, wild goat (serow), bear, leopard, rabbit, otter, mole, vole, and several types of squirrel. A unique species of wild buffalo (*Bubalus mindorensis*) is found on the Island of Mindoro in the central Philippines.

As various writers have observed, both the rat and various species of bats have achieved a far wider distribution (reaching Polynesia) than other non-domesticated animals, the former undoubtedly owing to its successes as a stowaway, and the latter due to its power of flight. In general, however, there is a steady decrease in the number and supra-species level variety of life-forms (particularly terrestrial mammals) as one moves from the great land masses of Asia and Australia into the realm of true oceanic Islands, culminating in the highly depauperate native biota of such isolated biological outposts as Hawai'i and Easter Island. Because they offered a wide range of virtually unoccupied habitats for the few organisms that were able to reach them before Western contact, true oceanic Islands such as those in the Hawaiian chain presented a striking contrast between numerous unique species that had arisen by adaptive radiation, but relatively few genera and families.

Widespread birds of some cultural prominence, as reflected in cognate names, include two doves (genus *Ducula*, genus *Treron*), the hornbill (in Southeast Asia and the western Pacific), white egret, woodpecker, wild duck, owl, and a quail or partridge. Within the Pacific the albatross, frigate bird and various terns or gulls are prominent.

Among reptiles the crocodile is common from the northern Philippines to the Solomons, although individual animals have been found as far east as the Marquesas in eastern Polynesia (Darlington 1980:229). Various species of snakes occur in the western Pacific and as far east as Fiji, Tonga, Samoa and Futuna in western Polynesia, but are generally absent in Micronesia, and are completely unknown in central and eastern Polynesia. In several of the languages of the Philippines and Indonesia the reconstructed term for 'python' (*sawa) has become the generic term for 'snake', attesting to the psychological prominence of this genus in the region. The distribution of the monitor lizard (genus *Varanus*) approximates that of the saltwater crocodile. One species, the 'Komodo dragon' (*Varanus komodoensis*), confined to the western tip of the Island of Flores and a few smaller Islands in the Lesser Sunda chain, is the largest extant lizard, sometimes reaching an adult length of three and one half meters.

Only in the interior of the larger Islands or on the Asian mainland is one ever far from the sea. Most AN-speaking societies are thus not only acquainted with a locally distinct terrestrial fauna, but also with the far less localised wealth of life that swarms in tropical seas. This includes such acquatic mammals as the whale (hunted in only a few isolated locations), dolphin, and in the western Pacific the dugong, as well as eels, sea snakes, sea turtles, the giant clam (genus *Tridacna*), conches (the shells of which are widely used as signal horns), octopus and squid, lobsters, various types of crabs, sharks and rays, and a dazzling variety of other fish noteworthy for their food value (Spanish mackerel, various tuna, mullet), danger on the reef (stonefish), or striking appearance (butterfly fish, parrot fish, puffer fish).

1.4   **Physical anthropology**

Likely survivors of a pre-AN population are seen in the short, dark-skinned, wooly-haired Negritos of the Philippines, who were traditionally (and in some cases still are) foragers living in cultural symbiosis with the dominant agricultural Filipinos (Garvan 1963). Some writers also distinguish a 'Dumagat' population on the east coast of northern Luzon, which is said to be 'Papuan-like', but others recognise no such distinction. Negrito groups are found in several parts of Luzon, in some of the Bisayan Islands (as Panay and Negros), and in Palawan and Mindanao. Outside the Philippines they are found in the interior of the Malay peninsula, where they speak Austroasiatic languages, although the number of loanwords from Malay is high, and apparently growing (Benjamin 1976). Other groups are found in the Andaman Islands, where they are linguistically distinct. The Negritos of Southeast Asia are presumed to represent the survivors of a population that reached this area during the Pleistocene at least 40,000 years ago. Today, all Negrito groups in the Philippines speak AN languages, but there must have been a time when this was not the case, and it has been claimed that many of the modern Negrito groups of Luzon share a pre-AN linguistic substratum (Reid 1987, 1994a). It is likely that the linguistic assimilation of Negrito bands in the Philippines and Malay Peninsula came about through trade contacts that led over time to an increasingly tighter economic interdependence of foragers and agriculturalists. The distinctiveness of the Negrito population is explicitly recognised in reflexes of the term *qaRta which appear in a number of Philippine languages as Agta, Alta, Arta, Ata, Atta, Ati, or Ayta (sometimes written Aeta). These words are often used by the dominant population of the Philippines to mean 'Negrito', but are sometimes self-appellations used by Negrito groups themselves. Reflexes of *qaRta are also found in both western and eastern Indonesia, and as far east as New Caledonia, where the meanings vary over 1. person, human being, 2. slave, and 3. outsider, alien person. Given Proto Austronesian *Cau, Proto Malayo-Polynesian *tau 'person, human being', and PWMP *qudip-en 'slave', PMP *qaRta probably meant 'outsider, alien person', an inference that is consistent with its application to the Negrito peoples of the Philippines.

In sharp contrast to the Philippines, there are no extant Negritos in Borneo, although the archaeology of Niah Cave in northern Sarawak has revealed a pre-Neolithic population extending back some 40,000 years. Given the broader ethnological picture for Southeast Asia the most likely bearers of the pre-Neolithic cultures at Niah Cave would have been ancestral Negritos. When and why these populations disappeared is unknown, but despite occasional claims of evidence for Negrito admixture in some groups (e.g. among the Muruts of Sabah), it would seem that AN and pre-AN populations in Borneo had little or no contact. Similarly, earlier accounts speak of 'Veddoid' physical characteristics among some Sumatran groups, but the population of Sumatra does not appear to differ markedly from that of Borneo or other parts of western Indonesia.

The situation in Taiwan is somewhat more complex. Dyen (1971d:171) stated that "there are reports of 'little black men', presumably Negritos, distributed widely on the west side of the Central Mountains, who disappeared about 100 years ago." Although discoveries in the Chang-pin caves and elsewhere have documented a pre-Neolithic (presumably Negrito) population on Taiwan for millennia before AN-speakers arrived, mythological references to a race of dwarfs that are current among Formosan aborigines do not indicate that they were black, and such stories are comparable to other tales of 'little people' that are widespread in the Pacific (Ferrell 1968, Luomala 1951).

Most AN-speakers in Taiwan, the Philippines, and western Indonesia are described as of 'modified Mongoloid' type. Skin color varies from olive to moderately dark brown. Hair is dark brown to black, and straight to wavy, with occasional crispness even in areas (like Java), where contact with earlier populations is not generally assumed. Eyes are dark brown to black, and ordinarily lack the Mongolian eye fold. Chai (1967), who distinguishes between an epicanthic fold and a Mongolian fold, reports frequencies of total absence ranging from 85% (Rukai men) to 61.1% (Amis women) for the former, and 96.3% (Tsou women) to 50.9% (Atayal women) for the latter among various Formosan aborigines. The Mongolian fold is thus unusually prominent among Atayal speakers, the northernmost mountain people on the Island. Chai found the mean stature of Formosan aboriginal men to vary from 164.6 cm (Amis) to 156.6 cm (Paiwan), and the mean stature of Formosan aboriginal women to vary from 155.9 cm (Amis) to 146.2 cm (Bunun). Setting aside the Amis, who appear to be unusually tall (and fair-skinned), these figures probably are representative, within fairly narrow limits, for most of the Philippines and western Indonesia. Although relatively short, the men of some groups are stocky and muscular, and obesity is rare.

The population of Madagascar is described by Murdock (1959:212) as "a complex mixture of physical types—Negroid, Mongoloid, and Caucasoid." The relatively light-skinned and straight-haired peoples of the Imerina Plateau conform to a general Southeast Asian type, while more African-influenced physical types predominate in the arid parts of the west coast (Sakalava, Bara, Mahafaly). The Caucasoid element is of limited distribution, and appears to be a product of intermarriage with Arab or European seafarers in relatively recent centuries.

Whereas in the Philippines the Negrito and southern Mongoloid populations are rather sharply distinguished populations, the physical anthropology of eastern Indonesia shows greater intergradation, ranging from western Indonesian to Papuan types. In the western Lesser Sundas, as in Sumbawa, Flores, Sawu or Sumba, physical type does not differ markedly from that in western Indonesia. Further east, in approaching New Guinea, phenotypes show much greater variation, sometimes diverging sharply from what is typical of western Indonesia. In general the most markedly Papuan physical types are found among speakers of non-AN languages, as on Alor, although the correlation between language affiliation and phenotype has been blurred by centuries of social and economic contact and gene flow on Islands such as Timor.

Mismatches of linguistic affiliation and physical type suggest that the northern Moluccas have had a complex history of human settlement. Both Papuan and AN languages are spoken on the Island of Halmahera. In general Papuan languages are spoken in northern Halmahera and on the adjacent Island of Morotai (the 'North Halmahera language family'), while AN languages are spoken in southern Halmahera. However, on the small Island of Makian off the west coast of Halmahera, Makian Dalam, or Taba, the language of the 'inside' of Makian (facing Halmahera) is AN, whereas Makian Luar, the language of the 'outside' of Makian (facing away from Halmahera) is Papuan. Surprisingly, the physical anthropology of Halmahera is the converse of the linguistic classification: many north Halmaheran speakers of Papuan languages such as Ternate, Tidore or Galela are physically Indonesian in type, while most south Halmahera speakers of AN languages exhibit a physical type more commonly associated with speakers of Papuan languages in the western Pacific. This skewing of physical type and linguistic affiliation suggests that language replacement has taken place in both northern and southern Halmahera, perhaps through centuries of jostling for control of the spice trade. In

the central and southern Moluccas physical type varies from the predominantly Indonesian type of areas such as Buru, Seram, Ambon, or Tanimbar, to the predominantly Papuan type found in the Aru Islands of the southeastern Moluccas.

Most AN speakers in New Guinea and the Bismarck Archipelago have dark brown skin and frizzy hair. However, this general description conceals a wealth of variation. Among peoples who are commonly characterised as Melanesian, skin color ranges from reddish-brown (Mekeo, Motu, Kilivila and similar peoples in Southeast New Guinea), to coal black (Buka, Bougainville, and other parts of the western Solomons).[1] Hair is naturally black to brown, reaching reddish-brown in some areas, and artificial bleaching with lime produces blond hair in parts of western Melanesia (as New Britain and New Ireland). Howells (1973) has noted that hair coils in Melanesia typically are looser than in African populations. As a result hair form is not frizzy, but ranges from wooly (Bismarck Archipelago) to bushy (Fiji). Eyes are dark brown, and the Mongolian fold occurs in some areas (as the north coast of New Guinea). Stature is variable, from relatively short in much of western and central Melanesia, to nearly the Polynesian norm in Fiji and New Caledonia. Physically, speakers of AN and non-AN languages in Melanesia appear to grade imperceptibly into one another. The attempts of some writers to distinguish a 'Melanesian' from a 'Papuan' physical type appear groundless, although there are clear somatic differences between highland and lowland populations in New Guinea that are independent of linguistic affiliation.

In addition, a few of the AN-speaking peoples of Melanesia are much closer in physical type to the populations of island Southeast Asia or Micronesia than they are to other populations in Melanesia. In some cases, as with the dozen or so Polynesian Outlier communities in the Solomon, Santa Cruz, Vanuatu and Loyalty Archipelagos, this variation can be explained as a product of back-migration. In other cases, however, the explanation must be different. The people of the tiny islands of Wuvulu and Aua, some 170 kilometres north of the mouth of the Sepik River in New Guinea and 375 kilometres due west of the island of Manus in the Admiralty group, have yellowish-brown skin with wavy to slightly frizzy hair, yet their home islands lie within Melanesia as it is usually defined. Even more significantly, Wuvulu-Aua subgroups with the languages of the dark-skinned, frizzy-haired peoples of the eastern Admiralties. A similar light-skinned, relatively straight-haired physical type appears to have been common in the now extinct population of the Kaniet Islands, some 170 kilometres northwest of Manus, and what is described as a 'mixed' Melanesia physical type is found on the tiny Island of Tench (or Tenis), 100 kilometres north of New Ireland and 65 kilometres east of the island of Emira in the St. Matthias Archipelago. It is noteworthy that where malaria is severe light skin and straight or wavy hair do not appear, but where it is mild or absent these physical traits sometimes are present. The German linguist Otto Dempwolff, who studied the problem of differential resistance to malaria during his earlier career as a medical doctor, regarded this partial correlation as a key to certain major features of the racial history of the Pacific. He concluded that early AN speakers were southern Mongoloids who had little resistance to malaria. In the western Pacific this latecoming wave of maritime immigrants encountered a long-established, dark-skinned, frizzy-haired population that had acquired malaria resistance through generations of exposure and selection. Those AN speakers who remained in severe malarial areas without intermarrying with the local population died out.

---

[1]   Andrew Pawley (p.c. 4/22/09) has pointed out that "clear Southeast Asian characteristics" are found even among the Koita, a Papuan-speaking group that has long been in intimate contact with the Motu, and has no doubt been strongly affected by gene flow from them.

Those who remained and intermarried survived, and in many cases passed on their language and culture, but were modified in physical type. Those AN speakers who moved on to non-malarial areas quickly enough were able to pass on their language and culture without modifying their physical type.

In general, Micronesians differ sharply in physical type from typical populations in Melanesia. They are sometimes described as intermediate between Southeast Asians and Polynesians, since they tend to be larger than most Southeast Asians, but are shorter and darker than most Polynesian groups. The Palauan phenotype shows possible contacts with western Melanesia, but this is not true of any other Micronesian population.

A similar physical type prevails throughout Triangle Polynesia, although some Polynesian Outlier populations show evidence of gene flow with neighboring Melanesian groups. As noted repeatedly by early European voyagers, Polynesians differ strikingly in physical type from most peoples of Melanesia in at least two respects: they are much taller, with lighter skins and straighter hair. Young men often have powerful builds, and not only do both sexes tend to corpulence as they age, but over much of Polynesia corpulence was institutionalised as a cultural value. Fijians are usually described as physically Melanesian, but they are culturally closer to Polynesians, vary considerably in skin color, and are far taller than most Melanesian populations further west. Rotumans are physically similar to Polynesians, but speak a non-Polynesian language.

1.5   Social and cultural background

AN-speaking societies cover a wide range of ecological adaptations and levels of control over their environment. Technologically, and in other ways as well, the simplest societies are those of the hunter-gatherers. Hunting and gathering groups have been known for some time in the Philippines and in Indonesia, as with the various Negrito groups of Luzon, the central Philippines and Mindanao, the Penan/Punan of Borneo, the Kubu and Lubu of Sumatra, the Toala of Sulawesi, and the Kadai of the Sula Archipelago in the central Moluccas.

In 1971 reports of a few previously uncontacted hunting and gathering families in the forested mountains of Mindanao caused a popular sensation. To some writers these simple people -- the Tasaday --, with their claimed ignorance of agriculture, their stone tools and lack of permanent habitations, represented the 'original Filipinos.' However, these views never sat well with the linguistic evidence, which showed that Tasaday and Blit Manobo, spoken by a sedentary agricultural population in Mindanao, shared a common linguistic ancestor some 500-750 years ago (Molony and Tuan 1976). The simplest explanation for Tasaday nomadism thus appeared to be reversion from an earlier sedentary lifestyle to a life of foraging. In 1987 the case of the Tasaday was asserted to be an elaborate hoax, contrived for political and monetary gain. After much, often acrimonious debate among social and cultural anthropologists, this view in turn has been overthrown, and it is now believed that the original reports, although not necessarily the interpretations that accompanied them, were accurate (Hemley 2005).

Heated debate has also characterised discussions of the origin of the forest nomads of Borneo. Hoffmann (1986) sees the Penan and Punan as earlier agriculturalists who reverted to forest nomadism as part of a supply system stimulated by the Chinese demand for forest products. Brosius (1988) and Sellato (1988) vigorously contest this view, arguing that the history of Penan/Punan nomadism is no different from that of the Negritos of the Philippines: nomadic groups in Borneo represent a pre-AN population that adopted AN

languages through contact, and historically have shown a pattern of sedentarisation and cultural assimilation to neighboring agricultural groups. This view is not inherently implausible, but the burden of proof certainly rests on those who argue that the forest nomads of Borneo, who are phenotypically, linguistically, and in some respects culturally very similar to their sedentary trade partners, have acquired their linguistic affiliation through language shift and their physical similarities to sedentary populations through chance. Rather than comparing them with the Negritos of the Philippines, who are biologically distinct from other Filipinos, a more revealing comparison of the Bornean nomads can perhaps be made with the Tasaday, who are also linguistically and phenotypically very similar to neighboring sedentary groups, and who appear to have abandoned an earlier sedentary lifestyle. Moreover, in at least two cases foraging is known to be historically secondary. The first of these is the Mikea, who follow a semi-nomadic lifestyle in the thorn forests of southwest Madagascar, but are descended from the agricultural founding population of Austronesian-speakers who migrated from Borneo (Kelly, Rabedimy and Poyer 1999). The second is the Moriori of the Chatham Islands. This Polynesian group, which reached the Chathams from New Zealand, was forced to adapt to a colder climate and a significantly impoverished natural environment in comparison with other Polynesian peoples. As a result, they lacked cultivated plants, domesticated animals and large trees for the construction of canoes or houses. When first encountered by Europeans they were largely migratory (Skinner 1923).

Like the Punan and Penan of Borneo which have received more attention in the recent literature, Sumatra has its own nomadic or formerly nomadic groups of hunter-gatherers. These go under a variety of names, including 'Orang Mamaq', 'Orang Ulu', 'Batin', 'Kubu' and 'Lubu'. Those found in the alluvial lowlands of southeast Sumatra between the Musi and Batang Hari Rivers are known in the literature as 'Kubu', while other groups in contact with the Mandailing Batak further north are called 'Lubu'. Essentially the same debate that has taken place during the past two decades with regard to the origin of the forest nomads of Borneo took place a century ago in the (mostly) German literature on the ethnology of Sumatra. In both cases one position holds that the nomads have undergone a 'devolution' from sedentary agriculturalist ancestors, while the other holds that they represent pockets of cultural conservatism. The cultural state of the Kubu is particularly striking, as these peoples are found near the site of ancient Śrīwijaya, a major center of Buddhist learning and maritime commerce in the seventh century. All Kubu groups appear to speak dialects of Malay, while the Lubu speak a dialect of Mandailing Batak. Such culturally conservative groups as the Tenggerese of east Java, or the Bali Aga of central Bali are sometimes said to show physical differences from the mainstream populations around them, but the idea that they represent relic populations with a radically different history is unsupported. Rather, like the Sundanese-speaking Badui of western Java, they appear to represent pockets of cultural conservatism of a type not unknown in western societies, as with the Amish of Pennsylvania.

Although the people of the isolated Mentawai and Enggano Islands west of Sumatra are sedentary horticulturalists, it is often remarked that at the time of western contact they possessed a very impoverished material culture. Early accounts maintained that the inhabitants of both groups lacked rice agriculture, weaving and metallurgy. Mentawai is said to have also lacked the manufacture of pottery and the use of betel. The occasional claim that these cultures are 'archaic' (Schefold 1979-80:13ff), like similar claims about the Tasaday, imply the preservation of a way of life once common to other speakers of AN languages. Comparative linguistic data relating to early AN culture, however, shows

unambiguously that, rather than being 'living fossils', these atypical cultures are products of reversion to a materially simpler way of life.

There is no known trace of a Negrito presence in the population of Sulawesi. A single group, the Toala (< *tau 'person, human being' + *halas 'forest', hence 'forest people') of the southwestern peninsula, were nomadic in the past, but reportedly were removed from rockshelters and settled in a single village through the intervention of the dominant Buginese of the region within historical times. Reports of other foraging groups sometimes surface, including claims that small segments of sedentary populations in the Gorontalo region adopted a nomadic lifestyle during the Dutch administration in order to find relief from the colonial taxation system, but in general the reported incidence of nomadism in Sulawesi is much lower than that in Borneo or Sumatra.

Unlike areas further west, hunter-gatherers have not been reported in the Lesser Sundas. The reasons for this difference between the Greater Sundas and Lesser Sundas are unclear, but two factors which distinguish the regions stand out. First, islands such as Borneo or Sumatra, or even Mindanao are considerably larger than Timor (the largest of the Lesser Sundas). To some extent, then, nomadism may correlate with quantity of available land for foraging populations to sustain themselves. Second, the relative abundance of edible forest products probably is a factor determining the possibility of maintaining a hunting and gathering lifestyle. The nomadic Punan of Borneo rely heavily on stands of wild sago, which they themselves help to propagate. By contrast, several of the Lesser Sunda Islands present a semi-arid savanna-like landscape of scattered trees and scrub growth that is far poorer in greenery than the rainforests of Borneo or Sumatra.

Most Austronesian speakers are agriculturalists. Village organisation varies from the dispersed hamlets of groups like the Saisiyat (Taiwan), Ibaloy (Luzon) and Subanen/Subanun (Mindanao), to the highly concentrated longhouse communities characteristic of much of central and western Borneo, parts of southern Sumatra, and the Chamic area of mainland Southeast Asia. Perhaps the most common type of village consists of a cluster of family dwellings arranged around a square, together with a communal building used for the conduct of social, political, and in some cases religious affairs. Bachelors' clubhouses were traditionally found in Taiwan, the northern Philippines, Kalimantan, Sumatra, western Micronesia (the Marianas, Palau) and most of Melanesia, and in central Micronesia large menstruation houses are found as well. In Island Southeast Asia rice granaries, and in Melanesia yam storage houses are common village structures. Traditional villages are often divided by a path or stream into two mutually supportive, mutually antagonistic halves—a type of dualistic organisation reported from traditional societies in many parts of the world. In some areas, such as Madagascar, New Zealand and the Island of Rapa in the Austral Islands, settlements were built on hilltops and fortified.

House types vary greatly, but frequently recurring features include 1) a gabled roof, 2) thatching of palm leaves, and 3) elevation on houseposts, and use of a (usually notched log) ladder to enter. Among the Atayal in the mountains of northern Taiwan, where winter temperatures can dip to freezing, traditional dwellings were semi-subterranean, and some excavation of the floor is also found in a few more southerly groups which lie in the Pacific typhoon belt, as with the Yami of Botel Tobago Island off the southeast coast of Taiwan, and the closely related Itbayaten and Ivatan in the Batanes Islands of the northernmost Philippines. In terms of sheer size the most imposing structures in the AN world are the longhouses of Borneo, some of which are said to reach a length of 400 meters. In the extreme case these constitute a single-structure village, although villages

may contain more than one longhouse. Like most single-family dwellings, longhouses are raised some two to three meters from the ground on wooden pillars. Typically, the structure is divided along its length into public and private portions, the former constituting a gallery where work, play and social contact take place, and the latter subdivided by walls into nuclear family units. Although less grand, the single (extended) family dwellings of some other western Indonesian peoples—particularly the Batak, Nias and Minangkabau of Sumatra, and the Toraja of central Sulawesi—are sometimes architecturally magnificent structures (Waterson 1990).

Almost everywhere in the AN world boats are important for transportation. Although evidence for the outrigger among Formosan aborigines is problematic, this distinctive stabilizing device is almost universal among AN-speaking populations outside Taiwan. Double-outrigger canoes are the norm in island Southeast Asia and Madagascar, and are found in parts of western Melanesia, while single-outrigger canoes are confined to the insular Pacific. On the larger islands of island Southeast Asia simple dugouts without an outrigger are poled or paddled on the rivers. Some riverine populations which traditionally have had no contact with the sea, such as the Kayan and Kenyah of central Borneo, are nonetheless skilled canoemen. Both boat construction and house construction were traditionally accomplished without the use of nails. Boat planks were joined by means of dowels and lashing, and house beams by mortise and tenon joints.

Most Micronesian islands are low coral atolls, and so are quite small, but these are punctuated by occasional high islands that form major population and political centers. The contrast between low and high islands in Micronesia is fundamental both in terms of population size (and hence political influence), and in terms of culture history. While it is clear that Micronesia could not have been settled without a sophisticated navigational technology, for example, the practice of long-distance voyaging has been lost on all of the high islands, where it is no longer of critical importance. On the atolls, however, open-sea voyaging is crucial to survival, since the seasonal typhoons can leave these islands temporarily uninhabitable, and under these conditions only those groups that could successfully relocate could pass on their culture and language.

The economy of most of the indigenous peoples of Taiwan, the Philippines, Indonesia, and Madagascar is based on rice, although millet is equally or even more important among some Formosan aboriginal groups. In the relatively arid islands around Timor maize, probably introduced by the Portuguese in the sixteenth century, has within historical times been the major crop. Sago is the staple over much of the Moluccas, as well as in isolated pockets to the west, most notably in the swampy Melanau coastal zone of Sarawak. Throughout island Southeast Asia yams, taro, and other root crops generally are of secondary importance, but in a few scattered parts of Southeast Asia, and in the Pacific as a whole, these plants have become central to the economy. The South American sweet potato—a subject of both ethnobotanical and ethnohistorical controversy—has assumed considerable importance in some areas (Yen 1974, Scaglion 2005).

Rice cultivation is of two types: dry and wet. Dry rice is grown by slash-and-burn agriculture in small lowland or hillside swidden plots, where it typically is intercropped with other cultigens. Soil fertility is rapidly depleted, and many plots eventually give way to sword grass (*Imperata cylindrica*), creating long-term or permanent 'green deserts'. Wet rice provides a much higher yield, but also requires far greater labor for construction and maintenance of the irrigation system. Among the most impressive achievements of traditional agriculture anywhere are the massive rice paddies of the Ifugao and their

neighbors in northern Luzon, where entire mountain slopes have been transformed into irrigated terraces descending for 500 meters or more (Conklin 1980).

Major refreshments include the nut of the betel palm (*Areca catechu*), which is wrapped in a leaf with powdered lime and chewed, and *kava*, a mildly intoxicating drink made from the fermented root of the *Piper methysticum* shrub. While the betel nut is chewed in much the same way that cigarettes are smoked in the modern world, *kava* has ritual and ceremonial associations in many Pacific Island cultures, and typically is drunk as part of a formal gathering rather than by individuals in isolation. In general the use of these refreshments is geographically complementary, the former being characteristic of insular Southeast Asia and the western Pacific, and the latter of Remote Oceania. Some early twentieth century ethnologists, as Friederici (1912-1913) and Rivers (1914), even spoke of betel and *kava* 'cultures' as historical strata in the settlement of the Pacific, although the two distributions overlap in some areas, as in the Admiralty Islands, where betel is widely used, but *kava* is also part of the traditional culture on the islands of Baluan and Lou.

The most widespread domesticated animals are the dog, pig, and fowl, all of which are eaten, the latter two more commonly than the former. Reflexes of *maŋ-asu 'to hunt using dogs' (< *asu 'dog'), are found in a number of languages in the Philippines and Indonesia, thus attesting to the traditional value of dogs as companions of the hunt. Over much of island Southeast Asia the word for dog reflects *asu, while in the Pacific it is extremely variable. This striking difference in lexical variability almost certainly is due to the impoverishment of land fauna on Pacific Islands: where hunting decreased in importance so did the economic importance of the dog, and in times of scarcity it became a competitor for food—or food itself. Under such circumstances dogs disappeared from many islands, only to be later reintroduced with new names. In the Philippines and Indonesia the carabao is an important work animal, used especially to plough rice paddies. Aging animals are slaughtered and eaten on important ritual occasions. Goats and horses are kept in some parts of Indonesia and the Philippines, and cattle are of great importance in Madagascar, where (following an African pattern) they are a measure of wealth.

Typical manufactures include pottery, made almost everywhere that suitable clays are available, the outrigger canoe and its associated paraphernalia, nets and traps of various kinds for fishing and capturing small game, the bow and arrow, the blowgun (common in insular Southeast Asia, but rare in the Pacific), bark cloth (most typical of the central and eastern Pacific, but also found in Taiwan and Indonesia), the back loom and woven fabrics made with it (widespread in insular Southeast Asia, scattered in the Pacific), such musical instruments as the bamboo nose flute and hollowed log slit-gong, and various household implements. In Taiwan, Indonesia, the Philippines, and Madagascar, metallurgy was traditionally important (Chen 1968, Marschall 1968). In at least the first three areas this included the smelting of iron ore in a charcoal furnace by means of a vertical wooden or bamboo bellows operated with a piston. Bronze casting by the lost wax method occurs in Indonesia and the Philippines, as does the working of silver, gold, tin, and tumbaga (an introduced copper-gold alloy).

Trade is the major form of relationship between most AN-speaking communities. In Indonesia hunting-gathering groups and their sedentary neighbors are said to have engaged in 'silent trade', the former leaving jungle produce and the latter salt and manufactured goods in a predetermined location. Far more sociable are the great trade networks of Melanesia, in which communities are linked through individual trade partnerships that may pass down from father to son. One of these, the *kula ring*, which encompassed the Trobriand, Amphlett, and other islands in the Massim area southeast of New Guinea, is

especially well-known through the work of the British anthropologist Bronislaw Malinowski. In this trade network two types of goods—long necklaces of red shell and bracelets of white shell—circulate in opposite directions through many hands over a roughly circular area more than 170 kilometres in diameter. Both Malinowski and his French contemporary Marcel Mauss stressed that in such systems the material dimension of trade is subsidiary to social, political, and magico-religious considerations. Another important traditional trading partnership linked Motu speakers in the region of modern Port Moresby with both AN-speaking and Papuan-speaking peoples around the Gulf of Papua. The trading voyages of the Motu in their large sailing canoes were called 'hiri', and the simplified form of Motu that was used as a medium of communication in the commercial transactions during these trading voyages came to be called 'Hiri Motu.' Unlike these systems of exchange, which were basically egalitarian, tributary systems based in part on inequality were found in both Micronesia and Polynesia. One of these linked Tahiti with certain of the Tuamotu atolls in southeastern Polynesia. Perhaps the most spectacular of these tributary systems, sometimes called the 'Yapese empire', linked Yap with other communities in the western Caroline Islands of Micronesia as far east as Nómwonweité (Namonuito) atoll, some 900 kilometres distant, and was driven by a belief that the Yapese could magically control the destructive typhoons that periodically sweep across this region.

Over much of eastern Indonesia and in parts of Sumatra at least some types of trade are closely linked with systems of kinship and marriage. Members of the Leiden School of ethnology in the 1930s noted that systems of kinship and marriage in eastern Indonesia are characterised by the widespread (but not universal) occurrence of three general features: 1) unilineal descent groups (corporate kin groups defined with reference to an 'apical' ancestor), 2) preferential matrilateral cousin marriage, and 3) 'circulating connubium', better known in the more recent literature as asymmetric exchange. Since it involves the circulation and counter-circulation of certain narrowly specified categories of symbolically 'male' and 'female' goods (the latter including wives), asymmetric exchange can be seen not only as a system for the regulation of marriage, but also as a system of political alliance with intriguing general resemblances to the non-marriage based trade networks of Melanesia. In the Philippines and western Indonesia apart from Sumatra descent groups are absent. Although descent groups are present in the great majority of Formosan and Pacific Island societies, matrilateral cousin marriage is far less common among them than in eastern Indonesia, and the kinds of alliance systems built upon it are rare, but may have once been more common (Hage and Harary 1996).

Many AN-speaking societies are characterised by marked social stratification. Many of the ethnic groups of central and western Borneo recognise hereditary classes of nobles, commoners, and slaves. Similar tripartite systems of hereditary rank are reported from Nias, Sulawesi, various groups on Sumba, Sawu, Flores, Roti and Timor in the Lesser Sundas, Kei and other parts of the central Moluccas, and in Yap and among the early contact Chamorros in Micronesia. Slaves usually were war captives, but could be debtors or serious violators of customary law in their natal communities. Perhaps the most striking manifestation of social stratification in the AN world occurs in Polynesia, where a high chief traditionally was seen as so charged with sacred power (*mana*) that contact with him or anything that he touched could jeopardise the life of a commoner. Micronesian paramount chiefs had great authoritiy, and often ruled over an extensive tributary domain, but do not appear to have been invested with a sacral character. Most Melanesian societies, by contrast, are characterised by a 'bigman' system of achieved status based on acquired wealth, although hereditary rank is found sporadically, as among the Mekeo of New

Guinea, and in various groups in the southeast Solomons, Vanuatu, New Caledonia and the Loyalties. Given the absence of strong centralised power, polities in Melanesia are typically smaller and more fragmentary than those in Micronesia or Polynesia, with the result that language communities are often much smaller as well.

Traditional religious ideas in the Austronesian world center about the placation of spirits. In non-Muslim and non-Christian communities in the Philippines and Indonesia disease is diagnosed and treated by a shaman who often is female, or a transvestite male. Scattered widely in island Southeast Asia is (or was) a belief in minor souls located in each of the shoulder joints, and a major soul located in the head. Souls were regarded as capable of taking flight, thereby causing faintness or even death to their possessor. The rice plant was regarded as having a soul (Malay: *səmangat padi*), the loss of which could prevent germination. Among Malays, Javanese, and some other western Indonesian peoples, rice intended for consumption could be harvested with a sickle, but seed rice was harvested with a small blade concealed in the palm of the hand so as to avoid startling and possibly causing the flight of the rice-soul.

Headhunting traditionally was important over much of island Southeast Asia. Major headhunting expeditions were often correlated with the agricultural cycle, and in some areas with an annual death feast. Independent ethnographic accounts state that this practice served not only to secure trophies in war, but also (and in the native mind perhaps more importantly) to renew the collective vitality of the human and agricultural community through the capture and ritual incorporation of extraneous soul-force in compensation for soul-force lost from the community during the previous growing cycle.

Finally, although the AN world can be defined by a sharp linguistic boundary in the west, physical and cultural boundaries seem more blurred. With the exception of the strongly sinicised Vietnamese, the physical type of mainland Southeast Asians covers much the same range as that of insular Southeast Asians of corresponding latitudes. Many of the cultural traits that have been described for island Southeast Asia occur as well among speakers of Austroasiatic, Tai-Kadai or Sino-Tibetan languages. In some cases cultural agreements are striking, as with the use of a distinctive headhunting tattoo among certain Naga (Sino-Tibetan) peoples of central Assam and the AN-speaking Atayal of northern Taiwan, Dusun of Sabah, and Mentawai of the Barrier Islands west of Sumatra, or the 'clapping bamboo' dance among the Hlai (Tai-Kadai) of Hainan Island in southern China, the Karen (Sino-Tibetan) of peninsular Burma, and AN-speaking groups ranging from at least the Philippines to eastern Indonesia. Physical and cultural similarities between the Karen, Nagas, and other Sino-Tibetan speaking peoples of the Asian mainland, and the Kayan and similar AN-speaking groups of central Borneo struck some early observers (as Hose and McDougall 1912:2:241) with such force that they imagined connections by migration, despite the absence of corroboratory linguistic evidence. While migration is now all but universally rejected as an explanation of cultural resemblances between mainland and insular Southeast Asian peoples unless supported by evidence of linguistic relationship (e.g. Chamic), the question of whether such resemblances are due to diffusion or to an ancient community of origin is a serious one to which we shall return in discussing the external relationships of the AN languages.

1.6   External contacts

Important external cultural and linguistic influences began to affect AN-speaking peoples about 2,000 years ago in insular Southeast Asia. These can be distinguished in

their historical order of appearance as 1. Indian, 2. Chinese, 3. Islamic, and 4. European (primarily Portuguese, Spanish, Dutch, and English). External influences on the AN-speaking peoples of the Pacific have been both shorter in duration and more fragmentary in their distribution than those of island Southeast Asia.

The most important early external contacts with AN-speaking peoples came from India. Around 2,000 years ago Hindu notions of divinity, kingship and the state, as well as Indic scripts, began to penetrate mainland Southeast Asia and western Indonesia. Hindu-Buddhist states arose in Sumatra and Java, producing numerous architectural structures, of which the most famous surviving examples are the stupa of Borobudur and the temple of Prambanan in central Java. Syllabaries derived from the ancient Brahmi script (via the southern Pallava script) form the basis of various indigenous traditions of syllabic writing on palm leaves or bamboo in Indonesia and the Philippines. Somewhat unexpectedly, these traditions appear in some groups, such as the Batak of Sumatra, and the Hanunóo of Mindoro Island in the Philippines, that have been relatively isolated in historic times.

By the late seventh century Hindu-Buddhist states based on Indian notions of kingship and world order had arisen in southern Sumatra. The most powerful of these was Śrīvijaya, which probably was Malay-speaking, an inference supported by a group of five short commemorative inscriptions on stone from southern Sumatra and the adjacent Island of Bangka, which are written in what is generally described as 'Old Malay', heavily interlaced with Sanskrit (Mahdi 2005). Three of these inscriptions bear dates ranging from 683 to 686 AD. Moreover, the Chinese Buddhist pilgrim I-ching, who journeyed from China to India in 671, stopped enroute for six months to study Sanskrit grammar at the port of (Shih-li) fo-shih, described as some twenty days passage from Canton. On his return voyage after ten years in India he spent four more years in the same location, copying Buddhist texts from Sanskrit into Chinese. Coedès (1971) identifies Shih-li-fo-shih with Śrīvijaya, and quotes I-ching as saying that in 671 the same region was called Mo-lo-yu, a name that can be identified with Melayu, the name of a historical state in southern Sumatra, and the ethnolinguistic self-designation for speakers of Malay.

It is probable that Indian cultural and linguistic influence extended beyond southern Sumatra during or even prior to the Śrīvijaya period, but the available evidence is fragmentary. Dahl (1951:368) noted that a stone inscription in Sanskrit from Muara Kaman in east Borneo records an apparently abortive attempt to establish an Indianised state in that area around 400 AD. In subsequent centuries the center of Hindu-Buddhist state formation in Indonesia shifted from southern Sumatra to Java, where it reached its apogee in the kingdom of Majapahit (1293-early sixteenth century). In time Majapahit was submerged under the rising tide of prosperous Muslim port cities that gradually increased in power after the beginning of the sixteenth century, and Indian religious, cultural, and linguistic influence ceased to exist in Java, although much of the earlier tradition was transported from eastern portions of Java to Bali, where it survives today.

The Indian period in western Indonesia lasted nearly a millennium, and left a major linguistic legacy. Although there is no evidence that the Philippines was directly exposed to Indian cultural or linguistic influence, Sanskrit loanwords are also found in a number of Philippine languages. The most likely explanation for this situation is that Malay speakers disseminated both native vocabulary and nativised Sanskrit words during the centuries immediately preceding western contact. Perhaps the strongest type of surviving cultural evidence that might be cited in support of direct Indian contact with the Philippines is the presence of nativised syllabaries of Indian origin in sixteenth century Tagalog and some relatively isolated modern ethnolinguistic groups both in Indonesia and the Philippines, but

not among Malays. However, even this distribution is most plausibly explained as a product of contact with Malays, since 1. Malay loanwords in Philippine languages indicate significant Malay contact with the Philippines in any case, and 2. the Indic script used in the Old Malay inscriptions of Śrīvijaya could have been widely disseminated by pre-Islamic Malay traders before it was replaced among Islamic Malays, who adopted the Arabic script, and Christianised lowland Filipinos, who adopted the Roman alphabet.

To show the extent of lexical borrowing from early Indian sources, about half of the more than 25,000 base entries in the Old Javanese dictionary of Zoetmulder (1982) are of Sanskrit origin. While this is an impressive record of contact, it must be kept in mind that the language of the Old Javanese texts was that of the courts, and hence reflects the linguistic world of the educated elite, not the peasantry. Moreover, despite a wealth of Sanskrit loanwords relating to religion, government, trade, and such material objects as pearls, silk, gemstones, glass and beads, the basic vocabulary of Old Javanese was almost untouched, the 200-item Swadesh basic vocabulary for Old Javanese having only two known Sanskrit loans: *gəni* (Skt. *agni*) 'fire' and *megha* (Skt. *megha*) 'cloud.'

Chinese contact with the Philippines began during the Northern Sung dynasty (960-1126), although sustained trade relations came later. Unlike Indian contact, which introduced writing, architectural styles, notions of the state and religious ideas, or Arabic contact, which introduced various religious and legal ideas, Chinese contact with Island Southeast Asia was largely commercial. Although the seventh-century Buddhist pilgrim I-ching was from Canton, and thus presumably spoke an earlier form of Cantonese, most Chinese settlers in Southeast Asia within historical times have been speakers of Southern Min (Minnan), or Hakka. Schurz (1959) notes that when the Spanish colonisers of the Philippines initiated the galleon trade between Acapulco and Manila in 1565 they found Chinese junks in Manila Bay already engaged in lively trade with the local population. Sung dynasty records suggest that trading contacts with parts of the Philippines had begun by the beginning of the eleventh century, and it evidently involved merchants sailing both from Canton and the Fujian coast. Probable Chinese loanwords that are widespread in insular Southeast Asia include *waŋkaŋ* 'Chinese junk', *uaŋ* 'money', *hupaw* 'money-belt', and *hunsuy* 'smoking pipe'. It is unlikely that most of these words spread into Island Southeast Asia earlier than the Ming dynasty (1368-1644).

Hall (1985:213) reports that the north coast of Sumatra was visited by Arab traders from "at least the tenth century A.D.," and Chinese records from the late thirteenth century indicate that Islam had begun to take hold in Jambi, southern Sumatra by that time. However, Islamic influence evidently was not yet strong or uniform, since Marco Polo visited the port of Samudra in northern Sumatra in 1292, and noted that its population had not yet converted to Islam. This situation soon changed, as Islam became firmly established in Sumatra by the beginning of the fourteenth century, and was exported from there by Malay-speaking missionaries. Islamic sultanates were founded at Brunei in northwest Borneo, at Ternate and Tidore in the northern Moluccas, and in the Sulu Archipelago in the southern Philippines. The Islamic penetration of the Philippines was accomplished by Malay missionary-traders from the sultanate of Brunei, and led to the introduction of numerous Malay, Sanskrit, and Arabic loanwords not only into the languages of the southern Philippines (where Islam survives today), but also into those of the central, and to some extent, the northern Philippines.

The Arabic vocabulary in Malay/Indonesian has been described by Jones (1978), who cites over 4,500 loanwords. Many of these are concentrated in the areas of religion and law, but they also include names of the days of the week, astronomical bodies, and the like.

Many of the same loanwords are found in Acehnese, Sundanese, Javanese, and other languages spoken by strongly Islamised ethnic groups in western Indonesia, as well as by the Muslim populations of the southern Philippines. As with the Sanskrit loans in the Philippines, Arabic loanwords are found in a number of Philippine languages spoken by populations that are not known to have ever been Islamic. Again, the diffusion of this vocabulary appears to have been mediated by Malay traders. The distinctive appearance of a final glottal stop for expected zero in words that originally ended with a vowel, points to Brunei Malay as the source for many of these forms. No AN language east of the Moluccas shows clear evidence of early loans from Sanskrit, Chinese, or Arabic.

Although European contact with languages of the AN family dates from at least 1292, when Marco Polo stopped in northern Sumatra on his return from China, linguistic materials were not collected until early in the sixteen century. In 1521, Antonio Pigafetta, the Italian chronicler of the Magellan expedition, recorded a vocabulary of about 160 words for the language of Cebu Island, in a region of the central Philippines where Magellan was shortly to meet his untimely death. Toward the end of the same year around 425 words were recorded for a Malay dialect from an unstated location, although given the known route of the expedition this probably was the northern Moluccas. The vast majority of AN languages were still undiscovered and their interrelationships unrecognised, yet Pigafetta's vocabularies mark the beginning of western scholarly interest in what we now know was the most widespread language family on Earth prior to the great European colonial expansions of the period 1500-1800 (Cachey 2007).

The Magellan expedition was the forerunner of a far heavier traffic on the world's seas in the century to follow. During this period (1600-1700) the Dutch wrested control of the lucrative spice trade from the Portuguese in the Moluccas, and established themselves for more than three centuries as colonial masters of the island world later to be known as 'Indonesia.' To the north of the Philippine Archipelago, which the Magellan expedition had claimed in the name of King Philip of Spain, Holland secured a second, smaller foothold in the AN world on the island of Formosa (Taiwan).[2]

As in Indonesia, the Dutch presence in Formosa (1624-1662) grew out of mercantile motives, and was initially limited to trade between the Dutch East Indies Company and local producers or distributors. In both areas there was a subsidiary interest in religious conversion, but this interest tended to play a larger role in Formosa than in Indonesia. There were no doubt several reasons for this difference of focus, but one seems especially important. When European contact began, Malay was widely used as a *lingua franca* in coastal areas of island Southeast Asia, and it was through this language that the Dutch conducted trade with the local populations. By the very fact of their acquaintance with Malay, however, these coastal populations had been exposed to other foreign influences. Christianizing efforts were made in Indonesia during the seventeenth century, but these appear to have been largely thwarted by the presence of Islam in most of the more accessible areas. Because no such obstacle existed in Formosa, the study of the local languages began at once in preparation for the translation of the scriptures. As a result, although practical vocabularies of Malay and Javanese (de Houtman 1603) and even a short Malay-Dutch dictionary (Wiltens-Danckaerts 1623) were published during this period, the major Dutch publications of linguistic interest in insular Southeast Asia during

---

[2]   Since Taiwanese is a dialect of Southern Min (one of the languages commonly included under the cover term 'Chinese'), it is customary to use 'Formosan' to refer to the aboriginal languages of Taiwan. I follow this practice, and use 'Formosa' as a geographical designation for the pre-modern period, but 'Taiwan' when referring to the island as a modern political entity.

the seventeenth century did not concern the languages of Indonesia, but rather the AN languages of Formosa (Happart 1650, Gravius 1661).

Dutch missionary activity, and with it the translation of the gospels into the local languages, was cut short by the expulsion of the Dutch from Formosa in 1662. Since penetration of the hinterland and consequent contact with ethnic groups that had not yet been reached by Islam was delayed for several centuries in Indonesia, few publications appeared on Indonesian languages other than Malay until the 1850s. A similar situation obtained in Malaysia under British rule, but the Spanish friars of the Philippine missions produced dictionaries, grammars and doctrinal materials in Tagalog, Bikol, the major Bisayan dialects, and some other local languages from the second half of the sixteenth century onward, and an important grammar of Chamorro, the native language of Guam and the other Mariana Islands, appeared very early (Sanvitores 1668).

Language barriers, together with national and ecclesiastical rivalries, did not favor the ready diffusion of linguistic information between the major European colonial powers. Opportunities for comparing Malay or other politically important languages of Indonesia with Tagalog or other politically important languages of the Philippines were thus limited during the colonial period. While the Spanish supply route to the Philippines crossed the open Pacific from Mexico with a single stop in Guam, Dutch voyages to Indonesia included a distant way-station after the sometimes difficult passage around the Cape of Good Hope: the great island of Madagascar off the coast of east Africa. It was thus perhaps inevitable that the relatively transparent relationship of Malagasy to Malay would be recognised by Dutch sailors who had been exposed to both languages. When this recognition came, early in the seventeenth century, the existence of a language family spanning at least the rim of the Indian Ocean was established.

Based on imperfect and mislabeled vocabularies collected by the Dutch voyager Jacob Le Maire in western Polynesia during the previous century (Engelbrecht and van Hervarden 1945:133-138, R.A. Kern 1948), Hadrian Reland (1708) further indicated the likelihood of an eastern extension of Malay-related languages to at least western Polynesia. The true geographical extent of this still unnamed language family was, however, only suspected. Surprisingly, the relationship of the Formosan languages to Malay—in some cases no less evident than that of Malagasy to Malay—apparently was not recognised, at least in print, until the nineteenth century. But a vast region between Southeast Asia and the Americas remained virtually unknown to Europe. In 1768 the Englishman James Cook began the first of three voyages of exploration in the Pacific Ocean. During the second of these voyages (1772-1775) vocabularies were collected on a number of Polynesian islands, on the large southerly Melanesian island of New Caledonia, and at several locations in the New Hebrides chain (modern Vanuatu).

In a book published in 1778 a Swiss member of this expedition, Johann Reinhold Förster, expressed what would become a persistent confusion of language and physical type in accounts of the linguistic relationships of Pacific peoples. He noted that the many widely scattered islands of the eastern Pacific ('Polynesia') were inhabited by a tall, well-built, relatively fair-skinned people of similar language, while the larger, often malarial islands of the western Pacific ('Melanesia') were home to a shorter, darker, frizzy-haired population speaking a babel of tongues. Förster observed that the 'Polynesian language' resembled Malay, but the languages of Melanesia, in accord with the differing physical type of their speakers, were not related to these, or even to one another.

Even though many details needed to be added and others corrected, sufficient data now existed to determine the east-west scope of the language family that included Malagasy,

Malay, and the Polynesian languages. With Förster's publication the territorial extent of the AN language family appears to have become general knowledge among Europe's intelligentsia: both the English scholar William Marsden (1783) and his Spanish contemporary Lorenzo Hervas y Panduro (1784) noted that Malay-related or Polynesian-related languages extended from Madagascar in the west to Easter Island in the east, an extraordinary 206 degrees of longitude. As further data were collected this general conclusion was confirmed many times over. Despite such progress, a belief that classifications based on linguistic criteria must correspond to those based on racial criteria continued to dominate thinking regarding the languages of Melanesia.

In 1834 William Marsden reasserted the existence of a far-flung language family that includes Malagasy, the languages of the Malay Archipelago, and those of the eastern Pacific. He called the former 'Hither Polynesian' and the latter 'Further Polynesian', conceiving of the two as divergent expressions of a single 'general language'—an oblique reference to what today would be called a proto language. Marsden's use of the name of a limited geographical region to designate a language family that extends well beyond it was inappropriate, and did not find general acceptance. Shortly thereafter the influential German scholar Wilhelm von Humboldt (1836-39) used the term 'Malayisch' to designate the same collection of languages. Again, the terminology was inappropriate, and did not take hold. At about the same time the German Indo-Europeanist Franz Bopp (1841), became convinced that Malay, Javanese and the Polynesian languages are related to Indo-European, and for convenience of reference he proposed that they be called by a compound term formed from the name of a western and an eastern member. Since Malay, now represented by the grammar and dictionary of Marsden (1812), was the best-known western language, the family was christened, somewhat belatedly, 'malayisch-polynesisch' (Ross 1996a).

During the second half of the nineteenth century scientific work on the Malayo-Polynesian languages began in earnest. The details of this work will be described later. For now it is enough to note that the name 'Malayo-Polynesian' became established by general usage. This term had the advantage of making the relationship of Malay to the Polynesian languages transparent, but it also tended to perpetuate the illusion that the languages of Melanesia belong elsewhere. Although he spoke of 'Malayo-Polynesian' languages von der Gabelentz (1861-73) concluded that the grammatical similarities of Melanesian and Polynesian languages are too numerous and basic to be due to borrowing. Codrington (1885), who like von der Gabelentz opposed the view that the Melanesian languages are unrelated to Malay and Polynesian, avoided the term 'Malayo-Polynesian' altogether, referring instead to the 'Ocean' family of languages.

It was not until the twentieth century that a name was found for the Malayo-Polynesian language family which avoided an implicit appeal to race. In 1906 the Austrian linguist and ethnologist Wilhelm Schmidt showed that the Mon-Khmer languages of mainland Southeast Asia are related to the Munda languages of India. He called this language family 'Austroasiatic' ('southern Asiatic'). At the same time he pointed to resemblances between the Austroasiatic and Malayo-Polynesian languages, and suggested that the two families form coordinate branches of a larger superfamily which he called 'Austric.' In keeping with the term 'Austroasiatic' and the established names 'Indonesia', 'Melanesia', 'Micronesia', and 'Polynesia', Schmidt renamed the Malayo-Polynesian family 'Austronesian' ('southern islands'). Although Schmidt's Austric hypothesis was not generally accepted, his terminological innovation was taken up by Jonker (1914), Blagden (1916), and more significantly by Otto Dempwolff, both in his major early papers (1920,

1924-25), and in his three-volume *Vergleichende Lautlehre des austronesischen Wortschatzes* (1934-1938), a work which laid the foundations for the modern comparative study of the AN languages. Some writers, as Stresemann (1927), and Dyen (1947a, 1951, 1953b, 1962) nonetheless continued to favor the older terminology. As a result the names 'Malayo-Polynesian' and 'Austronesian' were used in an equivalent sense from the early part of the twentieth century until quite recently.

The American linguist Isidore Dyen published a genetic classification of the AN languages (1965a) in which he suggested that 'Austronesian' be used for the entire language family, and 'Malayo-Polynesian' for a lexicostatistically-defined subset of it. From this point on the names 'Malayo-Polynesian' and 'Austronesian' have for many scholars ceased to be synonymous but, as will be seen, Dyen's definition of 'Malayo-Polynesian' never achieved wide currency. In the mid-1970s Mills (1975:2:581) and Blust (1977a) independently proposed that the term 'Malayo-Polynesian' be used for all non-Formosan AN languages, and this usage has since been generally adopted by other scholars in the field.

Before leaving the subject of terminology one other matter should be mentioned. A terminology with misleading implications was used by the Swiss linguist Renward Brandstetter (1916) who, despite making important contributions to comparative AN linguistics, excluded the languages of the Pacific. Long after the relationship of Malay to the languages of Polynesia had been clearly established he was thus able to speak of 'Common Indonesian' and 'Original Indonesian' as though the term 'Indonesian' designates a language family, or even a linguistically justified subgroup. As a step in the reconstruction of Proto Austronesian phonology Dempwolff (1934-1938) posited a 'Proto Indonesian' (PIN) sound system, but then explicitly acknowledged that his PIN could account for all historical developments in the languages of Melanesia, Micronesia, and Polynesia, and so was equivalent to Proto Austronesian. Similarly, the English linguist Sidney H. Ray (1926) and his Australian protégé Arthur Capell (1943) avoided the term 'Austronesian' on the (generally unshared) assumption that the AN languages of Melanesia descend from a prehistoric pidgin spoken by trader-colonists from island Southeast Asia. Although both writers refer to the 'Indonesian' origin of widespread vocabulary in the languages of Melanesia, they are silent on the relationship of the languages of Polynesia and Micronesia to those of Indonesia.

Finally, Dutch writers after Kern have sometimes spoken of 'Indonesian' languages not out of explicit opposition to the arguments offered for an AN language family, but rather out of the use of an accident of colonial history (Dutch control of Indonesia) to define a field of scholarly endeavor. This point is worth emphasizing for two reasons. First, as will be seen, there is no linguistic basis for recognizing an 'Indonesian' branch of the AN language family, since the languages of western Indonesia appear to be more closely related to those of the Philippines, Malaysia, Madagascar and mainland Southeast Asia, than they are to the languages of eastern Indonesia. Second, the tendency among Dutch scholars to isolate the languages of Indonesia as a self-enclosed field of study has shown signs of increase during the second half of the twentieth century. Thus, although Adriani (1893) spoke of 'Malayo-Polynesian' languages, and Esser (1938) referred to groupings of 'Malayo-Polynesian' languages in Indonesia, more recent writers, as Gonda (1947), and Teeuw (1965) refer to 'Indonesian languages', or even to an 'Indonesian family of languages'. This isolating tendency was resisted by Anceaux (1965), and Uhlenbeck (1971:59), and younger Dutch scholars who have worked abroad have in general come to

view the AN languages of Indonesia as a politically-defined subset of the AN language family rather than a natural unit of comparison.

1.7   Prehistory

Even in the absence of other lines of evidence, the wide distribution of grain agriculture, the cultivation of tubers, animal domestication, pottery manufacture, loom weaving, house construction, the outrigger canoe complex, and the like strongly suggest that the common ancestor of the AN-speaking peoples already possessed a culture of 'Neolithic' type, an impression that is supported by both archaeological and lexical data.

Although human history in insular Southeast Asia can be traced back over one million years, only the last few millennia are relevant to the AN diaspora. Very ancient remains of human ancestors have been found in Java. These include the celebrated fragments of Java Man (*Homo erectus*), discovered by Eugene Dubois in the bed of the Trinil River in 1891 and 1892, and initially dated to the Middle Pleistocene between 130,000 and 700,000 years ago. With improved dating techniques the age of these and of similar remains found subsequently in comparable geological contexts has now been recalculated to at least 1.2 million years BP. An apparently more advanced *Homo erectus* population which may have practiced cannibalism some 100,000 years ago, is represented by the Ngandong remains found in deposits of the Solo River. Later human remains from Southeast Asia include Upper Pleistocene cranial fragments found by Dubois at Wadjak, central Java, in 1890 (*Homo wadjakensis*), and one of the earliest skulls yet recovered of modern humans (*Homo sapiens sapiens*), from Niah Cave, Sarawak, dated at roughly 40,000 BP.

It has been suggested that Australia was initially colonised by a segment of the *Homo wadjakensis* population which crossed the sea barrier from the southern end of the now submerged Sunda shelf. Unambiguous signs of human presence in northern Australia are now placed minimally at 50,000 years BP, with some proposals suggesting far earlier (but not yet universally accepted) dates. In the Pacific pre-Neolithic populations have been dated to at least 50,000 BP on the north coast of New Guinea, and to earlier than 30,000 BP in some of the islands of the Bismarck Archipelago. There is little doubt that these remains represent an ancestral Papuan population. For at least New Guinea this population would not have been physically separated from that of Australia until the end of the last glaciation, when the Torres strait which lies between the two landmasses was flooded by rising sea levels.

Astonishingly, in the Fall of 2003 archaeological excavations on the island of Flores in the Lesser Sunda chain uncovered evidence of an entirely new human species, christened *Homo floresiensis*. Fossil evidence suggests that this dwarfed cousin of modern humans survived until at least 13,000 years ago, and local traditions of little people have fueled speculation that it may have coexisted with AN-speaking peoples who reached the island within the past 4,000 years. However, given the wider context of 'little people' stories in the AN world the use of oral tradition as evidence for the recency of *Homo floresiensis* survival must be treated with caution.

Paleolithic remains with dates as early as 47,000 BP are also known from the Tabon caves of the central Philippines, Niah Cave in northern Sarawak, the Changpin caves on the east coast of Taiwan, and the Leang Burung rockshelter on the island of Sulawesi. Some of these are found in areas (Taiwan, Borneo) which form part of continental shelves that were exposed as dry land during glacial maxima, and these pre-Neolithic populations could very well have reached their attested locations on foot. As noted already, the living

descendants of these prehistoric hunter-gatherers in insular Southeast Asia almost certainly are the widely-scattered populations of Negrito foragers. Australia-New Guinea and probably Sulawesi, on the other hand, could only have been reached by the use of very early watercraft, perhaps bamboo sailing rafts. By the time AN-speakers began to arrive in island Southeast Asia the physical environment was very different. Following the glacial retreat some 10,000 years ago sea levels rose, flooding many low-lying areas on the continental shelves, and so leaving the more elevated regions as a new world of islands.

The earliest Neolithic culture identified to date in island Southeast Asia is a cord-marked pottery tradition called 'Tapenkeng' by its discoverer, K.C. Chang. Tapenkeng pottery in association with quadrangular stone adzes, polished slate points, and stone net sinkers is widespread on the western plain of Taiwan, and marks the initial settlement of the island by sedentary Neolithic populations (Chang 1969). Although originally dated as early as 6,300 BP, Chang's chronology is now questioned by many prehistorians, and a consensus is emerging that the earliest firmly dated Neolithic sites in Taiwan cluster around 5,500 BP (Tsang 2005). Direct physical evidence of grain crops in Taiwan cannot yet be dated to the earliest levels, but it is clear from the linguistic evidence that rice and millet were cultivated at the time the AN language family began to divide into primary dialect regions. In Chang's interpretation, by about 4,500 BP the Tapenkeng culture had produced two descendants, the Lungshanoid culture in western and southern Taiwan, and the Yüanshan culture in northern and eastern Taiwan. The former shows similarities with contemporaneous archaeological cultures on the Chinese mainland, and the latter with Neolithic cultures in the Philippines and Indonesia. More recent archaeological evidence has documented an AN settlement of the northern Philippines between 4,000 and 4,500 BP, with somewhat later dates from most parts of Indonesia.

As in Taiwan and the Philippines, the arrival of Neolithic cultures in the western Pacific appears to have been abrupt. Because of its durability and great potential for stylistic and material variation, pottery is a key cultural marker in most archaeological assemblages. By far the most noteworthy pottery type in the Pacific, or the AN world as a whole, is Lapita ware, named from a type site first excavated in New Caledonia in 1952 (Gifford and Shutler 1956). Lapita pottery is not a drab utilitarian product, but an elaborately decorated ware that probably had important functions as an article of trade. Pacific archaeologists have been so captivated by the appeal of this ceramic tradition that they sometimes speak of a 'Lapita culture', a 'Lapita homeland', or even the 'Lapita peoples' (Kirch 1997). Lapita sites are characterised by a preference for coastal settlement, or settlement on small islands lying offshore from often much larger landmasses. The economy was based on fishing and horticulture, and included such cultigens as the yam, several types of taro, sugarcane, banana, breadfruit, and coconut, but no grain crops.

The earliest Neolithic site associated with Lapita pottery is the remains of a pile village near the island of Mussau in the St. Matthias Archipelago, some 160 kilometres northwest of New Ireland, and dated to about 3,500 BP. Within a few centuries cultures with very similar pottery had appeared in Fiji and western Polynesia. The rapid spread of Lapita pottery through Island Melanesia and into western Polynesia indicates a highly mobile population capable of open sea navigation, that probably was engaged in long-distance trade of both manufactured and natural products. Among the latter, obsidian from either of two traceable sources, Lou Island in the Admiralty group, and the Talasea Peninsula of New Britain, has been found in archaeological sites as far east as the southeast Solomons, and as far west as Sabah in northern Borneo (Bellwood 1997:224).

It is now clear that Papuan speakers preceded AN speakers in New Guinea and Island Melanesia by tens of millennia. Radiocarbon dates from Matenkupkum in New Ireland, and Kilu on the island of Buka in the western Solomons, show that stone age peoples managed to reach these islands with some type of watercraft more than 30,000 years ago. In addition, the cuscus was introduced to these islands from mainland New Guinea by human intervention around 9,000-10,000 BP, and the wallaby by about 7,000 BP, showing that there was continuing contact between New Guinea and the Bismarcks by means of some type of watercraft (Spriggs 1993). A similar settlement history almost certainly applies to islands in the western Solomons, which during the Pleistocene were part of the single united landmass of 'Greater Bougainville'.

Pawley and Green (1973) proposed the term 'Near Oceania' for the Pacific Islands from New Guinea through the Solomons, and 'Remote Oceania' for those that are further removed from insular Southeast Asia. To a large extent this distinction correlates with that part of the Pacific in which sailing involves intervisible islands as opposed to that part in which sailing requires at least an overnight voyage, and hence a more critical dependence on a navigational knowledge of the stars, winds, and tides. The Solomons chain thus appears to mark a critical boundary in the settlement of the Pacific. Although it was long thought that two Papuan languages reached the remote Santa Cruz Islands some 350 kilometres southeast of the Solomons, Ross and Næss (2007) have argued convincingly that these are highly aberrant AN languages that may form a primary branch of the Oceanic group. No Papuan languages are found further south or east, although some of the languages of southern Vanuatu, New Caledonia, and especially the Loyalty Islands are phonologically and lexically very divergent. Moreover, despite the absence of Papuan languages in southern Melanesia or of archaeological evidence of a pre-Lapita population in this region, the physical anthropology, distinctive cultural traits, and such linguistic features as the repeated innovation of non-decimal numeral systems and extensive use of serial verb constructions, are widespread in Vanuatu, New Caledonia and the Loyalty Islands, strongly suggesting a history of contact with Papuan speakers, although the details of how and where this contact occurred are yet to be reconciled with other types of evidence (Blust 2005a, 2008b, Pawley 2006).

The prehistory of New Caledonia may still hold some major surprises. The population is of a general Melanesian physical type, but some individuals -- particularly in the north -- show a striking phenotypic resemblance to aboriginal Australians. On the other hand, unlike most parts of Melanesia, in which a 'big man' system of acquired rank is prevalent, hereditary rank is important in many of the native cultures of New Caledonia and the Loyalties. Extensive prehistoric stoneworks on the Island of Maré in the Loyalties indicate that a centralised chiefly authority capable of summoning corvée labor for public works has existed in this area for some centuries.

Lapita pottery is found in the earliest levels in Tongan sites at about 3,000 BP, but it shows a gradual simplification in decorative motifs, and reduction in types of vessel forms, before disappearing entirely around 2,000 BP (Kirch 1997:68, 159ff). At a much later date pottery was reacquired from neighboring Fiji, where a ceramic tradition was maintained, although decorative styles understandably underwent many changes. Polynesian cultures might be characterised in archaeological terms as 'post-Lapita' traditions, since they derive from a culture that made this distinctive pottery, yet by historical times had evolved into descendants that were completely aceramic.

The archaeology of Micronesia lags behind Fiji and Polynesia, although much progress has been made in recent years. So far there are few radiocarbon dates of over 2,000 years,

despite a surprising cluster of questionable dates from around 3,500 BP in the Marshall Islands. Given the widespread subsidence of coral atolls it is possible that some of the earliest archaeological sites in Micronesia are now below water. As will be seen, Palauan and Chamorro have very different histories from most languages of Micronesia. The archaeology of Palau is still relatively undeveloped, but it is already clear that Marianas prehistory differs radically from that of other Pacific areas dominated by the Lapita pottery tradition. A large suite of radiocarbon dates have shown that the Marianas were settled by the ancestral Chamorros by at least 3,500 BP, an achievement which required an open sea voyage of about 2,200 kilometres—by far the longest successful open-sea voyage attested from this early period. Guam and some of the other Mariana Islands are notable for the large dolmen-like stone formations, called *latte* in Chamorro, which were erected in many places, perhaps as supports for community buildings or temples. In addition, rice was traditionally cultivated by the Chamorros, making the Marianas the only part of the Pacific in which grain crops formed a dietary staple.

   Table 1.1 presents a range of radiocarbon dates associated with Neolithic sites in insular Southeast Asia and the Pacific (Pacific dates from Kirch 2000:89, 94-95). These are chosen to highlight the earliest dated assemblages in each area, so as to clarify the relative chronology of the AN expansion out of Southeast Asia into the Pacific. It should be noted that some of these dates have been recalibrated since 2000, and there is now general agreement that secure dates for the appearance of Neolithic cultures in the Philippines cluster around 4,000 BP or slightly earlier. However, the overall pattern of a west-to-east cline of decreasing time-depths in the radiocarbon record remains unchanged.

Table 1.1 Dating of Neolithic cultures in insular Southeast Asia and the Pacific

Area	Location	Site	Date (BP)
Taiwan	Tainan	Industrial Park	5500
Philippines	northern Luzon	Rabel, Laurente	4800
Indonesia	Sangir Islands	Leang Tuwo Mane'e	4000
Indonesia	south Sulawesi	Ulu Leang 1	4000
Indonesia	Timor	several caves	4000
Melanesia	Mussau	Talepakemalai	3550-2700
Melanesia	Mussau	Etakosarai	3500-3300
Melanesia	Santa Cruz	Nanggu	3200-3100
Melanesia	Vanuatu	Malo Island	3100-3000
Melanesia	New Caledonia	Vatcha	2800
Central Pacific	Fiji	Natunuku	3200-3100
Polynesia	Samoa	Mulifanua	3000
Polynesia	Tonga	Moala's Mound	3000
Polynesia	Hawai'i	Halawa (Moloka'i)	1400

   Although this section describes the prehistory of areas where AN languages are currently spoken, or were historically spoken, it would be incomplete if it omitted areas

where AN languages may have once been spoken, but no longer are. Throughout their reconstructable history AN-speaking peoples have been a territorially expanding population; the settlement of Triangle Polynesia within the past three millennia is only the most recent expression of a much longer history of movement out of Asia. And just as Australoid or Negrito populations probably were widespread in insular Southeast Asia prior to the arrival of the Austronesians, so AN-speaking peoples probably were once found in areas that today are dominated by other groups.

The Pescadores, or Penghu (P'eng-hu) Islands are located in the Taiwan strait some 50 kilometres west of south-central Taiwan, and 150 kilometres from the coast of Fujian province in southern China. Chinese immigration to these islands began during the Sung dynasty (960-1279 AD), perhaps as early as the late eleventh century. Chinese records mention no earlier inhabitants of the Pescadores, but Tsang (1992) has shown that the material culture of these islands from about 4600 BP exhibits striking similarities to the contemporaneous cultures of southwestern Taiwan. These similarities include not only manufactured products such as pottery and artifacts of stone, bone, and shell, but also inferable cultural practices such as ritual tooth evulsion (common within the ethnographic present among many Formosan aboriginal groups).

Given their position between mainland China and Taiwan it is natural to ask whether the Pescadores might have been settled by Neolithic farmers as part of a series of population movements out of coastal southern China. Chang (1986) has argued that Dapenkeng (Ta-p'en-k'eng) is a regional variant of an archaeological culture that was widely distributed on the adjacent coast of southeast China as early as seven millennia ago. If the Pescadores once had an AN-speaking population that has disappeared without a linguistic or cultural trace, the same could be true for southern China.

Bellwood (1997:208-213) has suggested that the founding Neolithic culture of Taiwan can most plausibly be derived from the rice-growing archaeological cultures of the lower Yangzi River, which are well-attested prior to 7,000 BP. At the site of Hemudu on the south shore of Hangzhou bay, waterlogging created an anaerobic environment in which normally perishable materials were remarkably well preserved. The basal levels, radiocarbon dated to between 7,200 and 6,900 BP, contain evidence of pile dwellings with sophisticated mortise and tenon joints, boatbuilding, matting, loom weaving, abundant stores of rice, and domesticated animals including the dog, pig, chicken and carabao. One excavated pile building was seven meters in width and 23 meters in length, suggesting either a communal residential structure or a public building of some type. There are thus clear indications that the lower Yangzi River in the late sixth millennium BC was an area of abundant food resources which could have supported a substantial and probably continuously expanding population.

Much later, during the Han dynasty (206 BC to 220 AD) Chinese expansion out of the Yellow River basin initiated the lengthy historical process of the sinicisation of southern China. A few non-Chinese groups, as the Hmong-Mien (formerly: Miao-Yao) peoples of the Guizhou plateau and adjacent areas, and some Tai-speaking peoples, as the Zhuang of Guangdong, have survived among the engulfing majority, but there can be no doubt that many other non-Chinese minorities were culturally and linguistically absorbed during the centuries of Chinese expansion southward from the Yellow River valley. In addition to historical references in Chinese sources to the 'thousand Yueh', a term for the numerous non-Chinese minorities that once occupied China south of the Yangzi, recent genetic studies have suggested that 'Chinese' is a cover term for two genetically distinct populations, one which groups more closely with the non-Chinese peoples of the northern

steppelands, and another which groups more closely with Southeast Asians (Cavalli-Sforza, Menozzi and Piazza 1994).

These remarks take us outside the historically-defined AN world, but they are important as an indication of how much the distribution of human populations may have changed during the past several millennia. As recently as two thousand years ago a large part of Polynesia, including Hawai'i, Easter Island, New Zealand, and many other islands east of Samoa, still lay beyond the expanding eastern boundaries of the AN world. And, just as certainly, the Pescadores Islands and probably coastal portions of southern China, lay within the contracting western boundaries.

2 *A bird's eye view of the Austronesian language family*

2.0   Introduction

Because it will be necessary to mention many language names that may be unfamiliar to the general reader, an outline survey of the AN language family will be useful before examining data. This survey consists of five main parts: 1. the large-scale structure of the AN language family, 2. language and dialect, 3. national languages and *lingua franca*s, 4. language distribution by geographical region, and 5. overview of language size and descriptive coverage. Since a relatively complete list of AN languages along with information on classification and language size can be found in Lewis (2009), there is no need to duplicate this information here. Rather, the focus will be on the history of research, salient features of language distribution, and typology. In order to highlight both major languages and those that are most endangered, the ten largest and ten smallest languages are given in tables for each geographical region. For smaller regions in which the total number of languages is not much greater than twenty an exhaustive list of language names is provided.

2.1   The large-scale structure of the Austronesian language family

The history of debates concerning AN subgrouping will be reviewed in Chapter 10. Suffice it to say here that the following major divisions of the family are accepted by a number of the leading scholars in the field.

2.1.1 Austronesian

Austronesian divides into at least ten primary subgroups, of which nine are represented only in Taiwan. They are:

1. Atayalic (Taiwan)
2. East Formosan (Taiwan)
3. Puyuma (Taiwan)
4. Paiwan (Taiwan)
5. Rukai (Taiwan)
6. Tsouic (Taiwan)
7. Bunun (Taiwan)
8. Western Plains (Taiwan)
9. Northwest Formosan (Taiwan)
10. Malayo-Polynesian (extra-Formosan)

Atayalic consists of Atayal and Seediq in northern Taiwan, each with a number of distinct dialects; East Formosan includes Ketagalan, with dialects Basay and Trobiawan (extinct), Kavalan, and Amis, spoken along Taiwan's narrow east coast, plus Siraya,

spoken on the southwestern plain until the second half of the nineteenth century; Puyuma is an isolate spoken in coastal areas of southeast Taiwan that has had a longstanding borrowing relationship with neighboring Paiwan; Paiwan is an isolate spoken in coastal areas of southeast Taiwan, that has had a longstanding borrowing relationship with both Puyuma and Rukai; Rukai is a collection of quite divergent dialects or closely related languages spoken in the mountains of south-central Taiwan; Tsouic is a group of three languages (Tsou, Saaroa, Kanakanabu) spoken in the mountains of south-central Taiwan to the northwest of Rukai. This proposed genetic unit, which was first proposed by Ferrell (1969), and defended in greatest detail by Tsuchida (1976), has recently come under attack by Chang (2006) and Ross (2012), both of whom suggest that Tsou does not subgroup with the other two languages; Tsouic may therefore need to be abandoned; Bunun is a single language with three main dialects that occupies a large part of the montane interior of central and south-central Taiwan; Western Plains is a group of some five languages (Taokas, Favorlang/Babuza, Papora, Hoanya, Thao) formerly spoken on the western plain of Taiwan, of which only Thao survives in the region of Sun-Moon Lake, on the western edge of the Central Mountains; Northwest Formosan is a tentative group embracing Saisiyat, Pazeh and Kulon. No convincing evidence has yet been found that would enable us to reduce this collection of languages to a smaller number of primary branches. Despite a general typological similarity shared by most Formosan languages, the level of phylogenetic diversity among the AN languages of Taiwan is thus very high. Translated into Indo-European terms, the geographical distribution of primary AN subgroups is roughly equivalent to finding representatives of every branch of the Indo-European language family within the borders of the Netherlands.

All AN languages outside Taiwan, including Yami, on Botel Tobago Island off the southeast coast of Taiwan, belong to the Malayo-Polynesian (MP) subgroup of AN. Since MP includes all but about 25 members of the language family it is not surprising that Proto Austronesian and Proto Malayo-Polynesian were long confounded, as in the pioneering reconstruction of 'Uraustronesisch' (= PMP) by Dempwolff (1934-1938). Typologically the MP languages cover a wide range. As will be seen in a later chapter, the major clues to perceiving their historical unity are found in shared phonological mergers and in distinctive irregular changes that have affected particular lexical items.

2.1.2 Malayo-Polynesian

Malayo-Polynesian divides into two primary branches, Western Malayo-Polynesian (WMP) and Central-Eastern Malayo-Polynesian (CEMP). There are some 500-600 WMP languages reaching from Yami through the Philippines, the Greater Sunda Islands of Indonesia (including Sulawesi), and mainland Southeast Asia to Madagascar. Also counted among WMP are two languages spoken in western Micronesia: Palauan and Chamorro. It is possible that WMP is not a valid subgroup, but rather consists of those MP languages that do not belong to CEMP. Its chief unifying character is the presence of systems of nasal substitution in active verb forms, as in Malay *pukul* 'hit' (base form) : *mə-mukul* 'to hit' (active verb), or Chamorro *saga* 'stay' (base form) : *ma-ñaga* 'to stay' (active verb). Such systems appear to be fully functional only in languages that have been called 'WMP', but there are fragmentary indications that such a process may have operated in the verbal morphology of PMP or even PAN. If so, nasal substitution is not an innovation characteristic of WMP, but rather the retention of a process that has become fossilised

elsewhere in the AN family, and WMP should perhaps be renamed 'Residual Malayo-Polynesian'.

2.1.3 Western Malayo-Polynesian

Higher-level subgrouping within WMP has so far proved difficult. Among major branches of WMP that are widely accepted are 1. a Philippine group, which includes all languages of the Philippine Archipelago except the Sama-Bajaw (or Samalan) languages spoken by traditionally nomadic 'sea gypsies' of the central and southern Philippines and various parts of Indonesia-Malaysia, 2. a North Sarawak group which includes a number of the languages of northern Sarawak in Malaysian Borneo, 3. a Barito group which includes Ngaju Dayak and Ma'anyan of southeast Kalimantan, as well as Malagasy, 4. a Malayo-Chamic group which includes the Malayic languages of insular Southeast Asia, and the Chamic languages of mainland Southeast Asia, and 5. a Celebic subgroup which includes all languages of Sulawesi south of Gorontalic, except the South Sulawesi group (whose best-known members are Buginese and Makasarese). The North Sarawak group may be one of two coordinate branches of a North Borneo group that also includes the languages of Sabah. Recently it has been suggested that Malayo-Chamic is part of a larger collection of languages that includes Balinese, Sasak, and Madurese, but not Javanese (Adelaar 2005c). Finally, some preliminary evidence has been offered in support of the view that all non-Malayic languages of Sumatra with the possible exception of Enggano, share an immediate common ancestor (Nothofer 1986).

2.1.4 Central-Eastern Malayo-Polynesian

Central-Eastern Malayo-Polynesian includes nearly all AN languages of eastern Indonesia and the Pacific region. It contains two primary branches: Central Malayo-Polynesian, and Eastern Malayo-Polynesian.

2.1.4.1 *Central Malayo-Polynesian*

Central Malayo-Polynesian (CMP) includes about 120 languages in the Lesser Sunda Islands and the southern and central Moluccas of eastern Indonesia. Higher-order subgrouping within CMP has proven difficult to date. Esser (1938) divided most of the languages of eastern Indonesia between 'Ambon-Timor' and 'Bima-Sumba' groups, but presented no evidence for his classification, parts of which are not likely to withstand investigation (Blust 2008a). Many of the lexical and phonological innovations used to assign languages to CMP do not cover the entire set of languages, and so suggest that this grouping arose from an original dialect chain that served as a 'diffusion corridor' rather than through a series of 'clean' language splits. Some lower-order subgroups are relatively well-established, particularly in the central Moluccas.

2.1.4.2 *Eastern Malayo-Polynesian*

Eastern Malayo-Polynesian includes the languages of the northern Moluccas and the Pacific region. It contains two primary branches, South Halmahera-West New Guinea, and Oceanic.

2.1.4.2.1 *South Halmahera-West New Guinea*

South Halmahera-West New Guinea (SHWNG) is a collection of some 30-40 languages spoken in the northern Moluccas and adjacent parts of the north coast of the Bird's Head Peninsula of west New Guinea. The present languages appear to continue a prehistoric dialect chain in which sound changes diffused from opposite ends, and sometimes overlapped in the middle regions (Blust 1978b). Most languages in this group are small, and poorly described. The best-known SHWNG languages are Buli, spoken in southern Halmahera, Taba, spoken on Makian Island just west of southern Halmahera and Numfor, which has served as a local trade language in coastal regions of western New Guinea for some centuries.

2.1.4.2.2 *Oceanic*

Oceanic (OC) contains upwards of 450 languages spoken in Polynesia, Melanesia east of the Mamberamo River in Papua, and Micronesia (exclusive of Palauan and Chamorro). The first division within Oceanic appears to separate the languages of the Admiralty Islands in western Melanesia from all others. Ross (1988) has proposed several large primary branches of Oceanic in western Melanesia, but the articulation of Oceanic languages into nested subgroups leading up to Proto Oceanic remains to be worked out in detail. Part of the difficulty, as in eastern Indonesia, probably is due to the rapid movement of prehistoric AN-speaking settlers in this region. Reasonably well-established lower-level subgroups outside western Melanesia include Southeast Solomonic, New Caledonian, Nuclear Micronesian (the Oceanic languages of Micronesia with some uncertainty about Nauruan), and Central Pacific (Polynesian, Fijian and Rotuman). The Polynesian group, which occupies a core territory nearly twice the size of the continental USA, is divided into two primary branches: Tongic (Tongan and Niuean) and Nuclear Polynesian (the rest). Following recent revisions proposed by Marck (2000) the general view now is that Nuclear Polynesian divides into eleven branches (Pukapuka, East Uvea, East Futuna, West Uvea, West Futuna-Aniwa, Emae, Mele-Fila, Tikopia, Anuta, Rennell-Bellona and Ellicean), most of which are represented by single Polynesian Outlier languages spoken in Melanesia. The Ellicean group in turn divides into three branches: Samoan, the Ellicean Outliers (spoken in Micronesia and the Solomons), and Eastern Polynesian. Finally, for over four decades Eastern Polynesian was divided into three primary branches: Rapanui (Easter Island), Tahitic (Tahitian, Maori, etc.) and Marquesic (Marquesan, Mangarevan, Hawaiian). However, as will be explained more fully in sect. 10.3.3.1, recent revisions of the radiocarbon chronology for the settlement of eastern Polynesia in Wilmshurst et al. (2011) strongly suggest that the Tahitic/Marquesic division is illusory, a conclusion that is also reached by Walworth (n.d.) in revisiting the linguistic evidence presented for this division.

2.2   Language and dialect

How many AN languages are there? This simple question is surprisingly difficult to answer. First, the language/dialect distinction can be marked in more than one way. Do we group speech communities into languages on the basis of intelligibility, cognate percentages, structural similarity, some combination of these, or in some totally different way? Second, are languages discrete entities with clearly demarcated boundaries, or like

many features of nature are they, at least in some cases, natural continua upon which boundaries can be imposed only with an inescapable measure of arbitrariness?

If we choose intelligibility as a criterion for distinguishing language from dialect in the AN family we must confront the fact that intelligibility testing has been done so far only in certain areas, most notably by members of the Summer Institute of Linguistics. Moreover, even if such information were available we would have to deal with certain universal problems in using intelligibility tests to determine language boundaries: 1. intelligibility may be non-mutual, and influenced by social factors such as differential prestige, or by prior exposure, 2. intelligibility may be delayed, as often happens for Americans during their initial exposure to 'broad Australian.' The results of testing may differ, then, depending upon length of exposure during or prior to the test situation.

Quantitative measures of lexical similarity provide another means for determining language boundaries. In practice this has generally been carried out through use of a standard 100-word or 200-word lexicostatistical test-list. The composition of such lists varies slightly from user to user, with special adaptations made for individual language families (thus 'sun' and 'day', while perfectly good for Indo-European languages, are repetitive in AN, since many of these languages express the former by 'eye of the day', and both 'eye' and 'day' appear independently on the list). One of the difficulties that has emerged in using lexicostatistical percentages for locating language boundaries is that different researchers use different cut-off points, with obviously different results. Dyen (1965a) defines speech communities as dialects of the same language if at least 70% of their basic vocabulary is cognate, while Tryon (1976) uses 81% for the same purpose. Many AN languages share a highest cognate percentage with another language that lies somewhere between these figures, and the choice of which one to adopt will therefore have significant consequences for the results of a language count.

The term 'structural similarity' can serve as a cover term for 1. morphological similarity, 2. syntactic similarity, and 3. phonological similarity, including both similarity in phoneme inventory and similarity in word shape. Any one of these parameters may affect intelligibility, and in some cases may produce results that vary with the criterion that is given preference. For example, by a lexicostatistical definition Sa'ban of northern Sarawak is a Kelabit dialect, sharing some 82% of its basic vocabulary with the standard dialect of Bario. However, rapid sound change has drastically altered the phonetics, phonemics, and morphology of this dialect to the point that speakers of Bario view it as radically different from their own speech (cp. Bario *tərur*, Sa'ban *hrol* 'egg', Bario *tulaŋ*, Sa'ban *hloəŋ* 'bone', Bario *munəd*, Sa'ban *nnət* 'correct' or Bario *ŋ-ukab*, Sa'ban *m-wap* 'to open', all of which are cognate). In cases such as this cognate percentages seem less important than intelligibility as a criterion for distinguishing language from dialect.
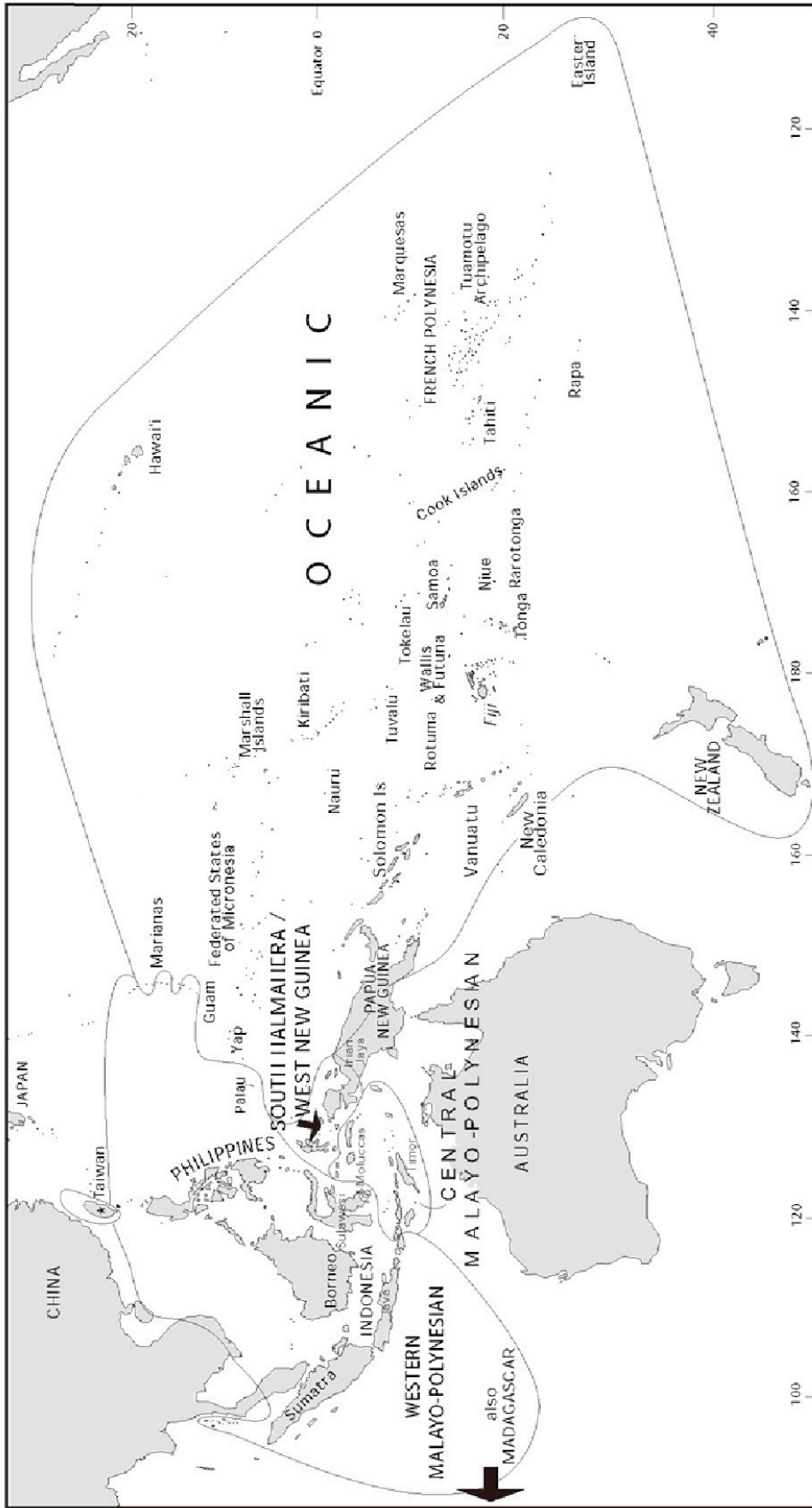
Map 1: The Austronesian language family and major subgroups

Dialect chaining presents another array of problems. In a dialect chain adjacent dialects share a high degree of mutual intelligibility, while speakers of non-adjacent dialects show a progressive decrease of intelligibility with geographical distance. If there is no sharp break in the chain such a situation creates conceptual difficulties in counting languages. A chain with members A-E, for example, may show clear dialect status for A/B, B/C, C/D and D/E, but decreased intelligibility between communities with one intervening dialect (A/C, etc.), and unambiguous language status for communities with two intervening dialects (A/D, etc.). How do we count languages in such a case? If A and D are separate languages, how can they share common dialects B and C? Since any division of such a continuum will be arbitrary, one solution is to treat the whole as one internally complex language. This solution probably is preferable to recognizing artificial internal boundaries, but it raises other problems. First, as just noted such a 'language' contains mutually unintelligible dialects. For purposes of developing literacy materials such a solution clearly won't work. Moreover, if dialects B and C were to disappear, the single language A-D would become two languages, A and D.

Dialect chains tend to develop along coastlines or on linear arrays of islands. The Melanau dialect chain, which stretches for some 230 kilometres along the coast of Sarawak, is an example of a coastal dialect chain, although its northernmost member, Balingian, stands sufficiently apart from the next language to the south to be considered a separate, but closely related language. The Chuukic (Trukic) dialect chain, which extends 2,500 kilometres from the high island of Chuuk (Truk) in the east to the atoll of Tobi in the west, is an example of an insular dialect chain, and is undoubtedly the longest dialect chain on earth. Marck (1986) has suggested that clear language breaks in Micronesia tend to correlate with distances greater than those that can be covered by an overnight voyage.

Dialect networks are similar to dialect chains, except that decreasing intelligibility associated with distance is not constrained in linear fashion, but is multidirectional. Hockett (1958:323ff) proposed the term 'L-complex' (language complex) for situations such as those exemplified in dialect chains and dialect networks without, however, distinguishing between them. It is sometimes useful to distinguish between dialect chains and dialect networks, but at other times it is useful to have a cover term that includes both, and 'L-complex' can serve that purpose.

An example of an extensive dialect network is the Bisayan complex of the central Philippines. In part because of geography (many small to medium-sized islands separated by narrow, easily-crossed seas), and in part because of history (comparatively recent and rapid expansion of a single prehistoric language at the expense of others that were previously in the region), the Bisayan region today is teeming with dialects that are sufficiently different to carry separate 'language' names, yet which are all interlinked in a complex pattern that shows few if any sharp breaks. A second area in which dialect networks complicate the counting of languages is Sabah. Three large and fairly distinct L-complexes are generally recognised: 1. Dusunic, 2. Murutic, and 3. Paitanic. Again, geography probably can be implicated in the creation of this situation. Unlike most parts of Borneo, where broad and gradually ascending rivers provide natural highways to the interior, most rivers in Sabah are short and swift, and the only way to penetrate the interior is by overland travel. In general, linguistic differences in Sabah are greatest between coastal communities on opposite sides of the island, while interior groups show less differentiation. This suggests that the AN settlement of northern Borneo initially favored the coastal zone and only began to move inland at a much later date.

In some respects it can be argued that the language/dialect distinction, difficult as it can be to determine in 'objective' terms, is not an objective phenomenon, but rather the complex result of human attitudes and treatment. Objectively, there is no natural language boundary between the Low German dialects of the north of Germany, and those of the Netherlands or Belgium. Yet many Dutch speakers are fiercely insistent that Dutch is an independent language, and most certainly not a dialect of German. In this case the fact that the speech varieties in question are spoken in politically independent nations, coupled with sentiments persisting from the military occupation of the Second World War, combine to create a strong emotional impetus to the claim for independent language status. Although the details are different, similar situations can be found in the AN family (that is, speakers claim independent language status for what linguists would be tempted to call dialects of a single language). Moreover, if two language communities have been treated in the literature as distinct it becomes difficult to combine them into a single language for survey purposes. Within the Bisayan dialect network Hiligaynon and Aklanon share some 85% of their basic vocabulary, yet there are separate grammars and dictionaries for each which cannot easily be combined into a listing for a single language. In such a case it seems almost unavoidable that the two varieties will be treated as distinct languages, whether this is 'objectively' justified or not.

The fact that similar or dissimilar names are used for closely related speech communities may also sometimes influence a decision. On the basis of their names alone Gaddang and Ga'dang of northern Luzon might be regarded as dialects of a single language, but Lewis (2009) treats them as two languages based on lexicostatistical percentages and intelligibility testing by fieldworkers from the Summer Institute of Linguistics. On the other hand, many Bisayan communities that are closely related are commonly identified by very distinct names, tending to influence the observer toward recognizing language distinctions that are largely inspired by terminological distinctness. It should also be clear that ethnic groups and languages are not coterminous. Although the *Ethnologue* recognises such Sumatran groups as Batin, Muko-Muko, and Pekal, or such Javan groups as Badui and Tenggerese as separate 'languages' these groups are almost universally described in the relevant literature as speaking dialects of well-established languages (Minangkabau or Kerinci for the first three, Sundanese for Badui, and Javanese for Tenggerese). What sets them apart is either a pattern of fairly heavy borrowing (Batin, Muko-Muko, Pekal), or a distinct religious or cultural pattern. Unlike most other Sundanese or Javanese speakers who are Muslims, the Badui and Tenggerese have unique religious systems that contain features of the pre-Islamic Hindu-Buddhist religion of Java, together with various animistic beliefs. Moreover, the Badui (divided into 'Inner Badui' and 'Outer Badui') inhabit what they regard as a sacred territory that is strictly taboo to outsiders. These cultural differences clearly show that such groups have a distinct ethnic identity, but they do not in themselves constitute evidence that they speak distinct languages, and all firsthand reports suggest that they do not.

Finally, creole languages present special problems for genetic classification since, in a sense, they have two 'parents.' In general, the issue of how to assign creoles to genetic groupings is decided on the basis of basic vocabulary. Tok Pisin, the national language of Papua New Guinea, for example, is regarded as a creolised form of English, since the great bulk of its vocabulary (both basic and non-basic) is of English origin, despite the presence of many structural features that are typical of Oceanic languages. Similarly, Chavacano/Chabakano in the southern Philippines, is regarded as a creolised form of Spanish, since the great bulk of its vocabulary is Spanish-derived.

Keeping these caveats in mind, we can now proceed to a survey of AN languages region-by-region. To ensure general comparability I have closely followed the decisions of Lewis (2009), although in cases for which I have direct knowledge it is clear that the *Ethnologue* tends toward overdifferentiation. Cases in point include the Kayan dialects of Borneo, which the *Ethnologue* represents as seven distinct languages, although only two (Kayan and Murik) seem warranted by the evidence, and Javanese, which the *Ethnologue* represents as five languages (Javanese, Caribbean, or Surinam Javanese, New Caledonian Javanese, Osing, Tenggerese) although, despite some striking and socially significant differences of dialect (particularly as regards Osing), only one seems warranted. Where variable language names are given I have tried to choose the one which is best-known through its use in the published literature, even though this may not be the preferred form in the *Ethnologue*. Some speaker numbers that reportedly are more recent are drawn from sources other than Lewis (2009).

The following survey is divided into two parts. The first reviews those languages that have an officially recognised role in the governmental system of independent nations whose citizenry is predominantly AN-speaking, as well as important *lingua franca*s which lack such an official standing. Because it is concerned with official or national languages the presentation necessarily follows political boundaries. The second part is organised instead around natural features of geography (large individual islands, coherent island groups, etc.), since these were often more important in premodern times as determinants of language boundaries, culture areas and the like. In it the ten largest and ten smallest languages are given for each major geographical area to highlight both those languages that have the greatest political and social importance (even if this is not encoded in any official manner), and those languages that are most seriously endangered.

2.3   National languages and *lingua franca*s

Most national or official languages of political states whose members are predominantly AN-speaking are naturally AN, although English may be the sole official language, or one of the official languages of government and commercial affairs. Table 2.1 lists the national/official languages of all independent states whose citizenry is predominantly AN-speaking, in decreasing order of population. Because the international importance of any language is closely correlated with its size the accompanying discussions will be longest for the topmost parts of the table. Populations estimated in July, 2012 are from CIA-The World Factbook (www.cia.gov/cia/publications/factbook/geos/kr.html#People). FSM = Federated States of Micronesia.

Singapore distinguishes national and official languages (Malay is the national language, while Mandarin, English, Malay, and Tamil are all official), but this is unusual. In most cases one or more language is legally recognised as official, although statement and practice do not always coincide. In Nauru, for example, only Nauruan is official, but English is used in most governmental and commercial transactions, and in Fiji both English and Fijian are official, although the language of most government affairs and the media is Fijian, and English-language church services would be quite unusual. As a rule, English plays a greater role as an alternative official language in nations with smaller populations and greater economic dependence on the English-speaking world

Table 2.1 National/official languages in states with Austronesian-speaking majorities

No.	Nation	Area (kilometres^2)	Population	Language
01.	Republic of Indonesia	1,919,440	248,645,008	Bahasa Indonesia
02.	Republic of the Philippines	300,000	103,775,002	Filipino/English
03.	Federation of Malaysia	329,750	29,179,952	Bahasa Malaysia
04.	Malagasy Republic	587,040	22,005,222	Malagasy/French
05.	Singapore	693	6, 310,129	Bahasa Melayu
06.	Papua New Guinea	462,840	5,353,494	Tok Pisin
07.	Timor-Leste	15,007	1,143,667	Tetun/Portuguese
08.	Fiji	18,274	890,057	Fijian/English
09.	Solomon Islands	28,450	584,578	Pijin
10.	Brunei Darussalam	5,770	408,786	Bahasa Kebangsaan
11.	Vanuatu	12,200	256,155	Bislama
12.	Samoa	2,944	194,320	Samoan
13.	FSM	702	106,487	English
14.	Kingdom of Tonga	748	106,146	Tongan
15.	Kiribati	811	101,998	Gilbertese/English
16.	Marshall Islands	181	68,480	Marshallese/English
17.	Republic of Palau	458	21,032	Palauan/English
18.	Cook Islands	230	10,777	Rarotongan/English
19.	Tuvalu	26	10,619	Tuvaluan
20.	Republic of Nauru	21	9,378	Nauruan

2.3.1 The Republic of Indonesia, The Federation of Malaysia, Singapore, Brunei Darussalam

Since the political importance of a national language generally is connected with its size, the AN language of greatest political importance probably is Bahasa Indonesia (bahasa = 'language'). After trade and cultural exchange commenced between India and China some 2,000 years ago an overland route was impractical, and travel between the two areas was largely by sea. Because the outer route west of Sumatra is dangerous, most shipping passed through the narrow Strait of Malacca between Sumatra and the Malay Peninsula. Since Malay was spoken on both sides of this passage it is not surprising that Malay speakers became involved in the India-China trade from an early time. The spice trade, which linked eastern Indonesia with India and points west, followed a similar route, and Malay-speaking middlemen consequently came to play a significant role not only in their home territory along both sides of the Strait of Malacca, but in various commercially strategic locations elsewhere in the archipelago. As a result, Malay dialects are found today not only in southwest Borneo, Sumatra and the Malay Peninsula, where they appear to be native, but also in Brunei, Banjarmasin (southeast Borneo), Manado (northern Sulawesi), Ambon (central Moluccas) and Kupang (west Timor), where they are the linguistic residue of former commercial outposts, and so define the historical route through which spices and other goods moved from eastern Indonesia to the Strait of Malacca. In addition, during the period 1405-1433 the Chinese admiral Zheng-he, at the behest of the Ming emperor Yongle, led seven colossal maritime trade and diplomacy expeditions into the Indian

Ocean, and as far as the east coast of Africa, establishing a major entrepot in Malacca. Many of his Minnan-speaking crew settled along the Strait of Malacca and intermarried with the local Malay population, giving rise to the distinctive contact-induced variety of Malay known today as 'Baba Malay' in both Malacca and Singapore (Pakir 1986, E. Thurgood 1998). Finally, during the European colonial period some speakers of Malay were resettled on both the Cocos (Keeling) Islands in the eastern Indian Ocean, and on Sri Lanka, spawning other dialects of Malay. Adelaar (1996) notes that Cocos Malay originated from an initial population of about 100 slaves brought by Alexander Hare and John Clunies-Ross to the then uninhabited islands, subsequently supplemented by laborers from Banten in west Java, and Madura, giving rise to a Javanese-influenced form of Malay that Adelaar regards as a 'Pidgin Malay Derived variety,' or PMD. Sri Lanka Malay is generally thought to owe its origin to the settlement of Malay speakers in South Asia after the Dutch seized control of the spice trade from the Portuguese in 1656 (Adelaar 1991).[3]

When the Indonesian independence movement acquired an intellectual following in the early part of the twentieth century, it was natural that Malay should be chosen as the communication vehicle for a new national life. Not only was Malay widely used as a second language, but it was a first language for only a fraction of the population (at that time probably no more than three or four million people in Indonesia), and therefore did not pose the threat of ethnic domination that was perhaps inevitable with Javanese. With the Indonesian proclamation of independence on August 17, 1945 Malay was adopted and modified to become Bahasa Indonesia.

Malaya, and the Bornean territories of Sarawak, Brunei, and Sabah underwent a rather different development under British colonial rule. Independence came peacefully in 1957, and the Federation of Malaysia was born, with Malay as the national language. Brunei (officially Brunei Darussalam), possessing natural resources that guaranteed an exceptionally high per capita income, opted for a separate existence as an Islamic sultanate under British protection. In 1984 it became an independent state. Singapore, which was part of the original Federation of Malaysia, broke away in 1965. Under different names Malay remains the national language both of Brunei (Bahasa Kebangsaan = 'national language'), and of Singapore, although, as already noted, in the latter Mandarin, English, Malay, and Tamil all are official.

Figures in Wurm and Hattori (1981) suggest that by the beginning of the 1980s around 20 million Indonesians spoke Malay as a first language, half of them in Sumatra, about 6 million in greater Jakarta, 3.3 million in Kalimantan, and 'many thousands' in other parts of the country. According to estimates made in 2004 (CIA-World Factbook), the ethnic composition of the Malaysian populations in percentages was: Malays 50.4, Chinese 23.7, indigenous 11, Indian 7.1, others 7.8. Official estimates made in 1980 recognised only two ethnic categories, namely Malays (58%), and others, meaning primarily Chinese, Indians, and Pakistanis (42%).

From the standpoint of language policy implementation the role of Malay in Indonesia and Malaysia has never been the same. Whereas Malay offered a neutral alternative to the threat of Javanese ethnic domination in Indonesia, in Malaysia the use of Malay—the native language of more than half the present population—posed just such a threat,

---

[3]   Mahdi (1999a,b) argues that Austronesian-speaking peoples had begun to settle the South Asian continent between 1000 and 600 BC, and states (1999b:191) that there was an "Austronesian presence in Sri Lanka at least as early as 450-400 BC". Whatever the merits of this claim, however, it can have nothing to do with Sri Lankan Malay.

particularly to the industrious and economically successful Chinese community. Perhaps partly for this reason, and partly because of the worldwide ascendancy of English as an international language in the twentieth century, some fluency in English is common among Malaysians with at least a secondary school education, whereas a practical knowledge of Dutch has all but disappeared in the younger generation of Indonesians.

Given their differing colonial traditions, the national languages of Indonesia and Malaysia were subjected to divergent modernizing influences. This is seen especially in the orthographies introduced in the course of adopting a romanised script. The principal differences that distinguished the spelling of Bahasa Malaysia from that of Bahasa Indonesia at the time of independence were 1) *y:j* (palatal glide), 2) *j:dj* (voiced palatal affricate), 3) *c:tj* (voiceless palatal affricate), 4) *sh:sj* (voiceless palatal fricative), 5) *ny:nj* (palatal nasal), 6) *kh:ch* (voiceless velar fricative), 7) *u:oe* (high back rounded vowel), and *o:oe* before some final consonants (mid back ~ high back rounded vowel), 8) *ě:e* (mid-central vowel/schwa), and 9) *e:é* (mid-front vowel). Thus *chukor* 'shave, shaving', *pěnyu* 'sea turtle', and *ekor* 'tail' in Malay were written *tjoekoer*, *penjoe* and *ékor* in Indonesian. Because of the greater text frequency of vowels the use of *oe* for a unit phoneme (from the Dutch orthographic tradition) was perceived by some as burdensome, and there was thus a tendency to replace it with *u*, even when other Dutch orthographic conventions were maintained. This tendency was often resisted in personal names, as Soekarno, or Soeharto.

In 1973, after years of lobbying, especially by the Indonesian scholar S. Takdir Alisjahbana, who lived for many years in Malaysia, a uniform orthography was adopted with the following values: 1) y, 2) j, 3) c, 4) sh, 5) ny, 6) kh, 7) u/o, 8) e, and 9) é. In general the orthographic unification of Bahasa Indonesia and Bahasa Malaysia thus followed the less systematic, but somewhat more economical conventions of English, except that *tj/ch* was replaced by *c*. Brunei retains the older orthography, still writing e.g. *chukor*, where both Bahasa Indonesia and Bahasa Malaysia now write an initial *c*.

According to the 1980 census, some 17.6 million Indonesians (11 million of them urban) reported Bahasa Indonesia as their home language. More recent reports, as Lewis (2009) give the number of first language speakers as about 22,800,000 in the Republic of Indonesia, and 23,187,680 in all countries, with 'over 140,000,000 second language users with varying levels of speaking and reading proficiency.' In eastern Indonesia, including the western half of the island of New Guinea, Moluccan Malay is more commonly used in daily conversation than standard Indonesian, a continuation of the local importance of this language as a *lingua franca* dating from the period of the spice trade. Since the great majority of second language speakers of Bahasa Indonesia count Javanese as their first language, a phenomenon that is likely to increase over time is the growing influence of a Javanese substratum on the pronunciation and structure of Indonesian.

2.3.2 The Republic of the Philippines

The second largest AN national language is Filipino, a slightly modified form of standard Tagalog. At the time of initial Spanish contact in the first half of the sixteenth century Tagalog was spoken around Manila Bay and in areas to the south, on the northern Philippine island of Luzon. Like Malay, it was thus in a sense predestined by geography to become a language of major political importance. During the American colonial period (1898-1946) in preparation for eventual independence the Institute of National Language was founded, and in 1937 Tagalog—then spoken by some 25% of the population—was chosen as the communication vehicle for Philippine national life. Following independence

on July 4, 1946 the use of Tagalog, under the generally unaccepted name 'Pilipino', increased rapidly until, according to one projection (Gonzalez and Postrado 1976) in the mid 1970s it was spoken by about 70% of the population of the Philippines. More recently, as a result of a constitutional amendment in 1987 the spelling 'Pilipino', has been replaced by 'Filipino'. This has posed something of a psychological dilemma, since Tagalog lacks a phoneme /f/, and Filipinos have consequently been left in the awkward position of having to choose between a perceived rusticism and a foreignism in deciding the pronunciation of the name of their national language.

Islam was introduced to the Philippines by Brunei Malays sometime prior to the arrival of Magellan in 1521, and had begun to spread as far north as Manila Bay when the Spanish initiated the Manila Galleon trade in 1565. Spanish colonial influence over a period of more than three centuries resulted in the introduction of Christianity to almost all lowland Filipinos from the Bisayas northward. However, Mindanao was left largely untouched by Spanish missionizing influence, and Mindanao and the Sulu Archipelago today constitute the Islamic heartland of the southern Philippines, a region that in some ways has felt excluded from the rest of Philippine national life.

The languages of the Philippines are sometimes divided into 'major' and 'minor' categories. Constantino (1971) recognised eight major languages: Cebuano, Tagalog, Ilokano, Hiligaynon, Bikol, Waray (also called 'Samar-Leyte Bisayan'), Kapampangan, and Pangasinan. All of these are spoken by Christianised lowland groups, although some sizeable languages, as Maranao and Magindanao, are spoken by Muslim populations in Mindanao and may now qualify as major languages based on size. As in Indonesia, the national language of the Philippines is not based on the vernacular with the most native speakers. Similarly, there is a bias in the geographical distribution of major languages: seven of the eight generally recognised major languages are concentrated in a continuous zone reaching from the southernmost part of the Bisayan Islands (central Philippines) to the fertile rice plain of central Luzon. Ilokano, spoken primarily along the coastal plain of northwestern Luzon, with substantial immigrant communities in the Cagayan valley of northern Luzon, is connected with the other major languages by a relatively narrow corridor between the sea and the central Cordillera. Finally, it is noteworthy that five of the eight major languages (all but Ilokano, Pangasinan, and Kapampangan) form a near dialect chain extending more than half the length of the Philippines.

2.3.3 The Malagasy Republic

The fourth largest AN national language after Bahasa Indonesia, Tagalog and Bahasa Malaysia, is Malagasy (Malgache). All reports agree that there is considerable dialect variation. The official census classification of the Malagasy government recognises some twenty 'ethnies' or cultural groups, and a lexicostatistical study by Vérin, Kottak and Gorlin (1969) which used data from sixteen dialect groups shows that a number of these score below the 'language limit' of 70% cognate basic vocabulary which is often thought to mark the boundary between dialects of the same language and distinct languages.

Beginning in the sixteenth century speakers of Merina, a dialect of the high plateau of east-central Madagascar, established an expansionist kingdom that eventually subjugated much of the island. Because it was the language of the major population center and subsequent capitol Antananarivo (gallicised as Tananarive), as well as the dominant ethnic group, Merina (called 'Hova' by some earlier writers) was the form of Malagasy with which the French had most contact during the colonial period (1896-1958). Upon the

achievement of independence in 1958 it became the national language. According to government estimates Merina speakers numbered about 3.2 million in 1993. Other dialects with more than one million speakers include Betsimisaraka, spoken along much of the east coast between roughly 15 degrees and 20 degrees south latitude, and Betsileo, spoken on the plateau south of Antananarivo. In addition, fluency in French is still common among the educated.

2.3.4 Papua New Guinea

The nation of Papua New Guinea includes the eastern half of the large island of New Guinea (second in size only to Greenland), the Bismarck Archipelago, the island of Bougainville in the western part of the Solomons chain, and various smaller islands between these. In the latter part of the nineteenth century Papua (the southern part of the eastern half of New Guinea) was a British protectorate. Much of the rest of the country was under German colonial rule until the end of World War I, when it became an Australian trust territory.

Unlike most AN national languages, Tok Pisin is a creole. Its classification as AN thus depends on the criteria considered relevant to the genetic classification of pidgin and creole languages in general. Although the phonology is simplified (example: the eleven vowels of standard English are reduced to five), much of the lexicon of Tok Pisin is drawn from English, including such basic nouns as *man* 'man', *meri* 'woman', *pikinini* 'child' (ultimately from Portuguese), *tis* 'tooth', *gras* 'grass, hair', *blut* 'blood', *pik* 'pig', *snek* 'snake', *san* 'sun', *mun* 'moon', and *ren* 'rain', such basic verbs as *slip* 'sleep', *kam* 'come', *go* 'go', *kuk* 'cook', *tok* 'talk' (also 'language'), *lukim* 'look at, see', *bringim* 'bring, take', and *tingim* 'think about', as well as such grammatical formatives as *long* 'preposition of location or direction' (*Boroko i stap long Mosbi* 'Boroko is in Port Moresby'; *Ol meri i go long Boroko* 'The women went to Boroko'), and *biloŋ* 'genitive; in order to' (*kap biloŋ ti* 'teacup'; *mi go biloŋ kisim pe* 'I'm going to get my paycheck'). Somewhat surprisingly in view of this pattern, a few native (mostly Tolai) terms are used for introduced objects, as *balus* (native dove sp.) 'airplane'. Syntactically and semantically, however, the AN element in Tok Pisin is very strong. The pronouns, for example, although based on English morphemes, exhibit such typical AN structural features as an inclusive/exclusive distinction in the first person non-singular, and a dual number: *yumi* 'we, us (incl.)', *mipela* 'we, us (excl.)', *yu* 'you (sg.)', *yupela* 'you (pl.)', *yutupela* 'you (dual)'. Similarly, *brata*, though derived from English 'brother', actually means 'same sex sibling', hence: 'brother (male speaking); sister (female speaking).' Tok Pisin continues a form of English that was widely used during the nineteenth century among Pacific island plantation laborers, who were often drawn together from diverse locations and language groups. A similar form of pidgin English called 'Pijin' is used in the Solomon Islands, and in Vanuatu, where it is known as Bislama (Beach-la-mar).

Within the former territory of Papua until recent decades Tok Pisin competed with a second pidgin language—Police Motu—as the favored *lingua franca*. Unlike pidgin English, Police Motu is both lexically and structurally an AN language. Motu is spoken natively around the important harbor and population center of Port Moresby. When Europeans arrived the Motu were engaged in seasonal trading voyages (called *hiri*) across the 300 kilometres wide Gulf of Papua, and some knowledge of a simplified form of Motu, called 'Hiri Motu', was found among other language groups in the area. The name 'Police Motu' originates from the introduction during the British colonial period of a simplified

form of Motu as the language of the territorial constabulary. During the twentieth century this form of Motu achieved wide currency in coastal areas of Papua (though not in the former German territory of New Guinea), although it was commonly denigrated as a 'corrupt' form of Motu. Police Motu played a key role for Allied forces in mobilizing local Papuan resistance to the Japanese occupation during World War Two, and when Papua New Guinea became independent in 1975 "its status improved so much so that it is now regarded as one of two unofficial national languages of the country" (Dutton 1986:351).

The other politically noteworthy AN language of Papua New Guinea is Tolai (also called Kuanua, Raluana, and Tuna), spoken natively by some 65,000 persons around the harbor town of Rabaul, on the Gazelle Peninsula near the northern tip of the island of New Britain. Through its use by the Methodist mission Tolai has attained the status of a *lingua franca* throughout the New Britain-New Ireland region. It is, moreover, the principal contributor (perhaps together with other closely related languages of southern New Ireland) to the native vocabulary and structure of Tok Pisin (Mosel 1980).

2.3.5 Timor-Leste/Timor Lorosa'e

Perhaps better-known as East Timor, this is currently the youngest AN-speaking nation. When Europeans seized control of the lucrative Moluccan spice trade from local merchants in the sixteenth century the Portuguese and Spanish were the first arrivals on the scene, followed some 80 years later by the Dutch. While the Portuguese and Spanish reached a territorial settlement that left the Philippines in Spanish hands and parts of what is now Indonesia and Malaysia in Portuguese hands, the Dutch and English directly challenged Portuguese claims in island Southeast Asia. As a result, most of what is today Indonesia came under Dutch control, with Portuguese influence or control remaining only in isolated pockets. The largest of these was the eastern half of the island of Timor, which (together with the enclave of Oekusi Ambeno in western Timor) remained a Portuguese colony after Indonesia gained its independence in 1949. On November 28, 1975 East Timor declared itself a sovereign state, and was almost immediately invaded by Indonesia. In July, 1976 East Timor was incorporated into the Republic of Indonesia as the Province of East Timor, a move that met with massive resistance, and a bloody and prolonged struggle over the next 23 years. The new nation of East Timor (subsequently renamed Timor-Leste, or Timor Lorosa'e) was internationally recognised as an independent nation on May 20, 2002, with Tetun and Portuguese as official languages.

2.3.6 Fiji

The Fijian Archipelago contains four major, and hundreds of minor islands on both sides of the International Date Line from about 16 degrees to 20 degrees south latitude. The largest of these are Viti Levu and Vanua Levu. Although culturally and linguistically distinct, the tiny Rotuman Archipelago, some 465 kilometres north-northwest of the nearest Fijian island, is administratively also a part of Fiji.

As of July, 2005 the population of Fiji was estimated at 893,354, of which 51% was of Fijian ancestry, 44% of Indian ancestry (descended from Hindi-speaking plantation laborers brought to Fiji during the colonial period), and 5% of European, Chinese, or other Pacific Island ancestry. This represents a change from twenty years earlier, when the Indian segment reportedly constituted slightly over 50% and the Fijian segment a little under 45% of the total population.

Fijian has figured prominently in discussions regarding the treatment of dialect chains in counting languages. Traditionally only one language was recognised, in part due to neglect of the wider dialect picture. In 1941 Arthur Capell provided the first explicit statement of a major east-west dialect division within Fiji which can be taken as grounds for the recognition of two distinct languages, western Fijian and eastern Fijian (Capell 1968:407). This claim was developed more fully by f and Sayaba (1971), who named the western language 'Wayan'. A similar position appears in Schütz (1972), and Geraghty (1983).

The earliest European missionaries to reach Fiji (in 1835) landed in the Lau Islands and the Tongan Archipelago to the east. However, they soon discovered that the political center of Fiji was on Viti Levu. There, from the small offshore island of Bau, a militarily potent paramount chief named Cakobau ruled over much of eastern Viti Levu, parts of Vanua Levu, and various smaller islands. This circumstance, together with other considerations, persuaded Westerners to adopt Bauan as the dialect through which mission activities were to be conducted. Bauan became the language of the Fijian Bible, and in time the language of the British colonial government. It maintains its importance today vis-à-vis other Fijian dialects.

2.3.7 Solomon Islands

The relatively new nation of the Solomon Islands, which became an independent member of the British Commonwealth on July 7, 1978, includes all islands of the Solomons chain from the Shortland and Treasury Islands in the west through Santa Ana and Santa Catalina Islands to the southeast of San Cristobal (now called 'Makira'). In addition, it includes the Santa Cruz Archipelago, some 500 kilometres east of the southern Solomons, and the remote Polynesian Outlier islands of Ontong Java (Luangiua), Tikopia, and Anuta. It excludes the large island of Bougainville in the far west of the group, its major satellite Buka, and several smaller neighboring islands, which belong to Papua New Guinea. The largest of the Solomon Islands is Guadalcanal, but the most populous is Malaita. English is the official language, but pidgin English (written 'Pijin') is more widely used as a *lingua franca*. Pijin is structurally similar to New Guinea pidgin English, but differs from it somewhat in pronunciation and vocabulary. For reasons of colonial policy it did not earlier enjoy the wide popularity of Tok Pisin.

2.3.8 Vanuatu

Vanuatu ('The land') is a double chain of about 80 islands, located between 12 and 21 degrees south latitude, and 166 and 171 degrees east longitude, with a land area of approximately 12,200 square kilometres. The former Anglo-French Condominium of the New Hebrides, it became independent on July 30, 1980. The population as of July, 2003 was just under 200,000.

In the nineteenth century the London-based Melanesian Mission began scriptural translations into the local languages, and Mota, a language of the Banks Islands in the far north of the New Hebrides group, achieved the status of a mission-propagated *lingua franca* of some importance. However, it never achieved wide currency, and so was eclipsed by the introduced European languages of the plantation environments. Bislama, an English-based creole, is the national language, although Bislama, English, and French are all official.

2.3.9 Samoa

The Samoan Islands lie some 1000 kilometres northeast of Fiji. As a result of colonial rivalries in the nineteenth century the archipelago was divided in two. Germany gained control of the major western islands of Sava'i and Upolu (about 2,944 square kilometres), while in 1904 the United States acquired the smaller eastern islands (about 199 square kilometres). At the outbreak of World War I New Zealand assumed administration of the German possessions in Samoa, and in 1962 this part of the archipelago gained its independence under the name 'Western Samoa' ('Western' was dropped in 1997). American Samoa remains a territory of the United States. As in most Polynesian island groups, internal dialect differences are minor. Although Samoan is the language of all official business, English is widely understood in both Samoa and American Samoa.

2.3.10 The Federated States of Micronesia

At the end of World War II, as a consequence of military victory, a linguistically and culturally heterogenous collection of islands in Micronesia came to be administered by the United States government as the United Nations Trust Territory of the Pacific Islands. In 1979 this arrangement was modified, as several parts of the former Trust Territory sought national status, and the residual core of the old administrative entity adopted a constitution as the Federated States of Micronesia, an association of Pohnpei (Ponape), Chuuk (Truk), Yap, and Kosrae (Kosaie). In 1986 independence was achieved under a compact of free association with the United States. Given the linguistic diversity of the area English functions as the official and common language of all governmental operations.

2.3.11   The Kingdom of Tonga

At the time of Western contact in the seventeenth century a line of sacred chiefs ruled Tonga. Impressed by the power and dignity of their office, European voyagers from Cook onward designated these rulers as 'kings'. The Kingdom of Tonga adopted its first constitution in 1875, and soon thereafter signed treaties with Germany, the United Kingdom, and the United States guaranteeing its independence. However, due to internal difficulties, in 1900 Tonga negotiated a treaty with the United Kingdom which made it a British protectorate. Stability was restored, adjustments made to the modern world, and in June, 1970 the Kingdom of Tonga became an independent nation.

According to Biggs (1971:490) dialect differences are minor throughout most of the Tongan Archipelago, which measures only some 750 square kilometres. A distinct language, more closely related to Samoan, was spoken on the northern island of Niuatoputapu when a vocabulary was collected by the Dutch voyager Jacob Le Maire in 1616. Niuatoputapu reportedly now has a Tongan-speaking population, but in the absence of exact linguistic information it is possible that the Samoic language still survives. Tongan is the language of all official business.

2.3.12 Kiribati/Tuvalu

The former Gilbert Islands, linguistically and culturally a part of Micronesia, straddle the Equator just west of the International Date Line along a northwest-southeast axis from about three degrees north to about two and one half degrees south latitude. The former Ellice Islands, linguistically and culturally a part of Polynesia, lie along a northwest-

southeast axis slightly to the east of the Gilberts from about six to ten and one half degrees south latitude. Great Britain declared the Gilbert and Ellice Islands a protectorate in 1892, then made them a colony in 1916.

Both archipelagos have exceptionally high population densities. The Gilberts, consisting almost entirely of low coral atolls, are among the most seriously overpopulated Pacific islands, with a consequent pattern of outmigration to other areas. Such a contrived union in a part of the world where life on a subsistence basis remains a viable alternative to modern industrial society was perhaps doomed to failure from the start. Following a referendum in the Ellice Islands, which showed a majority in favor of separate political status, on October 1, 1975 the former Gilbert and Ellice Islands colony divided into two territories. On October 1, 1978 the former Ellice Islands became the independent nation of Tuvalu. Late in 1979 the former Gilbert Islands became the independent nation of Kiribati (the native rendering of 'Gilberts'). In a 1979 treaty of friendship the United States surrendered all claims to the sparsely inhabited Phoenix and Line Island groups, which now form part of the territorial domain of Kiribati. Both Tuvaluan and Gilbertese (also known as I-Kiribati) reportedly have distinctively different northern and southern dialects. Although Gilbertese remains vigorous as a home language, English is official.

2.3.13 Marshall Islands

The Marshall Islands consists of two parallel chains of coral atolls called the Ratak (sunrise) and Ralik (sunset) Islands, between 5 and 12 degrees north latitude. Each of these is associated with a distinct dialect of Marshallese. Like other breakaway states from the former United Nations Trust Territory of the Pacific Islands, the Marshall Islands attained independence in 1986 under a compact of free association with the United States. Because of an American army base which has been located on Kwajalein atoll since 1964, and a longstanding economic dependence on the United States, English is widely spoken as a second language, and both English and Marshallese have official status.

2.3.14 Republic of Palau

The islands of the Palau group, located at about eight degrees north latitude on a line running through the island of Shikoku, Japan, and the eastern Bird's Head Peninsula of New Guinea, have had a complex history of external contact and rule. This area was adminstered by Spain through the Philippines for over a century, but some of the islands were occupied by Germany in 1885. At the conclusion of the Spanish-American War of 1898 Spain sold the islands to Germany. During World War I Japan wrested control from Germany and ruled the islands until the end of World War II. Following three decades as part of the United Nations Trust Territory of the Pacific Islands under American guardianship, Palau opted for independence in 1978 rather than joining the Federated States of Micronesia. A compact of free association with the United States was approved in 1986 and ratified in 1993. This complex history of contact is reflected in the vocabulary, which contains large numbers of loanwords from both Spanish and Japanese, and a smaller number from English and German. In 1994 Palau became independent (initially under the name 'Belau'). The nation of Palau includes not only the islands of the Palau Archipelago in the western Caroline Islands, but also the outlying atolls of Sonsorol, Tobi, and Angaur, which are linguistically and culturally distinct, and constitute separate states within the nation. Both English and Palauan are official in the Palau Archipelago, but Sonsorolese

and English are official in Sonsorol, Tobi and English are official in Tobi, and Angaur, Japanese and English are official in Angaur.

2.3.15 Cook Islands

The fifteen Cook islands, located some 1,000 kilometres southwest of Tahiti, between about ten and twenty two degrees south latitude, fall into distinct, geographically separated northern and southern groups. The capital, and most of the population is located in the southern island of Rarotonga. Formerly a colony of New Zealand, the Cooks achieved self-government in 1965, but continue to maintain a relationship of free association with New Zealand. According to Lewis (2009), in 2008 there were about 7,300 speakers of Rarotongan in the Cook Islands, but 33,220 in all countries, the great majority of expatriots residing in New Zealand. These figures show a decrease in both figures from the 1979 census that was previously available, and suggest that the Cook Islands are undergoing a steady process of gradual depopulation.

2.3.16 Republic of Nauru

The Republic of Nauru, a single island of 21 square kilometres, became a German colony in the nineteenth century, but was occupied by Australia during World War I. Apart from the Japanese occupation of 1942-1945 Nauru was under Australian administration until it gained its independence on January 31, 1968. Although tiny, the island was originally rich in phosphate deposits which were mined for the most part by immigrant Pacific Island laborers (principally Gilbertese). During the 1970s this obscure island nation, only a fraction the size of the city of Honolulu, led the world in per capita income and operated its own airline, Air Nauru. As a result of fiscal mismanagement the government has since gone into bankruptcy. It has only one native language without dialect differences. Although Nauruan is the official language English is widely understood and spoken, and for obvious practical reasons is used for most government and commercial purposes.

2.4   Language distribution by geographical region

This section will briefly survey the distribution of AN languages, and provide a few salient facts about them. Unlike the previous section, which was organised around modern political states, and so took note only of languages that have an official status, this section is based on natural features of geography. The native languages of Taiwan, for example, which were not mentioned in §2.3, are covered here, and the island of Borneo is treated as a unit, rather than divided between the Republic of Indonesia, the Federation of Malaysia and Brunei. Typological observations are minimal, and are intended only to provide a global preliminary impression, as both phonological and morphosyntactic typology will be treated at greater length in subsequent chapters. In general the orthography of sources has been retained except as stated in the Preface. Unless stated otherwise glossing conventions are those of the sources.

The divisions adopted for this section are: 1. Taiwan, 2. the Philippines, 3. Borneo (and Madagascar), 4. Mainland Southeast Asia, 5. Sumatra-Java-Bali-Lombok, 6. Sulawesi, 7. the Lesser Sundas east of Lombok, 8. the Moluccas, 9. New Guinea (and immediate satellites), 10. the Bismarck Archipelago, 11. the Solomon and Santa Cruz Islands, 12.

Vanuatu, 13. New Caledonia and the Loyalties, 14. Micronesia, 15. Rotuma-Fiji-Polynesia. For reasons of space and because much of this information is available in Lewis (2009), the listing of languages is confined to the ten largest and ten smallest languages in each region unless the total slightly exceeds 20, in which case an exhaustive listing is given. This restriction also serves to highlight the politically most important languages on the one hand, and the most endangered languages on the other, and shows at a glance the striking variations in language size as one moves from Southeast Asia into the Pacific. Except in Taiwan, where language extinction has made major inroads, names of extinct languages or those that function primarily as second languages are excluded from the tables, and populations that are estimated within a range X-Y are represented by the midpoint (hence Rotinese 123,000-133,000 is represented as 128,000).

Needless to say, figures given for number of speakers are rough estimates. Where figures for ethnic population vs. figures for speakers are known these are distinguished, but this usually is not possible. Moreover, since not all population figures are based on counts made available in the same year one is invariably faced with comparing 'apples' and 'oranges', as where Urak Lawoi' and Cacgia Roglai both are listed with 3,000 speakers, but the first is based on a 1984 estimate and the second on a 2002 estimate. If we assume that the number of Urak Lawoi' speakers has been increasing over the past three decades then it should have a larger speaker population than Cacgia Roglai today. However, without more detailed information there is no way to be certain that the speaker population has been growing rather than declining.

2.4.1 The Austronesian languages of Taiwan

The island of Taiwan is about 380 kilometres north-south, with a maximum east-west distance of around 150 kilometres, and an area about that of Holland at roughly 36,000 square kilometres. To the north it is bounded by the East China Sea, to the northeast by the Ryukyu Islands of Japan, to the east by the Pacific Ocean, to the south by the Bashi Channel which separates it from the Philippines, and to the west by the Taiwan Strait which separates it from the mainland of China, some 160 km away. Politically Taiwan is part of the Republic of China, which also includes the 64 islands of the Pescadores, or Penghu (P'eng-hu) group in the Taiwan Strait, several smaller islands hugging the coast of Fujian province in mainland China, and several islands off its east and southeast coasts, of which the most important is Lan-yu, also called Botel Tobago, or Orchid Island. Although Chinese settlement of the Pescadores began as early as the Sung dynasty (960-1279), heavy settlement of the mainland of Taiwan by Minnan-speaking immigrants from the adjacent mainland of China did not commence until the end of the sixteenth century. Most early Chinese immigration was to the western plains, as this area afforded the best rice lands. Some 24 aboriginal languages are commonly recognised, although others probably existed when the Dutch arrived in Taiwan early in the seventeenth century. Nine of these are now extinct, and several others probably will disappear by 2025. All extinct languages were once found in the better agricultural lands of the western plains and in the Taipei or Ilan basins, and their extinction (or cultural absorption) can be attributed directly to competition for land between their speakers and the incoming Taiwanese, with most of the destructive consequences of contact taking place during the period 1660-1870.

Brief history of research

The earliest written materials for the Formosan languages are scriptural translations made by Spanish and Dutch missionaries during the seventeenth century. From 1626-1642 the Spanish maintained a precarious toehold near the northern tip of Taiwan. Despite the brevity of their stay they produced at least one vocabulary and a translation of portions of the Bible in 'la lengua de los indios Tamchui', published in Manila in 1630. This publication, now difficult to obtain, presumably relates to the Basay, whose territory at that time included the region of the Tamshui River to the north of modern Taipei. As a result of this contact some Spanish loanwords made their way into the native languages, most notably Basay and Kavalan. A random sampling of apparent Spanish loanwords in Kavalan includes *baka* < Spanish *vaca* 'cow', *byabas* < Spanish *guayaba-s* 'guava', *kebayu* < Spanish *caballo* 'horse', *paskua* 'New Year's Day' < Spanish *Pascua* 'Passover; Easter; Christmas', *prasku* 'bottle' < Spanish *frasco* 'flask, bottle', *sabun* < Spanish *jabón* 'soap', and *tabaku* 'to smoke' < Spanish *tabaco* 'tobacco'. In addition, some Philippine loanwords reached Kavalan (and probably Basay, which is less well-documented), apparently through Filipinos who accompanied the Spanish in their mission to northern Taiwan. Likely examples are Kavalan *Raq* 'wine' < Tagalog *álak* (Malay *arak*, from Arabic) 'wine', *baŋka* 'canoe' < Tagalog *baŋkáʔ* 'boat', and possibly *bilaŋ* < Tagalog *bílaŋ* 'to count'. The recognition of Filipino intermediaries can also help to explain why some Spanish nouns, such as *byabas* are borrowed in their plural forms, as this is common in Tagalog and other languages of the central Philippines.

During their somewhat longer stay in Taiwan, from 1624 to 1662, the Dutch also collected data on native languages, mostly from Siraya and Favorlang of the southwest plain. Like the Spanish materials, the Dutch materials on Formosan languages were obtained in connection with scriptural translation. The most extensive of these were a dictionary of Favorlang, later known as Babuza (Happart 1650), and a translation of the Gospels of Matthew and John into Siraya (Gravius 1661). A linguistic analysis of the 17th century Dutch texts on Siraya (representing two distinct dialects) has been carefully prepared by Adelaar (1997a, 2012).

Following the expulsion of the Dutch from Taiwan in 1662 no other language materials were collected until the Japanese colonial occupation of 1895-1945. At the end of the World War II a second period of scholarly neglect ensued, broken only by the important grammar of Tung (1964). Around 1970 this situation began to change as Chinese, Japanese, and American linguists carried out fieldwork on the surviving languages, driven in part by a growing awareness that these languages are critical to understanding the early history of the AN language family, and in part by a realisation that some of them would disappear in a matter of decades. Important landmarks in scholarship include Ogawa and Asai (1935) and Ferrell (1969), both of which provide general surveys of the Formosan aborigines with comparative vocabularies, Tsuchida (1976), and the collected papers of Paul Jen-kuei Li (2004b). Other major works that have appeared on Formosan languages in recent years are Szakos (1994) on Tsou, Holmer (1996) on Seediq, Chang (1997) on Seediq and Kavalan, Li and Tsuchida (2001, 2002) on Pazih/Pazeh, Blust (2003a) on Thao, Rau and Dong (2006) on Yami, Li and Tsuchida (2006) on Kavalan, Zeitoun (2007) on Mantauran Rukai, Teng (2008), Cauquelin (2008) and Cauquelin (to appear) on Puyuma, Li (2011) on Thao, Adelaar (2011) on Siraya, and Zeitoun (to appear) on Saisiyat. Also important is a description of the Formosan Language Archive (Zeitoun, Yu and Weng 2003), and a string of publications over the past decade that focus on particular features of morphosyntactic typology across a number of Formosan languages, and in which Elizabeth

Zeitoun has been the central figure (Zeitoun ed., 2002, Zeitoun, Huang, Yeh, Chang, and Wu. 1996, Zeitoun, Yeh, Huang, Chang, and Wu 1998, Zeitoun, Huang, Yeh, Chang, and Wu 1998, Zeitoun, Yeh, Huang, Chang, and Wu 1999, Zeitoun, Yeh, Huang, Wu, and Chang 1999). Many younger scholars are now actively involved in work on the Formosan languages, and much additional material has been published in Chinese, including a number of reference grammars.

Language distribution

Table 2.2 lists the known Formosan languages in order by number of speakers, their proposed subgrouping, and the estimated size of the ethnic communities in which they are embedded. Information on number of speakers is from Lewis (2009), information on size of ethnic groups is from the Council of Indigenous Peoples, Executive Yuan (Taiwan), 'Statistics of indigenous population in Taiwan and Fukien areas by tribes for townships, cities and districts'.

Table 2.2 The Austronesian languages of Taiwan

No.	Language	Subgroup	No of speakers
1.	Amis	East Formosan	165,579 (2004)
2.	Atayal	Atayalic	88,288 ethnic (2004)
			63,000 speakers (1993)
3.	Paiwan	Paiwan	77,882 ethnic (2004)
			53,000 speakers (1981)
4.	Bunun	Bunun	45,796 ethnic (2004)
			34,000 speakers (1993)
5.	Puyuma	Puyuma	9,817 ethnic (2004)
			7,225 speakers (1993)
6.	Seediq	Atayalic	24,000 ethnic (2008)
			far fewer speakers
7.	Tsou	Tsouic(?)	5,797 ethnic (2004)
			5,000 speakers (1982)
8.	Saisiyat	NW Formosan?	5,458 ethnic (2004)
			3,200 speakers (1978)
9.	Yami	Malayo-Polynesian, Bashiic	3,255 ethnic (2004) 3,000 speakers (1994)
10.	Rukai	Rukai	11,168 ethnic (2004)
			far fewer speakers
11.	Kavalan	East Formosan	732 (2004)
12.	Thao	Western Plains	530 ethnic (2004)
			15 speakers (1999)
13.	Kanakanabu	Tsouic(?)	250 ethnic (2000)
			6-8 speakers (2012)
14.	Saaroa	Tsouic(?)	300 ethnic (2000)
			5-6 speakers (2012)
15.	Pazeh	NW Formosan?	200 ethnic?
			last speaker died 2010
16.	Babuza/Favorlang	Western Plains	extinct
17.	Basay	East Formosan	extinct

No.	Language	Subgroup	No of speakers
18.	Trobiawan	East Formosan	extinct
19.	Hoanya	Western Plains	extinct
20.	Luilang	East Formosan?	extinct
21.	Kulon	NW Formosan?	extinct
22.	Papora	Western Plains	extinct
23.	Siraya	East Formosan	extinct
24.	Taokas	Western Plains	extinct
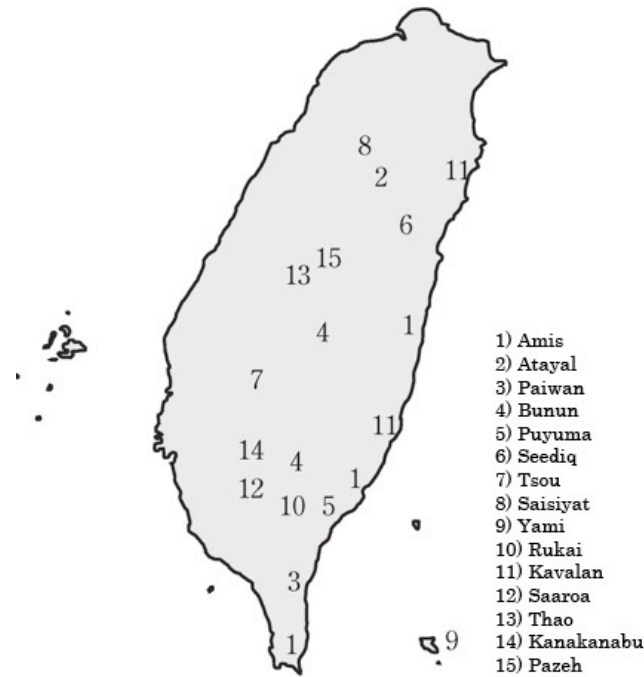
Map 2 The Austronesian languages of Taiwan

In general, the largest and most vigorous Formosan languages are those that are located in the least desirable lands. The Amis occupy a long and very narrow strip along Taiwan's east coast, where the mountains meet the sea with little level land between, the Atayal occupy the rugged mountains of northern Taiwan, and the Bunun the mountains of much of central Taiwan, where wet rice agriculture is difficult to practice (Chen 1988:17-18). Except for the Yami, smaller groups for the most part live in areas where they are in competition with the Taiwanese for local land and resources.

The notion of language extinction perhaps needs to be clarified. Although there are no fluent speakers of languages such as Basay or Taokas, speakers of other aboriginal languages may recall small numbers of lexical items or simple constructions from these socially defunct forms of speech. A few older Kavalan speakers, for example, can recall some Basay vocabulary, as this was the first language of a parent (but not one used actively in the community where they were raised). For this reason it is still possible to collect limited lexical data on some Formosan languages that are technically extinct.

There are conflicting reports on the vitality of Kavalan. Li (1982c:479) says that Kavalan "has only a few speakers," and "is still actively spoken only in Hsinshe, a very small speech community at a sea port in the east coast of Taiwan." Similarly, Grimes

(2000) listed Kavalan as "nearly extinct" in 1990, and Lewis (2009), citing a report from 2000 claims 24 speakers in a population of 200. Bareigts (1987:6), on the other hand, estimates that there are two to three thousand individuals who have either an active or a passive knowledge of Kavalan, and he names six villages in which the language is said to be still spoken by a majority of the inhabitants.

The name 'Ketangalan' (also 'Ketagalan') is potentially confusing, as it has two distinct referents in the literature. In 1944 the Japanese linguist Naoyoshi Ogawa used 'Ketangalan' and 'Luilang' for two extinct languages of the northern end of Taiwan, but Mabuchi (1953) suggested 'Basai' for 'Ketangalan' and 'Ketangalan' for 'Luilang', a suggestion that has been generally adopted by subsequent scholars. Trobiawan is represented by a single village within the Kavalan language area. Li (1995a:665) calls Basay and Trobiawan 'divergent varieties of the Ketagalan language,' and these are consequently listed separately in Table 2.2. Other names found in the literature which may refer to distinct languages are:

1. ***Qauqaut.*** Early twentieth century oral traditions hold that the Qauqaut were settled in the Taroko area of Hualien county on the northeast coast of Taiwan until about 1690. At that time they had a dispute with the Seediq and moved north along the coast to the region of Su'au. According to early Chinese documents for the Kavalan area, the Qauqaut were 'linguistically and culturally distinct from all other Formosan natives' and did not intermarry with other groups (Li 1995a:670). The Qauqaut numerals 1-10 were recorded in katakana script by a Japanese government official late in the nineteenth century. Although at least 1-9 reflect PAN reconstructions, the phonology of these forms appears to be quite distinct (*isa > *isu* 'one', *duSa > *zusu* 'two', *telu > *doru* 'three', *Sepat > *sopu* 'four', *lima > *rimu* 'five', *enem > *enu*, 'six', *pitu > *pi* 'seven', *walu > *aru* 'eight', *Siwa > *siwu* 'nine', *puluq > *toru* 'ten'). Li interprets the final -*u* of all forms except 'seven' as a requirement of the Japanese syllabary, and so posits forms *is*, *zus*, *dor*, etc. without a final vowel. He assumes that *pi* was *pit*, with failure to hear the final voiceless stop. The problems with reaching any firm conclusions about the linguistic position of Qauqaut are 1) the linguistic data is so limited, 2) the phonetic interpretation of the katakana script is unavoidably speculative, and 3) there is no way to determine whether errors were committed by a single linguistically untrained recorder of a very restricted set of forms from a single speaker.

2. ***Taivuan*** (Tevorang). Seventeenth century Dutch sources report that there were at that time three 'Tevorangian' villages, Tevorang, Taivuan and Tusigit, located in the Yüching basin in what is now eastern Tainan, Chiayi and Kaohsiung counties. These were said to be a day's journey on horseback from Fort Zeelandia (modern Tainan city). The village of Taivuan was visited by a Dutch delegation in January, 1636, and was described as a large settlement "situated in a beautiful valley about a day's journey from the central mountains" (Campbell 1903:112). Japanese sources generally consider the Tevorangians part of the Siraya, but Ferrell (1971:221ff) argues that they were culturally and linguistically distinct. He offers no linguistic data to support this conclusion, basing it entirely on seventeenth century and early eighteenth century reports of cultural differences between the populace of Taivuan and the Siraya.

3. ***Takaraian*** (Makatau). The Takaraian, or Makatau "were located in the plain to the southeast of the Siraya, in present-day eastern Kaohsiung and Pingtung counties" (Ferrell 1971:225). The number of villages is unknown, but included at least Takaraian and Tapulang. Japanese scholars have tended to classify the Takaraian as part of the Siraya, but the Dutch sources make it clear that Siraya was not understood in the Takaraian villages.

Based on materials culled from official colonial documents by the Japanese scholar Naoyoshi Ogawa (1869-1947), Tsuchida and Yamada (1991) were able to show that Siraya, Taivuan, and Takaraian probably were three distinct languages.

4. ***Pangsoia-Dolatok.*** In the 1630s seven populous villages of the Pangsoia were located near the coast around the mouth of Linpien creek, in present-day Pingtung County at the southern end of Taiwan. In addition, five villages of the Dolatok were located near the mouth of the Lower Tamshui River (not to be confused with the Tamshui River in northern Taiwan). Again, there are no linguistic data, and Ferrell is forced to conclude that "it is impossible to ascertain whether the Dolatok-Pangsoia may have been related to the Siraya, Takaraian or Longkiau, or whether they may have been a lowland group of Paiwan" (Ferrell 1971:229).

5. ***Longkiau.*** During the Dutch period there were fifteen to twenty Longkiau villages two days journey south of Pangsoia, in the lowlands of the Hengchun Peninsula (Ferrell 1971:231). Since some of these can be identified with known Paiwan villages, Ferrell speculates that the Longkiau belonged to the Sapdiq division of the Paiwan.

6. ***Lamay.*** The small island of Lamay, called Tugin in the native language, and Hsiao Liu-chiu (Little Ryukyu) in Mandarin, is located several miles from the coast near the present border between Kaohsiung and Pingtung counties. According to Ferrell (1971:232), in Dutch times the aboriginal population of the island fiercely resisted all outsiders who tried to land. Even in conducting trade with the Chinese the natives rowed out to the anchored junks rather than allowing them to come ashore. Because of continuing frictions in the 1630s the Dutch laid siege to the island, dispersing its inhabitants as slaves among the Siraya. No language material was recorded, and the linguistic affiliation of the population consequently remains completely unknown.

Typological overview

It is difficult to generalise about the typology of Formosan languages, particularly with regard to phonology or morpheme structure. Phoneme inventories usually have fifteen to twenty consonants and four vowels (the vowel 'triangle' plus schwa), but are otherwise quite variable. Segment types found in two or more Formosan subgroups that are not typical of AN languages include voiced and voiceless interdental fricatives (Saisiyat, Thao, most Rukai dialects), palatal fricatives (Atayal, Saisiyat, Thao), a voiceless alveolar affricate (Atayal, Seediq, Tsou, Kanakanabu, Saaroa, Rukai, Amis, Paiwan), voiceless laterals (Thao, Saaroa), retroflex stops or laterals (Rukai, Paiwan, Puyuma), a voiceless velar fricative (Seediq, Pazeh, the Ishbukun dialect of Bunun), and uvular stops (Atayal, Seediq, Thao, Bunun, Paiwan). In addition, Thao, Bunun and Tsou share the preglottalisation of *b* and *d* as an areal feature, although—as will be shown—preglottalised and imploded stops have a scattered distribution in the AN languages of insular Southeast Asia. In addition to unusual segment types, Formosan languages generally have more fricative consonants than is typical of AN languages. While most languages in the Philippines have just *s* and *h* (or even just *s*), for example, several of the Formosan languages have four or more fricatives. Thao, with seven fricative phonemes (f, c [θ], z [ð], s, lh [ɬ], sh [ʃ], h), and one fricative allophone ([v] as an allophone of *w*) appears to have the record, but is followed closely by Saisiyat, with five (s [θ], z [ð], ʃ, b [β], h).

Morpheme structure is also quite variable in Formosan languages. Most languages allow no consonant clusters, although sequences of unlike vowels occur. The canonical shape of non-reduplicated bases can thus be schematised as CVCVC, where all consonants are optional. Both Tsou (Wright 1999) and Thao (Blust 2003) are exceptional in permitting

a wide range of consonant clusters in initial and medial position, as in the following Thao examples: *cpiq* 'thresh grain by beating the stalks', *lhfaz* 'a belch', *pruq* 'earth, ground', *psaq* 'kick', *qtilha* 'salt', *shdu* 'suitable, sufficient'; *antu* 'marker of negation', *ma-dishlum* 'green', *fuczash* 'a steamed mixture of grains and sweet potatoes', *mi-lhuŋqu* 'sit down'. Although Amis is also sometimes written with a variety of initial consonant clusters, this is simply an orthographic convention by which the schwa is given zero representation, as in *ccay* (Fey 1986) for [tsətsáiʔ] 'one', *spat* for [səpát] 'four', or *hmot* 'rectum' for [həm:ót] 'coccyx'.

Syntactically almost all Formosan languages are predicate-initial. Since some recent analyses have suggested that several of these languages are both morphologically and syntactically ergative, it is perhaps inappropriate to characterise word order in terms of the common practice of typologists, using the symbols S, V, and O. This is, however, a convenient way of representing certain types of information in shorthand fashion, and will be followed here despite the theoretical objections that can be raised against it. Almost all Formosan languages, then, are VSO or VOS. The principal exceptions are Saisiyat and Thao, which are SVO, evidently as a result of intensive contact with SVO Taiwanese over much of the past century. The case of Thao is particularly instructive, since verb-initial constructions appear more frequently in texts, and in spontaneous sentences during elicitation, particularly after working for some hours with a speaker.

Most Formosan languages have complex systems of verbal affixation that allow a range of nominal arguments, including actor, patient, location, instrument and benefactor to be morphologically encoded as having a special relationship to the verb. These kinds of verb systems are most commonly called voice systems or focus systems, where the term 'focus' diverges from its common meaning in general linguistics. In the interest of general comparability between descriptions that are often framed in terms of very different working assumptions and expressed in divergent terminologies, I will usually identify the relationships between verbs and these specially marked nominal arguments in both Philippine-type languages and Western Indonesian-type languages as 'voices'. However, quotations from the literature often use the term 'focus' for these relationships, and it is thus necessary in a survey volume such as this to switch back and forth between the two terminological usages. Finally, I have made minor changes in the original sources for some languages, both in grammatical terminology and in orthography. The following sentences or phrases from Atayal illustrate:

Atayal (Huang 1993)

1.　　sayun sakuʔ 'I am Sayun'
　　　Sayun **1sg.bn**

2.　　tayan yabu 'Yabu is Atayal'
　　　Atayal Yabu

3.　　yat tayan tali   'Tali is not Atayal'
　　　neg Atayal Tali

4.　　m-qwas qutux knerin 'A woman is singing'
　　　(**av**-sing one woman)

5.　　s-qwas-mu qwas qutux knerin 'I sang a song for a girl'
　　　bv-sing-**1sg.gen** song one female

6.   m-ʔabi tali 'Tali is asleep'
        av-sleep Tali

7.   ʔby-an tali 'the place where Tali has slept'
        sleep-**lv** Tali

8.   ʔby-un tali 'the place where Tali will sleep/what Tali will sleep on'
        sleep-**pv** Tali

9.   nanuʔ ʔby-an tali 'Where is the place Tali wants to sleep?'
        what sleep-**lv** Tali

10.   nanuʔ ʔby-un tali 'What is the thing Tali wants to sleep on?
        what sleep-**pv** Tali

Sentences 1-3 have nominal predicates, and sentences 4-10 verbal predicates. Auxiliaries may precede a verbal predicate, and negatives may precede any predicate. In sentences 4-10 the verb carries an affix that marks its relationship to one of the nominal arguments. In some languages the argument so singled out is preceded by a grammatical morpheme. In the variety of Atayal reported here this relationship is marked instead by word order, the last argument having the participant role encoded by the verbal affix. Sentences 4 and 5 differ in verbal affixation, *m-* marking the last nominal as actor, and *s-* marking it as benefactor. Sentences 6-10 also differ in verbal affixation. The affix *m-* again marks the actor. In many languages cognates of *-an* mark locative relationships, while cognates of *-un* mark patients; this terminology has been used here, but the contrast of *-an* and *-un* is subtle in Atayal and the terms 'locative voice' and 'patient voice' must be taken as conveniences rather than semantically precise descriptions. Of more general interest is the observation that verbs with the affixes *s-*, *-an*, or *-un* take genitive agents, and as a result many of these affixed forms function as verbs or as nouns, depending upon the larger syntactic context in which they are found. The proper characterisation of such verb systems has been the subject of longstanding controversy, and will be addressed in a later chapter. What is important to observe here is that the verbal typology of most Formosan languages is very similar to that of most Philippine languages. In other respects, as in historical phonology and lexicon, Formosan languages differ sharply both from one another and from all other AN languages.

2.4.2 The languages of the Philippines

The Philippine Archipelago contains more than 7,000 islands, and occupies a total area (including territorial waters) of over 300,000 square kilometres. It can be divided into three geographical regions of roughly equal size: the large northern island of Luzon (104,700 square kilometres), Palawan and the numerous Bisayan Islands of the central Philippines, of which the largest is Samar (13,100 square kilometres), and the large southern island of Mindanao (101,500 square kilometres).

To the north the Philippine Archipelago borders Taiwan, and to the south it borders the large islands of Borneo and Sulawesi. The distance between Taiwan and northern Luzon is bridged in part by the Babuyan and Batanes Islands, a chain of 25-30 small, windswept grassy islands which reportedly are intervisible in clear weather. Together they extend like stepping stones from near the northern end of Luzon (Fuga Island) to Y'ami Island, roughly 110 kilometres southeast of Lan-yu, or Botel Tobago Island, off the southeast

coast of Taiwan. The distance between the Philippines and Borneo is similarly bridged by the many islands of the Sulu Archipelago, extending from Basilan Island just off the southern tip of the Zamboanga Peninsula in southwest Mindanao, to Sibutu and other smaller islets near the mouth of Darvel Bay in eastern Sabah, and the sea gap between the Philippines and Sulawesi is partly filled by the Sangir and Talaud Islands. Lewis (2009) lists 175 languages for the Philippines, four of which are extinct.

Brief history of research

Spanish missionaries began the compilation of materials on Philippine languages in the seventeenth century, at which time contact was confined to the major lowland populations, which were physically more accessible. Ethnic minorities in the more remote mountain fastnesses or on smaller and less commonly visited islands were exposed to western influences only much later. Spanish work on Philippine languages included not only the translation of Christian doctrinal materials, but also Latin-based grammars and dictionaries. During the seventeenth century grammars were written for Bikol, Ilokano, Pangasinan, Tagalog and Waray-Waray, and during the first half of the eighteenth century Ibanag and Kapampangan were added to those languages which received the attention of the Spanish clergy. Work of a similar nature, extended to some of the minority languages, continued to the close of the nineteenth century, with some Spanish publications on Philippine languages appearing in the first two decades of the twentieth century.

With the advent of American control in 1898, studies of Philippine languages began to appear in English. Most of these were written by Americans, but contributions by Filipino scholars began to appear in the 1930s, and following Philippine independence at the end of the Second World War this trend accelerated. The Linguistic Society of the Philippines was founded in 1970 and has published its official organ, *The Philippine Journal of Linguistics*, since that time. Nearly all articles in this journal are in English. A smaller, but growing number of linguistic studies is being written in Tagalog, particularly when Tagalog is the object of description. Landmark publications include Bloomfield (1917), the first attempt to describe a Philippine language in terms of a model not derived from classical Latin grammar, Reid (1971), which for the first time made a large amount of comparable lexical data available for the many minor languages of the Philippines, Wolff (1972), perhaps the most detailed dictionary available for any Philippine language, Schachter and Otanes (1972), the most comprehensive reference grammar available for Tagalog, Zorc (1977), the most fine-grained dialect study done to date on any part of the Philippines, McFarland (1980), which provides the clearest mapping of language boundaries available in any publication, and Madulid (2001), which is the most comprehensive dictionary done for any specialised semantic domain. Other major contributions to Philippine linguistics in recent years are Rubino (2000), Collins, Collins, and Hashim (2001), Wolfenden (2001), Behrens (2002), Awed, Underwood, and van Wynen (2004), Lobel and Riwarung (2009, 2011), Lobel and Hall (2010), Ameda, Tigo, Mesa, and Ballard (2011), Maree and Tomas (2012), and Lobel (2013).

Language distribution

Table 2.3. lists the ten largest and smallest languages of the Philippines in number of first-language speakers (CP = Central Philippine, NC = Northern Cordilleran, CL = Central Luzon, SC = Southern Cordilleran, DN = Danao, P = Palawanic, MAN = Manobo):

Table 2.3 The ten largest and ten smallest languages of the Philippines[4]

Language	Location	Subgroup	No. of speakers
1. Tagalog	central Luzon	CP	23,853,200 (2000)
2. Cebuano	Bisayas	CP	15,807,260 (2000)
3. Ilokano	NW Luzon	NC	6,996,600 (2000)
4. Ilonggo/Hiligaynon	Bisayas	CP	5,770,000 (2000)
5. Bikol	southern Luzon	CP	4,842,303 (2000)
6. Waray(-Waray)	Bisayas	CP	2,510,000 (2000)
7. Kapampangan	central Luzon	CL	2,312,870 (2000)
8. Pangasinan	N. central Luzon	SC	1,362,142 (2000)
9. Kinaray-a	Bisayas	CP	1,051,968 (2000)
10. Maranao	SW Mindanao	DN	1,035,966 (2000)
1. Pudtol Atta	northern Luzon	NC	711 (2000)
2. Bataan Ayta	western Luzon	CL	500 (2000)
3. Faire Atta	northern Luzon	NC	300 (2000)
4. Northern Alta	northeastern Luzon	SC	200 (2000)
5. Batak	Palawan	P	200 (2000)
6. Inagta Alabat	northeast Luzon	NC	60? (2010)
7. Tasaday	southern Mindanao	MAN	25 (2005)
8. Isarog Agta	southeast Luzon	CP	5-6 (2000)
9. Ata	Negros	CP	2-5 (2000)
10. Ratagnon	Mindoro	CP	2-3 (2000)

---

[4]   Based on extensive fieldwork conducted from 2005 to 2007, Jason Lobel (p.c.) believes that Mt. Iraya Agta (listed with 150 speakers in 1979 by Lewis 2009), Sorsogon Ayta (listed with 18 speakers in 2000 by Lewis 2009) and Arta (listed with 15 speakers in 2000 by Lewis 2009) probably are extinct. He adds that Manide/Agta Camarines Norte (listed with 15 speakers in 2000 by Lewis 2009) should not be included on the list of most endangered languages because it has at least a thousand speakers.
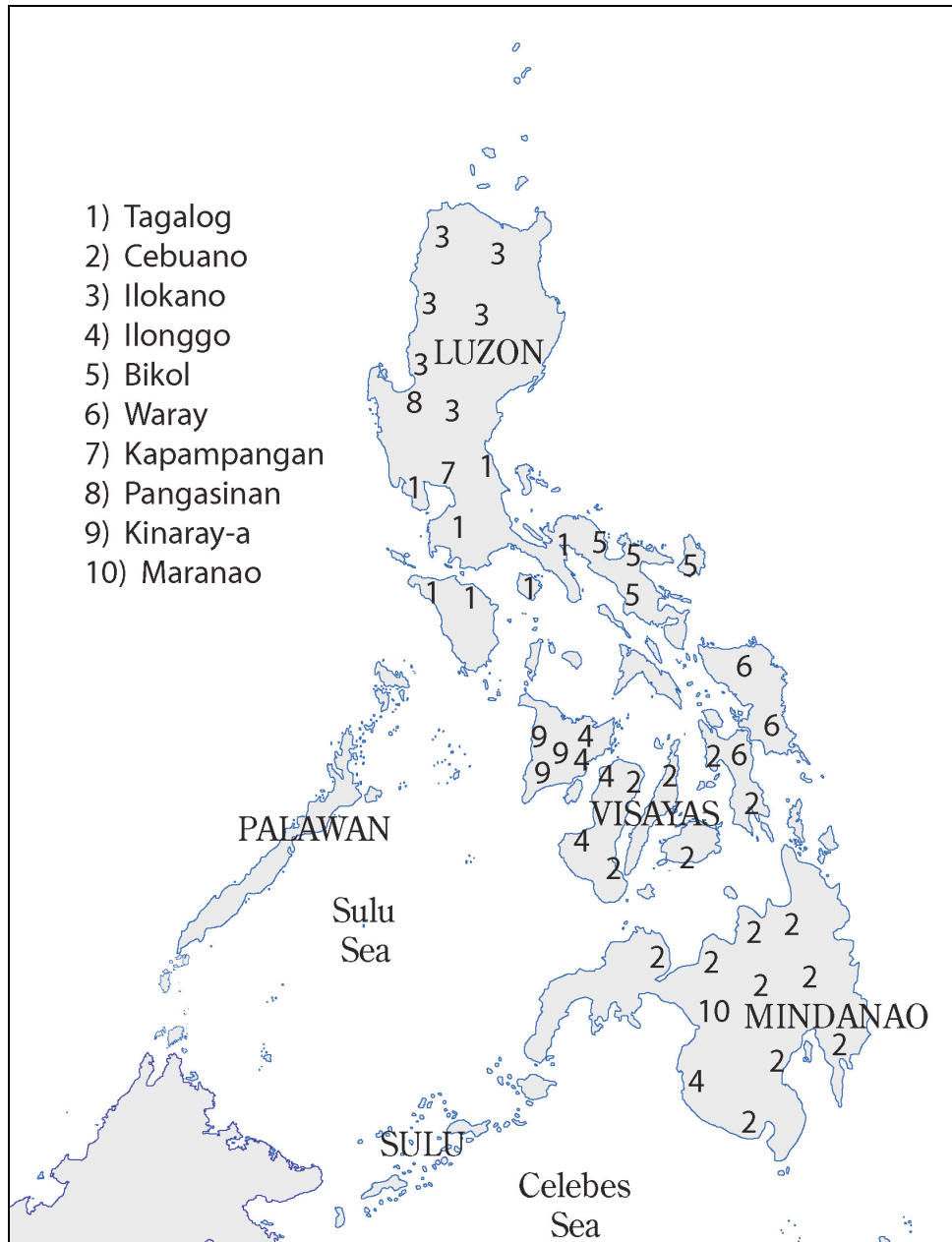
Map 3 The ten largest languages of the Philippines

For the smallest languages I have used Headland (2003), which has more detailed information than the *Ethnologue*. Some of this information clearly is provisional, given the difficulty of obtaining exact numbers for populations that are mobile, and sometimes wary of contact with outsiders. Lewis (2009), for example, cites a 1991 estimate for the Pudtol Agta of 500-700, or 171 families, but Headland gives a population of just 100 during the same time period for the same group. The pattern that emerges most clearly is that the largest languages of the Philippines are those spoken by lowland populations that are predominantly Muslim (Maranao, Magindanao) or Christian (the rest). The smallest languages are those spoken by interior populations which in general have been less affected by extraneous cultural or religious influences, at least until recent times. All but one of the ten smallest languages of the Philippines are spoken by Negritos, testifying to the traditionally small membership of foraging groups everywhere, and to the

marginalisation and endangerment of most Negrito populations in the Philippines today. Not coincidentally, the one non-Negrito group in this category is Tasaday, spoken by a population of feral foragers. Among sedentary groups, only a handful of languages have fewer than 5,000 speakers. These include Ibatan, a Bashiic language spoken in the Babuyan Islands immediately north of Luzon (1,000 speakers in 1996), Karao, a South Cordilleran language spoken in central Luzon (1,400 speakers in 1998), and I-wak, a South Cordilleran language spoken in central Luzon (2,000-3,000 speakers in 1987).

In general, language size correlates closely with type of terrain. Although spectacular results have been achieved in terracing steep mountain slopes for irrigated rice in some areas (most notably the Ifugao-speaking region of Banaue), on the whole the rugged uplands are not capable of supporting population densities of the order found in lowland areas. As a result language groups in the mountains are far smaller in average size than those in the lowlands. This is particularly true of the traditionally foraging populations of Negritos, since nomadic bands have a low upper population limit in any case, but it is also true of sedentary populations. In addition, the physical environment of the uplands reduces communication over large areas, further contributing to linguistic fragmentation. As might be expected, small linguistic groups in less accessible locations more easily escape notice. Some languages spoken by Negrito populations, for example, have been discovered only recently, as with Arta, and Northern and Southern Alta, first described by Lawrence A. Reid in the summer of 1987. Similarly, with rare exceptions the amount of scholarly attention that languages receive correlates closely with their size. Since the days of the Spanish friars several dictionaries and over 50 grammars have been written for Tagalog. In stark contrast, only one of the ten smallest languages (Casiguran Dumagat) is represented by a dictionary, and none by a grammar.

Two dialects of Ilokano, Northern and Southern, are commonly recognised. The earliest Spanish grammar of the language dates from 1627, and a number of others were composed during the 1860s and 1870s. As a result of population expansion in recent times Ilokano speakers have now settled large parts of the Cagayan valley, the Babuyan Islands, and many other parts of northern and north-central Luzon. They also form by far the largest part of the Filipino immigrant community in Hawai'i.

Kapampangan is spoken in the agricultural heartland of Luzon, to the north of Manila Bay. The earliest Spanish grammar dates from 1729, with a dictionary following in 1732. Kapampangan has long been in contact with Tagalog, and although the two languages have substantially different histories, they have borrowed liberally from one another. In earlier (probably prehistoric) times the direction of borrowing appears to have been predominantly from Kapampangan into Tagalog, but in more recent times this direction has been reversed.

Pangasinan is spoken in another rich agricultural region, just south of Lingayen gulf, where it is completely surrounded by Ilokano speakers except for a small part of its southern border, where it abuts Sambal. The earliest work on the language appears to be a Spanish manuscript grammar composed in 1690.

Tagalog is spoken in a large number of dialects centering on Manila Bay. In addition to the standard dialect of Manila, the Marinduque and Pagsanghan dialects have figured in various linguistic studies. As noted above, well over 50 grammars of Tagalog have been written in Spanish and English. The first of these dates from 1610, only 38 years after the transfer of the Manila Galleon from Cebu to Manila.

Mintz and Britanico (1985) recognise elevent dialects of Bikol, although chaining complicates the picture in some areas. The Northern Sorsogon, Southern Sorsogon, and

Masbate dialects appear to be transitional between Bikol and various forms of Bisayan (Waray in northern Samar, Ilonggo or Hiligaynon of Panay). There is thus no sharp language boundary between Bikol and the Bisayan L-complex of the central Philippines. Bikol is one of the best-described Philippine languages, represented by five Spanish grammars published between 1647 and 1904, and by one grammar written in English.

Without question the most difficult problem in listing the languages of the central Philippines is how to treat the extensive network of interconnections that make up the Bisayan dialect region. The most detailed study of this problem to date is that of Zorc (1977), who used a variety of methods, including intelligibility testing, lexicostatistical percentages, and functor analysis to determine linguistic boundaries. In general, similarity as measured in any of these ways decreases with geographical distance, making it impossible to draw sharp language boundaries. Based on lexicostatistical counts Zorc (1977:178) recognised West, Central and South Bisayan subgroups which "are linked together by transitional dialects." He does not call these dialect groups 'languages', and indeed it appears impossible to do so in any non-arbitrary way. Moreover, as noted above, central Bisayan dialects such as Waray appear to be transitional to the Bikol dialects of southeast Luzon. The Bisayan region thus constitutes a classic L-complex. Two factors probably underlie this situation. First, unlike Luzon and Mindanao, which are divided by rugged mountain masses that hinder transportation and communication, the Bisayan region presents fewer natural barriers to contact. Some of the islands have hilly or moderately mountainous interiors, but most settlement is in the coastal zone, and water transport is easy over the relatively narrow inland seas. Second, the present relatively low level of linguistic diversity in the Bisayan Islands almost certainly is a product of linguistic leveling which resulted from the expansion of a single prehistoric speech community at the expense of others that have left no descendants. The most recent expression of this expansive tendency has been the spread of Cebuano, originally confined to Cebu and Bohol Islands, into most parts of Mindanao within the past century, where it now functions as a regional *lingua franca*. The territorial expansion of Cebuano speakers in the southern Philippines can be seen as comparable in general terms to the expansion of Ilokano in the northern Philippines.

The major Bisayan dialects are Cebuano, Hiligaynon, and Waray, or Waray-Waray (also known as Samar-Leyte Bisayan), each with two million or more speakers. As recently as 1975 Cebuano surpassed Tagalog in number of first-language speakers, and has been overtaken in recent years primarily because of the position of Tagalog as the national language, and its consequent success in recruiting first-language speakers from other language groups. Of the ten largest languages of the Philippines five are members of the Central Philippine group (Tagalog, Bikol, Cebuano, Hiligaynon, Waray). It thus appears likely that the complex dialect network of the central Philippines, the apparent leveling of linguistic diversity in the same area, and the relatively large sizes of Central Philippine languages are all interrelated: as a result of some factor that triggered rapid population growth, the ancestral Central Philippine language community expanded rapidly over the Bisayan Islands and southern Luzon, reducing the linguistic diversity of the area, and giving rise to the historically attested Bisayan dialect network.

In general the languages of the southern Philippines show less evidence of language leveling than those of the central Philippines. The largest languages indigenous to Mindanao are Maranao and Magindanao, both spoken by predominantly Muslim populations. The Bilic languages (Tiruray, Bilaan, Tboli), spoken by small interior populations that have remained culturally conservative, are typologically rather different

from most Philippine languages. This observation has occasionally raised doubts about the validity of including them in a Philippine group, but close attention to diagnostic innovations supports their inclusion (Zorc 1986, Blust 1991a, 1992). The Manobo languages, which cover a large part of the island of Mindanao, present a problem of continuous variation which is reminiscent of that found in the Bisayan Islands of the central Philippines, although in general the Manobo languages appear to be somewhat more sharply divided into dialect groups.

Typological overview

Phoneme inventories in the Philippines tend to be fairly simple, typically containing 15-16 consonants and the vowels *i, u, i̵, a*. Some languages have only the three maximally distinct vowels of the 'vowel triangle', while a few others have developed larger systems of up to seven or eight vowels (Reid 1973a). A palatal series, fairly common in the languages of western Indonesia, is absent from most Philippine languages, with the notable exceptions of the Bashiic languages (where it is historically secondary), Kapampangan (where it is historically conservative), and some languages of the southern Philippines, as the Sama-Bajaw languages and Tausug (where it is, at least in part, a product of borrowing from Malay).

The most distinctive characteristic of the phonologies of Philippine languages is found in prosody: many languages of the northern and central Philippines have phonemic stress, as in Ilokano *búrik* 'to carve, engrave' vs. *burík* 'kind of bird with variegated plumage', Kapampangan *ápiʔ* 'lime (calcium carbonate)' vs. *apíʔ* 'fire', or Tagalog *tábon* 'banked or piled cover of earth, rubbish or the like' vs. *tabón* 'a megapode bird'. As will be seen, the history of stress in Philippine languages remains a major unsolved problem.

Philippine word structure also tends to differ from that of most other AN languages in allowing a large number of medial heterorganic consonant clusters in non-reduplicated bases. Examples from languages in the northern and central Philippines are Isneg *alnád* 'go back, return', *bugsóŋ* 'put into a bag', *xirgáy* 'name of a spirit', and Bikol *bikrát* 'to open something wide', *kadlóm* 'k.o. plant', *saŋláy* 'to cook.'[5] Both phonemic stress and the tolerance of heterorganic consonant clusters are lost almost everywhere in the southern Philippines, although the Bilic languages of southern Mindanao have developed additional consonant clusters in word-initial position.

Syntactically, most Philippine languages are predicate-initial, and have complex systems of verb morphology that allow a range of nominal arguments, including actor, patient, location, instrument and benefactor to be morphologically encoded as having a special relationship to the verb. In this respect they resemble most Formosan languages, not only in general typology, but also in the cognation of affixes that are central to the voice system. Tagalog can serve to illustrate Philippine-type verb systems (focused arguments are in bold italic; in the morpheme glosses infixes are placed after the stem):[6]

---

[5]   The preferred spelling for 'Isneg' is now 'Isnag', but since the major published source on the language uses   the first spelling I adhere to it here and throughout the book.

[6]   Some details peculiar to Tagalog, or to Tagalog and some of its closest relatives, have been omitted in the interest of illustrating features that are more widely-shared with other Philippine-type languages.

Tagalog (composite sources)

1.   b<um>ilí naŋ kotse ***ay lalake*** '***The man*** bought a car'
      buy-**av gen** car **nom** man

2.   b<um>ilí naŋ kotse ***si Juan*** '***Juan*** bought a car'
      buy-**av gen** car **nom** Juan

3.   bi-bilh-ín naŋ lalake ***ay kotse*** 'A/The man will buy ***the car***'
      fut-buy-**pv gen** man **nom** car

4.   bi-bilh-ín ni Juan ***ay kotse*** 'Juan will buy ***the car***'
      fut-buy-**pv gen** Juan **nom** car

5.   b<in>ilí naŋ lalake ***ay kotse*** 'A/The man bought ***the car***'
      buy-**pv.perf gen** man **nom** car

6.   b<in>ilh-án naŋ lalake naŋ isdáʔ ***ay bataʔ***
      buy-**perf-lv gen** man **gen** fish **nom** child 'A/The man bought some fish ***from the child***' (-an marks source)

7.   b<in>igy-án naŋ lalake naŋ libro ***ay bataʔ***
      give-**perf-lv gen** man **gen** book **nom** child 'A/The man gave a book ***to the child***' (-an marks goal)

8.   t<in>amn-án naŋ lalake naŋ damó ***ay lupaʔ***
      plant-**perf-lv gen** man **gen** grass **nom** ground) 'A/The man planted grass ***in the ground***' (-an marks location)

9.   i-b<in>ilí naŋ lalake naŋ isdáʔ ***ay pera***
      iv-buy-**perf gen** man **gen** fish **nom** money   'A/The man bought a fish ***with the money***' (i- marks instrument)

10.   i-b<in>ilí naŋ lalake naŋ isdáʔ ***ay bataʔ***
      bv-buy-**perf gen** man **gen** fish **nom** child 'A/The man bought some fish ***for the child***' (i- marks beneficiary)

Highlighted sentence constituents (*ay* phrases) have a morphologically specified relationship to the verb. The verbal affixes that specify this relationship are the infix -*um*-, the suffixes -*in* and -*an* and the prefix *i*-. In addition, perfective -*in*- functions as a portmanteau affix in verbs that take suffixal -*in* in the non-perfective form (that is, -*in*- requires a zero allomorph of -*in*). Although most Philippine languages have similar verb systems, the Samalan and Bilic languages deviate rather sharply from this pattern.

2.4.3 The languages of Borneo (and Madagascar)

Borneo, generally considered the world's third largest island at about 744,360 square kilometres, is divided politically between three countries. The largest part of the island, Kalimantan, belongs to the Republic of Indonesia. Two smaller parts, Sabah and Sarawak, are states in the Federation of Malaysia, and the smallest part is the independent nation of Brunei Darussalam. Despite some excellent early descriptions linguistic work in Borneo has generally lagged far behind that in the Philippines, or other parts of western Indonesia. To some extent this is because most language groups are small, and many are found in

fairly remote interior regions. Lewis (2009) lists 200 languages for Borneo (Brunei 17, Kalimantan 82, Sabah 54, Sarawak 47), of which one is extinct.

Brief history of research

In 1858 the Swiss missionary August Hardeland published a 374 page grammar of Ngaju Dayak plus texts, and the following year he published a 638 page dictionary in ten point double columns. Hardeland's grammar is, alongside the Timugon Murut grammar of Prentice (1971), one of the two most detailed grammars available for any of the languages of Borneo, and his dictionary is rivaled only by the recent *Kadazan Dusun–Malay-English dictionary*, compiled by a team of Kadazan speakers under the rubric of The Kadazan Dusun Cultural Association of Sabah in 1995. Dutch scholars, who were active in other parts of what was then the Netherlands East Indies, paid surprisingly little attention to the languages of Borneo. The most plausible explanation for this neglect is that Dutch scholarship on the languages of Indonesia traditionally had a strongly philological orientation, and so tended to gravitate toward those languages with a written literature going back at least several centuries. In this respect Borneo was seen as a backwater, an area occupied mostly by animist groups which had experienced little if any of the cultural impact of Indianisation, and which has been Islamised only in coastal pockets, most notably in Brunei (which lay outside the Dutch colonial sphere, as it was part of British Malaysia).

The most significant scholarly achievement concerning the languages of Borneo during the twentieth century arguably was the discovery by the Norwegian missionary and linguist Otto Christian Dahl that Malagasy is closely related to Ma'anyan of southeast Kalimantan (Dahl 1951). Based on Sanskrit loanwords which could only have entered Malagasy before it reached the coast of east Africa, Dahl concluded that the Malagasy migration took place no earlier than the fifth century, at a time when Indian cultural influence was spreading through the more accessible parts of western Indonesia. However, this view left important questions unanswered. In particular, since southeast Kalimantan was never Indianised, the manner in which Malagasy acquired Sanskrit loanwords was unclear. It is now known that Malagasy subgroups with the Southeast Barito languages as a group rather than with Ma'anyan in particular, and it is generally agreed that the Malagasy migration took place between the seventh and thirteenth centuries. More recent work by Adelaar (1989) has shown that the movement from Kalimantan to Madagascar almost certainly followed a period of contact with Śrīvijayan Malays in southern Sumatra, and may have involved the mediation of Malay seafarers, since the Malagasy themselves derive from an interior riverine population that would have had little experience or skill at open sea voyaging. Major recent publications on languages of Borneo include Kadazan Dusun Cultural Association (2004), Goudswaard (2005), Adelaar (2005a), Rensch, Rensch, Noeb and Ridu (2006), and Soriente (2006). Major publications on Malagasy include Abinal and Malzac (1888), Dahl (1951, 1991), and Beaujard (1998).

Language distribution

Table 2.4. lists the ten largest and ten smallest languages of Borneo. Since the figures available for Siang and Coastal Kadazan are identical, eleven languages are given in the first group. Malagasy appears separately at the end (M-C = Malayo-Chamic, BTO = Barito, SBH = Sabahan, KMM = Kayan-Murik-Modang), M-K = Melanau-Kajang, NS = North Sarawak):

Table 2.4 The ten largest and ten smallest languages of Borneo

Language	Location	Subgroup	No. of speakers
1. Banjarese	SE Kalimantan	M-C	3,502,300 (2000)
2. Ngaju Dayak	SE Kalimantan	BTO	890,000 (2003)
3. Iban	southern Sarawak	M-C	694,000 (2004)
4. Brunei Malay	Brunei	M-C	304,000 (1984)
5. Kendayan Dayak	SW Kalimantan	M-C	290,700 (2007)
6. Ma'anyan	SE Kalimantan	BTO	150,000 (2003)
7. Central Dusun	western Sabah	SBH	141,000 (1991)
8. Lawangan	SE Kalimantan	BTO	100,000 (1981)
9. Dohoi	SE Kalimantan	BTO	80,000 (1981)
10a. Coastal Kadazan	western Sabah	SBH	60,000 (1986)
10b. Siang	SE Kalimantan	BTO	60,000 (1981)
1. Punan Aput	east Kalimantan	KMM	370 (1981)
2. Lahanan	southern Sarawak	M-K	350 (1981)
3a. Punan Merap	east Kalimantan	?	200 (1981)
3b. Kanowit	central Sarawak	M-K	200 (2000)
4. Punan Merah	central Kalimantan	KMM?	140 (1981)
5. Ukit	central Sarawak	?	120 (1981)
6. Tanjong	central Sarawak	?	100 (1981)
7. Sian	central Sarawak	?	50 (2000)
8. Punan Batu	central Sarawak	?	30 (2000)
9. Lengilu	NW Kalimantan	NS	4 (2000)

As seen in Table 2.4, most large languages in Borneo are found in southeast Kalimantan, with Banjarese Malay nearly four times the size of any other language on the island. Banjarese may be exceptionally large because it is spoken in a major port city (Banjarmasin) that has served as a hub of trade with the outside world for centuries. The relatively large size of several of the Barito languages is more difficult to explain, but some of these communities have long had close commercial contacts with Banjarmasin, and it is possible that this relationship has contributed to their growth. Although Banjarese is considered a dialect of Malay, lexicostatistical percentages indicate that it is no closer to standard Malay or to various forms of Sumatran Malay than is Iban, which is commonly regarded as a distinct language. Undoubtedly, non-linguistic factors play a part in this perception, since the Banjarese are Islamic traders associated with the sea, while the Iban are interior animists associated with the cultivation of hill rice. Although it is a Malayic language it has many Javanese loanwords, probably acquired during the Majapahit period from the late thirteenth century until the fifteenth century. The question whether Banjarese was located at the mouth of the Barito River at the time of the Malagasy migration, and what (if any) role the Banjarese might have played in facilitating this migration remain unanswered, although there is evidence that both Malagasy and the Sama Bajaw languages originate from the basin of the Barito River, and that the Banjarese may be descendants of Sriwijayan traders who drew these indigenous groups out into a wider network of external contacts (Blust 2007d).

1) Banjarese
2) Ngaju Dayak
3) Iban
4) Brunei Malay
5) Kendayan Dayak
6) Ma'anyan
7) Central Dusun
8) Lawangan
9) Dohoi
10a) Coastal Kadazan
10b) Siang
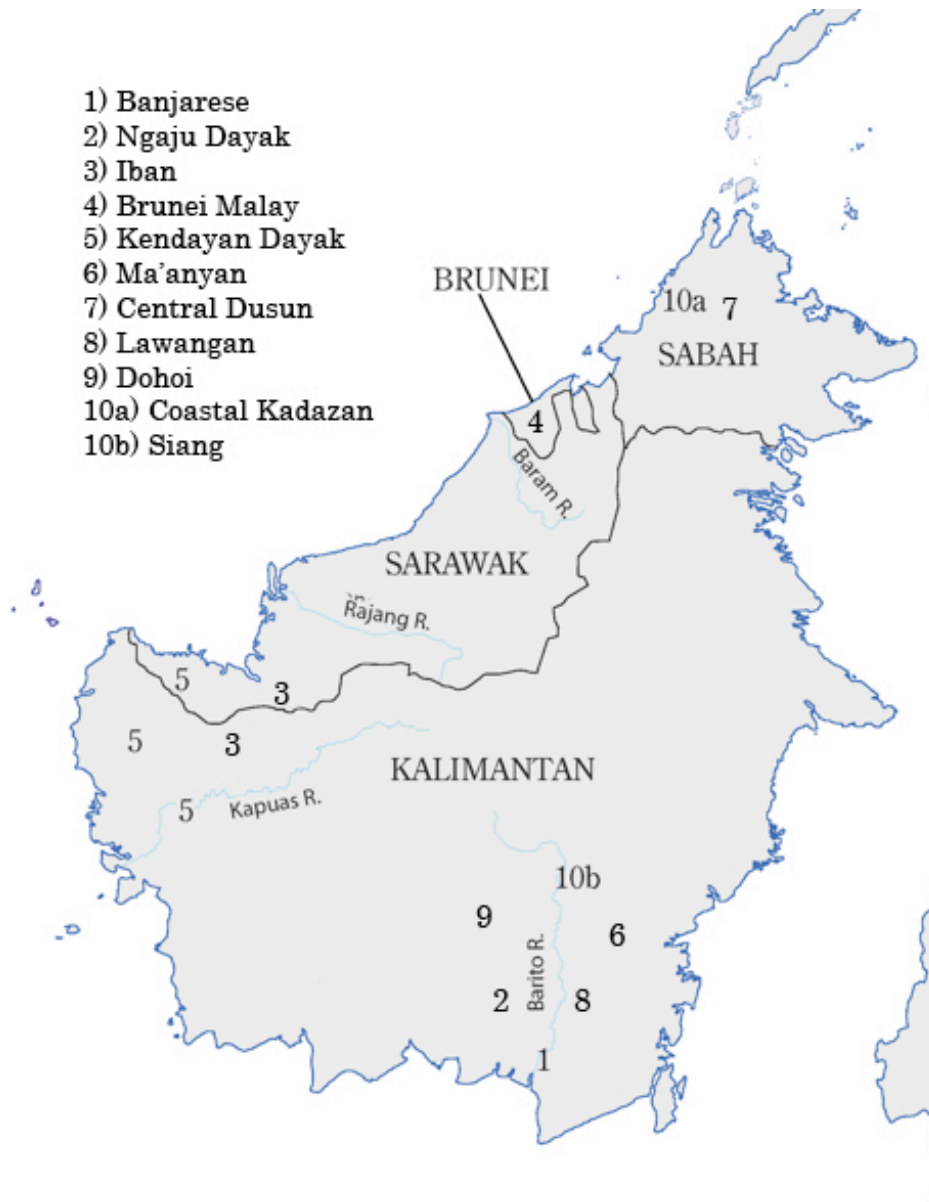
Map 4 The ten largest languages of Borneo

The relatively large number of Brunei Malay speakers also suggests special historical circumstances favoring the growth of larger populations. Brunei Malay is not native to northwest Borneo, and differs substantially from the native Dusunic and North Sarawak languages around it. As in several other parts of the Indonesian Archipelago, Malay is thus an imported language in this region. Good anchorage at the harbor of Bandar Seri Begawan clearly played a role in establishing Brunei as an early commercial center, and Śrīvijayan Malays presumably secured a foothold in this area fairly early to facilitate control of the spice trade leading from the Moluccas around the north coast of Borneo. Unlike Banjarmasin, however, which sits at the mouth of a major river leading inland, Brunei Bay has little connection with the interior. As a result, the influence of Brunei Malay on neighboring languages has historically been confined to the coastal zone of Brunei and Sarawak, with little penetration inland.

The territorial growth of Iban apparently is recent, as the Iban were largely concentrated in the upper Kapuas basin on the Sarawak-Kalimantan border until about 200 years ago.

Some groups then crossed over into Sarawak, and around the middle of the nineteenth century they began an explosive expansion into the territory of other groups in southern Sarawak, such as the now extinct Seru, and the moribund Ukit, Tanjong and Sian (Sandin 1967, Sutlive 1978). Whether this expansion was a cause or a consequence of rapid population growth is unclear, as the two cannot easily be separated.

The smallest language groups in Borneo include several nomadic 'Punan' bands, as well as several groups that may have been nomadic in the fairly recent past, such as the Ukit, Tanjong and Sian. Most of these are found in central and southern Sarawak, an area that lay in the path of the Iban expansion during the nineteenth century. The most endangered language in Borneo, according to the Ethnologue, is Lengilu, reportedly an aberrant North Sarawak language similar to Sa'ban, although no published data of any kind are available to test this claim.

Typological overview

The languages of Borneo are typologically more varied than those of the Philippines. Many languages south of Sabah have a palatal series that includes at least *j* and *ñ*. In Malayic Dayak languages such as Iban, and in Malay this series also includes a voiceless palatal affricate written *c*. A number of North Sarawak languages have unusual phonological systems. Several dialects of Kelabit, including the standard dialect of Bario, have phonemic voiced aspirates b^h, d^h, g^h, which until recently were unreported in any other language (Blust 2006a). What appears to be a similar series of stops is found in Ida'an of the eastern Sabah, although the corresponding elements in these languages are consonant clusters rather than unit phonemes (Goudswaard 2005). Other North Sarawak languages have unusual consonant alternations, as with *b* and *s* in Kiput. Word structure in the languages of Borneo differs from that of the Philippines in not allowing a wide range of medial consonant clusters and in generally disallowing the low vowel *a* in pre-penultimate syllables (where it is neutralised with schwa, or with a historical reflex of schwa).

Two features of pronominal systems that are common in Oceanic languages and in some of the languages of eastern Indonesia are also found in many of the languages of central and western Borneo, but are otherwise unknown in Taiwan, the Philippines or western Indonesia. The first of these is obligatory possessive marking on many nouns denoting body parts or kinship terms. The second is a fully developed system of marking dual and paucal numbers alongside singular and plural forms of pronouns. In some languages, such as Kenyah, a true quadral number also occurs in the pronoun system. Differences of detail suggest that these developments in Borneo are historically independent from those of AN languages further to the east.

From the standpoint of syntactic typology Borneo is a transition area leading from Philippine-type verb systems in Sabah, to morphologically reduced Western Indonesian-type verb systems as one moves southward into Sarawak and Kalimantan. Most languages of Borneo are SVO, but word order may be sensitive to voice. In Kelabit, for example, passive constructions are generally predicate-initial, while active constructions are SVO. Although some of the affixes that figure prominently in the verb systems of Formosan and Philippine languages have cognates in the languages of Sarawak or Kalimantan, they do not always function in the same way. Bario Kelabit of northern Sarawak can be used to illustrate:

Bario Kelabit (Blust 1993a)

1.   ŋudəh iko madil kuyad inəh 'Why did you shoot that monkey?'
      why **2sg.nom av**-shoot monkey that (verb base: *badil* 'gun')

2.   bədil-ən muh kənun kuyad inəh 'Why did you shoot that monkey?'
      shoot-**pv 2sg.gen** why monkey that

3.   b<ən>adil muh idan kuyad inih 'When did you shoot this monkey?'
      (shoot-**perf.pv 2sg.gen** when monkey this)

4.   ŋi iəh m-irup əbʰaʔ inəh 'He is drinking the water'
      dem 3sg.nom av-drink water that)

5.   n-irup iəh əbʰaʔ inəh 'He drank the water'
      perf.pv-drink **3sg.nom** water that

6.   rup-ən muh kənun nəh idih   'Why did you drink it?'
      drink-**pv 2sg.gen** why already it

Kelabit allows only actors and patients to be morphologically encoded in the verb, marking locative, instrumental, and benefactive relationships by prepositions. This is true even though the morphology used to encode locative, instrumental or benefactive arguments in Philippine-type languages is sometimes retained in Kelabit (and other Western-Indonesian type languages). For example, Kelabit *-an* is cognate with the similar suffix in many Philippine languages, where it is attached to both verbs and nouns, but in Kelabit it is used almost exclusively in nominalisations: *irup* 'to drink' : *rup-an* 'water hole (place where animals go to drink)', *dalan* 'path' : *nalan* 'to walk' : *dəlan-an* 'path made by walking, as through the grass', etc. In addition, a number of languages in Sarawak have systems of verbal ablaut similar to the vocalic alternations in English *sing* : *sang* : *sung*, but derived historically from reduction of the infixes *-um- and *-in- in certain stateable environments. These systems are most fully developed in languages of the Melanau dialect chain, as in Mukah *ləpək* 'a fold' : *lupək* 'to fold' : *lipək* 'was folded', *səput* 'blowpipe' : *suput* 'shoot with a blowpipe' : *siput* 'was shot with a blowpipe', *bəbəd* 'tie' : *mubəd* 'to tie' : *bibəd* 'was tied'.

Malagasy

With an area of about 587,040 square kilometres, Madagascar is the world's fourth largest island. Malagasy is included here because of the widely accepted argument of Dahl (1951) that it subgroups immediately with Ma'anyan and other languages of the Barito River basin in southeast Kalimantan. Its removal from the Bornean context and introduction into another large island without previous inhabitants, however, provided obvious opportunities for rapid population growth and dialect diversification. As a result, it is now larger than any language of Borneo, and has some twenty generally recognised dialects (Vérin, Kottak and Gorlin 1969). The standard dialect of Malagasy is Merina, spoken on the Imerina plateau of central Madagascar. Populations of major Malagasy dialect groups in 1993 include Merina (3,200,000), Betsimisaraka (1,800,000), Betsileo (1,400,000), Antandroy (635,000), Tañala (473,000), and Antaimoro (422,000). Together with a number of smaller groups these total some 9,390,000. Most Malagasy speech communities appear to participate in a chaining relationship, with cognate percentages dropping as low as 52-56% between non-adjacent communities (Sakalava with

Antambahoaka, Betsimisaraka, or Tsimihety, Antandroy with Antambahoaka, Antankarana, or Tsimihety). Under such conditions it would be possible to select random points in the continuum as sufficiently distinct to justify recognizing more than one Malagasy language. This is essentially how the Bisayan L-complex of the central Philippines and the Dusunic and Murutic L-complexes of Sabah have been treated, and the different treatment of Malagasy simply highlights the inconsistencies that characterise much work in distinguishing language from dialect both within the AN language family and between different language families.

In general typology Malagasy differs sharply from the Barito languages with which it subgroups. Phonologically it has a larger consonant inventory than is typical of most languages of western Indonesia, and it allows no final consonants, apparently a product of contact influence from the Bantu languages with which it came into contact on the coast of east Africa before reaching Madagascar.[7] Perhaps the most striking typological fact about Malagasy, however, is its close structural similarity to Philippine-type languages in Taiwan, the Philippines, northern Borneo and elsewhere. Like these languages, Malagasy is predicate-initial, and it allows a wide range of nominal arguments to be morphologically encoded on the verb as 'subjects' (in some analyses). The following sentences serve to illustrate:[8]

Malagasy (Keenan 1976)

1. mi-vidy mofo hoʔan ny ankizy **aho** 'I am buying bread for the child'
    av-buy bread for **art** child **1sg.nom**

2. i-vidi-an-ko mofo **ny ankizy** 'I am buying bread for the child'
    cv-buy-**cv**-**1sg.gen** bread **art** child

3. ma-nolotra ny vary ny vahiny **aho** 'I offer the rice to the guests'
    av-offer **art** rice **art** guest **1sg.nom**

4. a-tolo-ko ny vahiny **ny vary** 'I offer the rice to the guests'
    intv-offer-**1sg.gen art** guest **art** rice

5. tolor-an-ko ny vary **ny vahiny** 'I offer the rice to the guests'
    offer-**cv-1sg.gen art** rice **art** guest

This discrepancy between subgroup affiliation and typology is important in shedding light on the history of typological change in western Indonesian verb systems. The close agreements in verbal morphology that Malagasy shares with Formosan and Philippine languages can only be explained as a retention of features that were present in the common

---

[7] Dahl (1954) described this as substrate influence, on the assumption that Madagascar had a Bantu-speaking population that unwent language shift after the Malagasy arrived. However, as Adelaar notes (2010:165, 2012:148ff), Bantu influence on Malagasy was more likely due to contact on the Mozambique coast prior to the settlement of Madagascar, a process that probably continued after the settlement had commenced. Murdock (1959:215) was perhaps the first to suggest that the Malagasy settled the coast of East Africa west of the Mozambique Channel before populating Madagascar itself, and Blust (1994a:61ff) reached a similar conclusion based on the distribution of the outrigger canoe, which can be used to a considerable extent to trace the migration route followed by the Malagasy from southeast Borneo to East Africa.

[8] Keenan gives *i-vidi-ana-ko* in sentence 2, and *tolor-ana-ko* in sentence 5. According to K. Alexander Adelaar (p.c. 8/8/2012) the versions given here are the correct forms.

ancestor of all these languages (hence Proto Austronesian). It follows that the simplified verbal morphology of the other Barito languages, and of most other languages of western Indonesia, is a product of change that affected a very large geographical area in western Indonesia. If the Malagasy migration took place no earlier than the seventh century it would follow that many languages in southern Borneo that now have simplified verb systems similar to that of Malay had Philippine-type verb systems as recently as 1,300 years ago. Since Malagasy was geographically removed from the contact area in which this simplification took place it escaped the changes, and has retained many features from an earlier period which its closest relatives have lost.

2.4.4 The Austronesian languages of mainland Southeast Asia

AN languages are spoken in four general regions of mainland Southeast Asia: 1. the Malay Peninsula, 2. the coastal islands of peninsular Burma and Thailand, 3. interior areas of Vietnam, Laos and Cambodia, and 4. Hainan Island in southern China. This region is a geographical catch-all, consisting of islands, mainland coastal zones, and interior mountain zones. Lewis (2009) lists 21 Austronesian languages spoken in mainland Southeast Asia, although some of those in the Malay Peninsula clearly are dialects of Malay.

Brief history of research

The history of scholarship on Malay (and more recently, Bahasa Indonesia) greatly exceeds that on any other language in this section. As early as 1623 the newly-arrived Dutch compiled a Malay dictionary with grammatical remarks, and during the British colonial period in Malaya many other works on Malay were published, a trend that has continued to the present. As a result, Malay today is one of the best-studied of all AN languages. Early French work on Chamic was limited to Cham proper, with a grammar of Cham appearing in 1889, and a substantial dictionary in 1906. During the second half of the twentieth century other Chamic languages such as Jarai and Rhade began to receive scholarly attention. Published material on Moken-Moklen remains limited, although a substantial doctoral dissertation on this language was defended at the University of Hawai'i in the late 1990s (Larish 1999). Landmark publications include Aymonier and Cabaton (1906) for Cham, Wilkinson (1959), by far the most thorough dictionary of traditional Malay before extensive modernisation (Moeliono et al. 1989) for Bahasa Indonesia, and G. Thurgood (1999) for the history of Chamic.

Language distribution

Table 2.5. lists the AN languages of mainland Southeast Asia in decreasing order of size. Since the number of Moken speakers is unknown, this language is listed last. If the number were known, however, Moken probably would be roughly equivalent in size to Moklen (MAL = Malayic, CMC = Chamic; part of Malayo-Chamic):

Table 2.5 The Austronesian languages of mainland Southeast Asia

No.	Language	Subgroup	No. of speakers
1.	Standard Malay	MAL	10,296,000 (2004?)
2.	Pattani Malay	MAL	1,000,000 (2006)
3.	Negeri Sembilan Malay	MAL	507,500 (2004)
4.	Jarai	CMC	338,206 (2006)[9]
5.	Western Cham	CMC	321,020 (2006)
6.	Rhade/Rade	CMC	270,000 (1999)
7.	Eastern Cham	CMC	73,820 (2002)
8.	Northern Roglai	CMC	52,900 (2002)
9.	Southern Roglai/Rai	CMC	41,000 (1999)
10.	Haroi	CMC	35,000 (1998)
11.	Jakun	MAL	27,448 (2004?)
12.	Temuan	MAL	22,162 (2003)
13.	Duano'	MAL	19,000 (2007)[10]
14.	Chru	CMC	15,000 (1999)
15.	Baba Malay	MAL	5,000 ethnic (1979)
16.	Tsat	CMC	3,800 (1999)
17a.	Urak Lawoi'	MAL	3,000 (1984)
17b.	Cacgia Roglai	CMC	3,000 (2002)
18.	Moklen/Chau Pok	?	1,500 (1984)
19.	Orang Seletar	MAL	880 (2003)
20.	Malaccan Creole Malay	MAL	300 (2004?)
21.	Orang Kanaq	MAL	83 (2003)
22.	Moken/Selung	?	?

---

[9] Lewis (2009) lists 318,000 speakers of Jarai in Vietnam as of 1999, and 20,206 in Cambodia as of 2006. I have listed the more recent date in the table, although the total number of speakers in both countries probably exceeded the figure given here by that date.

[10] Includes speakers in Sumatra.

Map 5 The ten largest Austronesian languages of Mainland Southeast Asia

Malay is by far the most dominant language of this area. Eastern, or Phan Rang Cham, and Western Cham are dialects of a single language, but differ significantly in location, contact influences and number of speakers, and so are listed separately. Duano', Jakun, Orang Kanaq, Orang Seletar, Temuan and Urak Lawoi' should probably all be considered dialects of Malay, although at least the last of these is phonologically very divergent. Baba Malay is spoken by the Chinese immigrant community in Malaysia, and has roots going back several centuries (Pakir 1986, E. Thurgood 1998). What Lewis (2009) calls Negeri Sembilan Malay probably is a form of Minangkabau, and Pattani Malay is a northern Malay dialect that has been heavily influenced by contact with Thai (Tadmor 1995).

All AN languages of mainland Southeast Asia except Moken-Moklen belong to a single subgroup, Malayo-Chamic, which Adelaar (2005c) subsumed under 'Malayo-Sumbawan', a proposed grouping that in many ways appears to be problematic (Blust 2010). Malay communities tend to be coastal, while the deeper interior rainforest was historically the preserve of the Orang Asli (Malay for 'Original people'), who speak languages belonging to the South Mon-Khmer branch of the Austroasiatic language family. The Orang Asli themselves fall into two distinct groups: the nomadic Negrito hunter-gatherers (called 'Semang' in the older literature), and the sedentary farming southern Mongoloids (called 'Sakai' in the older literature). Both of these groups have been in contact with Malays for centuries, and some have begun to adopt substantial numbers of Malay loanwords. In addition to the Orang Asli the Malay Peninsula is home to small sedentary groups of

animists who live in interior areas and speak non-standard Malay dialects. These groups, often called 'Jakun', are generally described as 'aboriginal Malays', but may derive from earlier Orang Asli groups that have partly assimilated culturally and linguistically to the dominant Malay population.

On the east (South China Sea) side of the Malay Peninsula Malay communities extend northward along the coast and the lower reaches of the river systems until they meet southern dialects of Thai. In this region the two languages have been in contact for centuries, and some Malay dialects have acquired traits characteristic of mainland Southeast Asian languages such as lexical tone. On the west side of the Malay Peninsula Malay dialects extend much further north before they yield to southern Thai (Lebar, Hickey and Musgrave 1964, end map).

The west coast of the Malay Peninsula, especially as one moves north, has many more islands than the east coast. These islands, extending from the Mergui Archipelago in Burma, to Phuket Island in Thailand, and Langkawi Island off the coast of northern Malaya, are home to nomadic boat people. Although they are often called 'sea gypsies' like the Samalan speakers of the Philippines and Indonesia, linguistically they are very distinct from the Sama-Bajaw. Moreover, there are two different populations of sea gypsies in the islands to the west of the Malay Peninsula. In the southern zone are the Urak Lawoi' (Standard Malay 'Orang Laut'), who range between roughly Langkawi Island and the southern end of Phuket Island, and who speak a phonologically aberrant dialect of Malay. North of the Urak Lawoi', in the Mergui Archipelago, are the Moken, a group that appears to have no close linguistic ties with other AN languages. Settled Moken have come into contact with Thai, and like northern dialects of peninsular Malay have begun to acquire lexical tone. These dialects are known as Moklen to distinguish them from the less contact-influenced language of their relatives who continue to maintain a somewhat more isolated nomadic life in houseboat communities in the Mergui Archipelago. Unlike the Samalan peoples, who have historically been very active in trade and interethnic relations, the Moken are described as retreating from contact with outsiders, a behavior that has made it difficult to estimate their numbers (White 1922).

The Chamic group contains nine closely related languages in Vietnam, with one dialect of Cham (Western Cham) spoken around the Tonle Sap lake in central Cambodia. Western Cham, which is in contact with Cambodian, has acquired breathy consonants like those of many Mon-Khmer languages in Vietnam, Cambodia, and Laos. Some studies suggest that Eastern Cham, which is in contact with Vietnamese, has begun to acquire lexical tone, but this has been questioned (Brunelle 2005). The areal adaptations of the Chamic languages are so extensive that these languages were misclassified by Schmidt (1906) as 'Austroasiatic mixed-languages', a confusion that persisted in some quarters until at least 1942. Finally, a single Chamic language, Tsat, which was first recognised as Chamic by Benedict (1941), is spoken on Hainan Island in southern China.

Typological overview

Typologically, the languages of mainland Southeast Asia can be said to fall into two main groups: those that are relatively free from contact-induced structural influences, and those that show evidence of areal adaptation to their non-AN neighbors. The first category includes Moken and most dialects of Malay that are not in contact with Thai. The second category includes Moklen, northern dialects of Malay which have begun to acquire monosyllabism and tone as a result of prolonged contact with southern Thai, and the Chamic languages as a group. While the contact influences that have affected Moklen

appear to be relatively recent, and those that have affected Pattani Malay probably no more than 500-600 years old, the Chamic languages have had a much longer history of adaptation to mainland Southeast Asian linguistic structures. Thurgood (1999) provides a detailed documentation of the areal adaptations of Chamic languages to their Mon-Khmer neighbors over a 2,000 year period. As a result, Chamic languages are now typologically more similar to Mon-Khmer languages than to most languages of western Indonesia. Acquired traits include preglottalised and 'aspirated' stops (probably clusters of voiceless stop + *h*), breathy voice following voiced obstruents, a sesquisyllabic word shape (Cə)CVC, many word-initial consonant clusters, including stops or nasals followed by the liquids *l* or *r,* an expanded inventory of vowels, contrastive vowel length, and the loss of almost all affixes except *pə-* 'causative' (which could be of Mon-Khmer origin), and a rare verbal prefix *mə-*. In addition, Eastern Cham has begun to acquire lexical tone as a result of contact with Vietnamese, and Tsat of Hainan Island has developed five contrasting tones as a result of roughly a millennium of contact with Tai-Kadai and southern Chinese languages. Some of these contact-induced developments can be seen in comparing the lexicons of Malay and Jarai in Table 2.6.

Standard Malay, which derives from the Riau-Johore dialect at the southern tip of the Malay Peninsula, has no typologically unusual features, but some non-standard dialects of Malay depart from this basic groundplan. Kedah Malay, in northwest Malaya along the Thai border, has word-final uvular stops corresponding to uvular *r* in standard Malay, and these affect the quality of preceding high vowels. In some subdialects of Trengganu Malay, spoken in northeast Malaya near the Thai border, original final high vowels have become diphthongised and the diphthongal coda has become an obstruent *k, kx* or *h*, as in Kampung Peneh *kakɨy^{kx}* (Standard Malay *kaki*) 'foot, leg', or Kampung Peneh *tuw^{kx}* (SM *itu*) 'that' (Collins 1983b). As noted earlier, in Urak Lawoi' divergent phonological developments have reduced intelligibility enough to justify calling this a separate language, although probably more than 90 percent of its basic vocabulary is cognate with that of Standard Malay.

Table 2.6 Malay-Jarai cognates, showing areal adaptations in Jarai

PMP	Malay	Jarai	English
*mata	mata	məta	eye
*batu	batu	pətəw	stone
*puluq	puluh	pluh	unit of ten
*bulan	bulan	blaan	moon, month
*m-alem	malam	mlam	night
*qudaŋ	udaŋ	hədaaŋ	shrimp
*qulun	(oraŋ)	hlun	person, human being
*duRi	duri	drəy	thorn
*zaRum	jarum	jrum	needle
*epat	əmpat	paʔ	four
*paqet	pahat	phaʔ	chisel
*buhek	(rambut)	ɓok	head hair

The word-order typology of Malay and the Chamic languages is SVO. Although Malay verb morphology is moderately complex it does not approach the complexity of the verb systems of Philippine-type languages. The Chamic languages, which probably developed from a type very similar to Malay, use prepositions to express many of the kinds of

grammatical relationships encoded by verb morphology in Malay and other languages of western Indonesia. The following data serve to illustrate:

Malay/Bahasa Indonesia (composite sources)

1. si Ahmad pandai bə-rənaŋ 'Ahmad is good at swimming'
    pm Ahmad clever **intr-**swim

2. mǝreka bǝr-lihat-lihat 'They looked at each other'
    3pl intr-see.**recip**

3. saya mǝ-lihat tikus di-makan kuciŋ 'I saw a mouse being eaten by a cat'
    1sg av-see mouse **pass-**eat cat

4. anak yaŋ mǝ-ñǝmbuñi-kan diri itu masih tǝr-lihat
    child **rel act-**hide-**tr** self that still **ia-**see 'That child who is hiding is still visible'

5. pǝtani daerah ini umum-ña bǝr-tanam ubi kayu
    farmer area this usually **intr-**plant cassava 'Farmers in this area usually plant cassava'

6. pǝtani mǝ-nanam-kan padi di-ladaŋ-ña
    farmer **av-**plant-**tr** rice **loc-**field-**3sg.poss** 'The farmer planted rice in his fields'

7. pǝtani mǝ-nanam-i ladaŋ-ña dǝŋan padi 'The farmer planted his fields with rice'
    farmer **av-**plant-**ltr** field-**3sg.poss** with rice

Jarai (Blust n.d. f)

1. cǝdeh huiʔ 'The child is afraid'
    child fear

2. asǝw pǝ-huiʔ kǝ cǝdeh 'A dog frightened the child'
    dog **caus-**fear to child

3. dapanay tap pǝday hoŋ hǝlǝw 'The woman pounded the rice with a pestle'
    woman pound rice with pestle

4. cǝday ji hraʔ hoŋ gayji 'The boy wrote a letter with a pen'
    boy write letter with pen

5. asǝw hǝmǝw diay 'The dog died/has died'
    dog already die

6. mǝnuih pǝ-diay asǝw 'A man killed the dog'
    man **caus-**die dog

7. ñu hǝmǝw naw rǝgaw sa blaan laeh 'He went away a month ago'
    3sg already leave **past** one month **comp**

2.4.5 The languages of Sumatra, Java, Madura, Bali and Lombok

Sumatra is the sixth largest island on Earth, with a total area of about 473,481 square kilometres. It extends some 1,600 kilometres from northwest to southeast, and is divided along much of its length by the Barisan Mountains, which approach very close to the west coast in the southern half of the island. In the north the most prominent feature is Lake

Toba, a large body of water in the Batak highlands enclosing the island of Samosir. West of Sumatra is the chain of Barrier Islands, extending from Simeulue in the north to Enggano in the south. The languages and cultures of the Barrier Islands are characterised by extreme diversity.

Java is smaller than Sumatra, at about 126,700 square kilometres. It contains 32 volcanoes. At least in part because of its fertility Java has acquired one of the highest population densities on earth; in 1990 it was home to over 100 million people living in an area less than that of New York State. Bali, a much smaller, though equally volcanic island at 5,600 square kilometres, is geologically an extension of Java, being separated from it by a strait only a mile or two in width. Lombok lies just east of Bali, but differs radically in its plant and animal life, as the Wallace Line is perhaps most sharply defined at this point. Finally, the island of Madura, which is only slightly smaller than Bali, lies off the north coast of east Java. It contrasts sharply with the main island in being flat, dry and dusty. Lewis (2009) lists 38 languages for Sumatra, and 10 AN languages for Java-Bali. Together with Sasak of Lombok and Sumbawanese of western Sumbawa, this comes to 50 languages for the region defined here.

Brief history of research

The first important researcher on these languages was H.N. van der Tuuk, who produced one of the most original grammars prior to the rise of modern linguistics. Van der Tuuk's Toba-Batak Grammar (*Tobasche Spraakkunst*), published in two volumes in 1864 and 1867, attempted to describe Toba Batak free from the Eurocentric preconceptions that plagued most descriptions of non-Western languages at that time. His grammar remains today a classic that can be mined for much useful information. In addition to his Toba Batak grammar and pioneering work on sound correspondences between the languages of western Indonesia and the Philippines, in 1897 van der Tuuk began the publication of a gigantic semi-comparative project, the 'Kawi-Balineesch woordenboek', which attempted to bring the resources of historical linguistics to bear on the decipherment of Old Javanese (Kawi) texts. Although the Kawi-Balinese dictionary was finished shortly before his death, its idiosyncratic organisation made it difficult to use, since it was intended primarily as a tool for the interpretation of Old Javanese manuscripts. The appearance of Zoetmulder (1982) has at last put these problems to rest.

A considerable amount of published literature exists on modern Javanese, and somewhat less on Sundanese, Madurese, Balinese and Sasak. In addition, substantial works have appeared on Gayō and the Batak languages of northern Sumatra, and shorter works on some of the other languages of Sumatra. Most of the earlier linguistic literature on this area is in Dutch, although a smaller amount on the languages of Sumatra is in German. Landmark publications include van der Tuuk (1864-67) and (1897-1912), Snouck Hurgronje (1900), Hazeu (1907), Djajadiningrat (1934), Zoetmulder (1982), and the pioneering Javanese dialect atlases of Nothofer (1980, 1981).

Language distribution

Table 2.7. lists the languages of Sumatra, Java, Madura, Bali and Lombok (M-C = Malayo-Chamic, BI-B = Barrier Islands-Batak (Nothofer 1986), LPG = Lampungic):

Table 2.7 The ten largest and ten smallest languages from Sumatra to Lombok

No.	Language	Subgroup	No. of speakers
1.	Javanese	?	90,000,000 (2004)
2.	Indonesian	M-C	22,800,000 (2000)
3.	Sundanese	?	34,000,000 (2000)
4.	Madurese	M-C?	13,600,900 (2000)
5.	Minangkabau	M-C	5,530,000 (2007)
6.	Betawi/Jakarta Malay	M-C	5,000,000 (2000)
7.	Balinese	?	3,330,000 (2000)
8.	Acehnese	M-C	3,500,000 (2000)
9.	Sasak	?	2,600,000 (2000)
10.	Toba Batak	BI-B	2,000,000 (1991)
1a.	Lubu	BI-B	30,000 (1981)
1b.	Pekal	M-C	30,000 (2000)
1c.	Simeulue	BI-B	30,000 (?)
2.	Krui	LPG	25,000 (1985)
3b.	Penesak	?	20,000 (1989)
3c.	Sichule	BI-B	20,000 (??)
4.	Kubu	M-C	10,000 (1989)
5	Nasal	?	6,000 (2008)
6.	Enggano	?	1,500 (2000)
7.	Lom/Bangka Malay	M-C	900 (1981)
8.	Loncong	M-C	420 (2000)
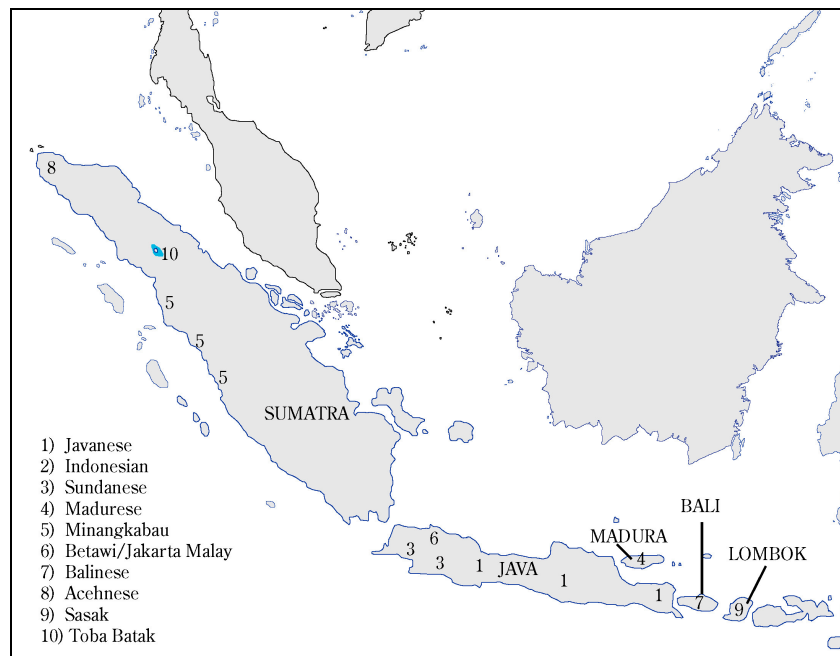
Map 6 The ten largest languages of Sumatra, Java and Bali

The distribution of languages in Sumatra shows an interesting pattern. Although AN speakers probably reached the island from Borneo, the area of greatest linguistic diversity

is the Barrier Islands, and to a lesser extent northern Sumatra, where Acehnese, Gayō and the Batak languages represent three quite distinct groups. By contrast, in most of southeast Sumatra, where the greatest linguistic diversity would be expected (as a reflection of settlement time) we find a network of Malay dialects and closely related Malayic languages such as Minangkabau and Kerinci. This distribution suggests that the linguistic history of Sumatra has not been one of continuous undisturbed differentiation, but probably included a major episode of language leveling in which incoming Malayic speakers replaced earlier languages and so 'reset the clock' of linguistic evolution.

Without question, the most salient general fact about the languages of this area is their size. Indonesian is the national language of the Republic of Indonesia, a nation with over 235 million people in 2003, and so owes its size at least in part to second-language learners. This is not true of Javanese, Sundanese, or Madurese which, extrapolating from the 1989 census data, would have numbered approximately 90 million, 33 million, and 16.5 million speakers in 2003. If we discount Indonesian because of the difficulty of distinguishing first-language speakers from second-language speakers, Javanese is nearly three times the size of the next largest AN language, Sundanese. Moreover, Javanese and Sundanese are spoken in contiguous areas on the island of Java. Of the roughly 360 million speakers of more than 1,000 AN languages, then, nearly two-thirds live in the Republic of Indonesia, about one-third live on the island of Java, and one quarter speak Javanese as a first language. Apart from Javanese only Mandarin Chinese occupies such a numerically dominant position in any language family, and this is in part because (unlike Javanese) it is a national language.

Typological overview

The languages in this area show a wide range of typological variation, making generalisation difficult. Acehnese has features typical of the Mon-Khmer languages of mainland Southeast Asia, a result of its former status as part of the early Chamic dialect continuum on the coast of Vietnam. Unlike most languages of western Indonesia, which have either the four vowels *a, ə, i, u*, or these plus the mid vowels *e* and *o*, Acehnese has at least ten contrasting vowels, and as many distinct diphthongs, as well as voiceless aspirated stops and murmured voiced stops. The Batak languages are predicate-initial, and have verb systems reminiscent of Philippine-type languages, although they differ from these in many details. Nias has typologically unusual prenasalised bilabial and alveolar trills of a kind reported elsewhere in the AN family only from Oceanic languages (Catford 1988).

Javanese differs from most other languages of western Indonesia in contrasting dental and retroflex stops, and in the feature of breathy voice or murmur as a phonetic property of its voiced obstruents. Balinese also has some retroflex consonants, but these almost certainly have been acquired from Javanese. Javanese also differs from most languages of the Philippines and western Indonesia in allowing a number of word-initial consonant clusters. Several of the languages of this area, including Karo and Toba Batak, permit medial preconsonantal liquids, and Nias and Enggano are the only AN languages anywhere in western Indonesia or the Philippines which allows no word-final consonants.

Although the Batak languages of northern Sumatra are predicate-initial and have some resemblance to Philippine-type languages, most languages of this area are SVO and have verb systems that can be characterised as Malay-like. That an SVO typology is probably innovative in western Indonesia is suggested by differences between modern Javanese (SVO) and Old Javanese from roughly 600-1,200 years ago, which was predicate-initial

with a verbal morphosyntax more like that of Philippine-type languages than is true of its modern descendant. The following data from Toba Batak and modern Javanese illustrate the range of variation (circumfixes are glossed on both sides of the base they bracket):

Toba Batak (Nababan 1981, following the orthography of Warneck 1977)

1.   halak na hatop mar-dalan   'a man who walks fast'
     man **lig** fast **intr**-walk

2.   hipas jala mokmok do ibana   'He is healthy and robust'
     healthy and robust **aff 3sg**

3.   taŋis do ibana   'She is crying'
     cry **aff 3sg**

4.   halak na sabar do parawat   'A nurse is a patient person'
     person **lig** patient **aff** nurse

5.   iboto-ŋku do si Tio   'Tio is my sister (male speaking)'
     female.cross.sibling-**1sg.gen aff pm** Tio

6.   modom do ibana di bilut   'He is sleeping in the room'
     av-sleep **aff 3sg** in room   (verb base: podom)

7.   di-garar ibana do utaŋ-ŋa i tu nasida   'He paid his debt to them'
     pv-pay **3sg aff** debt-**3sg.gen** that to them

8.   mar-hua ibana di-si   'What is he doing there?'
     what **3sg** at-there

Javanese (Horne 1974, with *tj* replaced by *c* and *dj* by *j*)

1.   alun-alun di-rəŋga-rəŋga gəndera   'The square was decorated with flags'
     public.square **pv**-decorate.**mult** flag

2.   kuciŋ-ku məntas m-anak   'My/our cat has just given birth'
     cat-**1pr.gen** just.now **av**-child

3.   ana siŋ manḍuŋ bəras, uga ana siŋ kuraŋ
     exist which enough rice, also exist which lacking   'Some have plenty of rice and some do not have enough'

4.   gorokan-ku kə-sərət-an salak   'I got some *salak* fruit stuck in my throat'
     throat-**1pr.gen ap**-wedged-**ap** salak

5.   kaṭok-e kə-cənḍak-ən; tuluŋ di-dawak-ake
     pants-**3pr ap**-short-**ap** help **pv**-make.longer-**caus**   'The/his trousers are too short; please lengthen them'

6.   piriŋ-e di-undur-undur-ake kana, banjur di-asah pisan
     plate-**3pr pv**-withdraw.**mult-ben** there, and.then **pv**-wash once   'Clear the table and wash the dishes'

7.   kancil k-onaŋ-an maŋan timun-e Pak Tani
     mousedeer **ap**-caught-**ap av**-eat cucumber-**3pr resp** farmer   'Mousedeer got caught eating Farmer's cucumbers' (verb base for 'eat': paŋan)

2.4.6 The languages of Sulawesi

Sulawesi, known earlier as 'the Celebes', consists of four long peninsulas radiating out from a central mountain massif, the result of inundations which flooded the lower-lying valleys between radiating mountain chains during the glacial meltdown at the beginning of the Holocene period. Sometimes called 'the orchid of the Equator' from its peculiar shape, it is the eleventh largest island on Earth, with a total area of about 189,000 square kilometres. For its size Sulawesi has an extremely long coastline, although most ethnic groups on the island live in inland areas and were traditionally known by the generic label 'Toraja' (< PMP *tau 'person, human being' + *daya 'toward the interior'). Lewis (2009) lists 114 languages for Sulawesi.

Brief history of research

Apart from the labors of Matthes (1858, 1859, 1874, 1875) on Makasarese and Buginese, little work was done on the languages of Sulawesi until near the end of the nineteenth century, when the Dutch Indonesianist Nikolaas Adriani wrote a grammar of Sangir which was in some ways inspired by van der Tuuk's earlier grammar of Toba Batak. Adriani continued to make valuable contributions to the study of these languages for about four decades. In more recent years many more researchers have become involved with these languages, as in the important descriptive and comparative studies of South Sulawesi languages by Mills (1975) and Sirk (1983, 1988), of Minahasan and Sangiric languages by Sneddon (1978, 1984), of Tomini-Tolitoli languages by Himmelmann (2001) and Quick (2007), of Bungku-Tolaki languages by Mead (1998, 1999), and of the languages of Buton, Muna, and the Tukang Besi Islands by Anceaux (1952, 1987), van den Berg (1989, 1996a), and Donohue (1999). Landmark publications include Adriani (1893, 1928), and the previously mentioned works.

Language distribution

Table 2.8. lists the ten largest and ten smallest languages of Sulawesi (SSul = South Sulawesi, GOR = Gorontalic, B-T = Bungku-Tolaki, K-P = Kaili-Pamona, M-B = Muna-Buton, SAN = Sangiric, T-T = Tomini-Tolitoli):

Table 2.8 The ten largest and ten smallest languages of Sulawesi[11]

No.	Language	Subgroup	No. of speakers
1.	Buginese	SSul	3,310,000 (2000)
2.	Makasarese	SSul	1,600,000 (1989)
3.	Gorontalo	GOR	900,000 (1989)
4.	Sa'dan Toraja	SSul	500,000 (1990)
5.	Tolaki	B-T	281,000 (1991)
6a.	Tae'	K-P	250,000 (1992)
6b.	Mandar	SSul	250,000 (2003)
7.	Ledo Kaili	K-P	233,500 (1979)
8.	Muna	M-B	227,000 (1989)
9a.	Sangir	SAN	200,000 (1995)
9b.	Mongondow	GOR	200,000 (Lobel p.c., 2008)

---

[11]   Jason Lobel (p.c., August 3, 2007) notes that Ponosakan probably is the smallest language in Sulawesi, with only three remaining competent speakers.

No.	Language	Subgroup	No. of speakers
1a.	Kalao	M-B	500 (1988)
1b.	Koroni	B-T	500 (1991)
1c.	Taloki	B-T	500 (1995)
2.	Talondo'	SSul	400 (2004)
3a.	Taje	T-T	350 (2001)
3b.	Waru	B-T	350 (1991)
4a.	Baras	K-P	250 (1987)
4b.	Kumbewaha	B-T	250 (1993)
5.	Bahonsuai	B-T	200 (1991)
6.	Dampal	T-T	100 (1990)
7	Liabuku	M-B	75 (2004)
8.	Budong-Budong	SSul	70 (1988)

1) Buginese
2) Makasarese
3) Gorontalo
4) Sa'dan Toraja
5) Tolaki
6a) Tae'/Southern Toraja
6b) Mandar
7) Ledo Kaili
8) Muna
9a) Sangir
9b) Mongondow
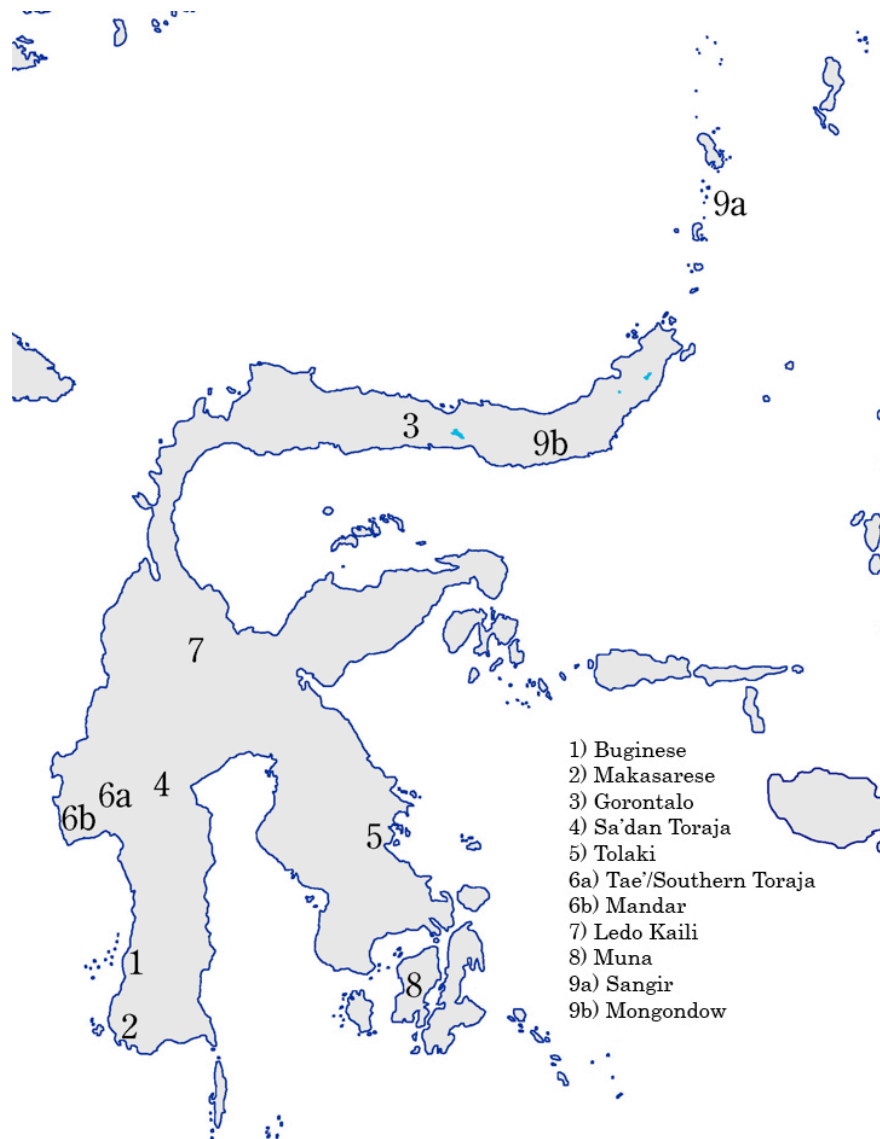
Map 7 The ten largest languages of Sulawesi

Eleven linguistic 'microgroups' are recognised in Sulawesi: 1. Sangiric in the Sangir and Talaud Islands of the far north, with additional toeholds on the northern tip of the Sulawesi mainland and on Mindanao in the southern Philippines, 2. Minahasan, around and just south of the city of Manado on the northern peninsula, 3. Gorontalic, dominating most of the east-west extent of the northern peninsula, 4. Tomini-Tolitoli, spread around the curve of the northern peninsula, 5. Saluan, occupying most of the eastern peninsula of Sulawesi, some of the Togian Islands in the Gulf of Tomini, and the Banggai Islands south of the tip of the eastern peninsula, 6. Kaili-Pamona, in the central mountain massif together with adjoining portions of the northern and eastern peninsulas, 7. Bungku-Tolaki, distributed over virtually the entire southeastern peninsula, 8. Muna-Buton, on Muna, Buton, and the southern tip of Selayar Island, 9. Wotu-Wolio, consisting of Wolio of Buton Island, and Wotu, a tiny enclave at the head of the Gulf of Bone, 10. Tukang Besi, spoken in the Tukang Besi Islands, stretching southeast of the larger island of Buton, and 11. South Sulawesi, occupying the entire southwestern peninsula of Sulawesi, and the northern two-thirds of Selayar Island. The first two collections of languages are primary branches of the Philippine subgroup, while Gorontalic is part of Greater Central Philippines, a branch that reaches from Tagalog in the north to Gorontalo and Mongondow in the south. The position of the Tomini-Tolitoli languages remains problematic, but there is some evidence for a 'Celebic supergroup' that contains groups 4-10, and excludes the South Sulawesi languages (Mead 2003b). The implications that this classification has for the AN settlement of Sulawesi remain largely unexplored.

Historically the two most dominant ethnic groups in Sulawesi have been the Islamic Buginese and Makasarese of the southwestern peninsula. Both of these groups have had wide-ranging trade relationships throughout the Indonesian Archipelago for centuries. The English naturalist Alfred Russel Wallace reported the presence of Bugis (Buginese) trading vessels in the Seram-Laut Archipelago of eastern Indonesia in the 1860s, and for some time the Makasarese Kingdom of Goa controlled the Moluccan spice trade. Under the general name 'Makasans' one or both of these groups was also long involved in the trepang (sea cucumber) trade for the Chinese market, sailing to the Aru Islands of eastern Indonesia and back with the seasonal monsoon winds, and even collecting on the coast of Arnhem Land in northern Australia, where they had sporadic contacts with the local population, and left some loanwords in the aboriginal languages of the area (Walker and Zorc 1981). In addition to the Buginese and Makasarese, highly mobile Bajau communities are scattered along the coastal regions of central and southern Sulawesi, where they have traditionally engaged in trade with the local populations.

Typological overview

Except for occasional retroflex laterals, the phoneme inventories of Sulawesian languages are generally unremarkable. A number of languages have transformed the PAN four-vowel system (*i, *u, *a, *e) to a symmetrical five-vowel system with matching front and back vowels at high and mid positions, a trait shared with many of the languages of eastern Indonesia and the Pacific.

While their phoneme inventories are not especially noteworthy, the languages of Sulawesi differ from most languages of western Indonesia and the Philippines in certain features of morpheme structure. The phonotactic trait that has attracted most attention is the tendency to reduce the number of consonant contrasts word-finally, or to lose final consonants entirely (Sneddon 1993). This development has affected all microgroups except 2-4, and contrasts strikingly with the historical development of most other languages of

western Indonesia or the Philippines, where final consonants are usually well-preserved. Although most languages that show advanced attrition of final consonants are in the 'Celebic supergroup' it is clear that final consonant attrition in the languages of Sulawesi has been a recurrent process. First, weakening of finals has also taken place in the Sangiric and South Sulawesi languages, which are not part of the 'Celebic supergroup'. Second, reconstruction within Celebic groups (as Bungku-Tolaki) shows that many final consonants were present in lower-level proto languages, but were lost in the separate histories of the modern languages (Sneddon 1993, Mead 1998). A second noteworthy feature of morpheme structure found in several parts of the island is the presence of prenasalised initial stops, often corresponding to simple stops in cognate words elsewhere, as in Bare'e (Pamona) *mbaju* (< *bayu) 'pound rice', *mbawu* (< *babuy) 'pig', *ndundu* (< *duRduR) 'thunder', or Muna *mbali* (< *baliw) 'side, half (of things coming in pairs)', *ndawu* (< *dabuq) 'fall'.

Syntactically the languages of Sulawesi can be divided into two groups: those of the northern peninsula, which have Philippine-type verb systems, and those further south that have altered an earlier Philippine-type system while still retaining a rich system of verbal affixation. The following sample sentences from Tondano in the north and Muna in the southeast can be used to give a general impression of the range of typological variation in Sulawesi verb systems (Sneddon 1970:35, fn. 4 calls these 'voice affixes', but labels them 'actor focus', 'patient focus' and the like):

Tondano (Sneddon 1970)
'The man will pull the cart with the rope to the market' (in each of the four voices):

1.   si tuama k<um>eoŋ roda wo ntali waki pasar
      nom man pull-**af** cart with rope to market

2.   roda keoŋ-ən ni tuama wo ntali waki pasar
      cart pull-**pf gen** man with rope to market

3.   tali i-keoŋ ni tuama roda waki pasar
      rope **if**-pull **gen** man cart to market

4.   pasar keoŋ-an ni tuama roda wo ntali
      market pull-**rf gen** man cart with rope

Alternatively, one might gloss these sentences respectively: 1. 'The man will pull the cart with the rope to the market', 2. 'The cart will be pulled by the man to the market with a rope', 3. 'The rope will be used by the man to pull the cart to the market', and 4. 'The market is the place to which the man will pull the cart with a rope'. Philippine-type verb systems show a strong statistical association with predicate-initial word order. Tondano deviates from this, since it allows four morphologically-marked voices, but requires the focused noun phrase to be initial. Muna is representative of many languages in Southeast Sulawesi which have a rather different type of verb system, one that van den Berg (1996b) characterises as based on a system of 'conjugation' (si-X-ha = 'together, collectively'):

Muna (van den Berg 1989)

1.   paka-mate-no no-bhari kahanda 'When he had just died there were many ghosts'
(**first**-death-his **3sg.real**-many ghost)

2.   no-ala-mo kapulu-no maka no-lobhi wughu-no
(**3sg.real**-take-**perf** machete-his then **3sg.real**-hit neck-his) 'He took his machete and hit him on the neck'

3.   naewine da-k<um>ala tora we kaghotia 'Tomorrow we will go to the beach again'
tomorrow **1pl.incl-act**-go again **loc** beach

4.   ne tatu naando se-gulu ghule 'There is a snake over there'
loc that be one-**clas** snake

5.   naewine da-si-kala-ha dae-kabua we tehi
tomorrow **1pl.incl-si**-go-**ha 1pl.incl**-fish **loc** sea 'Tomorrow we will go fishing together in the sea'

6.   kenta ka-ghawa-no sadhia mina na-bhari-a 'He never caught many fish'
fish **nom**-get-his always not **3sg**-many-**cl**

7.   a-leni-fi simbi-ku mo-ndawu-no 'I am swimming to find the bracelet I dropped'
1sg.real-swim-**tr** bracelet-my **act**-fall-**act.part**

Languages such as Muna differ from Philippine-type languages in lacking the correlation of verbal affixes with the marking of particular nominal arguments, and in allowing numerous cliticised pronouns, prepositions and the like to be built into single phonological words. Whereas the richly developed agglutinative morphology of Philippine-type languages was reduced and simplified in most of western Indonesia, then, (the Batak languages and Malagasy being notable exceptions), in Sulawesi there was a general tendency to replace this agglutinative heritage with a quasi-polysynthetic system of perhaps equal complexity.

2.4.7 The Austronesian languages of the Lesser Sundas east of Lombok

The Lesser Sunda chain extends eastward from Bali through Timor, where it overlaps with the southern Moluccas (the entire island of Wetar, in the southern Moluccas, lies west of the easternmost tip of Timor). In the administrative divisions used by the Indonesian government Bali, Lombok and Sumbawa are assigned to Nusa Tenggara Barat (Western Lesser Sundas), while Flores, the Solor Archipelago, Alor, Sawu, Roti and Timor are assigned to Nusa Tenggara Timur (Eastern Lesser Sundas). The presentation chosen here separates Sumbawa into western and eastern halves, since the major linguistic division separating WMP from CMP languages runs through this island. All languages listed in Table 2.9 except Kupang Malay are thus members of the Central Malayo-Polynesian branch of Austronesian. Although Timor is politically divided beween the Republic of Indonesia and Timor-Leste, this division has no linguistic significance, and will be ignored. Lewis (2009) lists 68 languages in Nusa Tenggara, of which 49 are AN; two of these (Sasak and Sumbawanese) are treated in section 2.4.5., leaving 47 in the Lesser Sunda Islands considered here.

Brief history of research

The Lesser Sundas differ from western Indonesia in being home to both AN and non-AN languages. The latter are a distinct minority, although they outnumber the AN languages on the islands of Alor and Pantar, and form important enclaves in central and eastern Timor (Bunak, Makasae, Fataluku). All non-AN languages of the Lesser Sundas have been assigned to the Trans-New Guinea Phylum of Papuan languages, and evidently represent a prehistoric back-migration from the western end of New Guinea to the islands of Pantar, Alor and adjacent parts of Timor.

Few languages of the Lesser Sundas east of Lombok have been well-described. Among early Dutch researchers in this area J.C.G. Jonker can be singled out as the most important. During the period 1893-1896 Jonker produced grammatical descriptions, texts and a dictionary for Bimanese of eastern Sumbawa, and then moved on to his lifework—the study of the language of the small island of Roti. From 1905 to 1915 Jonker published a number of Rotinese texts, an 806 page Rotinese-Dutch dictionary, and a Rotinese grammar (Jonker 1908, 1915). He followed this some years later with an extensive collection of texts from Kambera, spoken in eastern Sumba.

Apart from the work of Jonker, which was distinguished more by its industry than its linguistic insight, the most detailed descriptions available to date have been done by Dutch and German missionaries who devoted their lives to one or another ethnic group. The languages of western Flores have been particularly well served by missionary-linguists. Most notable among these are the grammar of Manggarai by Burger (1946), the 1,041 page Manggarai-Indonesian dictionary of Verheijen (1967-1970), and the 646 page Ngadha-German dictionary of Arndt (1961). Other major works on the languages of this area include the 628 page Kambera-Dutch dictionary of Onvlee (1984), the 646 page Lamaholot-Indonesian-German dictionary (with associated texts) by Pampus (1999), and most recently the Kambera grammar of Klamer (1998), and several Tetun grammars, of which the most important is that of van Klinken (1999). Despite these grammars, and some publications by Geoffrey Hull, the languages of Timor have received relatively little attention. This is surprising, since Atoni and Tetun are among the largest languages in the Lesser Sundas, yet there is neither a dictionary nor a grammar of Atoni, and the most extensive dictionaries of Tetun (Morris 1984, Hull 2002, Williams-van Klinken 2011) are a fraction the size of Verheijen's dictionary of Manggarai.

Language distribution

Table 2.9. lists the ten largest and ten smallest AN languages of the Lesser Sunda Islands east of Lombok (WT = West Timor, WF = West Flores, CT = Central Timor, M-C = Malayo-Chamic, EF = East Flores, S-H = Sumba-Hawu). The exceptionally large number of round figures and ties in these estimates is clearly a reflection of how poorly the cultural demographics of this area have been recorded:

Table 2.9 The ten largest and ten smallest AN languages of the Lesser Sundas

No.	Language	Subgroup	No. of speakers
1.	Atoni	WT	586,000 (1997)
2a.	Bimanese	?	500,000 (1989)
2b.	Manggarai	WF	500,000 (1989)
3.	Tetun	CT	400,000 (2004)
4a.	Kupang Malay	M-C	300,000 (2000)
4b.	Sumbawanese	?	300,000 (1989)
5.	SW Lamaholot	EF	289,000 (2000)
6	Kambera	S-H	234,574 (2000)
7.	Sika	EF	175,000 (1995)
8.	Lamaholot	EF	150,000 (1997)
1a.	Laura	S-H	10,000 (1997)
1b.	Palu'e	WF?	10,000 (1997)
1c.	So'a	?	10,000 (1994)
1d.	Wanukaka	S-H	10,000 (1981)
2a.	Dhao/Ndao	S-H	5,000 (1997)
2b.	Ngadha, Eastern	WF	5,000 (1994)
3a.	Rajong		4,240 (2000)
3b.	Wae Rana		4,240 (2000)
4a.	Rembong	WF	2,140 (2000)
4b.	Rongga	WF	2,140 (2000)
5.	Komodo	?	700 (2000)
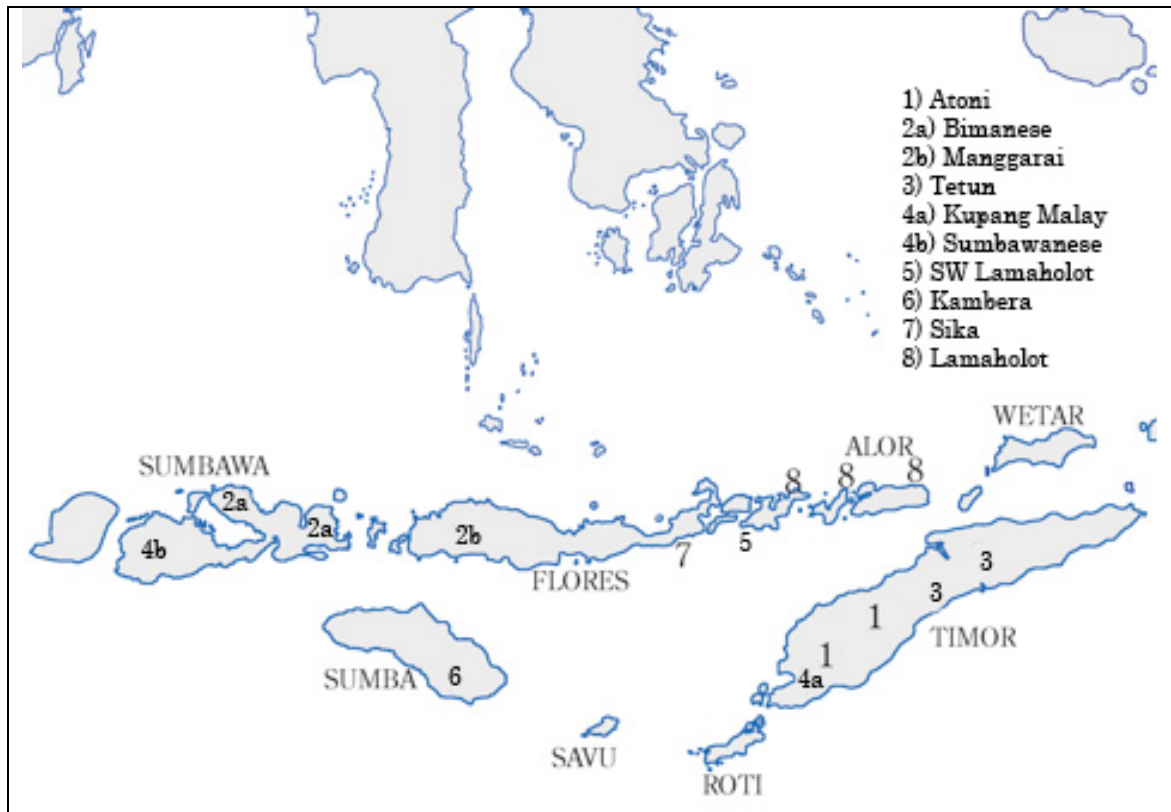
Map 8 The ten largest languages of the Lesser Sundas islands

As seen in Table 2.9, two of the four largest languages in the Lesser Sundas are spoken in western and central Timor. To some extent this may reflect the relatively larger landmass that Timor offers for languages to occupy. The two other largest languages in this area are spoken in eastern Sumbawa and western Flores, which face one another across the Sape Strait over the intervening islands of Rinca and Komodo. The smallest language in the Lesser Sundas appears to be Komodo, spoken on the tiny island of the same name, which is perhaps better known for its unique species of monitor lizard (the 'Komodo dragon') than for its human inhabitants.

Typological overview

Descriptive work on the languages of the Lesser Sundas remains spotty, and an overview of typological characteristics must begin with an admission that the picture may change substantially as better data become available.

In general the vowel systems of Lesser Sunda languages are richer than those of languages in Taiwan, the Philippines, or most of western Indonesia. Many languages lack the schwa, but have developed a five vowel system *a e o i u*; in others this has further evolved into a seven-vowel system with tense and lax vowels at both mid-front and mid-back positions. Some of the languages of Timor are typologically unusual in having non-laryngeal consonants which are co-articulated with a laryngeal gesture. Atoni of west Timor, for example, has what appear to be glottalised or pharyngeally constricted liquids and nasals, and Waima'a (Waimaha, Waimoa) of east Timor is reported to have phonemically aspirated and glottalised stops (Belo, Bowden, Hajek and Himmelmann 2005). The incidence of languages with phonemically imploded stops is also higher in this

area than in some others. Bimanese and Ngadha have two implosives, one labial and the other retroflexed; the Bimanese retroflex implosive is apico-domal. Hawu (called 'Savu' or 'Sawu' in the older literature) is reported to have labial, alveolar, palatal and velar implosives (Walker 1982, Grimes 2010).[12]

One of the most striking typological features of languages in this area is the erosion of word-final contrasts. Although some of the languages of Flores and Timor allow virtually all consonants to occur in final position, several others allow no final consonants. Languages that permit only open final syllables include Bimanese of eastern Sumbawa, Ngadha, Keo, Palu'e and Ende of Flores, Hawu, Dhao, and the languages of Sumba with the exception of Anakalangu. Rotinese is typologically unusual in allowing only -*k* and -*s*, both of which appear (at least historically) to be suffixes.

Another typological feature shared by many of the languages of eastern Indonesia apart from those further to the west, is the so-called 'reversed genitive', a feature that led Brandes (1884) to propose that the languages of Indonesia divide into eastern and western groups. The boundary between these typological categories was described by a line drawn between Hawu and Roti, Flores and the Solor Archipelago, east of Buton, west of the Sula Archipelago, east of Minahasa, Sangir-Talaud and the Philippines. This division, which came to be called the 'Brandes Line' can be illustrated by the contrast in genitive constructions between Malay *ékor babi* (tail-Ø-pig), and Dawan (west Timor) *fafi in iku-n*

(pig gen tail-3sg) 'tail of a pig'. In time it gradually became apparent that the Brandes Line is not a line at all, but rather a rough circle enclosing many of the languages of eastern Indonesia and western Melanesia that appear to have undergone structural change through earlier contact with Papuan languages.

Almost all languages of the Lesser Sundas are SVO, and verbal morphology is in general much simpler than that of Philippine-type or Western Indonesian-type languages. Manggarai, of west Flores, is reported as having no affixes at all (Verheijen 1977). A prominent feature of most languages of this region is the phonological attachment of a subject agreement morpheme or proclitic pronoun on the verb. In other languages both subject and object pronouns are cliticised to the verb, a phenomenon reminiscent of the verbal morphology of many Sulawesian languages. Some languages also show at least weakly developed verb serialisation, a phenomenon that is rare in the AN languages of Taiwan, the Philippines and western Indonesia. The following sentences from Kambera (eastern Sumba) and Tetun (central Timor) serve to illustrate:

Kambera (Klamer 1998, with some modifications of presentation)

1.   nda ku-hili beli-ma-ña-pa   'I am not going back to him again'
      neg 1sg.nom-again return-**emp-3sg.dat-impf**

2.   na-hoba-ya ìu nú
      3sg.nom-swallow-**3sg.acc** shark **deic** 'He was swallowed by a shark there' (lit. 'It swallowed him, shark there')

3.   lalu mbana-na na lodu   'The sun is too hot'
      excess hot-**3sg art** sun

---

4. nda ku-pi-a-ña na ŋandi-mu ru kuta hi hili kei-ŋgu-ña kawai
    neg 1sg.nom-know-just-**3sg.dat art** bring-**2sg.gen** leaf sirih **conj** again buy-**1sg.gen-3sg.dat** recently)   'I didn't know you brought sirih, so I also bought some just now'

5. nda ku-ŋaŋu-a iyaŋ   'I don't eat fish'
    neg 1sg.nom-eat-**mod** fish

6. na-ita-ya na hurat la pinu nulaŋ   'He saw the letter on the pillow'
    3sg.nom-see-**3sg.acc art** letter **loc** top pillow

7. na-unu mema-ña na wai mbana   'He drank the coffee immediately'
    3sg.nom-drink immediately-**3sg.dat art** water hot

Tetun (van Klinken 1999, with some modifications of presentation)

1. nia n-alai tiʔan   'She has run away'
    3sg 3sg-run already

2. tán nia n-aklelek haʔu foin haʔu fota nia
    because **3sg 3sg**-speak.abuse **1sg** then **1sg** hit **3sg**   'Because she verbally abused me, then I hit her'

3. haʔu kopi k-emu haʔi kaŋkuŋ k-á haʔi   'I don't drink coffee or eat watercress'
    1sg coffee **1sg**-drink not watercress **1sg**-eat **neg**

4. m-ola tais ó-k á té ní nia   'Take your sarong, as (yours) is that one'
    2sg-take sarong **2sg-poss** that because which **3sg**

5. nia dadi bá fahi   'He turned into a pig'
    3sg become go pig

6. ita hoʔo manu, rán kona ita, manu rán é dadi bá kfuti
    1pl.incl kill chicken blood touch **1pl.incl** chicken blood this become go mole 'If we kill a chicken, and blood touches us, the chicken blood becomes a (skin) mole'

7. hofonin mane sia mai labu iha h-oa sia né mama n-ó
    last.night man **p** come go.out **loc 1sg.poss**-child **p** this betel **3sg-** exist 'Last night men came courting my children, (and so) there will be betel (at the house)'

8. sia at bá r-afaho r-akawak   'They were going to go and help each other weed'
    3pl irr go **3pl**-weed **3pl**-mutual.aid

2.4.8 The languages of the Moluccas

The famous 'spice islands' of history, the Moluccas are a region of small to very small volcanic islands scattered through eastern Indonesia between Sulawesi and the Lesser Sundas on the west, and New Guinea on the east. The largest of these is Seram (Ceram), with a land area of about 18,700 square kilometres. Historically and politically, however, the most important islands in the Moluccas are the tiny islands of Ternate and Tidore in the northern Moluccas (the major source of cloves), and Ambon and Banda in the central Moluccas (the latter the sole original location of nutmeg). For the most part the southern Moluccas played little part in the history of the spice trade, and so were left out of the resulting networks of political and economic alliances that characterised the central and

northern Moluccas. Archaeological evidence of cloves which must have originated in the northern Moluccas has been claimed from sites as far afield as Terqa, in Syria, dating to about 3,700 BP. If accurate, this date implies that integrated trade networks linked the Moluccas with the Malay Peninsula within a few centuries after the AN penetration of insular Southeast Asia. However, the identification of the Terqa clove as deriving from the Moluccas has been questioned by Middle Eastern archaeologists, and the whole matter seems very much in limbo at present (O'Connor, Spriggs, and Veth 2007:16). Lewis (2009) lists 131 languages for the Moluccas, of which 18 are non-AN and one (Ternateño) is a Portuguese-based creole with Spanish relexification, leaving 112 AN languages.

Brief history of research

Little systematic work was done on the languages of the Moluccas until the Freiburg Moluccan Expedition visited the area from 1910-1912. Remarkably, Erwin Stresemann, an ornithologist on the expedition with an avocational interest in linguistics, collected a substantial quantity of accurate comparative data on the languages of Ambon, Seram, and the islands of the Goram (or Gorong) Archipelago east of Seram. On the basis of this material, unpublished data supplied by K. Deninger, and material from earlier published sources, Stresemann (1927) worked out the historical phonology and the subgrouping of more than 30 languages. In many ways this book still stands as the single most important contribution to the comparative study of these languages, although important additional studies, particularly in the area of subgrouping, have been done more recently by James T. Collins (1982, 1983a).

In addition to his comparative work, Stresemann published a grammar of Paulohi, a language of western Seram that was extinct by the time Collins conducted his fieldwork in the area in the mid 1970s. Ironically, since it was written by an ornithologist, Stresemann's grammar of Paulohi remains today "the only grammatical study of any length of a language of Central Maluku" (Collins 1982:109). Members of the Summer Institute of Linguistics began work in this area in the 1980s, and have contributed a number of article-length publications on the languages of both the central and southern Moluccas, as well as an unpublished grammar of Buru (Grimes 1991).

Language distribution

Table 2.10. lists the ten largest and ten smallest Austronesian languages of the Moluccas (M-C = Malayo-Chamic, K-F = Kei-Fordata, ECM = East Central Maluku, , S-B = Sula-Bacan, Y-S = Yamdena-Sekar, L-K = Luangic-Kisaric, SHWNG = South Halmahera-West New Guinea). As in the Lesser Sundas, the exceptionally large number of round figures and ties in these estimates is clearly a reflection of how poorly the cultural demographics of this area have been recorded:

Table 2.10 The ten largest and ten smallest Austronesian languages of the Moluccas

No.	Language	Subgroup	No. of speakers
1.	Ternate Malay	(WMP/M-C)	700,000 (2001)
2.	Ambon Malay	(WMP/M-C)	200,000 (1987)
3.	Kei	K-F	85,000 (2000)
4.	Fordata	K-F	50,000 2000)
5.	Geser-Goram	ECM	36,500 (1989)
6.	Buruese	S-B	32,980 (1989)
7.	Taba/Makian Dalam	SHWNG	30-40,000 (2001)
8.	Yamdena	Y-S	25,000 (1991)
9a.	Kisar	L-K	20,000 (1995)
9b.	Luang	L-K	20,000 (1995)
9c.	Sula	S-B	20,000 (1983)
9d.	Soboyo	S-B	20,000 (1983)
1a.	Amahai	ECM	50 (1987)
1b.	Paulohi	ECM	50 (1982)
1c.	Salas	ECM?	50 (1989)
2.	Loun	ECM	20 (?)
3a.	Hoti	ECM	10 (1987)
3b.	Hulung	ECM	10 (1991)
3c.	Kamarian	ECM	10? (1987)
3d.	Nusa Laut	ECM	10 (1989)
3e.	Piru	ECM	10 (1985)
4.	Naka'ela	ECM	5 (1985)
5.	Kayeli	ECM	3 (1995)
6.	Hukumina	ECM	1 (1989)

Map 9 The ten largest languages of the Moluccas

As seen in Table 2.10, the Moluccas have a rather large number of seriously endangered languages. In addition to Kayeli and Hukumina, with fewer than five speakers each in 1989, Loun, spoken in northwest Seram, is said to have 'a few' speakers (date of estimate not given). Most the smallest languages in the Moluccas are spoken on Seram, or nearby islands (Nusa Laut, on an island of the same name, Kayeli, on the eastern coast of Buru). The largest language in the Moluccas is Ambonese Malay, spoken primarily in the district capital of Ambon City, but also elsewhere on Ambon Island, and as a second language throughout the entire region. As noted elsewhere, the anomalous presence of a dialect of Malay in eastern Indonesia is a historical consequence of the spice trade, which once linked the source islands of the Moluccas with Malay-speaking middlemen along the Strait of Malacca. Three of the largest indigenous languages of the Moluccas (Kei, Fordata, Geser-Goram) are found in the small islands of the neighboring Geser-Goram and Kei Archipelagos extending southeast of Seram. The reason for such relatively large language groups in this region is unclear.

Typological overview

Most phoneme inventories of the central and southern Moluccan languages for which data has been published are not characterised by any unusual features. Some languages, as Alune of western Seram, have a labiovelar consonant k^w, but labiovelars have a scattered distribution throughout much of the AN family.[13] Most Moluccan languages have at least the vowels *a e o i u*, and some also have a schwa. As in the Lesser Sundas, there is a somewhat higher incidence of languages that lack final consonants than is true of Taiwan, the Philippines or western Indonesia, but many Moluccan languages retain original final consonants. Heterorganic consonant clusters are not unusual in the Moluccas, and as a result of historical vowel syncope a few languages have developed atypical word-initial clusters, as with Yamdena (Tanimbar Archipelago) *kmp^wean* 'I like', *tndiŋan* 'kind of fish', or *kb^watar* 'root-destroying grub'. In the northern Moluccas several languages, including at least Ma'ya and Matbat, have phonemic tone (Remijsen 2001). The most likely source for this feature is contact with Papuan languages, many of which reportedly have tone or pitch-accent systems, but such contact situations do not currently exist, and this explanation assumes past language distributions that must have differed from those found now.

Syntactically the languages of the central and southern Moluccas share a number of features with those of the Lesser Sundas. In both areas the order of major sentence constituents is almost universally SVO, verb morphology is relatively simple, and verbs commonly carry a subject agreement morpheme or proclitic pronoun. Many languages in both areas also have a 'reversed genitive' construction ('reversed' from the perspective of most AN languages), with possessor preceding possessed ('bird's tail' rather than 'tail of bird'). The following sentences from Larike (Ambon) and Taba, also called 'Makian Dalam' (south Halmahera), illustrate general features of the syntax of CMP and SHWNG languages respectively:

Larike (Laidig 1993)

1. Abu mana-pala elau 'Abu has a lot of nutmeg'
    Abu **3sg.gen**-nutmeg much

2. au-leka aku-ba laku aku-ina matir-mau
    1sg.nom-follow **1sg.gen**-father and **1sg.gen**-mother **3pl.gen**-wish 'I followed the wishes of my father and mother'

3. ana hi Husein laku Kalsum matuar-ana 'This child is Husein and Kalsum's child'
    child this Husein and Kalsum **3dl.gen**-child

4. Sait mana-ba mana-duma i-koʔi 'Sait's father's house is small'
    Sait **3sg.gen**-father **3sg.gen**-house **3sg.nom**-small

5. aʔu aku-iŋine au-na-anu-imi 'As for me, my desire is to eat you all'
    1sg.top **1sg.gen**-desire **1sg.nom-irr**-eat-**2pl.acc**

---

[13]   The term 'labiovelar' is used in much of the AN literature for consonants with a primary closure at either labial or velar places, and secondary lip rounding. More recent work in general phonetics, as Ladefoged and Maddieson (1996:356) calls these articulation types labialized or rounded (hence rounded labials and rounded velars). In the interest of comparability with the earlier literature I retain the term 'labiovelar' for both.

6.   mati matir-ure tau iri-ape 'Their bananas are not yet ripe'
        3pl 3pl.gen-banana not.yet **3pl.acc**-ripe

7.   mati-tunu iʔanu 'They baked fish'
        3pl.nom-bake fish

8.   hima Abu na doma 'That is Abu's house'
        that Abu **gen** house

9.   mana-rupae na hutua tahi i-na sanaŋ 'His wife was not happy'
        3sg.gen-wife **gen** heart not **3sg-gen** happiness

Taba (Bowden 2001)

1.   n-pun bobay pake sandal 'He killed the mosquito with a thong/slipper'
        3sg-kill mosquito use thong

2.   n-han n-hait te-su 'S/he hasn't yet gone up'
        3sg-go **3sg**-ascend **neg-pot**

3.   we ha-lu da e de yak k-on 'The second mango is for me to eat'
        mango **clas**-two **dist foc res 1sg 1sg**-eat

4.   k-yat coat lu ak-le tapin li 'I'm carrying two bundles (of firewood) to the kitchen'
        1sg-carry **clas.**bundle two **all**-land kitchen **loc**

5.   ŋan i-so yak k-wom nak 'One day I'll come back'
        day **clas**-one **1sg 1sg**-come again

6.   k-tala yan banden nyoa beit-utin co 'I got almost one hundred milkfish'
        1sg-meet fish milkfish almost **clas**-hundred one

7.   polo Taba ne mdudi cilaka 'If Makian had sunk, it would have been a disaster'
        if Makian **prox** sunk disaster

8.   n-yol calana de n-ha-totas (> natotas) 'She took the trousers and washed them'
        3sg-take trousers **de 3sg-caus**-wash

9.   yak k-goras-o (> kgorco) kapaya ni kowo bbuk
        1sg 1sg-scrape-**appl** papaya **3sg.poss** seed book 'I'm scraping papaya seeds onto
        the book'

2.4.9 The Austronesian languages of New Guinea and satellites

New Guinea, the world's second largest island with a land area of about 821,000 square kilometres, is divided roughly in half between the Republic of Indonesia and the nation of Papua New Guinea. In the nineteenth century the western half of the island formed part of the Netherlands East Indies, while the eastern half was divided between German New Guinea and the British colony of Papua. When Indonesia achieved its independence in 1949 Dutch New Guinea remained the last colonial stronghold of the Netherlands East Indies until it was incorporated into the Republic of Indonesia in 1963, under the name 'Irian', a name popularly believed in Indonesia to be an acronym for 'Ikut Republic Indonesia Anti-Nederland' (follow the Republic of Indonesia against the Netherlands), later augmented to 'Irian Jaya' (Great Irian). As a result of pressure from Papuan nationalists the former Irian Jaya has recently been renamed 'Papua'. To avoid confusion

between these two distinct uses of the term 'Papua'—the first referring to the southeastern quarter of the island during the early part of the twentieth century, and the second more recently to its western half—the more familiar name 'Irian Jaya' (or just 'Irian') is retained here. The shape of New Guinea is sometimes compared to a Bird of Paradise, with the head pointed northwest and the tail southeast, hence the name 'Bird's Head'(Dutch: Vogelkop) for the western extremity of the island. Because it lies just south of the Equator, New Guinea has a largely tropical climate. However, it is also extremely mountainous, and so affords a range of climates correlated with altitude, even housing a small permanent glacier (Puncak Jaya) at 4,884 meters elevation in the interior of Irian.

Radiocarbon dates show that New Guinea was settled at least 40,000 years ago, and it is now well-established that horticulture developed in the eastern highlands by 6,500-7,000 BP, and was widespread by the time AN speakers arrived along the coasts sometime during the second millennium BCE (Golson 2005). Although New Guinea highlanders show physical differences from coastal populations, the physical type of AN and Papuan speakers in coastal New Guinea is generally indistinguishable as a result of millennia of contact. Unlike the sparse hunter-gatherer populations in the Philippines, which assimilated linguistically to the incoming Austronesians with little gene flow until recent times, the horticultural 'Papuan' populations more than held their own, at least in part because of their superior numbers. As a result, many of the AN-speaking peoples of New Guinea show significant contact influences in both physical type and culture.

New Guinea is home to some 750 languages in an area roughly one and one half times that of France. The great majority of these are classified as 'Papuan', a cover term that appears to encompass a number of distinct language families. AN languages are mainly confined to the coastal regions of the mainland and to smaller offshore islands. The most significant penetration of AN languages into the interior of New Guinea occurs in the broad Markham valley on the northeast coast of the island, where relatively easy access to higher elevations is possible due to flat and gradually ascending terrain. In reaching New Guinea we leave insular Southeast Asia and enter the island world of Melanesia. Although the term 'Melanesia' is reasonably well-defined, the adjective 'Melanesian' is not, as it can refer to the AN languages of Melanesia (which do not form a subgroup), or to a relatively dark-skinned frizzy-haired physical type found among both AN-speaking and non-AN speaking coastal populations. To compound the potential confusion, non-AN speaking 'Melanesians' are said to speak Papuan languages, but with regard to physical type the term 'Papuan' is generally reserved for inland (or inland-oriented) populations. Lewis (2009) gives 120 AN languages for the island of New Guinea and its immediate satellites (exclusive of the Bismarck Archipelago).

Brief history of research

Although virtually all parts of the island are now in sustained contact with the outside world, New Guinea has been one of the last frontiers of Western scientific exploration. Some parts of the eastern highlands with large sedentary populations were only discovered by Westerners in the 1930s, and linguistic research in this area has lagged behind many others because of the number of languages, and the small size of many language communities. Some descriptive and comparative studies of AN languages in what was then German New Guinea were published during the first decades of the twentieth century, most notably Schmidt (1900-1901), Dempwolff (1905, 1939), and Friederici (1912-1913). The first substantial contribution to the description of an AN language in what was then British New Guinea is the Motu grammar and dictionary of Lister-Turner and Clark

(1930). The first broader attempts to survey the AN languages of New Guinea from a historical and typological perspective were carried out by Capell (1943, 1971). The quality of scholarly work on these languages reached an entirely new level with the Manam grammar of Lichtenberk (1983), and the landmark comparative study of Ross (1988). During the Dutch colonial period little work was done on the AN languages of Irian, the most notable exception being the survey of Anceaux (1961). Even with these contributions, however, few grammars or dictionaries exist for any of the AN languages of New Guinea.

Language distribution

Table 2.11. lists the ten largest and ten smallest AN languages of New Guinea and its immediate satellites exclusive of the islands of the Bismarck Archipelago (M-C = Malayo-Chamic, NNG = North New Guinea Cluster, PT = Papuan Tip Cluster, SHWNG = South Halmahera-West New Guinea, MC = Micronesian; all subgroups in this and subsequent tables of Chapter 2, except SHWNG are branches of Oceanic):

Table 2.11 The ten largest and ten smallest Austronesian languages of New Guinea

No.	Language	Subgroup	No. of speakers
1.	Papuan Malay	(M-C)	500,000 (?)
2.	Takia	NNG	40,000 (?)
3.	Motu	PT	39,000 (2008)
4.	Numfor/Biak	SHWNG	30,000 (2000)
5.	Adzera	NNG	28,900 (2000)
6a.	Tawala	PT	20,000 (2000)
6b.	Kilivila	PT	20,000 (2000)
7.	Keapara	PT	19,400 (2000)
8.	Mekeo	PT	19,000 (2003)
9a	Sinaugoro	PT	18,000 (2000)
9b.	Misima-Paneati	PT	18,000 (2002)
1.	Mindiri	NNG	80 (2000)
2a.	Iresim	SHWNG	70 (2000)
2b.	Vehes	NNG	70 (2000)
3.	Kayupulau	NNG	50 (2000)
4.	Gweda/Garuwahi	PT	26 (2001)
5.	Liki	NNG	11 (2005)
6.	Masimasi	NNG	10 (2005)
7.	Dusner	SHWNG	6 (1978)
8.	Tandia	SHWNG	2 (1991)
9.	Mapia	MC	1(?)

1) Papuan Malay
2) Takia
3) Motu
4) Numfor/Biak
5) Adzera
6a) Tawala
6b) Kilivila
7) Keapara
8) Mekeo
9a) Sinaugoro
9b) Misima-Paneati

Map 10 The ten largest Austronesian languages of New Guinea and satellites

Lewis (2009) states that there were 120,000 speakers of Hiri Motu (the former Police Motu, not the original Hiri Motu) in Papua New Guinea in 1989, but almost all of these were second-language speakers. Apart from Hiri Motu, then, the largest AN language in New Guinea appears to be Biak (sometimes called Numfor-Biak, from the two main dialects spoken on islands of the same name in Cenderawasih Bay, Irian). Biak speakers have traditionally controlled local trade networks in the large island-studded region of Cenderawasih Bay, and Biak has consequently served as a local *lingua franca* of some importance in this part of the Bird's Head Peninsula for many generations. Four of the other major AN languages of New Guinea (Mekeo, Keapara, Sinaugoro, Motu) are spoken in the Central District of Papua in southeast New Guinea. The most endangered AN languages in New Guinea do not show a geographical bias. Two of these (Bina and Yoba) belong to the same subgroup in southeasp t New Guinea, but the others are scattered from the Bird's Head (Anus, Dusner) to the north coast in Papua New Guinea (Terebu, Wab, Vehes, Mindiri), with one (Mapia), the westernmost member of the Chuukic dialect chain, located on a small island about 180 kilometres north of the Bird's Head Peninsula.

Typological overview

Three typological traits stand out sharply in the AN languages of New Guinea. First, although some South Halmahera-West New Guinea languages preserve final consonants, almost all of the Oceanic languages of New Guinea allow only CV syllables. Second, several languages, including at least Mor, spoken on a small island of the same name in

Cenderawasih Bay, Bird's Head, and Yabem and Bukawa, spoken far to the east in the Huon Peninsula, have lexical pitch distinctions. Third, and perhaps most strikingly, the AN languages of New Guinea fall into two typological classes with respect to the order of major sentence constituents: SVO and SOV. Verb-final languages are confined to the north and south coasts of the eastern part of New Guinea, and to a small area in the east of Bougainville Island. As in other parts of the world, a verb-final typology implies various other structural properties, most notably the use of postpositions. The unique SOV typology of AN languages in New Guinea almost certainly arose through contact, as nearly all Papuan languages are verb-final. The following sentences from Numbami (Huon Peninsula), and Motu (eastern Gulf of Papua) illustrate both SVO and SOV typologies:

Numbami (Bradshaw 1982)

1.   ti-lapa bola uni 'They killed the pig'
        3pl-hit pig die

2.   i-tala ai tomu 'He chopped down the tree'
        3sg-chop tree broken

3.   ti-ki biŋa de lawa manu ai-ndi waŋga i-tatala na
        3pl-send word to people **wh 3pl-gen.pl** canoe **3sg**-sink **rel** 'They sent word to the people whose boat sank'

4.   ma-ki bani manu ma-yaki na su ulaŋa 'We put the food we've pared into the pot'
        1pl.ex-put food **wh 1pl.ex**-pare **rel** into pot

5.   manu bembena-ma i-ma teteu na, i-loŋon-i biŋa numbami kote
        when at.first-**adv 3sg**-come village **rel 3sg**-hear-**tr** talk Numbami **neg** 'When he first came to the village he didn't understand Numbami'

6.   iŋgo ta-tala kundu tomu na, a kole lua mo toli i-na-wasa i-na-tala tomu
        sub ('say') **1pl.in**-chop sago broken **rel** perhaps man two or three **3pl-irr**-go **3pl-irr** chop broken 'When we chop down a sago palm, perhaps two or three men will go chop it down'

Motu (Lister-Turner and Clark 1930)

1.   Hanuabada amo na-ma 'I came from Hanuabada'
        Hanuabada from **1sg**-come

2.   lau na tau, ia be hahine
        1sg.nom art man **3sg.nom art** woman 'I am a man, she is a woman' (**na** = more definite, **be** = less definite)

3.   asi gini diba-gu 'I cannot stand'
        neg know stand-**1sg.poss**

4.   sisia ese boroma e kori-a 'The dog bit the pig'
        dog **dem** pig **3sg**   bite-it

5.   hahine ese natu-na e ubu-dia 'The woman fed her children'
        woman **dem** child-**3sg.poss 3sg**   feed-**3pl.acc**

6.   mahuta-gu ai natu-gu e mase 'While I slept my child died'
        sleep-**1sg.poss** when child-**1sg.poss 3sg**   die

7.   boroma e ala-ia tau-na na vada e-ma 'The man who killed the pig has come'
       pig **3sg**   kill-it man-the **dem perf 3sg**-come

8.   tua ese au-na imea bogaragi-na-i vada e hado
       man **dem** tree-the garden middle-its-in **perf 3sg**   plant 'The man planted the tree in
       the middle of the garden'

2.4.10 The Austronesian languages of the Bismarck Archipelago

The Bismarck Archipelago contains the relatively large island of New Britain (37,800 square kilometres), and the smaller islands of New Ireland, the Admiralty group (of which Manus is the largest), and a few smaller island groups, such as the French Islands, New Hanover and the St. Mathias group. Its name derives from nineteenth century colonial times, when it formed part of German New Guinea. The islands are volcanic, and two of the major sources of obsidian which was widely traded for stone tools are located in this region, namely the Talasea Peninsula of New Britain, and Lou Island in the Admiralty group (words for 'obsidian' in many of the languages of Manus reflect *patu i low 'stone of Lou). Lewis (2009) gives 63 AN languages in the Bismarck Archipelago.

Brief history of research

Much of the descriptive and comparative material on languages of the Bismarck Archipelago has appeared within the past three decades. Some comparative data for the entire area can be found in Ross (1988), and survey volumes that focus on New Britain or New Ireland include Johnston (1980), Thurston (1987), and Ross (1996d). Despite recent progress, however, the coverage of this area is still spotty. Hamel (1994) is the only full-length grammar for any language of the Admiralties, and a handful of mostly very short and preliminary grammars exist for languages of New Ireland and New Britain. With regard to lexicography, a fairly substantial manuscript dictionary exists for Lakalai of New Britain, but as yet the only published dictionary for any language of the entire area is the short and very preliminary *Tanga-English dictionary* of Bell (1977), representing the language of one of the small islands which run parallel to the east coast of New Ireland.

Language distribution

Table 2.12. lists the ten largest and ten smallest Austronesian languages of the Bismarck Archipelago (MM = Meso-Melanesian, NNG = North New Guinea, ADM = Admiralties Family, SM = St. Matthias Family):

Table 2.12 The ten largest and ten smallest languages of the Bismarck Archipelago

No.	Language	Subgroup	No. of speakers
1.	Tolai/Kuanua	MM	61,000 (1991)
2.	Bola	MM	13,700 (2000)
3.	Nakanai/Lakalai	MM	13,000 (1981)
4.	Lihir	MM	12,600 (2000)
5.	Lavongai/Tungag	MM	12,000 (1990)
6.	Duke of York/Ramoaaina	MM	10,300 (2000)
7.	Bali/Uneapa	MM	10,000 (1998)
8.	Vitu/Muduapa	MM	8,800 (1991)
9.	Mengen	NNG	8,400 (1982)
10.	Patpatar	MM	7,000 (1998)
1.	Lenkau	ADM	250 (1982)
2.	Elu	ADM	220 (1983)
3a.	Amara	NNG	200 (1998)
3b.	Gasmata	NNG	200 (1981)
3c.	Mokerang	ADM	200 (1981)
4.	Label	MM	144 (1979)
5.	Nauna	ADM	100 (2000)
6.	Likum	ADM	80 (2000)
7.	Tenis	SM	30 (2000)
8.	Guramalum	NNG	3 (1987)

Map 11 The ten largest Austronesian languages of the Bismarck Archipelago

The largest AN language of the Bismarck Archipelago is Tolai (also known as Kuanua, or Tuna), native to the area of Rabaul harbor near the northeastern end of New Britain. At least in part because of its location Tolai has long been an important *lingua franca* in the region, and is the principal lexical contributor to Tok Pisin (Mosel 1980) The language of Duke of York Island is a neighbor and close relative. The smallest AN languages show no particular geographical concentration; five (Mondropolon, Lenkau, Mokerang, Nauna, Likum) are found in the Admiralty Islands, three (Amara, Gasmata, Getmata) are found in New Britain, one (Label) is found in New Ireland, and one (Tenis) is found on an isolated island in the St. Mathias Archipelago north of New Ireland.

Typological overview

The AN languages of the Bismarck Archipelago are extremely varied, making typological generalisation difficult. All languages of the eastern Admiralties have lost not only original final consonants, but also the vowels that preceded them, and so allow many final consonants in contemporary word forms. A number of the languages of Manus Island have voiced prenasalised trills at both bilabial and alveolar points of articulation. While alveolar trills are well-known from languages such as Fijian, bilabial trills are rare, and of considerable interest to general phonetic theory (Ladefoged and Maddieson 1996:129ff). In addition, (Hajek 1995) has drawn attention to reports of lexically contrastive pitch in three languages of New Ireland that for unexplained reasons were neglected and virtually forgotten for over two decades.

All AN languages of the Bismarck Archipelago are SVO. As in other parts of Melanesia, serial verb constructions are fairly common, perhaps ultimately a product of contact with Papuan languages during the initial period of settlement of the area. Loniu of eastern Manus, and Kaliai-Kove of New Britain can be used to illustrate:

Loniu (Hamel 1994)

1. iy amat itiyɛn iy amat a kaw 'That man is a sorcerer'
   3sg man **dem 3sg** man **poss** sorcery

2. wɔw a-la tah ɛ wɔw ɛ-li yaw 'Are you there, or have you gone away?'
   2sg 2sg-go **loc** or **2sg 2sg-perf** go

3. sɛh amat masih sɛh musih epwe 'All men are alike'
   3pl man all **3pl** alike only

4. mwat itɔ yɛni lɛŋɛʔi ñanɛ suʔu 'The snakes would eat the way their mother did'
   snake **3sg.stat** eat like mother **3dl**

5. hɔmɔw pasa ŋaʔa-n pwe 'No one knows her name'
   one knowledge name-**3sg neg**

6. he tɔ takɛni pat 'Who is throwing stones?'
   who **cont** throw stone

7. pɛti cah iy i-tɛŋ cɛlɛwan 'Why does she cry so much?
   for what **3sg 3sg**-cry much

8. an macɛhɛ ta ɛtɛ wɔw 'How much water do you have?'
   water how:much **loc ag 2sg**

Kaliai-Kove (Counts 1969)

1.   tanta ti-watai 'The men know'
       man **3pl**-know

2.   i-kinani tamine ɣa ti-la 'He lets the women go'
       3sg-permit woman and **3pl**-go

3.   u-ndumu-ɣao mao 'Don't lie to me'
       2sg-lie-**1sg neg**

4.   i-sasio-ri ɣa mamara 'He sends them far away'
       3sg-send-**3pl** and distant

5.   ta-ɣali iha salai 'We speared many fish'
       1pl.incl-spear fish many

6.   ŋa-ani moi salai tao 'I eat a lot of taro'
       1sg-eat taro much very

7.   waɣa ɣane sei e-le 'Whose canoe is this?'
       canoe here who **3sg-poss**

8.   ti-moro pa-ni ɣaβu aisali 'They live on the slopes of Gavu (a mountain)'
       3pl-reside on-it ɣavu slope

9.   tanta i-naɣe tuaŋa ai-tama 'The man is the 'father of the village' (bigman)'
       man **3sg**-be village its-father

2.4.11 The Austronesian languages of the Solomon and Santa Cruz Islands

The Solomons chain extends northwest-southeast from the atoll of Nehan (or Nissan) to the small island of Santa Ana. Politically it is divided between Papua New Guinea, which controls Buka, Bougainville, and the smaller islands to the north and west of these, and Solomon Islands, which controls the rest. Although smaller than New Britain, most islands of the Solomons are large by Pacific standards, and allow a contrast between 'sea people' and 'hill people'. Lewis (2009) lists 74 languages for the Solomons, of which five are extinct. However, 14 of these are non-AN, or are otherwise irrelevant to counting indigenous AN languages. Another 14 AN languages are spoken in islands of the Solomons chain which form part of Papua New Guinea, bringing the total number of extant AN languages in the Solomons chain to 69.

Brief history of research

The Solomons were included in the early survey volumes of Codrington (1885), and Ray (1926), and were surveyed exhaustively by Tryon and Hackman (1983), who provide a vocabulary of 324 words for 111 language communities. Basic information on phoneme inventories, common vocabulary and historical change is thus available for most Solomons languages. More in-depth studies were initiated during the early twentieth century by the Australian scholar W.G. Ivens, and the missionary C.E. Fox, who worked in the southeast Solomons. Together with more recent work there are now perhaps a dozen dictionaries for languages of this area. These tend to favor the languages of the southeast Solomons ('Āre'āre, Arosi, Bugotu, Nggela, Kwaio, Lau, Sa'a), and the Polynesian Outliers (Rennell-Bellona, Tikopia, Anuta, Pileni). To date there are no more than three or four full

grammars of Solomons languages, although a number of shorter grammatical sketches have been published. Recent contributions of some importance are White, Kokhonigita and Pulomana (1988) for Cheke Holo, Corston (1996) for Roviana, Davis (2003) for Hoava, and Palmer (2009) for Kokota.

Language distribution

Table 2.13. lists the ten largest and ten smallest AN languages of the Solomons chain (SES = Southeast Solomonic, MM = Meso-Melanesian, PN = Polynesian, UV = Utupua-Vanikoro):
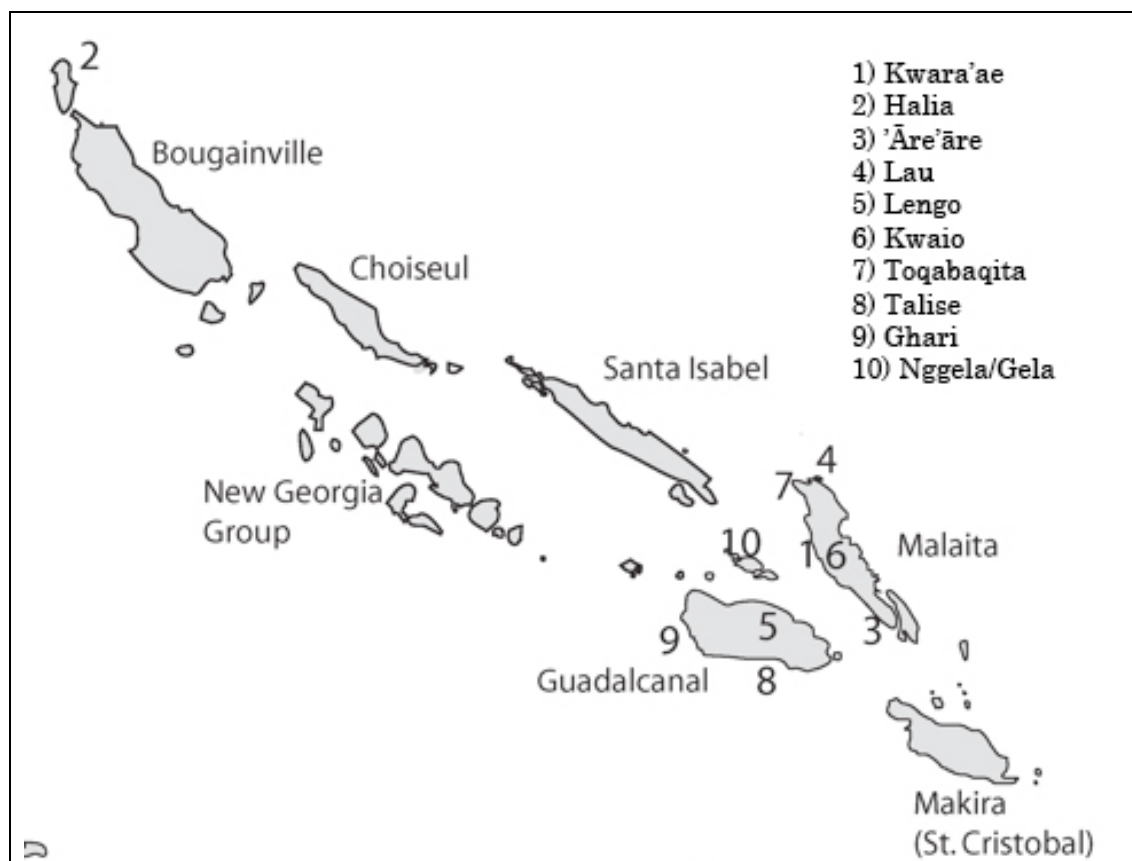
Map 12 The ten largest Austronesian languages of the Solomon Islands

Table 2.13 The ten largest and ten smallest languages of the
Solomons and Santa Cruz

No.	Language	Subgroup	No. of speakers
1.	Kwara'ae	SES	32,433 (1999)
2.	Halia	MM	20,000 (1994)
3.	'Āre'āre	SES	17,800 (1999)
4.	Lau	SES	16,937 (1999)
5.	Lengo	SES	13,752 (1999)
6.	Kwaio	SES	13,249 (1999)
7.	Toqabaqita	SES	12,572 (1999)
8.	Talise	SES	12,525 (1999)
9.	Ghari	SES	12,119 (1999)
10.	Nggela	SES	11,876 (1999)
1.	Anuta	PN	267 (1999)
2.	Nanggu	?	210 (1999)
3.	Ririo	MM	79 (1999)
4.	Oroha	SES	38 (1999)
5a.	Laghu	MM	15 (1999)
5b.	Tanimbili	UV	15 (1999)
6a.	Zazao/Kilokaka	MM	10 (1999)
6b.	Asumboa	UV	10 (1999)
7.	Vano	UV	5 (2007)
8.	Tanema	UV	4 (2007)

The largest languages of the Solomons chain are heavily concentrated in the southeast Solomons (all but Halia). The most endangered languages show a wider distribution, including two Polynesian Outliers (Nuguria and Nukumanu), four languages of the southern Santa Cruz Archipelago (Tanema, Tanimbili, Vanikoro and Asumboa), and others in both the western and the southeastern Solomons.

Typological overview

The AN languages of the Solomons have some phonetic traits that are widespread in Melanesia, such as automatic prenasalisation of voiced stops, or the presence of labiovelar consonants, but there is little that is peculiar to them. Like many other Oceanic languages, they tend to permit only open final syllables, either through loss of final consonants or addition of echo vowels. Two AN languages in Bougainville (Torau and Uruava) are SOV, almost certainly due to contact with the larger neighboring Papuan language Nasioi, and some AN languages of the New Georgia Archipelago are predicate-initial, at least in some constructions, but elsewhere in the Solomons virtually all languages are SVO. Because a major genetic and typological boundary separates the western Solomons from the central and southeast Solomons (Ross 1988) sample sentences will be given for each area:

Roviana (Todd 1978; morpheme boundaries not found in the original have been added, together with some morpheme glosses)

1. vineki ziŋa-ziŋara-na si asa 'She is a light-skinned girl'
    girl red-reddish-**att cop 3sg**

2. maŋini hola sari na rane ba veluvelu si ibu
    hot very **pl art** day but evening **cop** cool 'The days are very hot, but the evenings are cool'

3. mami vetu-raro si pa mudi-na sa mami vetu
    1pl.excl.poss cook-house **cop** at behind-**att art 1pl.excl.poss** house 'Our cook-house is behind our house'

4. lopu tabar-i-go rau 'I won't pay you'
    neg pay-**om-2sg 1sg**

5. kote oki atu-n-i-a rau sa bolo koa goi 'I will throw the ball to you'
    will throw away-**n-om**-it I **art** ball to you

6. hiva lau si rau sapu niu sa vaka 'I feel sick when the boat rocks'
    feel sick **cop** I when rock **art** boat

7. heki-n-i-a sa tama-na pude lopu haba-i-a sa tie sana
    forbid-**n-om**-it **art** father-**3sg.poss** so.that **neg** marry-**om**-him **art** man that 'Her father forbid her to marry that man'

8. pana ruku sa kaqu koa pa korapa-na si gita 'If it rains we will stay inside'
    if/when rain it **fut** stay at inside-**att cop 1pl.incl**

Toqabaqita (Lichtenberk 1984)

1. nia ʔe θauŋani-a sui naʔa luma nia 'He has finished building his house'
    3sg 3sg make-**3om comp perf** house **3sg**

2. wane baa ki kera taa-tari-a boθo baa 'The men kept chasing the pig'
    man that **pl 3pl red**-chase-**3om** pig that

3. kini ʔe ŋali-a mai raboʔe faŋa 'The woman brought a bowl of food'
    woman **3sg** carry-**3om** hither bowl food

4. oli na-mu nena 'Are you going back now?'
    go.back **perf-2sg** now

5. θaʔaro ʔe lofo kali-a fafo-na ŋa luma lakoo ki
    bird **3sg** fly surround-**3om** above-**3sg.poss art** house that **pl** 'The bird flew in a circle above those houses over there'

6. ku uʔunu ʔi sa-na wela 'I told the child a story'
    1sg tell.story to **gol-3sg.poss** child

7. niu neʔe ki na ku ŋali-a mai 'It was these coconuts that I brought'
    coconut this **pl foc 1sg.ac** carry-**3om** hither

8. ŋali-a mai ta si ʔoko fasia kuka kani-a ʔana boθo neʔe
    carry-**3om** hither some **clas** rope **purp pl.incl.seq** tie-**3om res.pro** pig this 'Bring a piece of rope so that we may tie this pig with it'

2.4.12 The languages of Vanuatu

North-to-south Vanuatu extends from the Torres and Banks Islands to Anejom (= 'Aneityum' in the older literature). In the north it is a double chain, but beginning with Efate the islands continue southward like stepping stones in a single line. Vanuatu is shorter than the Solomons chain, has a more nearly north-south axis, and its islands are considerably smaller than those of the Solomons. Lewis (2009) lists 110 languages for Vanuatu, of which one is extinct. Since this includes French, English, and Bislama, the total number of AN languages is 107.

Brief history of research

Although Mota of the Banks Islands was described in the late nineteenth century (Codrington and Palmer 1896), little was known about most languages of Vanuatu until the mid 1970s. In 1976 the Australian linguist D.T. Tryon published a linguistic survey of Vanuatu (then the New Hebrides), which included a comparative vocabulary of 292 items for 179 communities. This study laid the foundations for the comparative study of the languages, which until then had been largely neglected. Since the mid 1970s the pace of scholarship on the languages of Vanuatu has quickened. Thanks largely to the efforts of the Australian linguists John Lynch and Terry Crowley, grammars now exist for Lenakel (Lynch 1978a), Sye (Crowley 1998), Ura (Crowley 1999), and Anejom (Lynch 2000), as well as dictionaries for Lenakel (Lynch 1977a) and Kwamera (Lindstrom 1986). In addition, Lynch (2001) gives a comprehensive account of the historical development of the languages of southern Vanuatu. As a result of these publications the languages of southern Vanuatu, which were among the most poorly described in the Pacific a quarter of a century ago, have become among the best documented. Major publications on the languages of central and northern Vanuatu over the past three decades include texts and a short grammar of Nguna (Schütz 1969a, 1969b), a short dictionary of Southeast Ambrym (Parker 1970), a short grammar and dictionary of Lonwolwol on Ambrym Island (Paton 1971, 1973), a short grammar of the northern dialect of Sakao (Guy 1974), a short grammar of Big Nambas (Fox 1979), a grammar and dictionary of Paamese of central Vanuatu (Crowley 1982, 1992), a very detailed grammar of the North-East Ambae language (Hyslop 2001), a grammar and short dictionary of Araki (François 2002), an extremely detailed account of verbal semantics in Mwotlap/Motlav (François 2003a,b), a technologically pioneering grammar of South Efate by Thieberger (2006), which includes all of the primary sound data on an accompanying compact disc, and a grammar of Mafea by Guérin (2011). In addition, following the sudden and untimely death of Terry Crowley, grammars of four languages of Malakula have appeared, posthumously edited by John Lynch (Crowley 2006a, 2006b, 2006c, 2006d). Together with several still unpublished doctoral dissertations, these comments show that Vanuatu is no longer the neglected backwater of linguistic scholarship that it once was. Not only has the pace of scholarship on the languages of this region quickened, but more recent publications tend to be longer, and of higher quality than the work of the 1970s. Nonetheless it is clear that a great deal remains to be done, and given the small size of the surviving language communities and the pace of social and linguistic change in the area, time is of the essence.

Language distribution

Lynch and Crowley (2001:17-19) provide a full list of the languages of Vanuatu that includes estimates of number of speakers. Table 2.14 gives the ten largest languages from

this source. The ten smallest languages are more difficult to determine, as the column for number of speakers is marked 'handful' or 'handful?' in 18 cases. For this reason Table 2.14 lists the ten smallest languages for which actual numbers are given, and those languages that appear to be on the verge of extinction (or possibly, but not certainly extinct) are listed separately after the table. Additional information for the far north of Vanuatu was provided in a personal communication from Alexandre François. SM = Southern Melanesian Family, SNE = Shepherds-North Efate, MC = Malakula Coastal, AM = Ambae-Maewo Family, PNT = Pentecost, AP = Ambrym-Paama, SE = South Efate, SWS = Southwest Santo, WS = West Santo, MI = Malakula Interior, EP = Epi, T-B = Torres-Banks:[14]

Table 2.14 The ten largest and ten smallest languages of Vanuatu

No.	Language	Subgroup	No. of speakers
1.	Lenakel	SM	11,500 (2001)
2.	North Efate/Nguna	SNE	9,500 (2001)
3.	NE Malakula	MC	9,000 (2001)
4.	West Ambae/Duidui	AM	8,700 (2001)
5.	Apma	PNT	7,800 (2001)
6.	Whitesands	SM	7,500 (2001)
7.	Raga	AM	6,500 (2001)
8a.	Paamese	AP	6,000 (1996)
8b.	Efate, South	SE	6,000 (2001)
9.	North Ambrym	AM	5,250 (2001)
1a.	Tambotalo	WS	50 (1981)
1b.	Dixon Reef	MI	50 (1982)
2.	Bieria	EP	25 (2001)
3a.	Litzlitz	MI	15 (2001)
3b.	Maragus/Tape	MI	15 (2001)
4.	Mwesen	T-B	10 (2012)
5.	Ura	SM	6 (1998)
6a.	Araki	SWS	5 (2012)
6b.	Nasarian	MI	5 (2001)
7.	Olrat	T-B	4 (2012)
8.	Lemerig	T-B	1 (2012)

Moribund languages that are indicated as having only a 'handful' of remaining speakers include 12 others on Malakula (Aveteian, Bangsa, Matanvat, Mbwenelang, Naman, Nahati/Nāti, Navwien, Nivat, Niviar, Sörsörian, Surua Hole, Umbrul), two on Epi (Iakanaga, Ianigi), one on Ambrym (Orkon).

---

[14]   Cf. Francois (2012) for number of remaining speakers of Mwesen, Araki, Olrat and Lemerig.
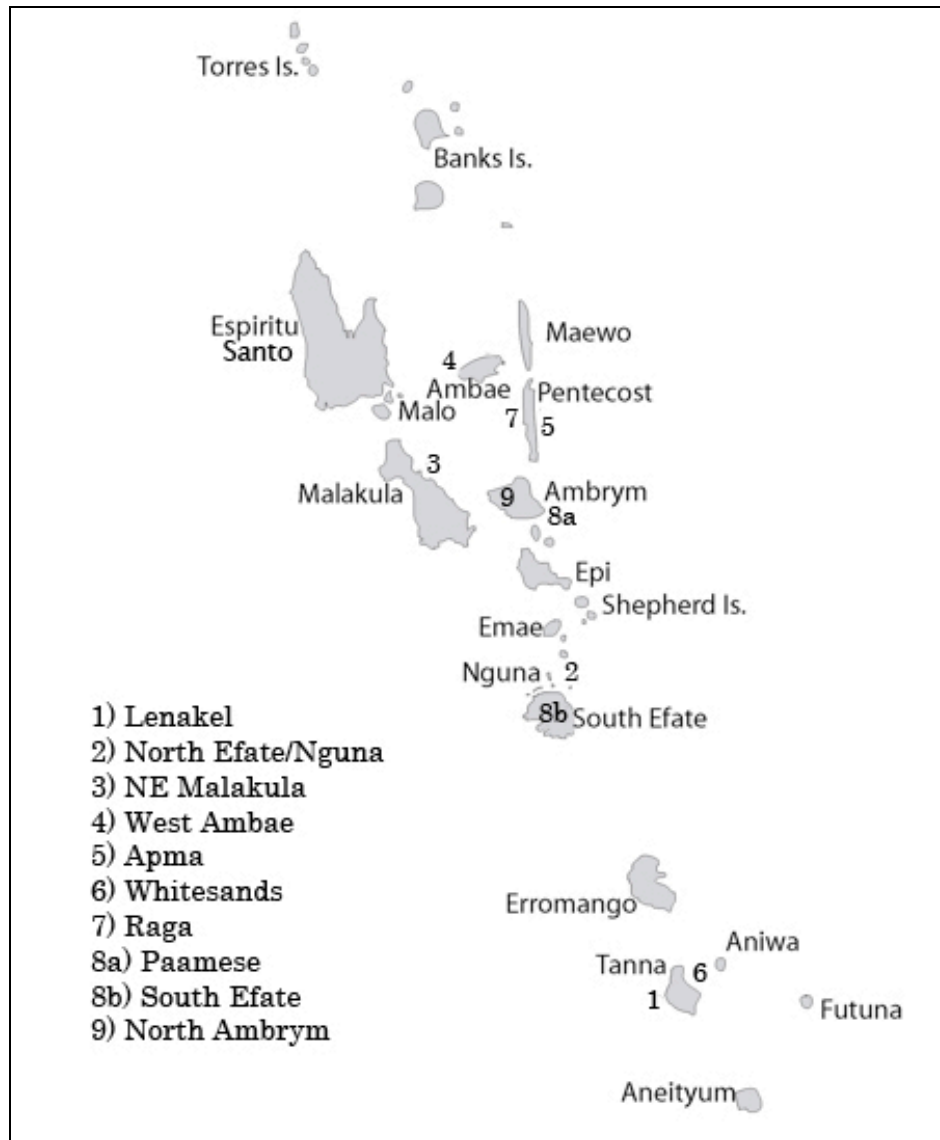
Map 13 The ten largest languages of Vanuatu

Perhaps the most striking general fact about the languages of Vanuatu is their extremely small size. In July, 2005, the population of Vanuatu was given as 205,754, or an average of about 1,923 speakers per language. This is far lower even than Papua New Guinea (population 5,545,268 as of July, 2005, with about 750 indigenous languages = 7,394 speakers per language), or the Solomon Islands (population 538,032, with 84 native AN and non-AN languages = 6,405 speakers per language). By world standards, or even Pacific Island standards, there are no large languages in Vanuatu. With some 11,500 speakers in 2001 Lenakel (Tanna Island) is the largest language in the country. There is no clear geographical bias for either the largest or smallest languages in Vanuatu.

Typological overview

The languages of Vanuatu are known for certain unusual phonetic traits. Some 15-18 languages of central Vanuatu have a series of apicolabial consonants produced by touching the tip of the tongue against the upper lip; these maximally include stops, nasals and

fricatives. In some languages apicolabials occur only before non-back vowels; in others they appear to be contrastive. Many languages in Vanuatu have lost POC *-VC, creating an atypically high frequency of monosyllabism. All languages of Vanuatu appear to be SVO, and serial verb constructions are common. The following sentences from Mota in the far north and Anejom in the far south illustrate aspects of the syntax. For typographical convenience the labiovelar nasal is written *m*^{*w*}:

Mwotlap (François 2003b)

1.   no-sot mino ne-mhay 'My shirt is torn'
      art-shirt my **stat**-tear

2.   na-mnē-k me-lem 'I have gotten my hands dirty'
      art-hand-**1sg perf**-dirty

3.   kēy m-il-il lawlaw n-ēm^w nonoy 'They have painted their house red'
      3pl perf-paint.**red** red **art**-house their

4.   ige su-su me-gen nō-mōmō na-kis 'The children have eaten my share of the fish!'
      hum small.**red rf**-eat **art**-fish **art-pc.1sg**

5.   kēy ma-galeg n-ēm^w mino vitwag 'They have built a house for me'
      3pl perf-build **art**-house my one

6.   ave na-gasel mino? Agōh, nok tēy tō agōh
           where **art**-knife my. **dx.1 1sg** hold **pr.st dx.1** 'Where is my knife? Here, I have it in my hand!'

7.   nēk te-mtiy tō en, togtō nēk te-mtewot vēste
      2sg cf1-sleep **cf2 coe**, then.**cf 2sg pot1**-wounded **pot2.neg** 'If you had slept you couldn't have gotten hurt'

8.   kē ta-van me qiyig 'I will go today'
      1sg fut-go **vtf** today

Anejom (Lynch 2000)

1.   is itiyi apan aan a naworitai iyenev 'S/he didn't go to the garden yesterday'
      3sg.p neg go **3sg loc** garden yesterday

2.   hal halav jek era amjeg 'There are some kids sleeping there'
      some children exist.**pl 3pl.aor** sleep

3.   Era ahagej a elpu-taketha 'The women are foraging for shellfish'
      3pl.aor forage.for.shellfish **as pl**-woman

4.   era m^wan atge-i pikad alp^was itii aara 'They killed that big pig'
      3pl.aor perf kill-**tr** pig big **dem.an.sg 3pl**

5.   is atce-ñ inhat aan 'S/he punched the rock'
      3sg.p punch-**tr** stone **3sg**

6.   ek pu idim apan añakVila a intah noup^wan 'I really must go to Vila some time'
      1sg.aor fut really go **1sg** Vila **loc indef** time

7.   is ika aan is pu apam plen imrañ 'He said that the plane would come tomorrow'
      3sg.p say **3sg 3sg.p fut** come plane tomorrow

8.   et amjeg a kuri itac a niom^{w} 'The dog is sleeping behind the house'
      3sg.aor sleep **as** dog behind **loc** house

2.4.13 The languages of New Caledonia and the Loyalty Islands

The relatively large island of New Caledonia, with an area of 19,100 square kilometres, lies some 1,600 kilometres due east of the Great Barrier Reef and the coast of Queensland, Australia. The three main islands of the Loyalty group (Uvea, Lifou, and Mare) lie about 160 kilometres due east of New Caledonia. The entire group, apart from the Belep Islands north of the main island, lies below 20 degrees south latitude, and the Isle of Pines, off the southern tip of New Caledonia, lies just within the Tropic of Capricorn. Lewis (2009) gives 40 languages for New Caledonia and the Loyalty Islands, of which two are extinct. However, this includes French, Javanese, Vietnamese, Bislama, and the French-based creole Tayo (also called Caldoche or Kaldosh), as well as the post-colonial Polynesian introductions Tahitian, East Futunan and Wallisian. The number of native AN languages is thus 34, of which two are extinct.

Brief history of research

New Caledonia and the Loyalties form an overseas territory of France, and French scholars have made most of the important contributions to our knowledge of the languages of this area. The first systematic linguistic survey of New Caledonia and the Loyalty Islands was that of the missionary Maurice Leenhardt (1946), who provided a comparative vocabulary of 1,165 words for 36 language communities, together with a 240-page overview of grammatical systems. The intent of this survey was to provide basic typological and lexical data on many languages that had previously not been described. This was followed by a number of articles by A.G. Haudricourt that were primarily concerned with historical change, and with determining the position of these very divergent languages within the AN language family, as well as by a short descriptive monograph (Haudricourt 1963). The work of preparing more detailed grammatical descriptions and dictionaries of the languages has continued in more recent years, largely (but not exclusively) through the efforts of French scholars trained by Haudricourt. Notable contributions include grammars of Xârâcùù (Moyse-Faurie 1995), Nyelâyu (Ozanne-Rivierre 1998), and Nêlêmwa (Bril 2002), dictionaries of Paicî (Rivierre 1983), Iaai (Ozanne-Rivierre 1984), Xârâcùù (Moyse-Faurie and Néchérö-Jorédié 1986), Cèmuhî (Rivierre 1994), and Nêlêmwa-Nixumwak (Bril 2000), and comparative studies by Rivierre (1973), and Haudricourt and Ozanne-Rivierre (1982). Contributions from outside the French tradition include grammars of Nengone (Tryon 1967a), Dehu (Tryon 1968a), Iaai (Tryon 1968b), and Tinrin (Tîrî) (Osumi 1995), and dictionaries of Dehu (Tryon 1967b), Nengone (Tryon and Dubois 1969), Canala/Xârâcùù (Grace 1975), and Grand Couli (Grace 1976). As a result of these contributions New Caledonia and the Loyalties have passed from an era of linguistic obscurity prior to 1946, and of drastic underdescription prior to the late 1960s, into one of relative richness of linguistic documentation.

Language distribution

Table 2.15. lists the ten largest and ten smallest languages of New Caledonia and the Loyalty Islands (PN = Polynesian, LI = Loyalty Islands Family, WMP = Western Malayo-Polynesian, NNC = North New Caledonia Family, SNC = South New Caledonia Family).

Table 2.15 Largest and smallest languages of New Caledonia and the Loyalty Islands

No.	Language	Subgroup	No. of speakers
1.	Wallisian/East Uvean	PN	19,376 (2000)
2.	Dehu	LI	11,338 (1996)
3.	New Caledonian Javanese	(WMP)	6,750 (1987)
4.	Nengone	LI	6,500 (2000)
5.	Paicî	NNC	5,498 (1996)
6.	Iaai	LI	4,078 (2009)
7.	Ajië	SNC	4,044 (1996)
8.	Xârâcùù	SNC	3,784 (1996)
9.	East Futunan	PN	3,000 (1986)
10.	West Uvean/Fagauvea	PN	2,219 (2009)
1.	Tîrî	SNC	264 (1996)
2.	Neku	SNC	221 (1996)
3.	Pwaamei	NNC	219 (1996)
4.	Pije	NNC	161 (1996)
5.	Vamale	NNC	150 (1982)
6.	Haeke	NNC	100 (1996)
7.	Arhö	SNC	62 (1996)
8.	Arhâ	SNC	35 (1996)
9.	Pwapwa	NNC	16 (1996)
10.	Zire	SNC	4 (1996)
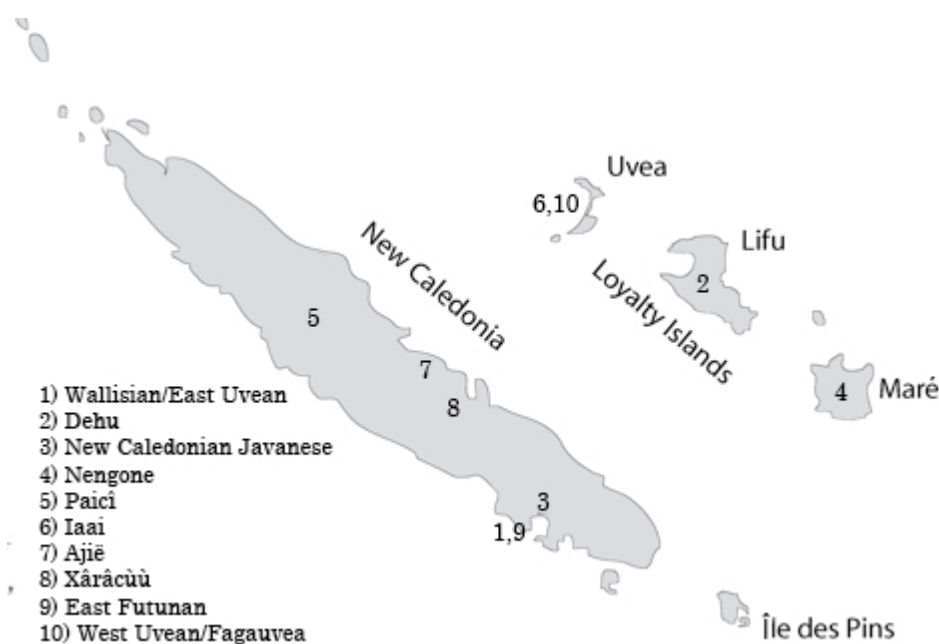
Map 14 The ten largest languages of New Caledonia and the Loyalty Islands

In general, the languages of this region are quite small. Two of the four largest languages are spoken in the Loyalties, and two others were introduced by immigrant laborers from French Polynesia (Wallisian), and Indonesia (New Caledonian Javanese). The largest native New Caledonian language is thus Paicî, with about 5,500 speakers in 1982. Unlike the Solomons, where 93 percent of the population is of Melanesian descent, or Vanuatu, where this figure reaches 98 percent, only 42.5% of the 216,494 residents of New Caledonia and the Loyalties in July, 2005 were of Melanesian extraction. This gives a native population of about 92,000 with 32 languages, hence an average language size of about 2,875, the smallest in the Pacific apart from Vanuatu.

Typological overview

The first unusual typological trait that is likely to be noticed in New Caledonia is the presence of phonemically contrastive tone in five languages spoken in the south of the island (Rivierre 1993). Other unusual typological traits are the occurrence of phonemically nasalised vowels, very large vowel inventories (Haudricourt 1971:377 reports nineteen contrastive vowels in Kunye, spoken on the Isle of Pines), a contrast of palatal, cacuminal and dental consonants, voiceless nasals, and very long word shapes in some languages, as in Nengone of the Loyalty Islands, where most lexical bases are four syllables or longer. All languages of New Caledonia and the Loyalties appear to be SVO. Sample sentences from Xârâcùù in southern New Caledonia and Nengone of Maré Island in the Loyalty group, follow:

Xârâcùù (Moyse-Faurie 1995)

1.   mʷââ-rè xöru 'His house is beautiful'
      house-**3sg.poss** beautiful

2.   è da na xöö-dö amû ŋê chêêdê 'Last night he ate some eggs'
      3sg eat **p** egg-chicken yesterday during evening

3.   (mè) bêêri bʷa nä toa rè ŋê xaa-mêgi
      irr old.man that **dur** come **dur** in season-hot 'The old man surely will return in the summer'

4.   anîî mè ke nä fè rè 'When will you leave?'
      when that **2sg dur** go **dur**

5.   è xwi bachéé daa mè péépé paii 'Bébé has been sick for three days'
      3sg make three day that Bebe sick

6.   dèèri nä sii kʷé rè töwâ xiti 'The people will not dance at the feast'
      people **dur neg** dance **dur** during feast

7.   wîna chaa ùbʷa nèxêê-sê röö! 'Leave a place for your sister!'
      leave one place **poss**-sister **dtr.2sg**

8.   nâ fadù dèèri ŋê ääda 'I'm dividing the food among the people'
      1sg divide people **rltr** food

Nengone (Tryon 1967a; however, I follow the orthography of Tryon and Dubois 1969, which is based on that of the London Mission Society Nengone Bible)

1. bon ha taŋo hnɛn ɔre du 'He was killed by the sun'
   3sg pres die by **art** sun

2. inu deko me ered 'I do not fight'
   1sg neg if fight

3. nidi nia kore retok 'The chief is very bad'
   very bad **art** chief

4. numu rue wanata 'There are two stories'
   exist two story

5. ci ule kore retok ore so wakoko 'The chief sees a heap of yams'
   pres see **art** chief **so** heap yam

6. buic ci eton wenore bone ha sic 'They are asking why he fled'
   3pl pres ask why **3sg past** flee

7. inu co ruaban ɔre mma bane so bɔn 'I shall clean the house for her'
   1sg fut arrange **art** house so.that for **3sg**

The morpheme glosses provided by Moyse-Faurie suggest that some Xârâcùù sentences are at least partly calqued on the French originals, and it is possible that some New Caledonian languages are converging toward French (in which virtually the entire population is fluent).

2.4.14 The languages of Micronesia

Micronesia was settled through at least four historically distinct migrations of AN speakers. Most speakers of Oceanic languages appear to have entered the area via Kiribati-Nauru, probably around 3,000 years ago, and from there they spread steadily westward until they reached Sonsorol, Tobi, and Mapia, nearly completing a circle back to the region of the Proto Oceanic homeland. Yapese may have reached Micronesia from the Admiralty Islands soon after the breakup of Proto Oceanic. Unlike these languages, Palauan and Chamorro reached their historical locations from the west as a result of separate migrations out of insular Southeast Asia. In the case of Chamorro the movement was almost certainly from the central or northern Philippines, and took place before 3,500 B.P (Blust 2000c). Lewis (2009) lists 25 languages for this region, most of them spoken in the Federated States of Micronesia.

Brief history of research

Lingua Mariana, a description of Chamorro in a classic Latin mold by the Spanish cleric Diego Luis de Sanvitores (1668), is the earliest grammar for any language of the Pacific, and one of the earliest grammars for any AN language. Despite this contribution little was known about most of the languages of Micronesia until after the conclusion of the Second World War, when the entire area became a Trust Territory of the United Nations under American administration.

The languages of Micronesia are now relatively well-studied. In large part this is due to a concerted effort made by scholars at the University of Hawai'i from roughly the late 1960s until the mid 1980s, when federal funding was available to support fieldwork and the writing of a number of grammars and dictionaries of previously undescribed or

underdescribed languages. These include pedagogical lessons and a large dictionary of Marshallese (Bender 1969b, Abo, Bender, Capelle and DeBrum 1976), pedagogical lessons, a reference grammar and a dictionary of Chamorro (Topping 1969, 1973, Topping, Ogo, and Dungca 1975), a reference grammar and a dictionary of Palauan (Josephs 1975, McManus and Josephs 1977), a reference grammar and dictionary of Kosraean (formerly called 'Kusaiean'; Lee 1975, 1976), a reference grammar and dictionary of Pohnpeian (formerly called 'Ponapean'; Rehg 1981, Rehg and Sohl 1979), a reference grammar and dictionary of Mokilese (Harrison 1976, Harrison and Albert 1977), a dictionary and grammar of Puluwat (Elbert 1972, 1974), a reference grammar and dictionary of Woleaian (Sohn 1975, Sohn and Tawerilmang 1976), a grammar of Ulithian (Sohn and Bender 1973), a reference grammar and dictionary of Yapese (Jensen 1977a, b), and a dictionary of Carolinian (Jackson and Marck 1991). A number of comparative studies also appear in Bender (1984), and a large set of Proto Micronesian reconstructions has recently been published in article form (Bender, et al. 2003a,b). Important work on Micronesian linguistics published outside the Hawai'i tradition includes a sketch grammar of Chuukese (formerly called 'Trukese'; Dyen 1965b), a large Chuukese-English dictionary (Goodenough and Sugita 1980, 1990), a grammar of Nauruan (Kayser 1993), a grammar of Marshallese (Zewen 1977), a dictionary of Gilbertese (Sabatier 1971), and two grammars of Chamorro, one in a traditional descriptive mold (Costenoble 1940), the other a highly theory-informed account (Chung 1998).

Table 2.16 The languages of Micronesia

No.	Language	Subgroup	No. of speakers
1.	Kiribati/Gilbertese	NMC	107,817 (2007)
2.	Chamorro	(WMP)	92,700 (2005)
3.	Marshallese	NMC	59,071 (2005)
4.	Chuukese/Trukese	NMC	48,170 (2000)
5.	Pohnpeian/Ponapean	NMC	29,000 (2001)
6.	Palauan	(WMP)	20,303 (2005)
7.	Nauruan	NMC?	7,568 (2005)
8.	Kosraean/Kusaiean	NMC	8,570 (2001)
9.	Yapese	OC	6,592 (1987)
10.	Mortlockese	NMC	6,911 (2000)
11.	Ulithian	NMC	3,000 (1987)
12.	Carolinian	NMC	3,000 (1990)
13.	Kapingamarangi	PN	3,000 (1995)
14.	Pingilapese	NMC	2,500 (?)
15.	Woleaian	NMC	1,631 (1987)
16.	Puluwat	NMC	1,707 (2000)
17.	Pááfang/Hall Islands	NMC	2,142 (2000)
18.	Mokilese	NMC	1,050 (1979)
19.	Nómwonweité/Namonuito	NMC	1,341 (2000)
20.	Nukuoro	PN	860 (1993)
21.	Sonsorol	NMC	600 (1981)
22.	Satawalese	NMC	458 (1987)
23.	Tobi	NMC	22 or more (1995)

Language distribution

Table 2.16. lists the languages of Micronesia (NMC = Nuclear Micronesian, WMP = Western Malayo-Polynesian, OC = Oceanic, without further specification, PN = Polynesian):

1) Kiribati
2) Chamorro
3) Marshallese
4) Chuukese/Trukese
5) Pohnpeian/Ponapean
6) Palauan
7) Nauruan
8) Kosraean/Kusaiean
9) Yapese
10) Mortlockese
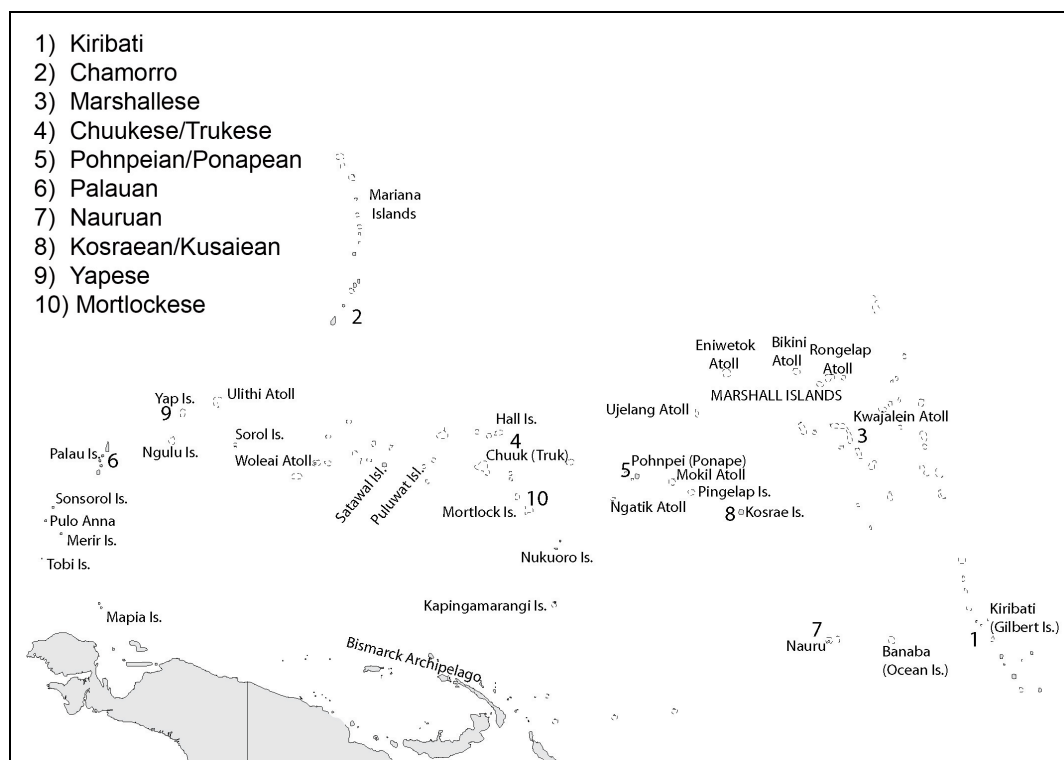
Map 15 The ten largest languages of Micronesia

Typological overview

In view of their disparate origins there is no reason to expect typological uniformity in the languages of Micronesia. Yapese, almost alone among AN languages, has a rich inventory of glottalised consonants. The Nuclear Micronesian (NMC) languages show far more surface vowels than is typical of Oceanic languages, where a five-vowel system is the norm. Chuukese and Puluwat are said to have nine vowel phonemes, and Pohnpeian seven. Marshallese has twelve surface vowels, although Bender (1968) has shown that these can be reduced to four underlying vowels specified only for height. Like many of the languages of Vanuatu, but unlike most other Oceanic languages, NMC languages have lost both final consonants and the vowel that preceded them. Many lexical bases are thus monosyllabic, although the remaining vowel is lengthened to maintain a bimoraic base, as in Chuukese iim^w (< *Rumaq) 'house', next to im^wa-n (< *Rumaq-na 'his/her house').

NMC languages often show complex vowel alternations. Palauan also has a complex phonology, but one that revolves around consonant alternations more than vowel alternations. In both cases the complexity of the synchronic phonology is directly related to types of historical change: NMC languages have complex vocalic histories, and Palauan a complex consonantal history. Syntactically, NMC languages and Palauan are SVO. Chamorro word order is complex, showing both predicate-initial and predicate-medial constructions. The verb morphology of Chamorro is typologically similar to that of Philippine languages and in some cases is historically connected through cognate affixes,

although Chamorro does not subgroup with any language of the Philippines. The following sentences from Pohnpeian, Chamorro and Palauan illustrate the range of typological variation within Micronesia. Rehg (1981) labels (V) an 'excrescent vowel,' and McManus and Josephs (1977:3) call Palauan *a* a 'contentless word which precedes nouns and verbs under various conditions.'

Pohnpeian (Rehg 1981)

1. me-n-(a)-kan o:la 'Those are broken'
    thing-there.by.you-(V)-**pl** broken

2. e wie dɔdɔ:k wasa:-t met 'He is working here now'
    3sg aux work place-this now

3. e pa:n taŋ-(a)-da 'He will run upward'
    3sg irr run-(V)-upward

4. i pʷai-n deŋki pʷɔt 'I bought a flashlight'
    1sg buy-**tr** flashlight **clas**

5. i pa:n lopuk-e:ŋ o:l-o se:u 'I will cut sugarcane for that man'
    1sg irr cut.crosswise-for man-that sugarcane

6. me:n Pohnpei inenen kadek 'Pohnpeians are very kind'
    one.of Pohnpei very kind/generous

7. e me-meir ansou me na: seri-o pʷupʷi-di-o
    3sg dur-sleep time that his child-that fall-down-there 'He was sleeping when his child fell down'

8. ke de:r wa:-la pʷu:k-e pa:-o 'Don't take this book down there!'
    2sg neg carry-there book-this down-there

Chamorro (Topping 1973)

1. dánkolo si Juan 'Juan is big'
    big **art** Juan

2. g<um>u-gupu i paluma 'The bird is flying'
    fly-**af.red art** bird

3. hu liʔeʔ i dánkolo na taotao 'I saw the big person'
    1sg see **art** big **lig** person

4. ha gimen si Juan i hanom 'Juan drank the water'
    3sg drink **art** Juan **art** water

5. todu i tiempo ma-cho-choʔchoʔ gueʔ duru 'All the time he works hard'
    all **art** time **vbl-red** work **3sg** hard

6. ha hatsa i lamasa chaddek 'He lifted the table quickly'
    3sg lift **art** table quick

7. matto gi gipot nigap 'He came to the party yesterday'
    come **loc** party yesterday

8. si Juan l<um>iʔeʔ i palaoʔan 'Juan is the one who saw the woman'
    art Juan see-**af art** woman

9.   l\<in\>iʔeʔ ni lahi ni palaoʔan 'the man saw the woman'
      see-**pf art** man **art** woman

10.   si Paul ha sangan-i si Rita ni estoria 'Paul told the story to Rita'
      art Paul **3sg** tell-**rf art** Rita **art** story

Palauan (Josephs 1975)

1.   a buík a r\<əm\>urt 'The boy is running'
      a boy **a** run-**av**

2.   a buík a r\<əm\>urt ər a serse-l a Droteo 'The boy is running in Droteo's garden'
      a boy **a av**-run in **a** garden-**3sg.poss a** Droteo

3.   a Toki a mə-latəʔ ər a ulaol 'Toki is cleaning the floor'
      a Toki **a av**-clean **def a** floor

4.   a sensei a ol-sebək ər a rəŋu-l a Droteo
      a teacher **a caus**-worry **def a** feeling-**3sg.poss a** Droteo 'The teacher is worrying
      Droteo (without meaning to)'

5.   ak ou-ʔad ər kəmiu e ak mo ʔəbuul
      1sg have-person because.of **2pl.emph** and **1sg fut** poor 'Having you as my
      relatives will reduce me to poverty'

6.   ak ou-ŋalək ər a səʔal 'I am the parent of a boy'
      1sg have-child **so a** male

7.   a blai el lurruul ər ŋii a Droteo a mle klou
      a house which build **so 3sg.emph a** Droteo **a** was big 'The house that Droteo built
      was big'

8.   tia kid a blsibs el ləti-lobəd ər ŋii a beab
      here **emph a** hole from.which emerge **so 3sg.emph a** mouse 'Here's the hole that
      the mouse came out of'

2.4.15   *The languages of Polynesia, Fiji and Rotuma*

The region considered here includes the Polynesian triangle, the Fijian Archipelago just to the west, and the tiny island of Rotuma, some 550 kilometres north of Vanua Levu in the Fiji group. Although this area is almost twice that of the continental United States, the interrelationship of Polynesian languages was recognised immediately by early European explorers of the Pacific, as during the three voyages of James Cook from 1768-1779. Even to the layman the common origin of the Polynesian languages is about as transparent as that of the Romance languages—a product of their relatively recent divergence from a common ancestral form within the past 2,000-2,500 years. Although most Polynesian languages are spoken within the Polynesian triangle, Kapingamarangi and Nukuoro are found in Micronesia, and some 13 other Polynesian Outliers in Melanesia. These include three languages in Papua New Guinea (Nuguria, which may be extinct, Nukumanu, Takuu), six languages in the Solomon and SantaCruz Islands (Rennell-Bellona, Luangiua/Ontong Java, Sikaiana, Tikopia, Anuta, Pileni), three in Vanuatu (Mae/Emae, Mele-Fila or Ifira-Mele, Futuna-Aniwa), and one in the Loyalty Islands (Faga Uvea/West Uvea). Biggs (1971:480-81) lists 48 Polynesian 'communalects', which he groups into 26 languages exclusive of the extinct Moriori of the Chatham Islands. This probably is the

most accurate count available. Fijian and Rotuman are somewhat more distant relatives of the Polynesian group.

Brief history of research

The earliest Polynesian vocabularies were collected during the voyage of the Dutch navigator Jacob LeMaire in 1616. It was not until the voyages of James Cook from 1768-1779, however, that material was collected from enough languages to recognise the geographical extent of the Polynesian subgroup. Because of its transparency even to untrained observers, the close relationship of the Polynesian languages was recognised before extensive descriptive material was available for any of the languages. Grammars and dictionaries of Polynesian languages began to appear by the middle of the nineteenth century, and today the Polynesian group is one of the best-studied in the AN family. In addition to such detailed grammars as Mosel and Hovdhaugen (1992) for Samoan, Bauer (1993) for Maori, Besnier (2000) for Tuvaluan, Elbert and Pukui (1979) for Hawaiian, and others following a more traditional approach, a number of languages are represented by large dictionaries, for example, Tongan (Churchward 1959), Niuean (Sperlich 1997), Samoan (Milner 1966), Maori (Williams 1971), Hawaiian (Pukui and Elbert 1971), Futunan (Moyse-Faurie 1993), Rarotongan (Savage 1980), Nukuoro (Carroll and Soulik 1973), Kapingamarangi (Lieber and Dikepa 1974), Rennellese (Elbert 1975), and Tikopia (Firth 1985). Although some earlier writers had commented on Fijian dialect diversity, a broad division into two distinct languages, Western Fijian and Eastern Fijian, was not clearly stated until Pawley and Sayaba (1971). Geraghty (1983) further refined the dialect picture in Fiji, and provided a pivotal theory of the history of Fijian dialects in relation to the Polynesian languages by appealing to a novel model of language split which Pawley (1996a) extended to Rotuman. Most work on Fijian has described the standard dialect of Bau. The major grammar of standard Fijian is Schütz (1985); the major bilingual dictionary is Capell (1968), but a far larger monolingual dictionary has recently appeared (Institute of Fijian Language and Culture 2007). Dixon (1988) is a major grammar of one (Eastern) non-standard dialect, and Pawley and Sayaba's (2003) dictionary of Wayan (Western Fijian) is larger and more detailed than any existing bilingual dictionary of the standard language. Although it has attracted considerable attention from general theoreticians, Rotuman is on the whole less well-described than Fijian or the larger Polynesian languages.

Language distribution

Table 2.17 lists the languages of Polynesia, Fiji, and Rotuma in descending order of size (EL = Ellicean, FJ = Fijian, TN = Tongic, EP = Eastern Polynesian, NP = Nuclear Polynesian, RT = Rotuman; the precise degrees of relationship between Fijian, Rotuman and Polynesian to one another remains uncertain, so it is assumed here that the three form coordinate branches of a Central Pacific group):

Table 2.17 The languages of Polynesia, Fiji, and Rotuma

No.	Language	Subgroup	No. speakers
01.	Fijian, Eastern	CP	330,441 (1996)
02.	Samoan	PN	217,938 (2005)
03.	Tahitian	PN	117,000 (1977)
04.	Tongan	PN	112,422 (2005)
05.	Maori	PN	60,000 (1991)
06.	Fijian, Western	CP	57,000 (1977)
07.	Rarotongan	PN	42,669 (1979)
08.	Wallisian	PN	29,768 (2000)
09.	Tuamotuan	PN	14,400 (1987)
10.	Tuvaluan	PN	11,636 (2005)
11.	Rotuman	CP	11,500 (1996)
12.	Marquesan, Northwest	PN	3,400 (1981)
13.	Rakahanga-Manihiki	PN	2,500 (1981)
14.	Rapanui/Easter Island	PN	2,450 (1982)
15.	Niuean	PN	2,166 (2005)
16.	Marquesan, Southeast	PN	2,100 (1981)
17.	Mangarevan	PN	1,600 (1987)
18.	Tokelauan	PN	1,405 (2005)
19.	Hawaiian	PN	1,000 (1995)
20.	Pukapukan	PN	840 (1997)
21.	Penrhyn/Tongareva	PN	600 (1981)
22.	Rapa	PN	521 (1998)
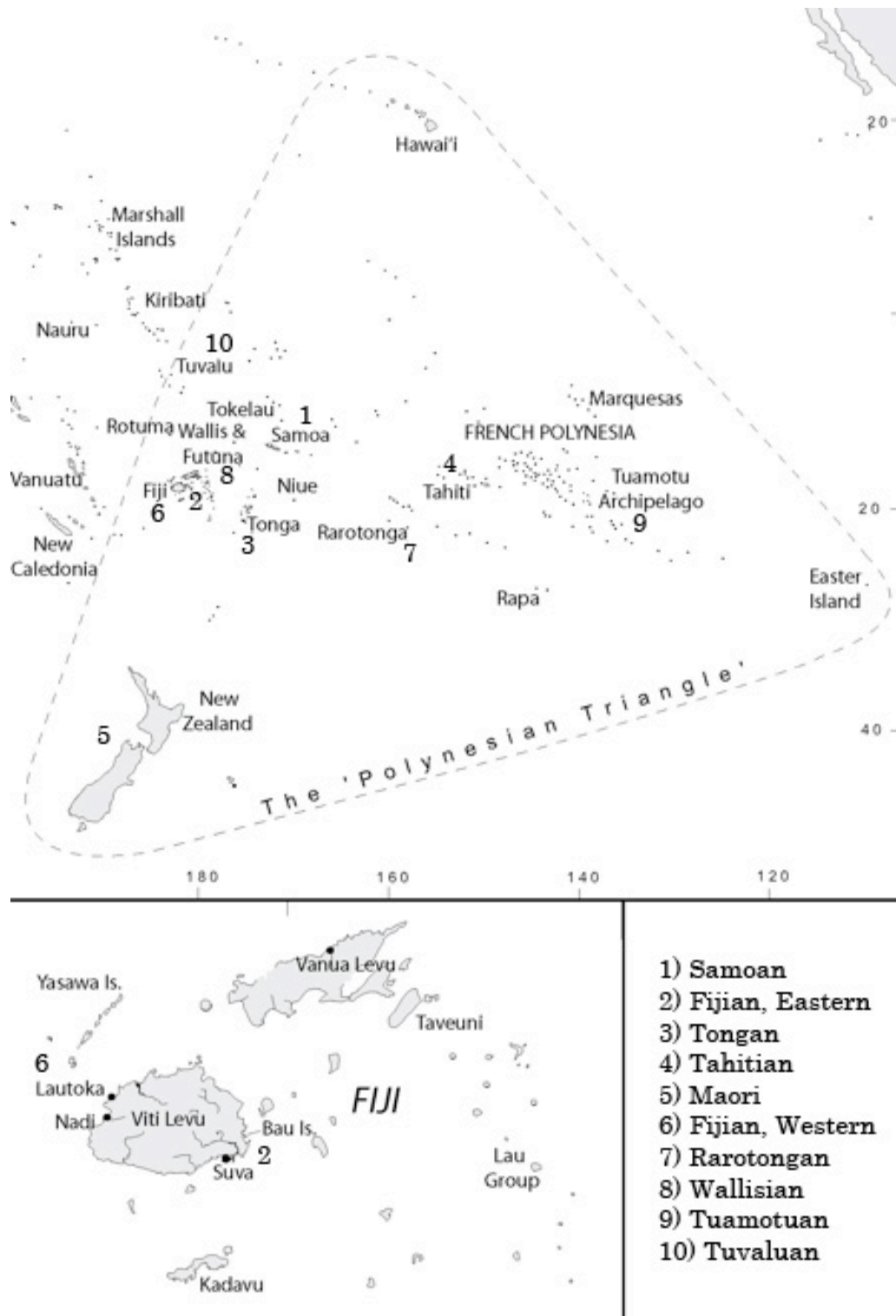
Map 16 The ten largest languages of Polynesia, Fiji, and Rotuma

The largest language in the Polynesia-Fiji-Rotuma region is Fijian. Estimates from mid-2005 placed the number of Fijian speakers at over 455,000. However, this number does not distinguish Eastern Fijian and Western Fijian, which generally are regarded as different languages. Lewis (2009) also recognises Gone Dau (500 speakers in 1977) and Lauan (16,000 speakers in 1981) as distinct Fijian languages, but most writers regard these as

dialects of Eastern Fijian. The number of Rotuman speakers is difficult to estimate, as they are dispersed. According to Lewis (2009) about 9,000 Rotumans resided in Fiji in 1991, while the population of the home island apparently has been relatively stable at around 2,500 over the past several decades.

Lewis (2009) gives 16,800 speakers of Rarotongan in the Cook Islands as of 1979, but 43,000 overall. The great majority of these, and hence probably the single largest community of Cook Islanders, live in New Zealand. As of 2003 the population of the Cook Islands was estimated at 21,008, with almost 90 percent of these living on the southern islands, and over half on the island of Rarotonga. These figures suggest that there has been little population increase in the Cook Islands over the past quarter of a century, probably as a result of extensive out-migration. The figure of 29,768 speakers of Wallisian does not distinguish Wallisian (East Uvean) from Futunan, as the two islands form part of a single administrative unit. Well over half of these live in New Caledonia. The figure of 1,000 speakers of Hawaiian in 1995 refers to 'mother tongue' speakers, most of whom would be residents of the small and privately owned island of Ni'ihau. Some 8,000 speakers are reported overall, most of whom have acquired Hawaiian in classroom settings as part of the 'Hawaiian renaissance,' a concerted effort made since the 1980s to revive the Hawaiian language through pedagogy, the media and periodic cultural events. Finally, Lewis (2009) gives 8,000 speakers of Tubuai-Rurutu (Austral Islands, French Polynesia) in 1987, but it is clear from remarks in Biggs (1971) that virtually all of these must be speakers of 'Neo-Tahitian', the *lingua franca* of much of southeastern French Polynesia. The linguistic situation on the isolated island of Rapa, the southernmost island in the Austral group, probably is the same as that in the northern Austral Islands, although—unlike the situation in Rurutu or Tubuai—the native language of Rapa was still spoken alongside Neo-Tahitian in 1921.

Typological overview

The Polynesian languages are well known for their small phoneme inventories, and exclusive use of open syllables, sometimes leading to sequences of up to four vowels within a word, as in Hawaiian *ʔaiau* 'look about with covetous or jealous eyes', or *iāia* 'him, her; to him, to her'.[15] Fijian is noteworthy for its prenasalised alveolar trill (written *dr*), and Rotuman for its use of systematic metathesis in grammatical paradigms. Syntactically, Fijian and the languages of Triangle Polynesia are predicate-initial, but Rotuman and some of the Polynesian Outlier languages are predicate-medial. The Polynesian languages divide almost equally into ergative and accusative types, yet use cognate grammatical morphemes in their verb systems. Even without further information this suggests that typological shift has occurred through reinterpretation of existing grammatical markers as a result of small changes in their distributional properties. The following sentences illustrate some general syntactic properties of Fijian and Hawaiian:

---

[15]   In further illustration of this point Andrew Pawley (p.c.) has pointed out that the late Bruce Biggs liked to cite the Maori sentence /I auee ai au i aa ia i aa ai i aua ao/ 'I lamented while he drove away those clouds'.

Fijian (Schütz 1985, with orthographic modifications suggested by the author)

1.   sā moce na ŋone 'The child is sleeping'
        asp sleep **def** child

2.   au na baci saba-k-a na mata-mu 'I'll slap your face'
        1sg fut itr slap-**tr-3sg def** face-**2sg**

3.   au sā ŋalu tū ŋā 'I'm keeping quiet (now)'
        1sg asp dumb **cont int**

4.   ni-u dau rai-ci koya, au taŋi 'Whenever I see her I cry'
        sub-1sg hab see-**tr 3sg 1sg** cry

5.   e tuku-n-a sara ko tata me keitou laki vaka-tā-kākana ki matāsawa
        3sg tell-**tr-3sg int prp** Father **sub dl.excl dir** picnic **abl** beach 'Father said the two
        of us (excl.) could go to the beach for a picnic'

6.   erau dau ŋgoli e na nō-drau tolo ni bāravi
        3dl hab fish **abl def poss-3dl** middle **poss** beach 'The two of them always fish in
        their stretch of beach'

7.   era cavutū mai na vei-vanua tani 'They have just visited different countries'
        3pl go.together **abl def dis**-land different

8.   e buta vinaka sara na lewe ni lovo
        3sg cooked good **int def** contents **poss** oven 'The contents of the earth oven were
        cooked thoroughly'

Hawaiian (Elbert and Pukui 1979)

1.   ua hele ke kanaka i Maui 'The man went to Maui'
        perf go **art** man to Maui

2.   ua kākau Pua i ka leka 'Pua wrote the letter'
        perf write Pua **obj art** letter

3.   ua kākau ʔia ka leka e Pua 'The letter was written by Pua'
        perf write **pass art** letter by Pua

4.   i ke ahiahi ka pāʔina 'The party is in the evening'
        in **art** evening **art** party

5.   ua noho ke kanaka i Hilo 'The man stayed at Hilo'
        perf stay **art** man at Hilo

6.   maikaʔi ka wahine i kāna mau hana pono
        good **art** woman **purp 3sg.poss** constant deed righteous 'The woman is good
        because of her righteous deeds'

7.   aia ke kumu ma Maui   'There is a teacher on Maui'
        exist **art** teacher on Maui

8.   he kaʔa ko-na 'He has a car'
        art car **psm-3sg**

9.   he kaʔa n-o-na   'There's a car for him'
        art car for-**psm-3sg**

2.5   Overview of language size and descriptive coverage

Table 2.18 provides a gross estimate of differences in language size across geographical areas, based on the ten largest and ten smallest languages as counted in the preceding survey. Madagascar is omitted, as it would distort the values for Borneo and cannot meaningfully be counted alone. A count for Taiwan is difficult, as only fifteen Formosan languages are still spoken. To approximate comparability with the other data I have divided these into two groups (of eight and seven respectively) based on reported number of speakers. Other complicating factors that cannot be controlled include: 1. a clear distinction is not always made between those claiming a given ethnic identity and those who are active speakers of the associated language, 2. the most recent dates for population estimates vary from one language to the next, and 3. since extinct languages are not included in the count, languages on the verge of extinction several years prior to the date of the population estimate may already have passed into an excluded category, thereby increasing the average size of the remaining languages. Numbers of speakers of the largest languages are given in thousands, those of the smallest languages in hundreds.

Table 2.18 Average size of the ten largest and ten smallest AN languages by region

Area		Largest languages	Smallest languages
1.	Taiwan	40.4	7.9
2.	Philippines	5,709	2.3
3.	Borneo	421	1.7
4.	Mainland SE Asia	1,140	37.6
5.	Sumatra, Java, etc.	16,629	201.9
6.	Sulawesi	799	3.95
7.	Lesser Sundas	319	92.5
8.	Moluccas	51	.2
9.	New Guinea	18.3	.6
10.	Bismarcks	12.2	1.6
11.	Solomons	15.9	1.5
12.	Vanuatu	5	.3
13.	New Caledonia, etc.	5.9	1.9
14.	Micronesia	34.3	10.8
15.	Polynesia, etc.	100.6	15.4

These figures confirm some expectations, but upset others. Not surprisingly, the largest languages are concentrated in western Indonesia and the Philippines, the nations with the largest AN-speaking populations. Mainland Southeast Asia also shows a concentration of relatively large languages, but the figures for this area are skewed by the inclusion of peninsular Malay, with over seven million speakers. Apart from Taiwan and Borneo there is a steady decrease in average language size from western Indonesia to Melanesia, and then a sharp upward spike in Micronesia and especially Polynesia-Fiji.

Although the matter cannot be pursued in any exact way here, it is clear that the amount of scholarly attention that a language is likely to receive correlates closely with its size. Areas of descriptive neglect are thus generally those where most languages are small. This is true of the Moluccas, New Guinea, Vanuatu, and the Solomons (where the largest and best-described languages are in the Malaita-Cristobal region). A more precise statement of this correlation would require quantitative data on the number of grammars and

dictionaries that have been published for AN languages in each of the fifteen geographical divisions recognised here. A notable exception is New Caledonia, where over the past two decades French scholars have contributed an unusual amount of high-quality descriptive work on languages that rarely have more than 5,000 speakers.

Considerations of language size also raise the issue of endangerment, a topic usefully surveyed by Florey (2005). Since many languages in Melanesia have fewer than 200 speakers, and apparently have continued at this level for generations, it cannot be argued that language endangerment correlates directly with size (Crowley 1995). Nonetheless, it is widely agreed that size is the most readily identifiable index of endangerment. From this perspective the greatest concentration of endangered AN languages is in the Moluccas, followed closely by Vanuatu and New Guinea. Taiwan, sometimes cited as an area with severely endangered languages has some languages nearing extinction (Pazeh and Thao), but the eight largest languages appear to be doing relatively well. One reason that it is difficult to determine a critical threshold below which a language cannot survive as a social entity is because the outcome of language contraction depends so much on the nature of the socio-economic matrix in which it is embedded. Speakers of Formosan languages are constantly exposed to Taiwanese and Mandarin, just as speakers of Maori or Hawaiian are constantly exposed to English, and in both cases the language of the dominant community confers social and economic advancement. In Vanuatu, where the ten largest languages average only about 5,000 speakers each, and the smallest only 30, there is far less pressure to assimilate to a dominant group. As a result, here and elsewhere in Melanesia it has been possible for many generations to maintain viable (multilingual) language communities with only 100-200 speakers. Austronesian languages with 100-200 speakers in Taiwan, New Zealand or Hawaii simply do not have the same chances of long-term survival that they would have as part of a more balanced multilingual society on some of the larger islands of Melanesia. Ultimately, however, even language communities that are marginalised by world languages such as Mandarin, English, or French will endure or disappear based on the strength of the sociocultural network that binds their speakers together.

3 *Language in society*

3.0 Introduction

Some general sociolinguistic information was given in the first two chapters, interspersed with other material. This chapter aims at a broad overview of the social context of language use in the AN-speaking world, without attempting to exhaust the range of possible topics. In organizing the discussion I have adopted a seven-way division as follows: 1) hierarchy-based speech differences, 2) gender-based speech differences, 3) vituperation and profanity, 4) secret languages, 5) ritual languages, 6) contact, and 7) determinants of language size.[16]

3.1 Hierarchy-based speech differences

Institutionalised differences of social hierarchy are found in many AN-speaking societies, and these are often correlated with specialisations of language use. This subsection is concerned with speech levels in island Southeast Asia, and with respect languages in Polynesia and Micronesia.

3.1.1 Speech levels in island Southeast Asia

Javanese culture is notorious, even among other Indonesians, for its complex patterns of social etiquette. Among the *priyayi*, the cultural elite or courtly noble class of Java, a traditional system governing acceptable patterns of personal interaction has gone beyond mere courtesy to acquire the status of an esthetic ideal. A fundamental distinction governing all judgements of behavior in Java is the contrast of *alus* and *kasar*. Behavior that is *alus* is refined, controlled, quiet and subdued; behavior that is *kasar* is the opposite. Abrupt movements or sudden changes in the pitch of one's voice are frowned upon as likely to startle: at least in the social ideal, everything is calculated to ensure smoothness, gradualness and harmony in one's behavior and social relations. As Geertz (1960:248) noted, "…the Javanese pattern their speech behavior in terms of the same *alus* to *kasar* axis around which they organize their social behavior generally."

The way that verbal behavior is made to satisfy the *alus* ideal in Java is through the use of proper 'speech levels' or, as they are sometimes called, 'speech styles'. As Geertz (1960:248) has pointed out, these are defined largely by lexical choice: "Thus, for 'house' we have three forms (*omah*, *grija*, and *daləm*), each connoting a progressively higher relative status of the listener with respect to the speaker." Some lexical categories, as those represented by pronouns, may have even more gradations of etiquette (four for the second person singular), while others may have fewer, as *di-* and *dipun-* marking the passive. Most of the vocabulary, however, is unmarked for distinctions of status-based speech etiquette,

---

[16]   Fox (2005) covers a number of the same topics that are discussed in this chapter, but with a different emphasis and from a different point of view.

many common nouns remaining unchanged regardless of the relative status of the interlocutors. Geertz notes that although status-sensitive terms form a small subset of the total lexicon, they tend to concentrate in the basic vocabulary.

Javanese speech levels are named. The basic distinction is between *ngoko* and *krama* ([krɔmɔ]), which some writers in addition to Geertz (for example Errington 1988) associate more or less directly with the *kasar/alus* distinction. Robson (2002), on the other hand, suggests that while *krama* speech corresponds to behavior that is *alus*, *ngoko* speech is not inherently crude. Instead, he describes it (Robson 2002:12) as a "basic stratum … the style in which one thinks to oneself and uses to one's intimate family and friends of the same age or younger. We can extend this to include those socially inferior to oneself and those (of whatever social status) whom one wishes to insult." Traditional Javanese social etiquette is complex, and its linguistic correlates reflect this complexity. The social dynamics of some interactional situations are ambiguous, and a choice between *ngoko* or *krama* may be difficult to make. In such situations a more neutral speech style called *madya* ('middle') can be used to avoid awkwardness. In addition to these three speech levels (low, middle, high) there is a special vocabulary called *krama inggil* which is inherently honorific, and is used with reference to anyone deserving of respect. According to Robson (2002:12) 'A second group of words, termed *krama andhap*, is deferential and is restricted to words for 'to accompany', 'to request', 'to offer', and 'to inform'. These are used of oneself in relation to the person for whom respect is due. *Krama inggil* and *krama andhap* are independent of speech style—that is, they can occur in either *ngoko* or *krama* as need be.' Horne (1974:xxxii) describes *krama inggil* in similar terms: "Regardless of which basic style one is using, he draws on a small (around 260-item) Krama Inggil (*ki*) or High Krama vocabulary to show special honor to the person he applies them to. Two schoolboys jabbering in Ngoko about a classmate getting mad would say *nəsu* 'angry,' but of the (respected) teacher losing his temper, they would use the *ki* word *duka* 'angry' (while still speaking Ngoko)." Robson (2002:13) gives the following sentences to illustrate differences in Javanese speech levels, or as he prefers to call them, 'speech styles' (*dh* = voiced retroflex stop):

Table 3.1 Examples of Javanese speech styles

Ngoko (Miyem, aged fifteen, to her younger sister)
Aku wis maŋan səgane 'I have eaten the rice'

Krama (Miyem to her uncle)
Kula sampun nədha səkulipun 'I have eaten the rice'

Krama with krama inggil (Miyem to her uncle about her father)
Bapak sampun dhahar səkulipun 'Father has eaten the rice'

Ngoko with krama inggil (Miyem to her sister about their father)
Bapak wis dhahar səgane 'Father has eaten the rice'

Madya (the old servant to Miyem)
Kula mpun nədha səkule 'I have eaten the rice'

The first pair of sentences, illustrating the basic *ngoko* : *krama* distinction in expressing the same semantic content, is striking in that there are no shared morphemes (*aku/kula* = 'I', *wis/sampun* = 'one, become', *maŋan/nədha* = 'eat', *səga/səkul* = 'rice', plus suffixes). The *madya* style used by the old family servant, on the other hand, appears to be an

abbreviated form of *krama* (*mpun* for *sampun*, *səkule* for *səkulipun*). As noted above, the socially specialised vocabulary of Javanese is much smaller than the general vocabulary. While the *ngoko* lexicon includes tens of thousands of items, Horne (1974:xxxi-xxxii) estimates that the *krama* vocabulary has around 850 terms, and the *madya* vocabulary about 35. In addition, as already noted, there are roughly 260 *krama inggil* terms, which cross-cut the *ngoko-krama* distinction.

In her modern Javanese dictionary Horne (1974) marks all *krama* and *krama inggil* forms, making it easy to compare general linguistic traits for the two styles. This material shows patterns of phonological correspondence between *ngoko* and *krama* for which the literature is generally silent, and suggests that some commonly repeated statements are in need of qualification.[17] Recurrent patterns that have been observed include the following:

(1) where *ngoko* (before the colon) has a back vowel *krama* (after the colon) uses a front vowel in forms that are otherwise usually identical: *agama : agami* 'religion', *akon : akɛn* 'tell someone to do something', *amarga : amargi* 'because', *aŋon : aŋen* 'someone's act of doing', *aŋo : aŋe* 'casual, for casual wear', *aŋon : aŋɛn* 'to herd or tend livestock', *bambu : bambət* 'bamboo', *bubar : bibar* 'disperse', *bubrah : bibrah* 'out of order, in disrepair', *buḍal : biḍal* 'leave, set out (as a group)', *bukak : bikak* 'opened, uncovered', *buŋah : biŋah* 'happy, glad', *ḍuga : ḍugi* 'judgment, common sense', *Jawa : Jawi* 'pertaining to Java', *gunəm : ginəm* 'speech', *kulon : kilɛn* 'west', *kuna : kina* 'old, old-fashioned', *mula : mila* 'originally', *rupa : rupi* 'appearance', *rusak : risak* 'damaged', *umpama : umpami* 'such as, for example', *unḍak : inḍak* 'a rise, increase'.

(2) where *ngoko* has -uCu- *krama* substitutes -əCa-: *buruh : bərah* 'laborer', *butuh : bətah* 'thing needed', *ḍuku : ḍəkah* 'small village', *kudu : kədah* 'ought to, must', *kukuh : kəkah* 'solid, strong', *kumpul : kəmpal* 'group, gathering', *muŋguh : məŋgah* 'in case, in the event that', *muŋsuh : məŋsah* 'opponent, enemy', *rusuh : rəsah* 'unruly', *tuduh : tədah* 'point to', *tuŋgu : təŋga* 'wait for', *tutuh : tətah* 'one who gets blamed'.

(3) where *ngoko* has -*i* or -*iC*, *krama* forms use -*os*: *adi : ados* 'fine, beautiful', *anti : antos* 'expect, wait for', *arti : artos* 'meaning', *ati : atos* 'careful', *batin : batos* 'inward feeling', *dadi : dados* 'be, become', *gati : gatos* 'serious, important', *gənti : gəntos* 'a change, replacement', *jati : jatos* 'teak', *kanti : kantos* 'patient, willing to wait', *kati : katos* 'catty'. Since all recorded examples involve -*di*(*C*) or -*ti*(*C*), it is possible that this pattern is restricted to final syllables that begin with a dental stop followed by *i*.

Other *ngoko : krama* pairs are phonologically unrelated, but patterns (1) to (3) are well-supported, and can be viewed as products of a set of templates that derive *krama* forms from their *ngoko* equivalents, since cognates in other languages always correspond to the *ngoko* form, as with Malay *agama* 'religion' (from Sanskrit), *bambu* 'bamboo', *bukak* 'open up or out', *duga* 'probing, fathoming', *rupa* 'appearance' (from Sanskrit), *rusak* 'damaged', *umpama* 'example, instance' (from Sanskrit), *kukoh* 'staunch, strongly built', *kumpul* 'gather', *musoh* 'enemy', etc. Pattern (1) contains at least two subpatterns. In one of these *ngoko* -*a* corresponds to *krama* -*i*. In another, *ngoko* -uCa- corresponds to *krama* -iCa-. These patterns compete for some forms and may show unpredictable variation, as

---

[17] Given the lengthy tradition of Dutch scholarship on Javanese this is surprising. Gonda (1948:371), for example, speaks only vaguely about "variability of the word-end," and even when he becomes most explicit about phonological templates he notes the -*a* : -*i* and CuCuC : CəCaC patterns only in passing, even though they form a major part of the mechanism for deriving *krama* forms from *ngoko*. Similarly, Uhlenbeck (1978a:288-93) notes that Krama-Ngoko pairs fall into two classes, those showing no similarity in shape, and those formed by a productive or semi-productive 'procédé' (derivational process). However, his discussion is also sketchy, and fails to clearly bring out the statistically dominant patterns.

seen in the differing treatment of *ḍuga* : *ḍugi* vs. *kuna* : *kina*. A third, less well-attested pattern is one in which *ngoko* -o(C) corresponds to *krama* -e(C) or -ɛ(C). Each of these subpatterns is generated by a process of vowel fronting, generally to *i*, less often to *e* or *ɛ*. In contrastive environments front vowels, especially high front vowels, are universally associated with diminutive, high-pitched, and refined, a constellation of qualities that connote *alus* as opposed to *kasar*, and it can therefore be assumed that pattern (1) is motivated by universals of sound symbolism. Patterns (2) and (3), on the other hand, appear to be language-particular developments.

Although many *krama* forms are transparently derived from their *ngoko* equivalents by one of these three templates, others show no phonological relationship, and so must be recruited from separate lexical items, as with *adoh* : *tǝbih* 'far, distant', *adol* : *sade* 'sell', *ayo* : *maŋga* 'come on!', *alas* : *wana* 'forest', or *bocah* : *lare* 'child'. *Krama* forms that are not derived by phonological template thus raise the question whether *ngoko* : *krama* might sometimes correspond to a native : borrowed distinction. Table 3.2 shows phonologically unrelated *ngoko-krama* pairs, together with etymologies where these are available. Javanese forms with an etymology are in italics:

Table 3.2 Phonologically unrelated *ngoko-krama* pairs with etymologies

Ngoko	Krama	Engish	Etymology
adoh	tǝbih	far	PMP *zauq
adol	sade	sell	Malay jual
ayo	maŋga	let's go!	Malay ayo
alas	wana	forest	PMP *halas
bǝnǝr	lǝrǝs	true, correct	PMP *bener
bǝŋi	dalu	night	PMP *beRŋi
bǝras	*wos*	husked rice	Malay bǝras/PAN *beRas
bocah	lare	child	none
dalan	margi	road	PAN *zalan
goḍoŋ	*ron*	leaf	PMP *dahun
kayu	kajǝŋ	wood, tree	PAN *kaSiw
lima	gaŋsal	five	PAN *lima
loro	kalih	two	PAN *duSa
panas	bǝnter	hot	PMP *panas
papat	(sǝ)kawan	four	PAN *Sepat
pitik	*ayam*	chicken	Malay ayam
puluh	(n)dasa	ten	PAN *puluq
si	pun	person marker	PAN *si
siji	*satuŋgal*	one	Malay tuŋgal
sukǝt	*rumput*	grass	Malay rumput
tǝlu	tiga	three	PAN *telu, Malay tiga
watu	sela	stone	PAN *batu

Most words in Table 3.2 that are likely to be native are *ngoko*. This is quite clear of the lower numerals: 1-5 and 10 have both *ngoko* and *krama* forms, although other numerals do not, and while Javanese *siji* is an innovation, all other *ngoko* forms of numerals reflect known etyma and appear to be native. Malay loanwords, on the other hand, are often found in the *krama* level. There are exceptions to these generalisations, but these appear to be departures from a more common pattern. Javanese *wos* 'husked rice' and *ron* 'leaf', for

example, evidently are native *krama* forms, while *bəras* almost certainly is a Malay loan. A few cases appear intractable, as with *abot : awrat* 'heavy.' Like *bəras : wos* 'husked rice' this *ngoko : krama* pair is a lexical doublet (< PMP *beReqat 'heavy'). Whereas *bəras : wos* easily resolves itself into a borrowed vs. native distinction, however, neither *abot* nor *awrat* can easily be explained as a loan. Finally, some *ngoko : krama* pairs are unambiguous loanwords in which the *krama* member is formed by phonological template: *agama : agami* 'religion', *nama : nami* 'name', *rupa : rupi* 'form, semblance', all from Sanskrit originals with final *-a*.

Three other observations merit brief notice. First, a number of *ngoko* forms have initial consonant clusters *bl-*, *br-*, *dr-*, *gl-*, *gr-*, *kl-*, *kr-*, *sl-*, *sr-* and *tr-*; *krama* forms with these clusters occur (*klumpuk : kləmpak* 'altogether', *srati : sratos* 'mahout'), but only where they are derived by phonological template. Second, the morpheme *pun* confers *krama* status on some grammatical morphemes when added to the *ngoko* form, as with *di- : dipun-* 'passive prefix', or *apa : punapa* 'what?'. Finally, although most writers recognise only a three-way distinction between *ngoko : madya* and *krama*, Horne (1974) lists a number of lexical bases that are marked as '*ng kr*' (used in both speech levels). These are invariably accompanied by a distinct *krama inggil* form, as with *abah : abah : kambil* 'saddle', *aḍi : aḍi : raji* 'younger sibling', *adus : adus : siram* 'bathe', *bapak : bapak : rama* 'father', or *bojo : bojo : garwa* 'spouse'. Since *ngoko* forms are used for all semantic categories and *krama* forms for few, it is not clear how sets like *bapak* (ng) : *bapak* (k) : *rama* (ki) 'father' differ from sets like *əmbok* (ng) : *ibu* (ki) 'mother', but the implication is that forms marked '*ng kr*' constitute a fourth category.

Similar, but less elaborate systems of speech levels are found in Sundanese, Madurese, Balinese and Sasak. Since these languages adjoin Javanese, and have been subjected directly or indirectly to strong Javanese cultural influence at various times, it is reasonable to suppose that the distribution of speech levels in western Indonesia is a product of diffusion. In a critique of earlier claims, Clynes (1994:141ff) has shown: 1) speech styles in Balinese almost certainly derive from Old Javanese (the language spoken until the end of the fifteenth century), 2) Javanese already had a complex system of speech styles by the fifteenth century, and 3) "the differences between spoken and written Javanese in the fifteenth century were so marked that a virtual state of 'diglossia' prevailed." While the first of these conclusions might be suspected simply on the basis of the distribution of AN languages with distinct 'speech styles' and the known cultural domination of Java in the area where this linguistic peculiarity exists, the second and third conclusions are less obvious and hence more interesting. Through comparison with Balinese Clynes argues that speech levels had already appeared in Old Javanese, even though written documents for the language do not reflect them. Such a diglossic situation would not be surprising: nearly half the language of the Old Javanese texts consists of Sanskrit loans, many of them esoteric terms that are unlikely to have formed part of the everyday vocabulary of the ordinary person. The written language was the language of the courts, and perhaps even a select few within the courts, and the circumstances of its use would no doubt have contributed to an increasing separation between spoken and written (or recited) language over time. More recently Nothofer (2000) has also shown convincingly that the speech strata of Sasak derive from contact with both Balinese and Javanese. By contrast Malay, which has also been subject to Javanese cultural and linguistic influence, has acquired no trace of the Javanese system of speech levels. This is true even of Jakarta Malay, which is surrounded by Javanese speakers. Socially sensitive lexical distinctions are made in a few cases, as with the etymologically humilific first person singular *saya* (etymologically

'servant; slave'), or the honorific third person singular *bəliau* (used, e.g. with reference to the President of the Republic of Indonesia and a few other distinguished personages), but these are isolated, and are not described in any published source as forming a stratum of the vocabulary. It is possible that Javanese speech levels did not diffuse to Malay because Malay had greater prestige than Javanese throughout most of the history of contact between these two language communities.

The history of Javanese speech levels has been the subject of some controversy. Robson (2002:11) points out that the *ngoko*/*krama* distinction can be traced to at least the early sixteenth century, and he speculates that "a vocabulary of respect seems to be a very old element in Austronesian languages. Its development in Javanese may have been stimulated by the existence of pairs of words of the same meaning, combined with the need to use synonyms in order to meet the requirements of poetry." He thus associates speech levels historically with ritual parallelism, a topic that will be addressed below. Nothofer (2000:57), on the other hand, holds that the system of speech levels "is not a feature of an older proto language (such as Proto Malayo-Polynesian) to which these five languages belong … but appears to have emerged in Javanese around 1000 AD." As noted by Clynes (1994), and Nothofer (2000), this system evidently then spread into adjacent language communities that came under Javanese domination at various times in their history.

Errington (1988) has documented the progressive breakdown of this system of status-sensitivity in the Javanese lexicon, a process that arguably was set in motion by the achievement of Indonesian independence. With modernisation came a new social order dictated by national and international politics, and hence the decline of the *priyayi* as a hereditary nobility with genuine political power. Since the traditional linguistic etiquette of the *priyayi* reflected a larger set of social norms, the transformation of traditional society into a modern nation state has had inescapable consequences for the system of speech levels in Javanese, which is likely to disappear in another generation or two.

3.1.2 Respect language in Micronesia and Polynesia

Some Pacific island societies have a special vocabulary used to address or converse with a high chief. Since hereditary status distinctions are common in Micronesia and Polynesia but rare in Melanesia, most reported cases are therefore found in the former areas.

Rehg (1981:359-75), summarizing earlier work by Garvin and Riesenberg (1952), but providing more linguistic detail, describes a system of 'honorific speech' or 'high language' in Pohnpeian that is intimately associated with the system of titles marking rank within the traditional political system. He notes that the major difference between speech levels is lexical, there being little difference between high language and common language in sound system or grammatical structure although, as noted below, this does not appear to be completely true .

The high island of Pohnpei is divided into five municipalities, each of which has two lines of chiefs. As Rehg (1981:360) puts it "The highest chief in one line is called the *Nahnmwarki* and the highest chief in the other the *Nahnken*. Below each of these chiefs there are numerous other title holders, the first eleven of which are considered to be the most important." According to Garvin and Riesenberg (1952:202) "Within the sub-clan each man is graded according to seniority of matrilineal descent, and titles are distributed roughly according to the same standard." Essentially everyone holds a title, and the relative ranking of two men with the same title is determined by seniority between the

larger sociopolitical units to which they belong. The basic distinction separates superiors in title and respected equals on the one hand, and familiar equals and inferiors in title on the other. Garvin and Riesenberg add that although respect behavior is part of daily life in Pohnpei, it is more elaborate and mandatory in ceremonial contexts. Failure to observe due respect is visited with supernatural punishment, as the *Nahnmwarki* is thought to travel everywhere with an invisible but potent spiritual partner, the *eni*, or ancestral ghost.

Rehg (1981:363) notes that 'Within the vocabulary of Ponapean, there are at least several hundred honorific morphemes.' Most of these are nouns and verbs, and they can occur in any of three patterns based on the separation or conflation of respect honorific (REH) and royal honorific (ROH) forms:

	Common	**Respect Honorific**	**Royal Honorific**
1)	A	B	C
2)	A	B	B
3)	A	C	C

Figure 3.1 Patterns of use with Pohnpeian respect vocabulary

In pattern 1) distinct forms are used for A) common speech, B) high title holders, and C) the highest chiefs. In pattern 2) B and C forms are conflated as respect honorific, and in pattern 3) they are conflated as royal honorific. Unlike Javanese, where many honorific forms are derived from common speech equivalents by use of phonological templates, most honorific forms in Pohnpeian appear to be unrelated to their semantic equivalents in common speech: *paliwar : ka:lap : erekiso* 'body', *mese : wasaile : si:leŋ* 'face', *kouru:r : kiparamat : rarenei* 'laugh', *u:pʷ : pe:n : pʷi:leŋ* 'drinking coconut', *meir : seimʷɔk : derir* 'sleep'. Where a relationship is apparent it is based on morphology rather than phonology, and generally derives an ROH form from its REH equivalent: *pe: : lime : limeiso* 'hand, arm', *sowe : pelikie : pelikiso* 'back', *sokon : irareileŋ* (Nahnmwarki), *irareiso* (Nahnken). In other cases common and respect forms share a base, but differ in affixation/compounding: *pa:nadi : pa:nkupʷur : mʷareiso* 'chest', *pa:npe:* (common) : *pa:npʷɔl* (REH/ROH) 'armpit'. Rehg (1981:366) points out that two morphemes are found in several respect forms, namely *-iso* 'lord' and *-leŋ*, a bound form of *la:ŋ* 'heaven, sky', thus marking the exalted status of its addressee.

In addition to honorific vocabulary Pohnpeian also uses humilific and honorific forms of possessive marking. The first of these is relatively simple, as seen in: *mɔŋei* 'my head' : *ei tuŋɔl mɔ:ŋ* 'my head (humilific)', or *ei se:t* 'my shirt' : *ei tuŋɔl se:t* 'my shirt (humilific)'. Although Pohnpeian distinguishes direct from indirect possession in common speech, humilific speech uses only indirect possession. According to Rehg (1981:371) "Honorific possessive constructions grammatically parallel common language possessive constructions, except that honorific classifiers, honorific pronouns, and honorific nouns are employed." Examples include *sawi-* (root) : *sawimʷ* (common) : *sawimʷi* (REH) : *sawi:r* (ROH) 'your clan', or *mʷare-* : *mʷaremʷ* : *mʷaremʷi* : *mʷara:r* 'your title'. Finally, Rehg (1981:374) notes a single phonological correlate of respect language, namely "the exaggerated prolongation of vowels in a few common words and expressions." Thus, in the common greeting *kasele:lie* (written *kaselehlie*) "the already long vowel *eh* may be

prolonged two to three times its normal length when addressing a superior or respected equal," and the first vowels in *ei* 'yes' and *i:yeŋ* 'excuse me' are similarly lengthened to show respect. Although honorific patterns of speech occur elsewhere, Pohnpeian appears to be the only Micronesian language that has developed this feature to any appreciable degree of complexity.

Pohnpeian respect language differs from Javanese respect language in a number of ways. First, it has three speech levels that are clearly correlated with social hierarchy, whereas Javanese essentially has two. Although one could argue that Javanese also has three speech levels, one of these (*madya*) is neutral rather than status-defined. It is thus designed to cope with situations of social ambiguity. By contrast, Pohnpeian respect language evidently tolerates no ambiguity. This follows from another difference between the structural frameworks of the two systems. Pohnpeian respect language is kinship-based, and so is more tightly bound to a formal sociopolitical system of hereditary titles than Javanese respect language, which reflects social class, however that is achieved. Third, whereas a large part of the respect language of Javanese is transparently derived from common speech through application of a small number of phonological templates, Pohnpeian makes very little use of phonological distinctions to mark status-sensitivity. Fourth, honorific speech in Pohnpeian may alter the normal grammatical structure of the language, whereas Javanese shows no such changes between levels. It is at least worth mentioning that the violations of linguistic structure permitted in honorific speech show intriguing parallels with the flagrant violations of social structure that Garvin and Riesenberg (1952:211) report for the traditional behavior of the Nahnmwarki, "who had sexual access to all women in his domain, married or not," and who could commit clan incest with impunity, a violation normally punishable by death. If one searches for the most compact description of the difference between the two systems, it is perhaps that the Pohnpeian system of respect language is ultimately based on fear, while the Javanese system is based on an esthetic ideal of *alus* behavior. Violation of the norms within the Pohnpeian system results in supernatural retribution; violation of the norms within the Javanese system of speech levels results in personal mortification.

Speech levels are also found in Samoan and Tongan of western Polynesia. As in Javanese and Pohnpeian, these levels are primarily or exclusively marked by lexical variation. Milner (1961), who provides the most comprehensive account of this phenomenon in Samoan, distinguishes five levels of lexically-marked courtesy: 1) vulgar (or indecent), 2) colloquial (or slang), 3) common (or ordinary), 4) polite (or respectful), and 5) most polite (or most respectful). He notes (1961:297) that when there is a polite equivalent or equivalents for a given ordinary word, "the use of an ordinary word is almost ruled out when a speaker addresses (and usually when he refers to) chiefs. If he is referring to himself, his kinsmen, or his possessions, then no matter how high his own rank may be, the use of ordinary words is, conversely, obligatory." Milner illustrates the difference between ordinary and polite speech with the following examples:

Ordinary	Polite	English
ʔua pē le maile	ʔua mate le taʔifau	The dog is dead
Fafano ou lima	Tatafi ou ʔaʔao	Wash your hands
ʔua tipi le maʔi	ʔua taʔoto le ŋaseŋase	The patient is having an operation

These examples suggest that most lexical variation in Samoan speech levels affects content morphemes rather than function morphemes. Milner's article concludes with a list of about 450 ordinary words and their polite equivalents. A sample of these appears in Table 3.3, together with Proto Oceanic etymologies where these are available:

Tabale 3.3 Some ordinary and polite lexical pairs in Samoan

Ordinary	Polite	POC	English
afi	mālaia	api	fire
afĩ ʔafato	afĩ līpoi	qapatoR	bundle of edible grubs
ʔai	talialo	kani	eat
aitu	saualiʔi	qanitu	ghost
ʔanae	āua	kanase	grey mullet
atu	iʔa mai moana ŋāʔoŋo	qatun	bonito
āvā	faletua	qasawa	wife
fafano	tatafi	paño	wash hands or feet
fafie	polata	papie	firewood
fafine	tamaʔitaʔi	papine	woman
fale	apisā	pale	house
fana	lāʔau mālosi	panaq	gun; shoot a weapon
fono	aofia	ponor	sit in council
ŋutu	fofoŋa	ŋusuq	mouth
inu	tāumafa	inum	drink
isu	fofoŋa	icuʔ	nose
lima	ʔaʔao	lima	hand, arm
mā	liliʔa	mayaq	embarrassed, ashamed
mata	fofoŋa	mata	eye, face
naifi	ʔoʔe/polo		knife
nifo	ʔoloa	nipon	tooth
niu	vailolo	niuR	coconut
susu	mau	susu	female breast
tae	otaota	taqe	excrement
taŋi	tutulu	taŋis	weep, cry
talo	fuāuli	talos	taro

The list from which this material is drawn permits some useful generalisations. First, unlike Javanese, where respect forms exist not only in the lexicon but also in the shapes of affixes, or Pohnpeian, where the system of status-sensitivity penetrates both the lexicon and some aspects of syntax, the respect language of Samoan appears to consist entirely of lexical substitutions. Even with regard to lexical substitution, however, Samoan respect language differs from the other languages considered so far, since in Javanese and Pohnpeian lexical substitution appears in nearly all cases to involve the interchange of one morpheme, or at least one word for another. In Samoan, on the other hand, respect forms sometimes are transparent circumlocutions, as in *iʔa mai moana* 'fish that comes from the deep sea' (= bonito), *lāʔau mālosi* 'strong instrument' (= gun), or *lasomili : ŋaseŋase o tāne* 'hydrocele' (the latter literally 'male sickness').

Second, Milner (1961:310) notes that "*naifi* is one of a number of words of English origin, the polite equivalent of which are Samoan words." This remark suggests that a

correlation exists between foreign = ordinary, and native = polite, but Milner's data shows only one other example that can be used to support this claim: *tāvini* (from English, through Tahitian) : *ʔauʔauna* 'servant'. As seen in Table 3.3, the pattern that stands out most saliently in comparing ordinary and polite forms in Samoan is one of inherited : ordinary, and innovative : polite. The recognition of this pattern is important, as it sheds far more light on the origin of the Samoan vocabulary of respect than the claim that non-native terms are commonly recruited for ordinary use and native terms for polite use. With only rare exceptions (e.g. *pē* : *mate* : POC *mate 'die'), where etymological information is available it shows that the ordinary vocabulary consists of forms inherited from Proto Polynesian, Proto Oceanic and more remote proto languages, while polite vocabulary is innovative. The nature of these innovations is not always clear, but in some cases it can be stated with a fair degree of confidence. As noted above, for example, the polite lexeme sometimes is a circumlocution. Given this pattern it seems clear that the Samoan vocabulary of respect is overwhelmingly innovative, and derives either from lexical coinage or from inherited lexical material that is used with semantic shifts, compounding, or descriptive phrases.

Third, Samoan respect vocabulary appears to differ from the other cases examined above in that a single respect term can sometimes substitute for a number of ordinary words that are semantically distinct. The clearest example of this in Table 3.3. is *fofoŋa*, used as a replacement for 'mouth', 'nose' and 'eye'. Milner (1966) gives this term with the meaning 'face (or any part of it such as nose, mouth, etc.)'. Finally, as in Javanese and Pohnpeian, the respect vocabulary of Samoan appears to be a marked register. This is apparent not only from the social or cultural circumstances in which it is used, but also from the formal properties of the system itself. Whereas words of ordinary language are usually monomorphemic, for instance, those of respect language are not infrequently analyzable either morphologically or syntactically or both. Circumlocutions are by their nature indirect, and the use of indirect reference or address is a hallmark of status-sensitivity in many languages. In addition, etymological information indicates that even where they are morphologically opaque respect forms appear to be secondary creations substituting for an ordinary word that must be avoided in given social circumstances.

Table 3.4 The three levels of Tongan respect vocabulary

Muomua	Lotoloto	Kakai	English
(Leading chiefs)	(Middle chiefs)	(People)	
hala	pekia	mate	dead
hoihoifua	katakata	kata	to laugh
laŋi	tauoluŋa	mata	eye/face
laŋi	fofoŋa	ulu	head
laŋi	faiaŋatoka	tano	to bury
tofa	toka	mohe	to sleep
tutulu	taŋi	taŋialoima	to cry
tamasiʔi	taŋata	siana	man
feitaumafa	halofia	fiekai	hungry
fekita	uma	uma	kiss
hoko	vala	fatei	sarong
houhau	tupotamaki	ita	angry
laŋilaŋi	fakavao	mumu	warm by a fire
fakamomoko	fakamomofi	fakamafana	warm oneself

In his general ethnography of Tonga, Gifford (1929:119-122) describes a three-way contrast in speech levels, distinguishing commoners, nobles and the king, as illustrated with selected examples in Table 3.4.

Philips (1991) provides more information about the use of Tongan speech levels, drawing attention to discrepancies between statement and practice, but without adding materially to the description of the basic system. As with the other languages examined above, inherited forms are strongly associated with ordinary language (*mate* < POC **mate* 'die, dead', *mata* < POC **mata* 'eye, face', *ulu* < POC **qulu* 'head', *tano* < POC tanom 'bury', *faka-mafana* < POC **mapanas* 'warm', etc.)   There are nonetheless a few exceptions to this pattern, as with *taŋi* < POC **taŋis* 'weep, cry', or *taŋata* < POC **tam^wataq* 'person, human being', both of which are listed as Lotoloto. One of the most striking features of the Tongan system is the use of *laŋi* (< POC **laŋit* 'sky, heaven') as a Muomua replacement for 'eye/face', 'head' and 'to bury', and *laŋilaŋi* as a similar replacement for 'warm by a fire'. In the first two meanings *laŋi* evidently is used in a very similar way to the suffixal *-leŋ* in Pohnpeian, which Rehg (1981) describes as a bound form of *la:ŋ* 'heaven, sky'. In both cases this appears to mark the exalted status of its addressee, and given the cognation of the term it must be wondered whether this might not point to a usage in a language ancestral both to Nuclear Micronesian and Polynesian languages. Since other cognate respect forms are lacking, however, this agreement probably is best treated as convergent. Near agreements between Tongan and Samoan, as in the Lotoloto term *fofoŋa* 'head' (Samoan *ulu* : *ao* 'head', but *fofoŋa* in the meanings 'mouth', 'nose' and 'eye/face') are best attributed to diffusion.

To summarize, despite claims to the contrary by some writers, Javanese and Samoan appear to lexically distinguish two levels of status-sensitivity, while Pohnpeian and Tongan distinguish three. All well-described systems of speech levels in Oceanic languages revolve around systems of hereditary titles, while in Javanese status-sensitivity is focused on the assignment of individuals to social strata that may be based on occupation as much as genealogy. In other words, the Javanese system appeals more to achieved status than ascribed status, although this may not always have been the case. In addition, although Milner (1961:301) states that Javanese and Samoan show a high degree of overlap in respect vocabulary relating to the human body, as well as actions and states closely associated with the body, the material available to me suggests that respect vocabulary in relation to the body is used much more in Micronesia and Polynesia than it is in Javanese, a disparity that almost certainly has to do with the perceived sacredness of the person of a high chief. Despite the generic similarities in respect behavior (including respect language) toward those of chiefly rank in Oceanic-speaking societies, there are important differences distinguishing Micronesia from Polynesia. In Polynesia violation of proper respect behavior ran the risk of contact with the *mana* of a chief, while in Pohnpei a similar violation was thought to be punished by the *eni* (< PAN **qaNiCu*), or ancestral ghost that accompanied a chief. While the consequences may have been similar, the Polynesian concept is depersonalised and automatic (almost like contact with electricity), while the Micronesian one is more akin to retribution.

These comparisons inevitably raise questions about the history of speech strata in AN languages: was such a system present in PMP, or are the attested systems products of convergent evolution? As noted above, Robson (2002) has speculated that the *ngoko* : *krama* distinction in Javanese may reflect an ancient AN use of ritual parallelism in which any given meaning could be represented by alternate terms to achieve a kind of declamatory effect in ritual recitations. In time, according to this view, synonymous forms

came to be associated with differences of social status. However, if this were the case we would expect to find comparative evidence allowing the reconstruction of at least fragments of such a system, and nothing of this kind has been found to date. Moreover, many *krama* forms are clearly derived from their *ngoko* equivalents by the application of a small number of phonological templates. If the Javanese speech levels evolved from a system of ritual parallelism they must initially have contained only *ngoko* : *krama* pairs that had no phonological relationship to one another (as *alas* : *wana* 'forest'), and only later expanded to include phonological mechanisms for the derivation of novel *krama* forms (as in *rupa* : *rupi*, *butuh* : *betah*, or *arti* : *artos*). The evidence to hand, then, suggests that speech levels in Javanese evolved within the separate history of that language. The Oceanic cases are perhaps more plausibly derived from a common ancestral form, but even here there appears to be no basis for reconstructing any details of such a system. Despite some superficial similarities, such as the honorific use of a reflex of POC *laŋit 'sky, heaven' in both Pohnpeian and Tongan, it is difficult to demonstrate a historical connection between respect languages in Micronesia and Polynesia. Rather, the evidence suggests that these linguistic elaborations arose from a common basis of hereditary chieftainship in Proto Oceanic society that was compromised in much of Melanesia as a result of heavy contact-induced change in both language and culture.

A comprehensive history of speech levels in AN languages cannot yet be written, but a likely scenario can be sketched as follows. Most languages show some lexical variation that is correlated with status-sensitivity, but this is rather loosely organised and enforced. In some societies these tendencies came to be more formalised and mandatory, perhaps initially on ceremonial occasions, and later in daily life. The cause or causes of this increase in formalisation must remain speculative, but changes in social structure that exacerbated status differences would presumably have preceded the appearance of fully developed speech levels. There is no evidence that any of this is ancient. Rather, the respect language of Java probably evolved near the end of the Hindu-Buddhist period, which led to the rise of courts and courtly language. It then diffused into the neighboring Sundanese, Madurese and Balinese languages, and from Balinese into Sasak. The respect language of Pohnpeian almost certainly arose *in situ* in a society that was already highly stratified along lines of hereditary status. The history of respect vocabulary in western Polynesia is more difficult to untangle. Because they share few cognates Milner (1961:300) suggests that the respect vocabularies of Tongan, Samoan, and the less well-known Wallis and Futuna have largely developed in the separate histories of these languages, but that "a few words especially those relating to high chiefs and royalty were perhaps in use before the Western Polynesians became separated." The problem with this proposal is that there is no 'Western Polynesian' subgroup, and an ancestral Western Polynesian linguistic community could, therefore, never have existed. Given the distribution of this feature over geographically contiguous but phylogenetically divided communities diffusion appears to offer the most plausible explanation of its history. Indeed, since western Polynesia is a well-known culture area that bridges a clear linguistic cleavage between Tongic and the rest of Polynesian, the use of speech levels in this region of the Pacific is almost certainly a product of stimulus diffusion from a single center of innovation, whether that was Tonga or Samoa. The direction of diffusion is unknown, but given the more sharply ramified system of status distinctions in Tongan speech levels and the former existence of the 'Tongan empire' (Geraghty 1994) which controlled not only Samoa but also Wallis and Futuna, a Tongan center of origin appears most likely.

3.2 Gender-based speech differences

It is not known how widespread male/female speech differences are in AN languages. Clear reports of such differences have been published for two language groups: the Cham of coastal Vietnam and the Atayal of interior northern Taiwan.

3.2.1 Men's and women's speech in Atayal

Li (1980b, 1982a) has described a system of pervasive lexical differences that distinguish men's and women's speech in two dialects of Atayal, spoken in northern Taiwan. These studies are very valuable, as they are the most complete analyses to date, and are arguably the only true cases of gender-based differences of speech style or register reported so far in any AN language.

Table 3.5 Male/female speech differences in Mayrinax Atayal

	Women	**Men**	**Skikun**	**PAN**
(1)	kahuy	kahuniq	qhuniq	*kaSiw 'wood; tree
	hapuy	hapuniq	puniq	*Sapuy 'fire'
	raʔan	raniq	ryaniq	*zalan 'path, road'
	paysan	pisaniq	psaniq	*paliSi-an 'taboo'
	matas	matiq	matas	*p<um>ataS 'to tattoo'
	k<um>aiʔ	k<um>aihuw	k<m>ehuy	*k<um>alih 'to dig'
	lataʔ	latanux	tanux	*NataD 'cleared ground'
	mataq	matiluq	mteluq	*mataq 'raw'
	ma-busuk	businuk	msinuk	*ma-buSuk 'drunk'
	q<um>alup	q<um>alwap	q<m>alup	*q<um>aNup 'to hunt'
	c<um>aqis	c<um>aqiŋ	c<m>aqis	*C<um>aqiS 'to sew'
	kucuʔ	kuhiŋ	kuhiŋ	*kuCu 'head louse'
	hawuŋ	hayriŋ	ciŋ	*saleŋ 'pine tree'
(2)	(raqis)	raqinas	rqinas	*daqiS 'face'
	(quway)	quwaniʔ	qwaniʔ	*quay 'rattan'
	(tulaʔ)	Tulaqiy	tlaqiy	*tuNa 'freshwater eel'
	(qabuʔ)	qabuliʔ	qbuliʔ	*qabu 'ash'
	(buhug)	buhinug	bhenux	*busuR 'hunting bow'
	(imaʔ)	imagal	ʔimagal	*lima 'five'
	(qaug)	qauag	-----	*qauR 'bamboo'
	(bual)	buatiŋ	byaliŋ	*bulaN 'moon'
	(ʔugat)	ʔuwiq	ʔugiq	*huRaC 'vein, tendon'
	(k<um>itaʔ)	k<um>itaal	ktayux	*k<um>ita 'to see'

Atayal divides into two large dialect groups, Squliq and Cʔuliʔ/Ts'ole'. Squliq dialects are relatively uniform, while Cʔuliʔ dialects are more heterogeneous. Distinct forms of men's and women's speech appear in Mayrinax, a Cʔuliʔ dialect (Li 1980b), and in Paʔkualiʔ, a dialect described only as 'Atayal' (Li 1982a). What is of most interest is that the men's speech forms in Mayrinax and Paʔkualiʔ generally correspond more closely to the cognate terms in other Atayal dialects, which do not distinguish between male and female speech. This can be seen in Table 3.5 (after Li 1982a), which illustrates gender-

based lexical differences in Mayrinax, and the uniform cognates in the CɁuliɁ dialect Skikun. Anticipated but non-attested female speech forms appear in parentheses.

Li documents these gender-based lexical differences in considerable detail. He classifies the derivational patterns that can be extracted from the data, and shows that female forms are more conservative than male forms. Based on a corpus of about 1,500 lexical bases he finds that 107 pairs of words are distinguished as 'male' vs. 'female'. These show no clear semantic bias, and although Li (1980b) takes pains to identify some patterns by which the male forms are derived from synonymous female forms, these patterns are extremely varied, each applying to only a small set of examples. Thus, to consider only the material in part (1) of Table 3.5, in *kahuniq* and *hapuniq* the male forms are derived from the female forms by dropping the diphthongal coda *-y* and adding *-niq*. In *raniq* and *pisaniq* the male forms are derived by contracting the sequence of like vowels (separated by automatic glottal stop) or vowel + glide and adding *-niq*, with further contraction of the sequence of like consonants produced by suffixation. These four examples can be regarded as forming a fairly consistent pattern, but it is the only pattern that is substantially exemplified in the material reported to date. In *matas* : *matiq* the final -VC is replaced by -*iq*, a derivational strategy that is otherwise unattested and only vaguely similar to the pattern used to derive *kahuniq*, *hapuniq*, *raniq* and *pisaniq*. In *k<um>aiɁ* : *k<um>aihuw* the sequence *-huw* is added to the female form (Li 1980b treats *-Ɂ* as phonemic, but its contrastive value is minimal). In *lataɁ* : *latanux* the male form is derived by adding *-nux*, and in *mataq* : *matiluq* it is derived by infixation with *-il-* and change of last-syllable *a* to *u*. The derivation of *ma-busuk* : *bus<in>uk* is similar to the preceding example, but not the same, as it uses the infix *-in-* with no change of the stem vowel. In

q<um>alup : *q<um>alwap* the male form is derived by infixing -*a*- after the last vowel, which then semivocalizes in prevocalic position. The last three forms show some general similarity, but again there are differences of detail in each pattern. The upshot of this brief glimpse is that it appears inappropriate to use the word 'pattern' to describe variation between male and female speech forms in Atayal. Unlike the *ngoko* : *krama* distinction in Javanese, where hundreds of *krama* forms can be generated by applying a handful of phonological templates, the derivation of male speech forms from female speech forms in Mayrinax Atayal requires nearly as many templates as there are forms. In short, the process of lexical derivation for male speech forms is largely unsystematic.

Li (1980b:10) states that "The men's forms are generally longer than women's by adding certain affixes." As evidence he cites 'the suffix *-nux*' in Mayrinax *βatu-nux*, Skikun, Mnawyan *βtu-nux*, Squliq *tu-nux* 'stone' (PAN *batu), 'the infix *-in-*' in Sakuxan, MaspaziɁ *rak<in>us*, Squliq *k<n>us*, Skikun, Mnawyan *rk<n>as* 'camphor laurel' (PAN *dakeS), 'the suffix *-riq*' in Mnawyan *yumu-riq* 'moss' (PAN *lumut), and so forth. Affixes, however, normally have two characteristics not found in these elements: they show clearly recurrent patterns, and they have identifiable meanings. Li states that these 'affixes' appear in a number of words in various Atayal dialects. As already seen, however, very few of these phonological elements are recurrent, and where they are their manner of attachment differs in particular details that renders most cases derivationally unique. In at least one case, that of *-in-*, an element used to derive male speech forms corresponds to a well-known PAN infix marking perfective aspect. This might be interpreted as evidence that male speech forms are sometimes derived from female forms by exploiting available morphological resources. However, it does not seem justified to regard these elements as identical, since the inserted element in e.g. *bus<in>uk* 'drunk' has no meaning, and unlike the inherited infix *-in-*, which is inserted before the first stem

vowel, the *-in-* of *bus<in>uk* is inserted before the last vowel of the stem. Not only do the 'affixes' in Atayal male speech forms constitute an open class, then, but where they are infixed the position of insertion is atypical. The only real generalisations that appear to be possible are that male speech forms generally are derived by adding material either through 'suffixation' or 'infixation', but never through 'prefixation', and that the net result is that the more conservative forms of women's speech are in many cases thoroughly disguised.

Since the male speech forms in Mayrinax and Paʔkualiʔ generally are more similar to the forms in other Atayal dialects that have no male/female speech distinction, there seems to be no choice but to attribute male speech forms to Proto Atayal. In fact, since Li (1982a) indicates that a few male speech forms are also found in Seediq a male/female speech distinction evidently must be posited for Proto Atayalic, ancestral to both Atayal and Seediq. The comparative evidence suggests, then, that the two speech registers existed side-by-side in an ancestral speech community. Modern Seediq dialects have tended to preserve the historically conservative female forms, while modern Atayal dialects have tended to preserve the innovative male forms. In Mayrinax and Paʔkualiʔ both registers are preserved, but not without some changes from the ancestral system. In the Tabilas (Cʔuliʔ) dialect of Atayal, according to Li (1982a) male and female speech registers are preserved in a gender-neutral form that has been transformed in some lexical bases into a distinction based on semantic features of the referent, as with *kahuy* (F) 'tree' : *kahu-niq* (M) 'firewood', *qaxaʔ* (F) 'large beads for necklace' : *bagiiq* (M) 'small beads for bracelet', or *t<um>inun* (F) 'to weave (cloth, silk)' : *t<um>inuq* (M) 'to weave (a mat)'.

The origin of Atayal men's speech remains unclear, although the most plausible hypothesis is that it originated as a secret language used by initiated males (Li 1983). This would explain the fact that the innovative forms of many lexical items are, or were originally the exclusive prerogative of male speakers. Li (1982a) hesitates to adopt this interpretation on the grounds that the derivation of male forms is highly irregular, while the derivation of most secret languages is rule-governed. But this objection seems pointless for two reasons. First, comparatively well-described secret languages such as Prokem in Jakarta (sect. 3.4.3.) show several patterns of word derivation, none of which is fully regular. Second, what is important in a secret language is that innovative forms are disguised so as to hinder intelligibility to those who have not learned them, and greater complexity/irregularity clearly serves this end rather than working against it.

3.2.2 Other cases?

The only other published report of gender-based speech differences in AN languages is that of Blood (1961), who noted male/female differences for speakers of Cham in Vietnam. Usually the differences are phonological, but they may be lexical. According to Blood, phonological differences between men's speech and women's speech in Cham are often due to unequal access to the traditional Indian-based Cham script, which is typically a province of knowledge open to men but not to women. In effect, the speech of men may contain conservatisms that are a direct consequence of literary pronunciations. Blood emphasizes that not all men have these traits, thus implying that 'women's speech' is simply the unmarked gender-neutral register, and that it is men's speech that is deviant. Presumably because this deviant style is associated with prestige, Blood has treated it as basic, and treated the basic style as marked. Unlike the Atayal case, there is no indication that these 'gender-based' speech differences (which are perhaps better treated as education-based), have ever had any function in concealing the content of messages.

Rather they raise an issue that cannot be pursued further here, namely the relationship of written language to spoken language.

Among the few remarks that suggest a further body of still undiscovered material on gender-based speech differences waiting to be studied is one by Voorhoeve (1955:21), who notes that in Rejang of southern Sumatra final nasals may be preploded in a syllable that begins with an oral consonant, as in *buleun/buleudn* 'moon'. Unlike nasal preplosion in other languages of western Indonesia, simple and preploded nasals evidently have acquired different social values, the latter occurring in chanting legends "and also when reviling people (especially by women, as I was told)." Jaspan (1984) gives Rejang *buleun* 'moon; month', *buleudn* 'month (in vituperative or irate expression)', but surprisingly this is the only 'vituperative' form that he gives in a dictionary of over 3,500 entries. Since he also fails to mention that the form *buleudn* is characteristically used by women, it is possible that Voorhoeve's comment was based on an offhand remark that has little bearing on systematic language behavior.

3.3   Vituperation and profanity

The example of Rejang *buleudn* touches on a point that is discussed at much greater length by Mintz (1991), who reports a special vocabulary of about fifty words in Bikol that are 'used only in anger, and sometimes jest, to replace standard vocabulary items.' According to Mintz speakers of Bikol use 'anger words' in circumstances where speakers of other languages might use profanity. The difference is that whereas the referent of profanity is generally an object of disgust, social delicacy, or reverence, the referent of 'anger words' is identical to that of the ordinary language equivalent that it replaces. In illustration he notes (1991:232) that 'If you were to spill rice and become annoyed, you might say *lasgás* or *lamasgás* instead of *bagás*. If an animal were to bother you, you might refer to it as *gadyá'* rather than *háyop*, or if someone were sleeping at a time when they were not supposed to be, you might show your displeasure by referring to this action as *tusmág* and not as *túrog*.' In each case the anger word is a substitute term of identical meaning to its ordinary language equivalent which differs only in being assigned to a different speech register.

Taable 3.6 Phonological derivation of 'anger words' in Bikol

Ordinary word	Anger word equivalent	English
pádiʔ	l(am)asdíʔ	priest
bagás	l(am)asgás	husked rice
tulák	lamasdák	stomach
burát	lusrát	drunk
buŋáw	lasŋáw	drunk
buŋóg	lusnóg	deaf
lubót	lusbót	buttocks
pirák	s(am)agták	money
túbig	kalʔíg	water
uríg	takríg/tukríg	pig
ídoʔ	d(am)ayóʔ	dog
insík	tugalsík	Chinese

Mintz points out that many anger words in Bikol are derived by a phonological template that replaces all but the final syllable, or the final -VC. This appears to show a higher degree of recurrence than the male speech forms of Atayal, but a lower degree than the *krama* register of Javanese. He cites seven cases of disyllabic bases in which the first syllable is replaced by *l(am)as-* or *lus-*. In two of these the resulting consonant clusters are changed by replacement of the second segment. Apart from this pattern other forms show a highly irregular derivational pattern in which the first syllable is replaced by one or more syllables that bear no phonetic similarity to it.

In general the derivational patterns of anger words in Bikol differ from those of male speech forms in Atayal in that Atayal innovates endings, while Bikol tends strongly to preserve them. Mintz does not comment on the fact, but it is noteworthy that all anger words that he cites carry final stress, although in most cases this is a consequence of the form having a medial consonant cluster. Another group of anger words is said to result from semantic change, as where *háyop* 'animal' is replaced by *gadyáʔ* (cf. *gadyá* 'elephant'), *bitís* 'foot' is replaced by *sikí* (cf. *sikí* 'foot of ungulate'), or *kakán* 'eat' is replaced by *habló?* (cf. *habló?* 'gulp, swallow without chewing'). Several of the anger words cited by Mintz refer to introduced concepts (priest, elephant, money, Chinese), suggesting that the angry speech register might be a fairly recent development.

Lobel (2005), who has taken up this problem again, specifically addresses the question of the origin of the angry speech register in Bikol. He notes (2005:154) that "It is unclear whether the angry register of the Bikol languages is an innovation unique to these languages or a retention from Proto Central Philippines." However, he has since discovered (2012) that similar 'angry registers' are found in other Central Philippine languages such as Waray-Waray, Asi/Bantoanon, Mandaya, and Kaagan, and in Manobo languages such as Binukid/Bukidnon of Mindanao. Moreover, a number of forms in this register can be reconstructed for at least Proto Central Philippines, suggesting that this aspect of the lexicon is not a recent development, but rather has a fairly long history in the central and southern Philippines. In addition, Lobel makes considerably more progress in identifying the derivation of lexical items in the angry register from items of ordinary vocabulary through such processes as infixation, partial replacement of the word, phoneme replacement, semantic shift, etc.

Because of the context in which they are uttered the anger words of Bikol can easily be confused with 'profanity'. However, vituperation and profanity appear to be distinct. First, vituperation is expressed through lexical substitutions for ordinary words rather than through ordinary words (however socially stigmatised) that relate to sex, the genitalia (especially of close relatives of the addressee), excrement, reviled animals, the supernatural, etc. Second, although both are maledictive, profanity typically is directed at an interlocutor (often with an associated pronoun) or is shouted in anger or frustration through displeasure with one's situation, whereas vituperation indicates displeasure with another person or one's situation through lexical selection from a marked speech register.

Data on profanity in AN languages is difficult to obtain, but with the help of colleagues I have collected some examples that permit a few limited generalisations.[18] Swearing can vary from mild to intense, in accordance with the degree of anger or irritation aroused, and can be directed at another person or undirected. Categories that are used to express

---

[18]   Data from Philippine languages are courtesy of Jason W. Lobel (p.c., March 3, 2006), data on Bahasa Indonesia courtesty of Uli Kozok (p.c., March 1, 2006), data from Marshallese courtesy of Alfred Capelle and Byron W. Bender (p.c., March 8, 2006), and data on Fijian courtesy of Paul A. Geraghty (p.c., March 1, 2006).

profanity include the following. In most cases these word are spoken or shouted at an interlocutor who has aroused the speaker's anger. Expressions in Philippine languages that seem clearly to be borrowed from or influenced by Spanish, such as Tagalog *taŋinamo* (< *puta aŋ ina-mo*) 'your mother is a whore', or variations on the Spanish *hijo de puta* 'son of a whore; whoreson' or *hija de puta* 'daughter of a whore' are mentioned only in passing, as there is no evidence that these are part of a native system of profanity:

1. Reference to the genitals and other body parts:

Ilonggo: *bilat saŋ bay* (vagina-**gen**-grandmother) 'your grandmother's vagina!'; Bahasa Indonesia: *puki emak* (often shortened to '*ki mak*' = vagina + mother) 'your mother's vagina!'. Said to be quite rude, but common in Sumatra; Marshallese *kōden jinōm*ʷ (vagina-**gen** mother-**2sg)**   'your mother's vagina!'; Fijian: *maŋa i tina-mu* (vagina of mother-**2sg)**   'your mother's vagina!'. Also *maŋa i bu-mu* (vagina of grandmother-**2sg)** 'your grandmother's vagina!'. This specific insult is thus attested in the central Philippines, western Indonesia, Micronesia and Fiji (and doubtless in other AN languages in between). Other expressions are directed at the one who has provoked the speaker's anger rather than at a close relative of that person, and may involve simple utterance of the name of an intimate body part, or suggest some socially undesirable or culturally embarrassing characteristic of that body part, as with Fijian: *boci* 'uncircumcised', *maŋa-levu* 'big vulva', *ŋgala-levu* 'big penis'.

2. Reference to sexual intercourse:

Bahasa Indonesia: *ŋ-entot* 'fuck!'; *ŋocok* 'mix it up!'; Fijian: *cai-si tama-mu* (fuck-**tr** father-**2sg)**   'fuck your father!'. Also *cai-si tuka-mu* (fuck-**tr** grandfather-**2sg)**   'fuck your grandfather!'. The Indonesian expression *ŋocok* is puzzling as an instance of profanity. Its literal senses are to shake (as a bottle of medicine, to mix the contents), to scramble, as eggs, to mix ingredients together'. In its profane use it may imply sexual intercourse as an expression of mixture. Profanity of this type could equally well be classified as cursing a person's lineage.

3. Reference to excrement:

Bahasa Indonesia: *tahi* 'shit!', *tahi kuciŋ* (shit-cat) 'bullshit!'. The former term is more likely to be used as an expression of undirected anger or frustration, and the latter in scoffing at something claimed by another. Fijian: *kani-a na de-mu* (eat-**tr art** shit-**2sg)** 'eat your shit!'

4. Reference to animals:

Tagalog *háyop ka* (animal **2sg)**   'you're an animal!' (said to be widespread in the central Philippines); Bahasa Indonesia: *anjiŋ* 'dog!', *babi* 'pig!', *kerbau* 'water buffalo!' (stupidity), *monyet* 'monkey!'. Shouted at a person who has aroused the speaker's wrath on account of some undesirable character trait that has just surfaced in his behavior.

5. Reference to stupidity:

Tagalog *gagu*, Bahasa Indonesia: *bodoh* 'stupid!'

6. Reference to unbalanced mental state:

Bahasa Indonesia: *gila*, *edan* 'crazy!', *sintiŋ* 'loony, 'off your rocker', nuts!'.

7. Swearing to oneself:

The following are said to be instances of swearing to oneself, or of undirected anger: Tagalog *púki* 'vagina!', used only in isolation, and "cannot be directed at a person as an insult" (Jason Lobel, p.c.); Tagalog *taŋina* (< *puta aŋ ina*) 'whore mother!' (without a possessive pronoun; cp. *taŋinamo* < *puta aŋ ina-mo* 'your mother is a whore', used as an insult); Rinconada Bikol *buray ni nanya* (vagina-**gen**-mother-**3sg)**   'his/her mother's

vagina' 'can be heard being uttered by young and old alike, even over the smallest of mistakes or surprises, and hardly catches anybody's attention when heard' (Jason Lobel, p.c.); Bahasa Indonesia: *pantat* 'buttocks!', *kontol* 'penis!'; Fijian: *cai-ta* (fuck-**tr**) 'fuck!', *maŋa-na* (vagina-**3sg)** 'her vagina!', *de-na* (shit-**3sg)** 'his/her shit!'. When a possessive pronoun is used these expressions differ from swearing at others in using the 3sg rather than 2sg possessive marker.

Both positive and negative conclusions can be drawn from this data. Expressions such as 'your (grand)mother's vagina!' appear to have a long history as profanity in the AN language family. Compared to English, what is conspicuously lacking is the use of expressions invoking divine wrath on the object of one's anger or scorn. Rather, in many AN speaking societies supernaturally-induced misfortune is believed to be an automatic result of the violation of taboos, and so the use of profanity to achieve or wish for the same result would be redundant. This in turn raises the question whether the term 'profanity' is even appropriate for cursing of this kind, since the opposition sacred : profane implies reference to divinity, and where this is lacking the vocabulary of insults is perhaps better characterised as 'obscenity' than 'profanity'.

3.4   Secret languages

As noted earlier, Li's description of the differences between men's and women's speech in Atayal suggests that male speech forms may have originated in attempts to produce a 'secret language', also called a 'ludling'. While this remains speculative, since no such function is known to exist in any modern Atayal dialect, deliberate distortion of the speech signal in order to conceal messages has been reported for other AN languages. Where we have fairly detailed reports these indicate a linguistic subcode similar in function to that of 'Pig Latin' in English.

3.4.1 Tagalog speech disguise

One of the earliest and formally most precise accounts of an AN secret language is Conklin's (1956) brief but highly informative description of 'Tagalog speech disguise.' According to Conklin (1956:136) "When a speaker in conversation attempts to conceal the identity and hence the interpretation of what he says, he may change the phonological structure of his utterances; this method of concealment I call *speech disguise*…The Tagalog term for this hog-Latin kind of speech is *baliktád*, which in other contexts means 'inside-out, upside-down, inverted, or backward." Conklin identifies eight different types of structural rearrangement or affixation used to form *baliktád* words, as follows (R = rearrangement, I = infixation):

R₁: (there is a complete reversal of the phonemic shape of the base): *salá:mat > tamá:las* 'thanks', *mag-simbá > mag-ʔabmís* 'attend Mass'. In the second of these examples a glottal stop is automatically inserted, in accordance with a process that applies to all vowel-initial bases following a consonant-final prefix. Note that only bases are affected by R₁, while affixes remain unchanged.

R₂: (there is partial reversal of the phonemic shape of the base, or metathesis): *dí:to > dó:ti* 'here', *gá:bi > bá:gi* 'taro'. Either the vowels or the consonants may be interchanged. This pattern is said to be limited to 'nonconsecutive disyllabic words.'

R₃: (there is complete transposition of the syllable shape of the base): *ʔi:tó > tó:ʔi* 'this', *pá:ŋit > ŋitpá* 'ugly', *kapatíd > tidpaká* 'sibling'. R₃ is the equivalent of R₁ on the

next higher level of phonological structure. Whereas R$_1$ copies a segment string in reverse without regard to its syllable structure, R$_3$ copies a segment string in reverse one syllable at a time, retaining the internal structure of syllables, but reversing the sequence of syllables in the base. Note that stress remains on the syllable to which it is originally assigned, moving within the base in R$_3$, but staying in place within the base in R$_2$.

R$_4$:   (there is partial transposition of the syllable shape of the base): *ma-gandá > damagán* 'beautiful', *kamá:tis > tiskamá* 'tomato'. Here, unlike R$_3$, the last syllable is removed to initial position, but the first two syllables remain in place. The placement of stress in this pattern is less clear than those discussed above: in *tiskamá* it remains on the syllable to which it is originally assigned, but in *damagán* it remains final and so changes syllables. Presumably the explanation for this difference is that Tagalog allows only penultimate and final stress, and that maintaining stress in final position is simpler than moving it to a syllable to which it was not originally assigned. Conklin notes that this type of morpheme distortion is particularly common, and goes by the iconically appropriate and undoubtedly humorous name of *tadbalík*. With disyllabic bases R$_3$ and R$_4$ would be indistinguishable.

I$_1$:   *tiná:pay > t<um>iná:pay* 'bread', *na > n<um>a* 'already'. This pattern simply inserts the very high frequency Actor Voice infix *-um-* (bolded) according to the normal pattern in productive verb morphology. The difference in this form of speech disguise is that the unmarked function of the affixed word is non-verbal.

I$_2$:   *sí:loʔ > s<ig>í:-l<o:g>óʔ* 'snare, trap', *ʔitlóg > ʔ<um>í:tl<om>óg* 'egg', *salá:mat póʔ > s<ag>á:l<ag>á:m<ag>át p<og>oʔ* 'thank you, Sir', *gá:liŋ > g<um>á:l<im>íŋ* 'comes from'. I$_2$ is more complex than I$_1$ in that it 1) requires a separate -VC- infix after all syllable-initial consonants, 2) the vowel of the infix -Vg- must be identical to the nucleus of the syllable to which the infix is attached, but the vowel of *-um-* agrees only with the nucleus of a word-final syllable, and 3) there are two separate infixes.

I$_3$:   *hindíʔ > h<um>ind<imí:p>iʔ* 'not', *puntá > p<um>ú:nt<amá:p>a* 'goes'. I$_3$ resembles I$_2$ in augmenting the base by double infixation, but here the second infix is -VCVC-, a shape that never occurs in ordinary language.

R$_3$I$_2$:   *hindíʔ > d<im>í:h<in>ín* 'no, not', *saglít > l<um>i:ts<am>ág* 'instant', *saʔán > ʔ<um>a:ns<am>á* 'where?'. In this pattern the original base, which has undergone both complete reversal of the syllable shape of the base and double infixation, is thoroughly disguised. Note that the non-phonemic glottal stop between identical vowels in *saʔán* is preserved in its disguised derivative word-initially, since all words that do not begin with another consonant must begin with glottal stop, while the phonemic glottal stop in *hindíʔ* is lost in its disguised derivative, since Tagalog disallows preconsonantal glottal stop.

Conklin describes some additional patterns, but those given above are sufficient to give a sense of the daunting complexity of this system. Indeed, the net result of having all these options available is a system of speech distortion which makes the conventions of English 'Pig Latin' (*Igpay Atinlay*) look like child's play. Conklin stresses that the range of speech disguise options he has described is more than a single individual normally commands, and that the system is necessarily always in a state of flux since only constant innovation can maintain the function of secrecy that it strives to achieve. With regard to the latter, he notes further (1956:139) that "Although the use of *baliktád* is not restricted to any age, sex, or social group, it is particularly popular among adolescents and unmarried teen-agers. The reasons for learning such a language are manifold: to prevent older relatives, nonrelatives, younger siblings, servants, vendors, and in general any nonmember of one's own small group from understanding one's conversation."   This appears to be particularly applicable

to courting behavior, which (following the Spanish *dueña* system) cannot be conducted without the presence of a female chaperone. Gil (1996) examines some of the formal properties of *baliktád* in greater detail, with particular reference to the
way in which Conklin's **R₁** and affixation interact in terms of ordering relations.

3.4.2 Malay back-slang

Evans (1923:276-277) provides a superficial one-page description of what he calls 'Malay back-slang' (*chakap balek* 'speak in reverse') which he obtained in Negri Sembilan, Malaya. The examples that he gives show little evidence of system, and it is unclear how these forms are learned. The first line of a pantun (traditional sung poem), for example, is given as *rioh rəndah bunyi-nya buroŋ* (clamor low-toned sound-**3sg** bird) 'the chirping of the birds makes a muffled clamor', and this is transformed to *yori yarah nubi nəruboŋ*. While *yori* can be seen as a product of syllable metathesis (with movement of the predictable glide in [rijoh] and deletion of preconsonantal *h*), neither *yarah* nor *nubi* can be derived in this way, and *nəruboŋ* appears to attach the 3sg possessive pronoun as a prefix to the following word, together with metathesis of the first two consonants. Other examples of ordinary Malay with back-slang equivalents show syllable metathesis (*aku* > *kua* '1sg', *pərgi* > *gipər* 'go'), but this is not observed consistently. Evans (1923:276) holds that *chakap balek* "is used by bad-mannered Malay children when they wish to talk secrets before their elders and betters, or before uninitiated companions."

3.4.3 Prokem

Dreyfuss (1983) has described a secret language that arose in the late 1970s among Jakarta youth who were influenced by the peace movement that had swept the United States and various European countries a decade earlier. Although this was mostly a spoken argot, in some cases it made its way into written fiction. He calls this 'the backward language of Jakarta youth' (JYBL). What is noteworthy in each of these three cases (Tagalog, traditional Malay, modern speakers of Indonesian in Jakarta) is the description of speech disguise as 'backward speech'.

Table 3.7 Standard Indonesian and JYBL equivalents

	SI	**meaning**	**JYBL**
1.	bisa	able	bokis
2.	bərapa	how much?	brokap
3.	cəlana	pants	cəlokan
4.	cina	Chinese	cokan[19]
5.	gila	crazy	gokil
6.	janda	divorced	jokan
7.	Jawa	Java	jokaw
8.	kita	we (incl.)	kokit
9.	lima	five	lokim
10.	Madura	Madura	madokur
11.	pənjara	Jail	pənjokar
12.	pesta	party	pokes

---

[19]   For expected **cokin. Possibly a typographical error in the original.

Most of the material described by Dreyfuss is internally more consistent than the *chakap balek* described by Evans. Moreover, it is far simpler than the patterns described by Conklin for Tagalog. Examples appear in Table 3.7 (SI = Standard Indonesian).

Although Dreyfuss proposes a needlessly complex analysis that appeals to substratum influence from Javanese and Sundanese, the rule here seems clearly to be 1. insert the sequence *-ok-* immediately before the penultimate vowel, 2. drop the final vowel, and 3. reduce a resulting word-final consonant cluster by dropping the second consonant. All words formed in this manner are derived from ordinary language forms that end with *-a*. Dreyfuss notes two examples of words with other final vowels that have disguised forms made in accordance with the same pattern: *bəli > bokəl* 'to buy', *bəgini > bəgokin* 'like this', and he very briefly cites 'more mundane backwards language backwardisms' such as *ribut > birut* 'noisy' and *habis > bais* 'finished'.

According to Prathama and Chambert-Loir (1990), as early as 1981 this argot was called 'Prokem', apparently a derivative of *preman*. The *Kamus Besar Bahasa Indonesia* (Moeliono et al. 1989) gives two entries for the latter: 1. private; civilian (not military); privately owned, and 2. epithet for thieves, pickpockets, etc. Prathama and Chamber-Loir derive the name 'Prokem' from the latter entry, and this association is repeated by Fox (2005:99). Since the purpose of an argot is to create a blanket of secrecy around orally transmitted messages, however, one would think that the first entry (cf. Javanese *preman* 'private, privately owned') would be a more plausible source for the name. In their dictionary of Prokem they cite somewhat over 1,000 forms with their Standard Indonesian (SI) equivalents, and it is clear from this material that the pattern of disguise described by Dreyfuss is only one of several that are employed in this register, although it may be the most frequent. Bases that begin with a vowel or *h-* almost never take *-ok-* infixation (the only example I have found is *əmpat > tokap* 'four', but this is irregular in a number of ways). Instead, VC- bases show 1. metathesis: V(N)C- > (N)CV- (where prenasalised stops move as a unit), 2. cluster reduction: simplification of an initial consonant cluster by loss of the *first* member, and 3. glottal stop insertion: breaking a resulting vowel cluster by insertion of a glottal stop: *atas > taʔas* 'above, on top', *enak > neʔak* 'tasty', *ibu > biʔu* 'mother', *utaŋ > tuʔaŋ* 'debt', *hari > raʔi* 'day', *ambil > baʔil* 'fetch, get', *əmpat > paʔat* 'four' (variant of *tokap*). Some consonant-initial bases also metathesise the first two consonants (*kawin > wakin* 'marry', *pərut > rəput* 'belly', *ribu > biru* 'thousand'), but this is a minority pattern. Finally, a very few bases show complete phonemic reversal, as with *ayam > maya* 'chicken, fowl'. Fox (2005:100) notes that "Prokem is also popularly referred to as *ngomong labik* 'backwards speech' because of its various forms of metathesis: thus, for example, *balik* 'return, reverse' becomes *labik*; *bikin* 'do, make' becomes *kibin*." He adds that "Prokem was the street fashion of the 1980s, especially in Jakarta, and it has since given way to—or developed into—other slang registers, one of which is known as *bahasa gaul*" ('mixed language'). The latter register reportedly makes heavy use of the infix *-in-*, sometimes more than once in the same word, as with *banci > binancini* 'transvestite'.

Other forms of speech disguise undoubtedly are present in various parts of western Indonesia. Dreyfuss (1983:56), for example, states that "In East Java there exists a variety of backwards language in which the backwards language word is the mirror image of its Javanese or SI counterpart." He illustrates this with Javanese, Standard Indonesian *gadis*, East Javanese backwards language *sidag* 'young girl', and similarly *manis > sinam* 'sweet'. Likewise, Fox (2005:99), referring to an unpublished University of Leiden doctoral dissertation written by Th. C. van der Meij in 1983 indicates the presence of a

"disguise register used by transvestites in Jakarta." In each of these cases, whether the participating group is defined by generation, urban upbringing and the influence of western pop culture, sex, sexual orientation or other socially distinguishing factors, the motive force behind such special speech registers appears to be the need not just to define group membership, but to provide a private means of communication among members of a social subgroup. This need probably is more pressing in socially crowded environments, where individual privacy is more difficult to achieve than in many parts of the West, and for this reason an artificially devised code like 'Pig Latin' may seem more like a parlor game to Westerners, but the similar speech registers of AN speakers carry a far more serious value as social 'curtains'. Gil (2002) provides a valuable survey of ludlings in Malayic languages that came to my attention too late to include in the first edition of this book. It covers many of the same topics, but with additional details of form and function.

One last point needs to be made in relation to general phonological theory. Some of the devices described for Tagalog and Indonesian speech disguise are processes that occur in ordinary language change or occasionally in synchronic grammars, as with Conklin's R_2, but others are unfamiliar from the standpoint of normal language change or synchronic phonological processes. This includes all other reversals in Tagalog (complete phoneme reversal, complete syllable reversal, partial syllable reversal), as well as various of the devices used in Prokem. The reasons for this difference are of interest to general linguistic theory, and cannot be pursued here. However, it seems clear that secret languages are a product of the conscious manipulation of a linguistic code, whereas at least most ordinary sound change is unconscious, and largely driven by articulatory or phonological demands that lie beyond the control of the speaker. Given these differences of mechanism it should not be surprising to find corresponding differences of form.

3.4.4 Hunting languages, fishing languages, and territorial languages

Special languages connected with hunting and fishing are reported from a number of AN-speaking societies. According to van Engelenhoven (2004:21), for example, "In the Luang language there is a special lexicon to refer to animals hunted at sea. Remnants of this secret language were found by accident during a fieldwork session … Here, an old woman spontaneously provided alternative names for fish only to be used on the reef."

The most thorough treatment of this topic appears in Grimes and Maryott (1994), who describe a number of speech registers used only in hunting and fishing contexts in various parts of Indonesia. Grimes points out that the most common taboos on Buru in the central Moluccas are territorially restricted. Although dialect differences in Buruese are also associated with a territory, territorially restricted taboos require that the speaker use alternative lexical items for particular referents while in a given part of the island. In the Wae Lupa area of south-central Buru, for example, *menjaŋan* 'deer' (borrowed from Malay, but assimilated by all dialects), must be replaced by *wadun*, a word that normally means 'back of the neck'. Other taboos are said to be associated more with an activity than a territory: 'For example, during particularly heavy east monsoons groups of men will organize extended hunting in the jungle for three to five months in which there are special behavioural and linguistic taboos that must be strictly adhered to.'

Grimes notes that word taboos on Buru are formed by lexical substitution rather than phonological deformation. This can be accomplished in one of several ways. One strategy is to use semantic shift. While most dialects of Buru say *manu-t* 'bird' (PMP *manuk) and *pani-n* 'wing' (PMP *panij), for example, the Lisela dialect says *pani* 'bird'. A second

strategy is to use circumlocutions, as with *inewet* 'snake (generic)', literally 'the living thing', and *isaleu* 'python', literally 'a thing that goes ahead.'

The most striking example of linguistic taboo found on Buru is said to be associated with an uninhabited zone called 'Garan' in the northwest quadrant of the island, and here any language except Buruese is acceptable. As a result of this taboo a 'language' that is phonologically and syntactically Buruese, but lexically distinct has come into being. This register, called 'Li Garan' has no native speakers, but is spoken by both men and women in the Rana region of central Buru, and is taught to their children from an early age. If someone should violate linguistic etiquette by speaking Li Garan incorrectly (or presumably by using normal Buruese) while in Garan, this may precipitate 'sudden and violent storms, wind, rain, thunder, lightning, branches breaking and trees falling over, or other disturbing consequences that may extend to future generations' (Grimes and Maryott 1994:281). Although Grimes does not make the connection, this association of linguistic behavior, territory, taboo, and supernatural retribution on Buru is strikingly similar to a variant of the 'thunder complex' reported among the Tawala of southeast Papua, where "There are certain places in the bush where a person is not allowed to speak a different language, or dance, sing, shout, take photos, or make jokes," lest these actions precipitate a destructive thunderstorm (Blust 1991b:520).

In the same article Maryott describes the 'sea speech' (called 'Sasahara') of the Sangirese in the Sangir Islands between Sulawesi and Mindanao. This speech register, first reported by the Dutch linguist Nicolaus Adriani in 1893, is not grammatically distinct, and contains many ordinary words of Sangirese. However, it also contains many novel words that the Sangir use at sea, reportedly to prevent the sea spirits from overhearing and possibly interfering with their plans or intentions. Sasahara is used mainly by men while fishing or engaged in other activities at sea, and in the past this included maritime warfare. Words in the Sasahara register are said to be created by 1. circumlocution, 2. semantic disguise, 3. phonological disguise, 4. use of loanwords, or 5. in other, unknown ways. Circumlocution takes various forms, in one of which 'dog' might be called 'that which is always barking,' but with the longer form encoded in a single morphologically complex word. Semantic disguise makes use of normal Sangir words, but with meanings altered from slightly to drastically (e.g. 'turn oneself around' > 'return', 'think, reflect' > 'sleep'). Phonological disguise appears to be minimal, and unsystematic.

It is unclear whether hunting languages and fishing languages exist independently of 'territorial languages', or whether all of these are ultimately territory-based. In hunting languages taboos are usually described as applying only within the territory of the prey, and it is clear from Maryott's description that fishing languages apply among speakers who are at sea, whether they are fishing or intend to fish, or not. The Buruese Li Garan clearly is territory-based, without further reference to type of activity performed there. Comparison with other cases such as the Tawala of Southeast Papua, however, suggests a taboo against 'confusion of categories' that was traditionally believed to precipitate a punitive thunderstorm, and one type of category confusion is use of a language foreign to the territory in which it is uttered.

3.4.5 Iban antonymy

One case of possible speech disguise through semantic reversal has been noted in the AN literature. In Iban of southwest Sarawak a number of lexical items have meanings that can be characterised as reversals of the reconstructed sense (Blust 1980c). Semantic

change is a natural part of language differentiation, but Iban shows what appears to be a systematic polarisation of meanings in some words, as seen in Table 3.8:

Table 3.8 Semantic reversals in Iban

	Malay	Iban
1.	(h)aŋit 'foul-smelling'	aŋit 'fresh or fragrant smell'
2.	ampul 'expand, be blown out'	ampul 'soft, light, flabby'
3.	bəlaŋ 'banded (in coloring)'	bəlaŋ 'whitened (as coat with flour)'
4.	bərahi 'sexually excited'	bərai 'languish, pine for, miss'
5.	boŋkar 'heave up'	boŋkar 'pull down' (Scott 1956)
6.	caloŋ 'bailer, ladle'	caloŋ 'bung, cork'
7.	daboŋ 'preliminary tooth-filing'	daboŋ 'tooth or notch in serration'
8.	(h)ibur 'solace, comfort'	ibur 'shocked, distressed'
9.	itek 'duck'	itik 'chicken'
10.	kampoŋ 'hamlet, village'	kampoŋ 'forest never yet cleared'
11.	kilau 'dazzled'	kilau 'dusk, just after sunset'
12.	ladaŋ 'dry field'	ladaŋ 'swamp or wet paddy field'
13.	liut 'lithe, leathery, tough'	liut 'soft, silky'
14.	ñam-an 'tasty, delicious'	ñamñam 'tasteless, insipid'

Without further information it would be difficult to determine the direction of semantic change in these comparisons, but external witnesses invariably show that Malay is the conservative language and Iban the innovator. Whatever function antonymy once had in Iban has been lost, and its former presence in the language must be inferred on the basis of these 16 words and perhaps a few others (e.g. Cebuano *lúbus* 'tied, wound around tight', Iban *ləbus* 'of knots, ropes, etc., become loose, slip', where a Malay cognate is unknown). Comparison with living systems of antonymy, such as that of the Warlpiri of central Australia (Hale 1971), however, suggests that semantic reversals may have been used as a means of speech disguise, and that some of these entered the ordinary language, where they survived as clearly patterned semantic changes.

3.5   Ritual languages

Another type of speech register that has often been noted in the literature on AN languages can be subsumed under the general heading of 'ritual languages'. Fox (1988), who has done perhaps the most extensive work in this area, describes the common feature of ritual languages as the use of 'formulaic frames' that generally take the form of canonical parallelism. He notes that parallelism was first defined in 1753 in relation to Hebrew verse, and has since been recorded in the poetry or formal ritual language of many traditional cultures. In these "tradition demands that certain compositions be given dual expression. Words, phrases, and lines must be paired for a composition to be defined as poetry, ritual language, or elevated speech. Moreover, many of these traditions also prescribe, always with varying degrees of freedom, what words, phrases, or other elements of language are to be paired in composition" (Fox 1988:3). He calls this 'canonical parallelism', and illustrates it with a passage from a Rotinese death chant that includes the following lines in English translation (1988:16):

My boat will not turn back
And my perahu will not return.
The earth demands a spouse
And the rocks require a mate.
Those who die, this includes everyone,
Those who perish, this includes all men.

Similar uses of reiteration with variation are described for the ritual languages of many eastern Indonesian peoples, as well as speakers of other AN languages, and it is clear that this poetic device is both widespread and presumably ancient in the AN world.

The lexical choices available in such paired lines are delimited by convention, and these are referred to as 'dyadic sets'. Fox (1993) is an invaluable sourcebook on the specific lexical pairings that make up the substance of canonical parallelism. It contains over 1,500 entries, each of which is cross-referenced to one or more 'links'— lexical items with which it must conventionally be paired in successive lines of verse. Illustrative examples include *alu(k)* 'rice pestle', link: *nesu(k)* 'rice mortar', *dala(s)* 'row, path, road', link: *eno(k)* 'way, path, road', *do(k)* 'leaf', links: *aba(s)* 'cotton', *ai* 'tree, stick, wood', *baʔek* 'branch of tree, antler', *bifa(k)* 'lip, mouth, edge, rim', *hu(k)* 'trunk, root; origin', *ndana(k)* 'limb, branch of a tree', *pena* 'open bulbs/bolls of cotton', *tea(k)* 'hard, strong, firm, as wood or stone'. As the last example, with its multiple links suggests, dyadic sets are not prescriptive, but restrictive. In other words, if *do(k)* 'leaf' appears in a line of ritual language the succeeding line may contain any term from the set defined by the stated links, but no given term is prescribed. Where there is a single link, as with *alu(k)* and *nesu(k)* the choice of dyadic mate may appear to be prescriptive, but simply selects from a smaller set of available forms.

The earliest known use of ritual parallelism in an AN language was noted in Ngaju Dayak of southeast Borneo by the Swiss missionary August Hardeland in 1858. In his 374-page grammar of Ngaju Dayak Hardeland included a 136-page transcription of a liturgical text recited in a special 'language' by a *balian* (shaman) at the time of a death feast. Although he did not give it a name, later writers have called this register *basa sangiang*, or the language of the ancestral spirits (Schärer 1963). According to Schärer (1963:10) the sacred literature of the Ngaju is sung by shamans in *basa sangiang*, "the language of the first ancestors, which was spoken by the gods and the first men and is still spoken today. All sacred stories are transmitted in the *basa sangiang*, and all profane stories are told in ordinary Ngaju." In addition to songs, the *auch oloh balian* or priestly chants are also recited in the *basa sangiang*, and a knowledge of these is transmitted orally from teacher to pupil through years of apprenticeship.

Fox (2005) cites references to similar liturgical registers that make use of canonical parallelism in a number of other AN languages, including Kendayan, Mualang and Iban in southwest Borneo, Berawan of northern Sarawak, Timugon Murut of western Sabah, Nias in the Barrier Islands west of Sumatra, Malagasy, Bolaang Mongondow in northern Sulawesi, Sa'dan Toraja of central Sulawesi, Buginese of south Sulawesi, some groups of Atoni in west Timor, and Puyuma of southeast Taiwan, among many others.

3.6   Contact

Most examples of language use discussed in this chapter involve speech registers, and appear to be motivated by a social subgroup wishing to distinguish itself, or to conceal

messages from others within the same language community. Language contact and contact-induced convergence are different from these in arising from the interaction of distinct linguistic communities. However, these interactions are also governed by important social considerations, and therefore will be considered here.

3.6.1 Ordinary borrowing

There is probably no language anywhere that has not borrowed some features of vocabulary or structure from others, and AN languages have been both recipients of loanwords from many other languages, and sources of loanwords in some languages as well. Languages with particularly large and conspicuous numbers of loanwords from non-AN sources are generally confined to western Indonesia, where important intensive cultural contacts with Indian civilisation began some 2,000 years ago.

Zoetmulder (1982) has compiled a 2,368 page dictionary of Old Javanese, and he notes (1982:ix) that "The influence of Sanskrit on Old Javanese has been enormous. Of the more than 25,500 entries in this dictionary, over 12,600, that is almost half the total number, go back, directly or indirectly, to a Sanskrit original." It is important to keep in mind that Old Javanese is known primarily through textual material that was composed in the courts of Hindu-Buddhist rulers. The form of this language that has survived was therefore a vehicle of literary expression, and much of the Sanskrit vocabulary found in Zoetmulder's voluminous dictionary probably was unknown to the mass of ordinary Javanese living at the time. This inference is supported by the far lower frequency of Sanskrit loanwords in modern Javanese, which is a continuation of the earlier spoken language. The proportion of Sanskrit loanwords in modern Javanese is unknown, but probably does not exceed 5%, and includes very few items of basic vocabulary (among the rare exceptions are *səgara* 'sea', *təlaga* 'lake', and *mega* 'cloud' in the Ngoko register of modern Yogyakarta Javanese), all of which undoubtedly derive from recitations of the Indian epics in the traditional shadow puppet theatre.

The percentage of Sanskrit loanwords in Malay apparently never approached the levels found in Old Javanese, and modern Malay probably is no richer in these forms than modern Javanese, although several hundred examples can be found in either language (Gonda 1952). Sanskrit loanwords are also found in many of the lowland languages of the Philippines. These almost certainly were acquired through contact with Malay, although in some cases the form of Sanskrit loans in Philippine languages is more conservative than in modern Malay, as with Tagalog, Aklanon *mukháʔ*, Malay *muka* 'face' < Skt. *mukha*, 'mouth, face, countenance' (cp. Old Javanese *mukha*, 'mouth, face, countenance'). In general, as one moves further away from Tagalog the number of Sanskrit loans decreases, a correlation that probably arose from the presence of a Malay-speaking trading colony in Manila Bay prior to the arrival of the Spanish (Wolff 1976).

Subsequent to the period of Indianisation in southern Sumatra and Java, Islam was introduced near the end of the thirteenth century. Although Arabic influence does not appear to have been as overwhelming in any one AN language as that of Sanskrit in Old Javanese, Arabic loanwords are in many ways more conspicuous in modern Malay and Javanese than are older Sanskrit loans. Jones (1978) lists well over 4,000 Arabic loanwords in Indonesian/Malay, noting that it is often unclear whether a given word of Arabic origin was borrowed directly from Arabic or through the medium of Persian. Arabic loans contain a number of sounds that are not native to Malay or other AN languages of western Indonesia, and these are commonly replaced by more familiar equivalents, as with

colloquial Malay *pikir*, formal *fikir* (Arabic *fikr*) 'to think', or colloquial Malay *kabar*, formal *khabar* ([xabaɾ]) (Arabic *khabar*) 'news'. Borrowing from Arabic in Malay produced a colloquial/formal speech diglossia which sometimes serves as the basis for hypercorrections of *p* to *f* in loanwords from other languages, as with English 'prostitute', pronounced 'frostitute' by some speakers of Malay.

Like Sanskrit loanwords, Arabic loans are found in many languages of the Philippines, particularly those in the Muslim-dominated areas of the south (Mindanao and the Sulu Archipelago). Most of these also appear to have been acquired through Malay. Malagasy likewise has a number of Sanskrit and Arabic loans that seem to derive primarily from early interaction with speakers of Sriwijayan Malay (Adelaar 1989), but this language was subject to additional subsequent influence from Arabic as a result of its geographical location. Loans from both of these sources are rare in eastern Indonesia, and are unknown in the Pacific.

Despite the rather lengthy period of contact between Minnan-speaking southern Chinese (generally called 'Hokkien' in Southeast Asia), and languages in the Philippines and western Indonesia, Chinese loanwords are rare in AN languages. Among examples that readily come to mind are Ilokano *bakiá* 'wooden shoe, clog', Malay *bakiak* 'wooden clogs, sabots', Tagalog *bámiʔ* 'shredded wheat or prawns with wheat flour or vermicelli and cress', Iban *mi* 'noodles', Javanese *bakmi* 'Chinese noodle dish', Tagalog, Cebuano *bibiŋká* 'rice cake with coconut milk', Malay *kueh biŋka* 'cake made of rice flour, coconut milk, egg and palm sugar', Malay *daciŋ* ' hand-held scales', Cebuano *hunsúy*, Maranao *onsoi* 'smoking pipe', and Tagalog *iŋkóŋ* 'appelation for grandfather or very old man', Malay *əŋkoŋ* 'grandfather' (Minnan *a goŋ*). Whereas both the Sanskrit and Arabic loans cover a wide semantic range, Chinese loans tend to be represented most heavily in the area of material culture and commerce. Most of these probably began to diffuse widely in insular Southeast Asia after the beginning of the Ming dynasty (1368-1644), although a few may have been introduced at an earlier period.

Loanwords from European languages are also plentiful in many AN languages, both in insular Southeast Asia and the Pacific. Philippine languages and Chamorro have borrowed heavily from Spanish (among other things, the native numerals of Chamorro have been completely replaced by their Spanish equivalents). As first noted by Lopez (1965), many Spanish nouns have been borrowed into Philippine languages in their plural forms. While this is perhaps understandable for nouns that have a pragmatically plural sense, such as shoes (Spanish *zapato* 'shoe' : *zapato-s* 'shoes', but Tagalog *sapatos* 'shoe' : *aŋ maŋa sapatos* 'shoes'), it is more difficult to understand for many other nouns, such as Spanish *arco* 'arch', Tagalog *alakós* 'bamboo arch or archway', Spanish *guayaba*, Tagalog *bayábas*, Western Bukidnon Manobo *bəyabas* 'guava', Spanish *papaya*, Bikol *tapáyas*, Western Bukidnon Manobo *kəpayas* 'papaya', Spanish *fresa*, Tagalog *presas* 'strawberry'. Malay was first exposed to Portuguese loans in the fifteenth century, acquiring such terms as *bəndera* 'flag', *jəndela* 'window', *məntega* 'butter', *nanas* 'pineapple', and *gubərnadur* 'governor'. Shortly thereafter peninsular Malay was exposed to English, and the Malay of Sumatra and Java to Dutch, a pattern that persisted for over three centuries. In more recent years English loans have been accepted into Bahasa Indonesia as well.

Names of some New World plants that were introduced during the Spanish occupation of the Philippines spread southward with their distinctively transformed pronunciations. Malay *biawas* 'guava', for example, is a regular development of Tagalog *bayábas*, but cannot regularly reflect either Spanish *guayaba* or Portuguese *goiaba*. The guava must, therefore, have been introduced to Southeast Asia by the Spanish in the Philippines, from

whence it spread southward through Borneo to the Malay Peninsula. However, words for 'pineapple' show the opposite trajectory. Whereas a number of Philippine languages have borrowed Spanish *piña* 'pineapple', as with Ilokano *pínia*, Pangasinan, Bikol *pínya*, Hanunóo *pinyá*, Western Bukidnon Manobo *ginya* 'pineapple', Maranao has *nanas*, a word that derives ultimately from Brazilian Portuguese *ananas* 'pineapple', and which must have spread from the south through Malay contact.

In many cases loanwords have been partially or fully adapted to the native phonology, thus transforming them in sometimes striking ways. This is particularly true of Chamorro, which does not permit syllable-final liquids, and replaces them with *t*, both in native words (*qipil > ifet* 'ironwood tree'), and in the large numbers of Spanish loans that have entered the language, as with Spanish *alba*, Chamorro *atba* 'dawn', Spanish *arma*, Chamorro *atmas* 'weapon, firearms', Spanish *legal*, Chamorro *ligát* 'legal, lawful' or Spanish *color*, Chamorro *kulót* 'color'. Undoubtedly the most striking case of phonological adaptation of European loanwords, however, is found in Hawaiian (Elbert and Pukui 1979:28). With only eight consonants and five vowels, and a CV syllable canon, Hawaiian was severely limited in the ways it could model the phonology of a language such as English, and the adaptations thus sometimes deviate radically from the originals, as with the well-known example of *Kalikimaka* 'Christmas', or various English given names (*Lopaka* 'Robert', *Keoki* 'George', *Kamuela* 'Samuel'). Table 3.9 charts the adaptations of the phonemes of English in the process of borrowing into Hawaiian:

Table 3.9 Phonological adaptations of English loanwords in Hawaiian

English	Hawaiian	Examples
p, b, f	p	*Pika* 'Peter', *pia* 'beer', *palaoa* 'flower'
v, w	w	*wekaweka* 'velvet', *waina* 'wine'
wh	hu, w, u	*huila* 'wheel', *wekekē* 'whiskey'
h, sh	h	*home* 'home', *Halaki* 'Charlotte'
l, r	l	*laki* 'lucky', *laiki* 'rice'
m	m	*mākeke* 'market'
n, ng	n	*Nolewai* 'Norway', *kini* 'king'
t, d, th, s, sh,		*kikiki* 'ticket', *kaimana* 'diamond',
z, ch, j, k, g	k	*kipikelia* 'diphtheria', *kopa* 'soap', *palaki* 'brush', *kokiaka* 'zodiac', *pika* 'pitcher', *Keoki* 'George', *kolokē* 'croquet', *Kilipaki* 'Gilbert'
j	i	Iesū 'Jesus'

Borrowing has naturally also occurred from AN source languages into languages belonging to other families. Although examples of AN loanwords in Sanskrit, Arabic or Chinese are unknown, European languages such as Dutch and English have acquired a number of lexical items through centuries-long contact with the Malay world. In English these include *orangutan* (Malay *oraŋ hutan* 'man of the woods'), *pandanus* (Malay *pandan*), *gutta-percha* (Malay *gətah pərca* 'tree sap + strip or piece, as of cloth'), *godown* (Malay *gudaŋ*; the term is ultimately of south Indian origin, but the phonological form of the English word suggests that it was borrowed directly from Malay), *cootie* (Malay *kutu* 'body louse'), and *boondocks* (Tagalog *bundók* 'mountain').

One of the most intriguing examples of borrowing from an AN source into non-AN languages is described by Walker and Zorc (1981). According to the account they present, speakers of one or more AN-speaking groups voyaging out of Indonesia visited the coast

of Arnhem Land in northern Australia from before 1800 until 1906. In the literature these people are commonly referred to as 'Macassan' traders, but most accounts fail to clearly indicate what language they spoke. Their purpose in visiting Arnhem Land was to collect trepang (sea cucumbers) for which there was a constant demand in the Chinese market, and the nature of the contact situation therefore was one in which AN-speaking traders dealt with local aboriginal suppliers. For this trade to work it was necessary to establish a means of communication, and it is clear that the local aboriginal suppliers learned the language of the foreign traders to a far greater extent than the traders learned any of the aboriginal languages with which they came in contact.

Walker and Zorc document in detail the heavy lexical borrowing that took place in one of the indigenous languages of Arnhem Land, Yolngu-Matha, a Pama-Nyungan language spoken by a hunting-gathering and fishing population on Elcho Island, off the coast of northern Australia, and at various other locations on the coast and interior of the Australian mainland. Distinctive phonological changes mark many of these loanwords as deriving from Makasarese, spoken in the major urban center of Ujung Pandang and surrounding locations of southern Sulawesi, but other loans are more difficult to source. Walker and Zorc divide the vocabulary of AN loanwords in Yolngu-Matha into four groups (the number of identified loanwords in each group appears in parentheses): Group 1: most probable Makasarese loanwords (99), Group 2: loanwords that are possibly Makasarese, or possibly from other Austronesian languages (59), Group 3: loanwords that are probably not Makasarese (21), and Group 4: possible Austronesian loanwords, but requiring further research (70). Recurrent phonological adaptations in the borrowing language are clearly stated, thus making the identification of cognates much more certain. A number of 'Macassan' loanwords in Yolngu-Matha are connected with voyaging or fishing, as with YM *baLaŋu*, Makasarese, Buginese *baláŋo* 'anchor', YM *ba:raʔ* 'west(wind)', Makasarese *báraʔ* 'west monsoon', YM *ḏimuru* 'northeast (wind)', Makasarese, Buginese *tímoroʔ* 'east monsoon' (YM *ḏ* = lamino-dental stop), YM *bi:kaŋ* 'fishhook', Makasarese *pékaŋ* 'fishhook, rod', YM *ḏu:mala*, Makasarese *sómbalaʔ* 'sail', YM *gapalaʔ* 'large boat; rudder', Makasarese, Buginese *káppalaʔ* 'boat', YM *garuru* 'sail', Makasarese, Buginese *karóroʔ* 'coarse cloth or leaves woven into sail', YM *gulawu* 'pearl', Makasarese *kúlau* 'any stony-hard substance, as mother-of-pearl', seeds in fruit, etc.', or YM *jalataŋ* 'south (wind)', Makasarese, Buginese *sallátaŋ* 'south wind, land wind'. With regard to YM *ba:raʔ* and *ḏimuru*, Walker and Zorc (1981:118, fn. 42) comment 'Informants state that *ba:raʔ* was the wind used by Macassans to sail from Ujung Pandang to Australia, and *ḏimuru* … was the wind used to return. This knowledge is borne out to be factual … and is yet another instance of the knowledge retained about Makassans so long after they have ceased coming to Australia.'

Wallace (1962:309) describes the annual voyages of the Makasarese to the Aru Islands north of Arnhem Land to secure a variety of goods, including pearls, mother-of-pearl, turtle shell, edible bird's nests and sea cucumbers for the Chinese and European markets. He reported that "The native vessels can only make the voyage once a year, owing to the monsoons. They leave Macassar in December or January, at the beginning of the west monsoon, and return in July or August with the full strength of the east monsoon." The loan vocabulary thus supports an inference that aboriginal men from Arnhem Land sometimes accompanied the visiting Makasarese on their ships in search of products for the Chinese and European markets. Other terms, such as YM *balaʔ* '(European style) house', Makasarese *ballaʔ* 'house', YM *baluŋa* 'pillow', Makasarese *paʔluŋaŋ* 'wooden headrest', YM *bi*(:)*mbi* '(young) sheep', Makasarese *bémbe*, Buginese *bémbeʔ* 'goat', YM

di:tuŋ, Makasarese, Buginese *tédoŋ* 'carabao, water buffalo', YM *du:ka*, Makasarese *tukaʔ* 'steps, ladder' and YM, Makasarese *jaraŋ* 'horse', make it virtually certain that some aboriginal crew members voyaged back to Ujung Pandang with the Makasarese, and brought a knowledge of the world of central Indonesia back to aboriginal northern Australia, since large domesticated animals could not be transported on Makasarese prahus, and ladders were used only in houses. Urry and Walsh (1981) cite historical sources dating as early as the late 1830s and early 1840s confirming this inference: "Nearly every prahu on leaving the coast takes two or three natives to Macassar, and brings them back the next season. The consequence is that many of the natives all along the coast speak the Macassar dialect of the Malayan language." Apart from the error of considering Makasarese a dialect of Malay this observation is fully consistent with the linguistic data cited by Walker and Zorc.

As already noted in connection with languages such as Mailu or Maisin in New Guinea, AN loanwords have also been adopted in a number of Papuan languages in western Melanesia. Far more difficult to detect are loanwords from one AN language into another, but as might be expected, these are very common. Because of its key role in early trade, Malay has been a major source of loanwords in island Southeast Asia for many centuries. This has greatly complicated the task of subgrouping the languages of this area, and led Dempwolff (1934-1938) to posit a number of erroneous 'Uraustronesisch' reconstructions. Among common Malay loanwords in coastal Sarawak are *ajar* 'learn', *arak* 'rice wine', *bagi* 'divide', *baju* 'shirt', *bawaŋ* 'onion', *bərani* 'brave, bold', *buŋa* 'flower', *caŋkul* 'hoe', *dagaŋ* 'trade', *guntiŋ* 'scissors', *harga* 'prince, cost', *janji* 'promise', *jala* 'casting net', *katil* 'bedsted', *kərja* 'work', *kuniŋ* 'yellow', *meja* 'table', *pakai* 'use', *rajin* 'hard-working, industrious', *ramai* 'bustling, busy', *rantai* 'chain', *rugi* 'lose (in a business transaction)', *sapi* 'cow', *səndiri* 'oneself', *səluar* 'trousers', *taji* 'metal cockspur', *tilam* 'mattress', and *toloŋ* 'to help'. Many of the same loanwords appear in Philippine languages, and elsewhere in insular Southeast Asia. Where they derive from non-AN sources and were spread through Malay, as with *arak* (Arabic), *baju* (Persian), *harga* (Sanskrit), or *meja* (Portuguese), they are easily detected as borrowings, but terms such as *bagi*, *buŋa*, or *taji* present greater comparative difficulties.

The motivations for borrowing are varied. In the case of many items of material culture borrowed terms (generally nouns) entered AN languages as a result of exposure to novel cultural products. Some of the most obvious of these are words such as Malay *bəndera* 'flag' (Portuguese *bandeira*), *məntega* 'butter' (Portuguese *manteiga*), or Malay *jəndela* 'window' (Portuguese *janela*). Some cases of this type, however, are more problematic. The appearance of Portuguese *roda* 'wheel' in many of the languages of western Indonesia (Malay, Toba Batak, Balinese, Sasak, Makasarese *roda*, Javanese *roḍa* 'wheel') suggests that wheeled vehicles were unknown until the Portuguese introduced horse-drawn or ox-drawn wheeled carts in the sixteenth century. The shape of the similar loanword in Philippine languages (Ilokano *ruéda*, Bikol *róyda*, Maranao *rida* < Spanish *rueda* 'wheel') suggests that the wheel was independently introduced into the Philippines during roughly the same time period by the Spanish. Surprisingly, however, apparent loans such as Old Javanese *padāti* 'cart (ox-cart)?', Malay *pədati* 'cart' are at variance with this inference, at least as regards western Indonesia.

It is often difficult to separate need from prestige (Clark 1982b). Words such as Malay *bəndera*, *məntega* and the like were borrowed because their referents were novel and needed to be named. At the same time, novel cultural introductions are associated with prestige, since those who acquire them first are likely to be regarded as more worldly and

privileged. Prestige alone, however, appears to have determined the direction of borrowing in cases where no novel cultural products were introduced. Why are Malay loanwords widely distributed in insular Southeast Asia, yet almost totally absent from languages of mainland Southeast Asia? This distribution can hardly be separated from the history of trade in the Indonesian Archipelago, and the central role of speakers of Malay in widespread trade networks. Those who arrived as traders were travelers, and occupied a superior position vis-à-vis the populations they encountered in their home territories. As a result many Malay lexical items were borrowed, even when they have no obvious or direct connection with trading activities: Ilokano *ádal* 'learning, education', Tagalog *áral* 'admonition, counsel' < Malay *ajar* 'instruction; learning', Tagalog *taŋháliʔ* 'late morning, noon' < Malay *təŋah hari* 'midday', Kayan *dian* 'candle' (< Malay *dian* 'candle', Kayan *lame ~ rame* 'noise and excitement of people enjoying themselves in a group' (< Malay *ramai* 'bustling, lively'), Kayan *təkjət* 'unexpected, startling' (< Malay *tər-kəjut* 'startled'). In much the same way, Tongan loanwords in languages such as East Futunan, East Uvean or Rotuman reflect the superior position of Tongan speakers vis-à-vis speakers of the recipient languages during periods of past military conquest and the formation of the 'Tongan empire' (Geraghty 1994). Prestige relationships, however, may change over time. The distinctive sound change *R > *y* marks some Tagalog words as old loans from Kapampangan (*baRani > *bayáni* 'hero', *zaRum > *ka-ráyom* 'needle', *taRum > *táyom* 'indigo'). During the recent past Tagalog has had far greater prestige than Kapampangan, and borrowing has been almost exclusively in the opposite direction, as seen in such likely Tagalog loans as Kapampangan *águs* 'current' (Tagalog *ágos* 'current' < *PMP qaRus), *bágyu* 'typhoon' (Tagalog *bagyó* 'typhoon' < PMP *baRiuh), or *gátas* 'milk' (Tagalog *gátas* 'milk' < PMP *Ratas). However, Tagalog words such as *bayáni* 'hero' show that Kapampangan must have enjoyed a higher status vis-à-vis Tagalog at some point in the past. Since Tagalog speakers are an intrusive population in the region of Manila Bay, while Kapampangan speakers represent a population that has been in central Luzon for a much longer time, borrowing from Kapampangan into Tagalog may have taken place primarily in the early phases of contact, when Tagalog speakers were still a minority in what is today the Tagalog heartland.

A somewhat different set of social circumstances seems to lie behind the appearance of Malay and Javanese loanword in Siraya, and evidently other Formosan languages. When the Dutch established commercial and missionary activities in southwest Taiwan from 1624-1662 they brought some speakers of Malay and Javanese with them, and contact between these individuals and the indigenous Siraya population led to the introduction of a number of loanwords into Siraya (Adelaar 1994b). Some of these then spread to other Formosan aboriginal groups, giving rise to comparisons that could be used to support what Mahdi (1994a,b) has called 'maverick protoforms', as with Malay *surat* 'thing written', Siraya *s<m>ulat*, Saaroa *s<um>a-suɬatə* 'to write', Kavalan *s<m>ulal* 'to write', *sulal-i* 'write it!', a comparison that is superficially plausible, but which contains multiple irregularities in sound correspondences.

Heavy borrowing may involve features of structure as well as vocabulary, and this leads to two different types of results. On the one hand, it may cause the boundaries of language families or of subgroups within a language family to become blurred, producing a linguistic area, or *Sprachbund*. On the other, it may affect only a single language, but produce distinct speech strata indicative of different historical traditions within a single modern language community. In general, structural borrowing is associated more with *Sprachbunde* and lexical borrowing more with speech strata.

3.6.2 Sprachbunde

Linguistic areas arise because structural features diffuse between languages that belong either to different families, or to different branches of the same family. Mere typological uniformity is not sufficient evidence that a collection of geographically contiguous languages forms a linguistic area. Most languages of the Philippines, for example, share a highly distinctive morphosyntactic typology, very similar phoneme inventories, and other structural features, yet the Philippines is not a linguistic area, since these shared traits are partly a product of common inheritance from PAN (the verb system), and partly a product of phonological mergers that probably happened independently (the loss of a palatal series from the phoneme inventories, etc.).

Perhaps the clearest large scale example of a linguistic area in the AN world is seen in the Chamic languages, which have assimilated typologically to their Mon-Khmer neighbors to such an extent that Schmidt (1906) misclassified them as 'Austroasiatic mixed languages'. Thurgood (1999) has detailed the changes that have led from a western Indonesian to a Mon-Khmer typology in languages that are genetically quite closely related to Malayic. The first change apparently was a shift of the primary stress from the penultimate to the final syllable. With final stress the penultimate vowel then tended to weaken. Apart from Acehnese, which evidently returned to insular Southeast Asia before this tendency had progressed far, and Northern Roglai, all other Chamic languages show vowel neutralisations in the penult. In Proto Chamic words that began with ʔ + vowel (which did not contrast with vowel-initial bases) the neutralised vowel in all daughter languages except Acehnese and Northern Roglai normally is *a*; in words that began with any other consonant it is schwa, or zero between consonants that form a pronounceable cluster (stop + liquid, nasal + liquid): *ʔasɔw > Acehnese *asɛə*, Rade, Jarai, Chru, Northern Roglai *asəw* 'dog', *ʔiduŋ > Acehnese *idoŋ*, Rade, Jarai, Chru *aduŋ*, Northern Roglai *iduk* 'nose', *ʔurat > Acehnese *urat*, Rade *aruat*, Jarai *arat*, Chru *araʔ*, Northern Roglai *uraʔ* 'vein', *huma > Acehnese *umʌŋ*, Rade *həma* 'swidden field', Jarai, Chru *həma*, Northern Roglai *huma* 'rice paddy', *lima > Acehnese *limʌŋ*, Rade *ema*, Jarai *rəma*, Chru *ləma*, Northern Roglai *luma* 'five'. Loss of the unstressed vowel triggered a series of structural consequences including: 1) the development of a variety of initial consonant clusters typical of Mon-Khmer languages, but not of AN languages (PMP *beli > Rade *blɛy*, Jarai, Chru, Northern Roglai *bləy* 'buy', PMP *beRas > Rade, Jarai *braih*, Chru *bra:h*, Northern Roglai *bra* 'pounded rice', PMP *duRi > Jarai *drəy*, Chru *druəy* 'thorn', Proto Malayo-Chamic *hulun > Rade, Jarai *hlun*, Chru *həlun*, Northern Roglai *hulut* 'slave, servant', PMP *malem > Rade, Jarai *mlam*, Chru *məlam* 'night'), 2) the development of a series of apparently aspirated consonants (PMP *paqit > Proto Chamic *phit 'bitter'; closer inspection shows that these are still clusters of a stop + h), and 3) the development of a series of preglottalised voiced stops: PMP *bahu > PC *ɓɔw 'stench', PMP *buhek > PC *ɓuk 'head hair' (Thurgood 1999:86 notes that there is some uncertainty whether these are preglottalised or imploded; he represents them with the symbols for implosive stops, but in my field data on Jarai I recorded preglottalised [ʔb] and [ʔd], with little evidence of an ingressive airstream).

Because it is large, and the typological features that define it spread across at least two distinct language families, the mainland Southeast Asian linguistic area to which the Chamic languages belong is relatively easy to recognize. Some AN languages, however, belong to smaller linguistic areas that are products of diffusion between adjacent languages that are related. One such area is south-central Taiwan, where preglottalised labial and dental consonants are found in languages that belong to three primary branches of AN:

Thao (Western Plains), Bunun (independent), and Tsou. This phonetic feature is distinctive for these languages, which form a geographically contiguous block, and it is therefore almost certainly a product of diffusion. The one other conspicuous linguistic area in which AN languages participate is Melanesia. On the whole this is much more diffuse than the mainland Southeast Asian linguistic area, since there is considerable variety within Melanesia. Nonetheless, some linguistic features, such as the high frequency of quinary numeral systems and serial verb constructions, which are unusual elsewhere in AN but common in Papuan languages, seem to have spread by contact. In western Melanesia, where contact between AN and Papuan languages is well-attested this is not problematic, but in Vanuatu, New Caledonia and the Loyalties the source of such apparent Papuan contact influences is far more problematic (Blust 2005a, Lynch 2009b).

3.6.3 Speech strata

When borrowing has been intensive over a period of time it may lead to stratification of the vocabulary. English, for example, has a distinct stratum of Romance loanwords that behaves differently than the native vocabulary with respect to some phonological processes. Several AN languages are known to have speech strata, but in general these can only be clearly distinguished in comparative perspective, and so require knowledge of reconstructed forms. Four such cases will briefly be described.

3.6.3.1   Ngaju Dayak

Speech strata in an AN language were first reported in Ngaju Dayak by Dempwolff (1922; further elaborated in Dempwolff 1937:52), where the majority of irregular phonological developments were attributed to an 'old speech stratum' (OSS). While Dempwolff's identification and classification of sound correspondences was sound, his notions of the Ngaju Dayak speech strata were not. Dyen (1956a) showed that Dempwolff's 'Old speech stratum' is in fact the native vocabulary, while what Dempwolff called 'regular' reflexes in this language consist of a loan stratum from Banjarese Malay, as shown in Table 3.10, where diagnostic reflexes of six PMP phonemes in the two speech strata are shown in relation to both PMP and to Standard Malay (SM):

Table 3.10 Ngaju Dayak speech strata

PMP	OSS	Regular	Malay
*e	e, ɛ	a	ə, a
*-a(h)	-ɛ	-a	-a
*R	h	r	r
*q	loss	h	h, loss
*D	-r-	d	d
*c	s	c	c

As Dyen pointed out, the 'regular' (= 'directly inherited') reflexes closely resemble the development of Banjarese Malay (in which SM *ĕ*, *a* = *a*, and SM *h*, loss = *h*), thus pointing to borrowing, while the 'Old speech stratum' is the native vocabulary, and not indicative of any extraneous source. This stratification of the lexicon is seen not only in recurrent differences in the reflexes of PMP phonemes, but also in lexical doublets such as PMP *beReqat > OSS *bəhat* 'heavy', 'Regular' *sa-barat* 'as heavy as' (Banjarese *barat*), PMP

*baRah > OSS *bahε* , 'Regular' *barah* 'ember' (Banjarese *bara, barah*), or PMP *hiRup 'sip' > OSS *ihop* 'drink', 'Regular' *hirup* 'sip' (Banjarese *hirup*). The methodological lesson that Dyen drew from this case of mistaken identity is that where heavy borrowing has occurred between related languages the native stratum is more likely to be richly exemplified in basic vocabulary. This is overwhelmingly the case in Ngaju Dayak, and although doublets occur, the native term is more likely to be basic ('heavy' as against the fixed expression 'as heavy as', 'drink' rather than 'sip', etc.).

In hindsight the presence of two speech strata in Ngaju Dayak is not surprising. This language is spoken in the Barito River basin, inland from the important port of Banjarmasin, which probably was established by Sriwijayan Malays as a trading station by the seventh or eighth century AD. Coastal Malay traders would have interacted frequently and fairly intensively with at least some interior populations that supplied jungle products in exchange for manufactured goods. Under these circumstances a heavy stratum of Malay vocabulary pointing specifically to Banjarese Malay would be expected in some of the indigenous languages of the area.

3.6.3.2 *Rotuman*

Like Ngaju Dayak, Rotuman, spoken in the tiny Rotuman Archipelago, some 500 km. northwest of Fiji, has two speech strata that Biggs (1965) labels 'stratum I' and 'stratum II.' Distinctive reflexes representing these strata are shown in Table 3.11 in relation to their source phonemes in a hypothetical language that Biggs calls 'Proto Eastern Oceanic' (PEO). The orthography is converted to modern views about Proto Oceanic:

Table 3.11 Rotuman speech strata

PEO	p	t	dr	k	l	q	s
I	h	f	t	ʔ	l	Ø	s
II	f	t	r	k	r	ʔ	s/h

Examples illustrating these differences are: *patu > *hɔfu* 'stone' (I) but *panaq > *fana* 'shoot' (II), *tolu > *folu* 'three' (I), but *tokon > *toko* 'staff, pole' (II), *dranum > *tɔnu* 'fresh water' (I), but *drano > *rano* 'lake; swamp' (II), *kulit > *ʔuli* 'skin' (I), but *toka > *toka* 'land, settle down' (II), *piliq > *hili* 'choose' (I), but *limut 'algae, moss' > *rimu* 'lichen sp.' (II), *taqun > *fau* 'year' (I), but *muqa > *muʔa* 'front' (II), *salan > *sala* 'path, road' (I), but *saqat > *haʔa* 'bad'). In dealing with data of this kind one would expect that some forms will be ambiguous for stratum assignment (e.g., POC *suRuq 'juice, sap, gravy' > Rotuman *su* 'coconut milk', *uRat > Rotuman *ua* 'vein, tendon'), and that strata will not be mixed when two diagnostic reflexes appear in the same morpheme.

Following Dyen (1956a), Biggs appealed to basic vocabulary to distinguish native from non-native strata. This established stratum I as native and stratum II as borrowed from an unspecified Polynesian source. He notes (1965:412) that "Of 328 Rotuman words with etymologies 124 (38%) are directly inherited, 107 (33%) are indeterminable, and ninety-seven (29%) are indirectly inherited." Extrapolating from both unambiguous loanwords and ambiguous forms Biggs concluded further that at least 18% of Rotuman basic vocabulary is borrowed, but this figure rises to 43% of the total lexicon for which diagnostic data is available. It is shown, then, that heavy borrowing of basic vocabulary is possible in some contact situations, but the proportion of loanwords will be greater for non-basic than for basic meanings. The Rotuman case further differs from that of Ngaju

Dayak in showing apparently productive bound morphology in the borrowed speech stratum, most notably the causative prefix *faka-* (< *paka-), and the transitive suffix *-ʔɔki*. Although Biggs is non-committal on the Polynesian source of Rotuman stratum II vocabulary, in most cases this looks like an earlier stage of Tongan, an inference that is consistent with the known history of Tongan conquest in western Polynesia.

3.6.3.3 *Tiruray*

Tiruray, spoken in the mountains of southwest Mindanao, has two distinct speech strata that were not recognised until comparatively recent times. Blust (1992) identified the borrowed stratum as deriving from a Danaw language, one of a group of languages that form part of the Greater Central Philippines group (Blust 1991a). Distinctive reflexes representing these strata are shown in Table 3.12 in relation to Proto Philippines (PPH) and to Maranao, the largest language in the Danaw group:

Table 3.12 Tiruray speech strata

	PPH	1	2	Maranao
01.	-CC-	C	CC	CC
02.	-k-	g	k	k
03.	-b-	w	b, w	b,w
04.	-d-	r	d, r	d, r
05.	-d	r	d	d
06.	-s(-)	h	s	s
07.	R	r	g	g
08.	i/uC	e/o	i/u	u/u
09.	-i/u	əy/əw	i/u	i/u
10.	-iw/uy	əy	iw/uy	iw/uy
11.	-ay/aw	əy/əw	ay/aw	ay/aw

In addition, the sequences *aCa and *aCe often appear as *oCo* in stratum 1 (*anak 'offspring' > *ʔonok* 'offspring; egg; fruit', *hawak > *ʔowok* 'waist', *tabeq > *towoʔ* 'animal fat', but this is not fully regular. Stratum 2 reflexes of these vowels show no change. Thus, *linduŋ > *diruŋ* (Maranao *lindoŋ*) 'seek shelter', *lakaw > *agəw* (Maranao *lakaw*) 'walk, go', *qabu > *awəw* 'ash', *qañud > *anur* (Maranao *anod*) 'drift', *hasek > *ohok* 'dibble', *busuR > *bohor* 'hunting bow', *laki > *lagəy* 'man, male', and *huRas > *urah* 'wash', belong to stratum 1, while *abas* 'pockmarks' (Maranao *abas* 'chicken pox'), *abay* 'side-by-side' (Maranao *abay* 'go by the side'), *ansəd* 'underarm odor' (Maranao *ansəd* 'offensive body odor'), *akuf* (Maranao *akop* 'scoop with both hands', *ubi* (Maranao *obi*) 'yam; sweet potato', *bisu* (Maranao *biso*) 'deaf', and *ʔəgas* 'slab of hard salt; hard interior of certain hardwood trees' (Maranao *gas* 'hard part of log'), *təgas* 'hard, obdurate' (Maranao *təgas* 'harden, solidify; salt obtained from boiling and evaporation; tough') belong to stratum 2. As in Ngaju Dayak lexical doublets reflect native and non-native forms of the same proto morpheme: *Ratas 'milk' > *ratah* 'human breast milk', *gatas* 'milk (purchased)', *sabuR > *sawər/sabug* 'sow seed by broadcasting', *tabaŋ > *towoŋ* 'help; to help', *tabaŋ* 'physical or material assistance rendered'.

While Biggs (1965) reported that at least 18% of Rotuman basic vocabulary is borrowed, as against 43% loans overall, the corresponding figures for Tiruray are at least 29% Danaw loans in basic vocabulary, and 47% overall (Blust 1992a:36). These figures

suggest that exclusive reliance on basic vocabulary to determine the native stratum can be more difficult than Dyen suggested in his study of speech strata in Ngaju Dayak. Tiruray, however, offers a second means of discriminating speech strata. Since the change *R > *r* is unknown in any other language of the southern Philippines, forms that show this change cannot be loans. By association, then, it can be determined that all other developments in stratum 1 are native: *busuR > *bohor* 'hunting bow', *ikuR > *igor* 'tail', *butiR > *buter* 'wart', *baRiw > *warəy* 'stale, tainted', etc. The use of *unique reflexes* thus provides a second control on inferences about direct vs. indirect inheritance in languages that have clearly demarcated speech strata. Although Biggs did not mention it, much the same could be done by use of the unique Rotuman development *t > *f*.

3.6.3.4   Thao

Thao, spoken on the shores of Sun-Moon Lake in the mountains of central Taiwan, has fewer than fifteen speakers, all born prior to 1938. This community is slowly disappearing by attrition, absorbed by the larger Chinese community of Taiwanese speakers in which it is embedded. One of the features of this language that soon becomes apparent in a comparative context is the presence of two distinct speech strata. As shown in Table 3.13, six Thao phonemes occur only in loanwords:

Table 3.13 Loan phonemes in Thao

PAN	Thao (native)	Thao (borrowed)
*b	f	b
*d	s	d
*l	r	l
*ŋ	n	(ŋ)
ʔ		ʔ
ʔ		h

All six of the segments in the rightmost column of Table 3.13 can be considered loan phonemes in Thao, but only the first three are well-attested, and the last two do not have clear historical sources. The identification of loanwords thus depends heavily on the appearance of *b, d* or *l*. The use of these six markers of indirect inheritance leads to the identification of a large number of Thao lexical items as borrowed, and the use of the first three points quite clearly to the source language as Bunun. While Bunun loanwords in Thao represent a wide semantic range, a surprisingly large number concentrate in the semantic domain of women, women's traditional activities and items of material culture associated with these activities. Examples include *bahat* 'pumpkin', *bailu* 'bean', *baruku* 'bowl', *binanauʔaz* 'woman, wife', *bulwa* 'cooking pan, wok', *hibur* 'to mix, stir things together', *hubuq* 'sprout or young shoot (used in expressions asking a woman how many children she has borne)', *kudun* 'clay cooking pot', *lishlish* 'to grate, of vegetables prior to cooking', *palanan* 'carrying basket', *paniaʔan* 'cooked vegetables', *pitʔia* 'to cook rice', *ma-qasbit* 'salty', *mun-sulan* 'to fetch water' and *tamuhun* 'round field hat (worn by women working in the fields)'.

Before the major influx of Chinese which transformed the indigenous language situation, the Thao were socially and linguistically interdependent with the much larger community of Bunun speakers, most of whom reside in villages located at higher altitudes in the mountains. Thao boys and Bunun girls traditionally were pledged in marriage by

their families at an early age, and at the time of marriage the families participated in a mutual exchange of fixed categories of goods. After the marriage Bunun wives resided in Thao villages, and in this setting Bunun was introduced to Thao-speaking children as a second language through their mothers. One of the consequences of this social arrangement is that many (not all) Bunun loanwords in Thao concentrate in areas relating to traditional 'women's work'—the kinds of activities in which Bunun-speaking women would have been heavily involved, and which they probably would have used their native language to describe.

3.6.4 Code-switching

Among the few studies of code switching between AN languages that has been done to date are those of Nivens (1998), and Syahdan (2000). Nivens (1998) argues that speakers of West Tarangan, in the Aru Islands of the southern Moluccas, switch between their language, Indonesian, and Dobo Malay, a local variety of Moluccan Malay, in both spoken and written discourse to achieve a variety of linguistic effects that cannot be achieved through the use of a single code in isolation. It thus suggests that code switching offers a way to enlarge one's linguistic repertoire. The second study describes how educated speakers of Sasak switch between 'high' Sasak or ordinary Sasak and Indonesian, depending upon the topic of the conversation, or the perceived relationship of status or familiarity between interlocutors.

Many Filipinos both in the Philippines and in the United States, reportedly code-switch rather freely between Tagalog and English, producing a style of speech that is humorously called 'Taglish'. An example, provided by Jason Lobel (with the Tagalog in italics and standard orthography), is < I think it's much better *na pupuntahan mo na lang sya* and you tell him *na* if he really wants the job, *e, kelangan nya talagang humarap sa akin* > 'I think it's much better if you go to him and tell him that if he really wants the job, then he needs to come talk to me face to face.' In some interpretations code-switching includes the integration of foreign words into native grammatical patterns, although this appears to stretch the meaning of the term beyond its normal use. The Wikipedia (http://en.wikipedia.org/wiki/Taglish), for example, cites examples such as *mag-da-drive* 'will drive', *mag-sya-shopping* 'will go shopping', or *na-print* 'printed', in which English bases *drive*, *shop* and *print* undergo tense-marking CV- reduplication and/or prefixation. While loanwords have long been incorporated into Philippine languages in this way (cp. Spanish *hora*, Bikol *óras* 'hour, time', *mag-óras* 'to time something', or Spanish *zapato*, Bikol *sapátos* 'shoe', *mag-sapátos* 'to wear shoes', where *horas* and *sapatos* are now part of the native lexicon) *drive*, *shopping* and *print* are clearly regarded as English, yet speakers can instantly form morphologically appropriate words that use them as bases.

3.6.5 Pidginisation and creolisation

As a result of plantation labor policies, several varieties of pidginised English arose in the AN world during the nineteenth century. All of these were in the Pacific region. The question whether there are any pidginised forms of AN languages is more problematic, and will be addressed below.

Probably the single best-known pidgin language with AN content is Tok Pisin, the national language of Papua New Guinea. The lexicon of Tok Pisin is mostly of English origin, while its grammar is typically Oceanic. Mosel (1980) has argued that the

grammatical structure of Tok Pisin and a minor part of its vocabulary derive from Tolai, a major *lingua franca* native to the Gazelle Peninsula of New Britain. Table 3.14 illustrates the vocabulary of Tok Pisin, showing how in many cases borrowed English morphemes have been both phonologically and semantically adapted to an Oceanic substrate:

Table 3.14 English-based vocabulary of Tok Pisin

Tok Pisin	English
man [man]	man
meri	woman
pikinini	child
het	head
gras biloŋ het	head hair
gras biloŋ skin	body hair
lek	leg
skru biloŋ lek	knee
wanpela	one
tupela	two
mitupela	we (dual excl.)
yumitupela	we (dual incl.)
b(a)rata	same sex sibling
susa	opposite sex sibling
tumbuna	grandfather, ancestor

The first example shows a more-or-less unmodified transfer from English to Tok Pisin (with vowel change [æ] to [a]). The second example illustrates semantic deviation in the form of English that is ancestral to Tok Pisin, since it is unlikely that AN speakers themselves would have chosen the proper name 'Mary' to mean 'woman/wife'. The third example is from Portuguese, filtered through plantation English. The fourth example again involves an only slightly modified transfer from English to Tok Pisin, but the next two examples show both the common genitive construction formed with *biloŋ* < English 'belong' and the almost universal AN lexical distinction between head hair and body hair. The seventh example shows replacement of [ɛ] by [e] and devoicing of final stops), and the eighth example again illustrates the use of the genitive construction with *biloŋ*, as well as use of the English word 'screw' in the sense of 'joint'. All Tok Pisin numerals are followed by *-pela* (< 'fellow'), and the pronouns reflect an inclusive/exclusive distinction and a dual number that are not morphologically marked in English. Finally, *brata* and *susa* are derived from English 'brother' and 'sister', but their meanings reflect the social categories of Oceanic kinship systems in which relative sex is a key parameter. Rather than meaning 'male sibling', then, Tok Pisin *brata* means 'same sex sibling' (brother of a man, sister of a woman), and rather than meaning 'female sibling', *susa* means 'opposite sex sibling' (sister of a man, brother of a woman). The kin term *tumbuna* 'grandfather, ancestor' is added to show that not all of the lexicon of Tok Pisin is of English origin.

What this small data set shows in microcosm is the marriage of two linguistic systems that remain largely separated in a single language: English has supplied most of the lexical content, but indigenous Oceanic languages have supplied the phonology, the semantic structure of the lexicon, and much of the syntactic structure on the level of the phrase and above. The general literature on pidgin and creole languages stresses the role of simplification in the formation of such languages: during the initial contact situation a

socially dominated community adopts 'the language' of the dominant group in a highly simplified form, while still retaining its native tongue to ensure that all communicative needs are met. Later generations that learn the pidginised language as a first language then expand the vocabulary, the inventory of grammatical devices, etc. so as to achieve a 'fully developed' creole. By the late 1970s some researchers in the Pacific began to feel reservations about this model. Examples of simplification can be seen in the phonology and phonotactics of Tok Pisin, where the much larger vowel inventory of English has been reduced to the five vowels (*a, e, o, i, u*) typical of Oceanic languages, typologically unusual segments such as [θ] have been replaced with more common substitutes (mouth > *maus*), and many consonant clusters are reduced or broken up, especially in final position. However, much of the complexity of the substrate languages is retained, such as the inclusive/exclusive distinction and dual number in the pronouns.

Mühlhäusler (1979) was perhaps the earliest writer to explicitly question whether the model of unmitigated simplification that was then current in views of pidginisation and creolisation is applicable to Tok Pisin, noting among other things that marked grammatical patterns, such as the use of reduplicated bases with intransitive verbs and of unreduplicated bases with their transitive counterparts, is found both in Oceanic languages and in Tok Pisin, where it was apparently calqued from Tolai (Table 3.15):

Table 3.15 The reduplication : transitivity correlation in Tok Pisin and Tolai

	transitive	intransitive
Tok Pisin	wasim 'wash something'	waswas 'bathe (oneself)'
	tingim 'remember, think of'	tingting 'think, ponder'
	lukim 'see'	lukluk 'look'
	tokim 'say something, speak to'	toktok 'to talk'
Tolai	iu 'wash'	iuiu 'bathe'
	tumu 'write down'	tútumu 'to write'
	kal 'dig up'	kakál 'to dig'
	tun 'burn, cook'	tutún 'to cook'

This point was further explored in detail by Keesing (1988) who, among other things, drew attention to the retained complexities in the pronoun system of all forms of Melanesian Pidgin English in the Pacific (inclusive/exclusive distinction in the first person non-singular, dual number), and their stability over time.

Crowley (1990) notes that Melanesian Pidgin English has three dialects: Tok Pisin in Papua New Guinea, Solomons Pijin in the Solomon Islands, and Bislama in Vanuatu. Bislama developed in situ, while the other two developed outside their home areas and were later brought in by returning plantation laborers, many of whom had learned Samoan Plantation Pidgin. All writers stress that Melanesian Pidgin English is an evolving language, with distinctions between e.g. rural, urban and 'bush' dialects of Tok Pisin, or generational differences between speakers of the same dialect. Crowley (1990) notes in addition that the lexicon of Bislama contains both English and French elements, sometimes in competition for the same meaning, as with *ariko* (French: *haricot*), or *bin* 'bean', *avoka* (French: *avocat*) 'avocado', or *pastek* (French *pastèque*) 'watermelon'. In some cases the French source of a Bislama lexical item is said to derive from Réunion creole, as with Bislama *pistas* (Provincial French *pistache*, possibly via Réunion creole) 'peanut', showing

how the vocabulary of creole languages from widely separated regions was sometimes brought together through a common colonial past.

The question of whether there are any AN lexifier languages for creoles is much more vexed. Malay has been an important *lingua franca* in island Southeast Asia for many centuries, and it has been suggested by some writers that 'Bazaar Malay' is a creole that became the basis for the national languages of both Indonesia and Malaysia. Collins (1980) has taken issue with this interpretation, arguing (1980:6) that the basis of these national languages is "not so-called bazaar Malay, but 'classical', literary Malay' into which colloquial elements have been deliberately inserted over a period of time in order 'to strengthen its psychological base."   Following earlier work by several social anthropologists, Collins characterizes Ambon as having a 'creole culture', with elements of Portuguese, Dutch, Malay, Javanese and indigenous Moluccan cultural practices, as well as elements of language. Among phonological features that distinguish Ambon Malay from Standard Malay are 1. the merger of schwa with *a*, 2. the loss of *h*, 3. the merger of all final nasals as *ŋ*, 4. the loss of nearly all final stops, 5. the irregular loss of unstressed syllables in a word that is followed by another word, as with SM *pərgi* : AM *pigi*, *pi* 'go', SM *sudah* : AM *suda*, *su* 'already', SM *jaŋan* : AM *jaŋaŋ*, *jaŋ* 'don't', or SM *punya* : AM *punya*, *puŋ* 'possess'. Most of the vocabulary of the two languages is cognate (including about 81% of the basic vocabulary on a Swadesh 200-word list), but their affix systems are strikingly different. According to Collins (1980:25) most AM affixes "seem to appear only in fixed (fossilised) forms and these in uses which sometimes differ from SM uses." The reasons that a simplified form of Malay has developed far from the home region of Malay speakers are inseparable from the history of the spice trade. Until the nineteenth century cloves and nutmegs were grown only in the central Moluccas, and from here they were traded as far as China on the one hand, and the Middle East and Europe on the other, for at least two millennia (Collins notes that Pliny discussed spices and the trade routes used in shipping them to Rome in 75 A.D.). Since this trade had to be funneled through the Strait of Malacca in order to reach India, the Middle East and Europe, the principal middlemen were Malays, and Malay thus came to function as a *lingua franca* in the spice trade at a very early time. From a *lingua franca* it became pidginised in certain communities connected with international trade, and from a pidgin it was later expanded. Despite this history Collins concludes that Ambonese Malay is not a creole. In part this conclusion rests on the observation that some non-standard dialects of peninsular Malay, such as Trengganu, show simplifications similar to those found in Ambonese Malay, yet the known social history of the Trengganu dialect provides no basis for the claim that it has been creolised. In Collins' words (1980:58-59) "The term creole has no predictive strength … Neither AM (Ambon Malay) nor TM (Trengganu Malay) are creoles. Rather they are linguistic reflections of processes far too complex for theories and labels which have been developed within a narrow framework."

Ambonese Malay as a contact language differs strikingly from Melanesian Pidgin English. In the latter the evidence of compromise between two linguistic systems and the establishment of a division of labor between them is very clear: most of the lexicon derives from English, while most of the phonemic system, phonotactics and grammar derives from an Oceanic substrate. In Ambonese Malay both the vocabulary and the grammar appear to be basically Malay (with clear loanwords from other languages), but a form of Malay that has been adapted to the phoneme systems and phonotactics of local indigenous languages, and morphologically simplified. These differences undoubtedly reflect differences in the nature of the contact situation in the two cases. Melanesian Pidgin English arose under

conditions of social inequality, with European overlords talking down to Melanesian laborers. Although Collins (1980:63) speaks of Malays 'talking down' to non-native speakers, who they assume cannot understand more complex forms of the language, Ambonese Malay apparently arose in a situation of relative social equality in which one AN-speaking group controlled the wider trade network, but did not have a dominant role in the social relations of the groups in contact. Unlike speakers of Melanesian Pidgin English, who were recruited from diverse groups and thoroughly mixed in the plantation work environment, speakers of Ambonese Malay presumably always had access to a native language from the time that Malay was introduced in Ambon as a *lingua franca*. The conditions for true pidginisation and hence creolisation thus may never have existed in the formation of this dialect of Malay.

A second example of a Malay-based creole is Sri Lanka Malay, which is reportedly spoken in at least five different communities that descend in part from laborers imported during Dutch and British colonial times. Malay has supplied the lexicon, while syntactic structure has been adapted to both Sinhalese and Tamil.

A third apparent example of a Malay-based creole which apparently is spurious is the dialect spoken by 'Peranakan' Chinese (overseas Chinese who were born and raised in the Malay-speaking world). This is commonly known in Malacca as 'Baba Malay', the Malay spoken by the Straits-born Chinese. E. Thurgood (1998) provides a thorough study of the nineteenth century form of Baba Malay (called 'Old Baba Malay'), and compares it with the rather different form of the language spoken today. Her conclusion is that Baba Malay arose through language shift among speakers of Hokkien (Minnan Chinese). She finds no evidence that Baba Malay was ever a creole. As with Ambon Malay, the Peranakan Chinese who use this dialect always had access to Hokkien, until they deliberately abandoned it through language shift. Since there never was a 'catastrophic' stage during which the transmission of language was interrupted, the precondition that is widely assumed as necessary to the formation of a pidgin, and hence a creole, was not met.

3.7   Determinants of language size

One of the most striking features that appears in the sociolinguistic profile of any language family is the sometimes enormous differences in size of speech communities. In the Indo-European family this can be seen in comparing the number of native (or second-language) speakers of Hindi, English, Russian or Spanish with that of languages such as Albanian or Greek. As noted already in Chapter 2, the size of AN languages varies dramatically, and shows a strong correlation with a west-to-east cline from insular Southeast Asia to the Pacific. Within the Pacific there is, moreover, a reversal of this tendency, with smaller languages generally being found in the western Pacific, and larger languages in Fiji and Triangle Polynesia. This observation, expressed in somewhat different terms, led to an important exchange between Pawley (1981) and Lynch (1981).

Pawley (1981) noted that in Melanesia one island or island group typically has many languages, while in Polynesia single islands or archipelagos almost always have one language. Since his discussion was concerned with AN languages this suggested some inherent difference between languages in Melanesia and Polynesia. Partly in response to earlier claims made to account for this observation he argued that the sole determinant of differences in such patterns was length of settlement: Melanesia was reached by AN speakers earlier, and so the languages found there had been differentiating in situ for a longer period of time than those in Polynesia. Lynch (1981) countered that, whereas this

observation may be true with regard to lexical differentiation, it does not account for the wide range of structural variation in the AN languages of Melanesia, some of which (such as SOV word order and the use of postpositions) is clearly due to contact.

This debate was useful at the time, but in retrospect it seems to frame the question too narrowly, and to overlook important determinants other than settlement time and possible contact influences. Rather than frame the debate in terms of the question 'What determines the number of languages per island/archipelago?' it seems more useful to ask 'What determines language size?' At least three non-social factors appear to be implicated in the answer to this question. First, the size of the available territory settled by a migrating group inevitably imposes an upper limit on language size, as in the atoll environment of Micronesia, where a language community may be located on less than three square kilometers of land. This in turn is related to isolation, since the same language could be spoken on neighboring atolls, but the likelihood of this decreases rapidly with distance. Second, the carrying capacity of the land, independent of its area, imposes an upper limit on population size, and hence on language size. Where food resources are poor language communities cannot be large even where territorial size would otherwise permit it. Third, length of settlement, as Pawley pointed out, clearly is an important factor in determining loss of homogeneity, which in turn reduces the size that language communities would otherwise have if language split had not occurred.

With reference to Melanesia, there are two rather clear social forces that bear on the size of language communities. First, as Lynch indicated, AN languages that were in intimate contact with Papuan speakers were far more likely to show increased rates of lexical replacement and structural change vis-à-vis those that had little or no contact of this type. But there is another well-known factor that distinguishes most of Melanesia from Polynesia. At the risk of oversimplification, it seems fair to say that many of the smallest AN languages in Melanesia are found where hereditary chieftainships do not exist. In 'big man' societies of this kind political power is achieved rather than ascribed, and generally extends no further than the village level. With a low level of sociopolitical integration, societies of this type will tend to show much greater linguistic fragmentation than those with paramount chiefs capable of ruling over large territories, commanding corvée labor, expanding the native domain by conquest, etc. To cite one of many possible examples, the island of Manus in western Melanesia is roughly the same area as the island of O'ahu in the Hawaiian chain, but the level of indigenous language diversity in the two cases is radically different. On O'ahu, as throughout the eight major islands of the Hawaiian chain, a single language was spoken, with minor dialectal variation, while in Manus some 25-30 languages were spoken both on the main island and on the many tiny satellite islands that ring it. To some extent this type of difference can be attributed to length of settlement: Manus has been settled by speakers of AN languages for perhaps 3,500 years, but the Hawaiian chain probably has not been settled for much more than a millennium. However, settlement time alone cannot be the entire story. Samoa, Tonga and Fiji have been settled for at least 3,000 years, yet the level of linguistic diversity in these archipelagos is more like that of Hawai'i than that of Manus or many other parts of Melanesia. To cite an example from insular Southeast Asia, the island of Java is far larger than Manus, and probably has been settled by AN speakers for an equally long period, yet has only three indigenous languages: Javanese, Sundanese and Jakarta Malay. But throughout much of its known history Java has been ruled by one centralised state or another, and the ability of leaders in such areas to integrate a large and geographically dispersed population under a single ruler has surely exerted a stronger retarding influence on language split than one can

expect to find where political power rarely exceeds the level of the natal village, as in much of Melanesia. Finally, in at least some parts of Melanesia where hereditary chieftainship exists, as in the Loyalty Islands, a one-island one-language pattern is also found. It can be concluded, then, that the level of sociopolitical integration in combination with the other factors considered here, has played a major role in the pattern of language fragmentation seen within the AN family, and hence in the size of language communities.

4   *Sound systems*

4.0   Introduction

In a language family as large and geographically dispersed as AN it can be expected that typology will show great diversity on all levels. PAN probably began to form dialect regions by 5,500-6,000 BP on the island of Taiwan. Even among the surviving Formosan languages there is wide variation in typology, despite the geographical proximity of language communities, and the absence until recent centuries of contact with languages belonging to other families. Outside Taiwan the size of the AN-speaking world increases enormously, and contact with languages belonging to other families, including at least Austroasiatic, Tai-Kadai, Sino-Tibetan, Niger-Congo, and 'Papuan,' introduces another factor to the typological spectrum.

Given this situation an adequate treatment of phonological typology in AN languages would require far more space than is available here. This chapter addresses the following topics: 1) phoneme inventories, 2) morpheme structure (phonotactics), 3) phonological processes, 4) conspiracies, 5) accidental complementation, 6) double complementation, and 7) free variation. The presentation is necessarily selective, but every effort has been made to minimise sampling bias through broad geographical, subgrouping, and typological representation. In the hope of approaching a balance between the general and the specific I have alternated between an areal and a topical organisation, first surveying phoneme inventories on a regional basis, then examining morpheme structure and phonological processes on a family-wide basis. In discussing sound systems the line between synchrony and diachrony can sometimes be hard to draw, and some historical information is consequently brought into the discussion, even though most matters relating to historical phonology are treated in later chapters.

4.1   Phoneme inventories

Phoneme inventories in AN languages are generally smaller than the world average. Based on a sample of 320 globally-distributed languages in the UCLA Phonological Segment Inventory Database, Maddieson (1984:7) concluded that "the typical size of an inventory lies between 20 and 37 segments," with 70% of the languages in his survey falling within this range. By contrast, probably 90% of all AN languages have 15-20 consonants and 4-5 vowels, hence total phoneme inventories that lie between 19 and 25 segments. Maddieson (1984:6ff) noted that inventory size depends on assumptions about whether complex articulations such as affricates or prenasalised stops are treated as units or sequences, and upon how suprasegmental features such as stress, or vowel nasality are counted (as single elements, or as components of individual segments). The course that he chooses for making these decisions is a complex mixture of different considerations, but in any case he does not count suprasegmental features such as stress or length as increasing the size of the segment inventory. Hawaiian is thus said to have 13 phonemes despite the

fact that in addition to its eight consonants and five short vowels it also has five long vowels.

To determine inventory size I count only consonants and vowels. Word-final diphthongs reflecting PAN *-ay, *-aw, *-uy and *-iw are also found in many attested AN languages, and traditionally these have been included in phoneme inventories, in part because they are often monophthongised in historical change. These -VC sequences will be treated as a separate segmental category in the discussion of sound change, but will not be counted in tabulating phoneme inventories. Stress, length and vowel nasality are also mentioned where relevant, but do not figure in counting the number of segments. Loan phonemes that are marginal to the system are marked as provisional by the use of parentheses. Keeping these qualifications in mind, the largest segment inventory reported for an AN language is that of Nemi (northeast New Caledonia), with 43 consonants and five vowels (which may occur either oral or nasalised) for an inventory size of 48 segments. Other segment inventories of nearly this size are found elsewhere in New Caledonia and the Loyalty Islands, and in some of the Chamic languages of mainland Southeast Asia. The smallest segment inventory is controversial, but appears to be that of Northwest Mekeo, spoken in southeast New Guinea, with 7 consonants and five vowels *i u e o a*, without contrastive vowel length (Jones 1998). This is followed closely by five Eastern Polynesian languages, each of which has eight consonants and five vowels that may occur short or long (*i u e o a, ī ū ē ō ā*). Table 4.1 shows the consonant inventories of these six languages, which differ in several details.

Table 4.1 Smallest consonant inventories found in Austronesian languages

Northwest Mekeo	p	b	k	g	m	n	ŋ	
South Island Maori	p	t	k	m	n	h	w	r
Rurutu (Austral Islands)	p	t	ʔ	m	n	ŋ	v	r
South Marquesan	p	t	ʔ	m	n	f	h	v
North Marquesan	p	t	k	ʔ	m	n	h	v
Hawaiian	p	k	ʔ	m	n	h	w	l

The constant features of these drastically reduced segment inventories are *p* and *m*, each of which has a single source in all six languages, and *n* which derives from PEP *n in South Island Maori, Rurutu, and North Marquesan but from the merger of PEP *n and *ŋ in South Marquesan and Hawaiian (PPN *ŋ > *k* in South Island Maori and North Marquesan). In addition, all of the Polynesian languages have an unmerged reflex of PEP *w which varies phonetically between a glide and a voiced fricative. Apart from these few points of agreement everything else in the typology of these inventories differs: Hawaiian has shifted *t to *k*, only Rurutu has a velar nasal, all languages except South Marquesan have shifted *f to *h* or glottal stop, and North and South Marquesan have shifted PEP *r to glottal stop (merging with the reflex *k in South Marquesan, but not North Marquesan). Maddieson (1984:7ff) gives only two phoneme inventories that are smaller than these: Rotokas (East Papuan, Bougainville Island, western Solomons), with six consonants and five vowels, and Mura, or Pirahã (isolate, northwest Brazil), with eight consonants and three vowels. He describes Hawaiian as having 13 segments, but as noted already, vowel length is phonemic, and inventory size will depend in part on how length is treated from a

segmental standpoint. Whatever the outcome of this calculation, it is noteworthy that the phoneme inventories of Eastern Polynesian languages show very little allophony. In Standard Hawaiian, for example, the only consonant which has more than one allophone is *w*, which varies between [w] and [v] (in the Ni'ihau dialect *l* also varies between [l] and [ɾ]). By contrast, consonants in Rotokas have many allophones, and hence an inventory of phones that considerably exceeds that of any of the languages in Table 4.1.

What is perhaps most intriguing about these very small phoneme inventories in Polynesian languages is that they represent the end-point of a continuous reduction in size over time as the prehistoric human settlement of the Pacific pressed ever further into the unknown (an observation which has led some specialists to quip that, had there been other islands east of Hawaii their Polynesian settlers would have ended up speechless!). Within the Central Pacific group of languages, consisting of Fijian, Rotuman, and Polynesian, there is a general west-to-east cline of decreasing inventory size, shown in Table 4.2:

Table 4.2 West-to-east cline in size of Central Pacific phoneme inventories

Language	consonants	vowel	total
Fijian	17	5 (+ length)	22
Rotuman	14	10	24
Tongan	11	5 (+ length)	16
Samoan	10	5 (+ length)	15
Rarotongan	9	5 (+ length)	14
Tahitian	9	5 (+ length)	14
Hawaiian	8	5 (+ length)	13
Rurutu	8	5 (+ length)	13

Why this cline should exist is unclear, since there is no known evidence from other language families that increased migration distance correlates with reduced phoneme inventories. The patterning, however, seems well enough established to merit notice.

4.1.1 Taiwan

The phoneme inventories of Formosan languages show considerable variation. The largest segment inventory is that of Tanan Rukai, with 23 consonants (two of them rare), and four vowels that occur either short or long. The smallest inventories are found in Kanakanabu and Saaroa, which have only 13 consonants and four vowels each (plus stress in the former), although Tsuchida (1976:59) notes that Saaroa has four additional consonants and two additional vowels that occur in loanwords. All other Formosan languages for which adequate descriptions are available fall within the 19-25 segment range stated for AN languages as a whole. The phoneme inventories of Tanan Rukai and Saaroa are given in Table 4.3:

Table 4.3 Largest and smallest phoneme inventories in Formosan languages

Tanan Rukai (Li 1977b)	**Saaroa** (Tsuchida 1976)
(p) t (ṭ) k ʔ	p t c k ʔ
b d ḍ g	m n ŋ
m n ŋ	s
c	ɬ
θ, s h	v
v ð	l
l ḷ	r
r	
w y	
vowels: i, o, ə, a, plus length	vowels: i, u, ə, a,
23 + 4 = 27	13 + 4 = 17

In AN languages generally voiceless stops are unaspirated, and the *t:d* contrast, which is signaled primarily by different values for the feature [voice], is redundantly signaled in many languages by a difference of place: *t* is postdental, while *d* (like *n* and *l*) is alveolar. Two distinctive traits found in many Formosan languages are: 1) the presence of a uvular stop *q*, which contrasts with *k* in most languages, and with both *k* and glottal stop in some, and 2) a large number of fricatives (compared with *s* and *h*, or just *s* in most languages of the Philippines). Thao, still kept alive around Sun-Moon Lake in central Taiwan by fourteen or fifteen elderly speakers (Blust 2003a), shows both features (in the practical orthography ʔ, θ, ʃ, ð and ɬ are written as ', c, sh, z and lh respectively):

Table 4.4 Consonant inventory of Thao

		Labial	**Dental**	**Palatal**	**Velar**	**Uvular**	**Glottal**
Stops:	vl.	p	t		k	q	ʔ
	vd.	b	d				
Nasals:		m	n		(ŋ)		
Fricatives:	vl.	f	θ, s	ʃ			h
	vd.		ð				
Laterals:	vl.		ɬ				
	vd.		l				
Flap:			r				
Glides:		w		y			

Together with the vowels /i u a/ this yields an inventory of 23 segments (/ŋ/ is found only in personal names and some Bunun loanwords). The voiced stops are preglottalised, an areal feature shared with Bunun (from which it probably spread), and with Tsou. The seven fricatives, constituting 35% of total consonants, probably is a record for any AN language. The fricatives /θ/, /ð/ , which are rare outside Taiwan, also show up in Rukai, and /ð/ alone is found in Bunun and Puyuma, /ʃ/, which is rare outside Taiwan, also shows up in Saisiyat, and /ɬ/, which is almost unknown outside Taiwan, also shows up in Saaroa. Other articulations found in Taiwan that are rare elsewhere in AN include the voiceless velar fricative (Pazeh), a pharyngeal fricative and an epiglotto-pharyngeal stop, found only word-finally in Amis (Edmondson, Esling, Harris and Huang 2005), retroflex consonants

(/ɖ/ in Rukai and Paiwan, /ʈ/ and /ɖ/ in Puyuma), a retroflex lateral (Rukai), a palatalised lateral (Paiwan), and a back-to-front flapped rhotic (Puyuma). It is striking, then, that the consonant inventories of Formosan languages contain a number of segments that are quite unusual elsewhere in Austronesian.

A number of Formosan vowel systems (Amis, Puyuma, Paiwan, most Rukai dialects, Saaroa, Kanakanabu, Pazeh) retain the PAN four-vowel system *i, *u, *a, *e (schwa). Some Atayal dialects have reduced this system to the vowel triangle, as have Bunun and Thao. Other dialects of Atayal and Seediq have a five-vowel system /i u e o a/. Tsou has developed a six-vowel system /i u e o ə a/, and Maga Rukai and Saisiyat are both reported as having seven contrastive vowels, /i u ɨ e o ə a/ for Maga, and an asymmetrical system with /i, (e), œ, ə, o, æ, a/ for Saisiyat. Stress is phonemic in Kanakanabu, and in all dialects of Rukai except Mantauran, but stress contrasts in cognate morphemes differ sufficiently between these languages that Li (1977b) omitted contrastive stress in his reconstruction of Proto Rukai phonology. In Thao stress has a very small functional load (about 98% of all base morphemes are stressed on the penult, and the remaining 2% on the final syllable). In this language it is clearly of secondary origin.

4.1.2 The Philippines

Philippine languages show far less phonemic variation than Formosan languages. Reid (1971) gives segment inventories for 43 minor languages of the Philippines (including Sangir of northern Sulawesi), which range from 21 consonants and four vowels (Itbayaten) to 14 consonants and four vowels (Kayapa Kallahan). None of the inventories of Philippine major languages falls outside this range. Formosan inventories thus have17-27 segments, while Philippine inventories have 18-25 segments (exclusive of stress or vowel length). Table 4.5 summarises the distribution of segment inventories across the 43 languages in Reid (1971) in descending order by size. L = vowel length, S = stress, and + indicates that the inventory would be made larger by the inclusion of contrastive suprasegmental features. Mansaka vowels are marked for shortness, and since length and stress almost always co-occur in those Philippine languages for which adequate descriptions are available, these symbols probably mark different perspectives on the same suprasegmental category:

Table 4.5 size of phoneme inventories in 43 Philippine minor languages

Consonants	Vowels	Other	Total	No. of languages
21	4		25	1
20	4	S	24+	1
18	6	S	24+	1
16	8	S	24+	1
19	4		23	2
17	6		23	1
17	5	L	22+	1
16	6	S	22+	1
15	7	S	22+	1
16	6		22	1
15	7		22	3
17	4	S	21+	1
16	5	L	21+	1
15	6	L	21+	1
18	3		21	1
17	4		21	2
15	6		21	2
16	4	S	20+	2
15	5	L	20+	1
14	6	L	20+	1
16	4		20	1
15	5		20	2
15	4	S	19+	2
15	4	L	19+	2
14	5	S	19+	1
14	5	L	19+	1
15	4		19	2
14	4	S	18+	5
14	4		18	1

Although the range of inventory size is only slightly smaller for Philippine languages than Formosan languages, Philippine languages show a much narrower range of variation in consonant type than the languages of Taiwan. Almost all Philippine languages have voiceless stops *p, t, k, ʔ*, voiced stops *b, d, g*, nasals *m, n, ŋ,* fricatives *s, s* and *h*, or rarely just *h,* and the glides *w* and *y.* In addition, several languages in northern Luzon, as well as Rinconada Bikol in southern Luzon, and Mansaka, Tausug, and the Samalan languages of the southern Philippines, have phonemically geminated consonants which, depending upon one's analysis, may count as segments or as clusters. Segments that are less consistent cross-linguistically are the liquids and palatals. With regard to the former, 23 of the 43 languages in Reid 1971 have just a lateral liquid (occasionally with a rhotic or other allophone in some environments), 18 have both a lateral and a rhotic, and two have no liquids. No language with a rhotic lacks a lateral. Some languages, such as Tagalog, have an *l : r* contrast that reflects restructured allophony (*r* was an intervocalic allophone of *d* until the adoption of Spanish loanwords made it contrastive). Unlike many of the languages of western Indonesia, very few Philippine languages have a series of palatal consonants. The Bashiic languages of the far north (Itbayaten, Ivatan) have developed

palatals *c, j,* and *ñ*. These began as allophones of *k, g, n* and *ŋ* adjacent to a high-front vowel, but a combination of sound change and borrowing have made them phonemic. Kapampangan has three palatal phonemes, *c* (written *tc*), *j* (written *dy*) and *ñ*. The palatal nasal is inherited, but the affricates are innovations, or are loan phonemes that may have been borrowed to satisfy a universal implicational hierarchy that the number of place features for nasals never exceeds that for stops. Other languages in the Philippines that have some palatal consonants include Inati of Panay Island, and the Sama-Bajaw languages and Tausug of the southern Philippines. Additional features of some interest include fairly rich systems of allophony for the voiced stops in some languages of northern Luzon (as Bontok), and automatically preglottalised *b* and *d* (but not *g*) in Sindangan Subanon of western Mindanao.

Like the Formosan languages, about half of all Philippine languages retain the PAN four-vowel system. Where this system has changed, the vowel inventory generally has been expanded, up to a maximum of eight vowels in Casiguran Dumagat (i e ɛ ɨ a u o ɔ). In some cases expansion has resulted from the reinterpretation of earlier allophones as new phonemes. An example is Tagalog *o*, which once occurred only as an allophone of *u* in final syllables, whether open or closed (even in reduplicated monosyllables such as *kuykóy* 'act of digging with bare hands or paws', or *tuktók* 'act of knocking, as on a door'). The introduction of Spanish loanwords broke this pattern of complementation, and [o] must now be considered a separate phoneme.

The most distinctive typological feature in the sound systems of Philippine languages is the widespread occurrence of phonemic stress. Because of their aberrant typological status within AN, and the problems they pose for historical reconstruction, Philippine stress systems merit special attention. Although contrastive stress is absent in the Bashiic languages and almost everywhere south of the Bisayas, it is found in most Cordilleran languages, in Kapampangan and Botolan Sambal of central Luzon, in virtually all Central Philippine languages, in South Mangyan languages such as Hanunóo, and in Palawan Batak. Languages such as Ilokano or Tagalog show numerous examples of stress contrasts, including many minimal pairs, as with Ilokano *ádas* 'to glean (in a harvest)' : *adás* 'faint through loss of blood', *símut* 'winged ant' : *simút* 'eat by dipping in sauce', or Tagalog *búlak* 'kapok tree' : *bulák* 'effervescence (of liquid beginning to boil)', *síloŋ* 'ground floor; downstairs' : *silóŋ* 'inferiority complex'. In the following discussion languages like Ilokano or Tagalog will be called 'stress languages'.

The relationship between stress and length in Philippine languages is somewhat more complicated than in many other languages. Figure 4.1 summarises this relationship in lexical bases for Tagalog, which presumably is representative of other languages of the region (1 = ultima, 2 = penult, 3 = antepenult):

	3	2	1
Stressed	--	--	--
Long	--	--	--
Stressed	--	+	+
Long	--	+	--

Figure 4.1 Relationship between stress and length in Tagalog

The nucleus of an unstressed syllable is never long. The nucleus of a stressed syllable, however, is long in the penult, but not in the ultima. Daniel Kaufman (p.c.) suggests that in Tagalog oxytone stress is the default pattern, and paroxytone stress is triggered by vowel length. This analysis requires the recognition of *both* phonemic stress (in the ultima) and phonemic length (in the penult). Traditionally, however, only stress has been recognised and lexically marked, and although there appears to be marginal evidence that stress and length are independent in affixed words, these prosodic features are not contrastive in lexical bases.

Most non-stress languages of the Philippines have penultimate stress, with qualifications to be noted below. However, even in stress languages such as Ilokano, Tagalog, or Cebuano, stress is rule-governed in some contexts. Rubino (2000:xxixff) lists a number of environments in which Ilokano stress can often be predicted, and some of these agree closely with Central Philippine languages. Perhaps the most reliable predictor of stress is that vowels preceding a consonant cluster will be unstressed (loanwords show some exceptions). Reduplicated monosyllables such as Ilokano *bakbák* 'fade, lose colour', *buŋbóŋ* 'explosion', *giwgíw* 'cobwebs', or Tagalog *liklík* 'act of digressing', *palpál* 'obstructed or filled (with dirt or weeds)', *tuktók* 'knocking (as on a door)' are thus invariably stressed on the last vowel, as are words that contain a medial geminate in Ilokano. In Tagalog the same is true of non-reduplicated bases which contain a medial consonant cluster, whether the segments are heterorganic (*agtá* 'Negrito', *gitlá* 'scare, fright, shock', *sukláy* 'comb'), or homorganic (*aŋkán* 'family lineage', *kumpís* 'deflated', *pandóŋ* 'head covering'). However, not all Philippine stress languages agree in these conditions. Bisayan languages such as Aklanon and Cebuano, for example, generally stress the vowel that *precedes* a consonant cluster, both in reduplicated monosyllables, and elsewhere (*búkbuk* 'bamboo weevil', *dáŋdaŋ* 'warm up, put near a fire', *hámpak* 'beat, slap', *sílʔot* 'push in, squeeze in', *tágsik* 'splash'). Most predictable patterns are oxytone (word-final), but sometimes a paroxytone (penultimate) pattern is also predictable in stress languages under specific phonological conditions.

Table 4.6 summarises the conditions under which stress is predictable in ten Philippine stress languages, as follows:

1 =   reduplicated CVC monosyllable (Ilokano *bakbák*),

2 =   non-reduplicated base with a medial heterorganic consonant cluster (Tagalog *sukláy*),

3 =   base with a homorganically prenasalised medial stop (Tagalog *kumpís*),

4 =   base with a medial geminate consonant (Ilokano *baggák* 'morning star'),

5 =   base with a consonant plus high vowel that semivocalises to yield a phonetic consonant cluster before a final vowel (Ilokano *sadiá* 'renowned', *bituén* 'star')[20],

6 =   base with identical vowels separated by glottal stop (Tagalog *baʔák* 'split, halved'),

7 =   base that contains the sequence *aʔa, eʔe, iʔe* or *uʔo*, if at least two consonants occur earlier (Ilokano *arináʔar* 'moonlight', *manabsúʔok* 'splash'),

---

[20]   Historically, it was final stress that triggered semivocalisation, but synchronically it is simpler to argue that final stress is predictable in words that contain a medial cluster, including derived clusters in which the second member is a derived glide.

8 =   base with a vowel that separates two identical CVC sequences (Ilokano *arimasámas* 'red skies at moonrise', *ŋurúŋor* 'cut throat').

Abbreviations for microgroups are C = Cordilleran, CL = Central Luzon, CP = Central Philippines, SM = South Mangyan, P = Palawanic; other abbreviations are O = oxytone, P = paroxytone, U = unpredictable, N = not applicable.

Table 4.6 Conditions for predictable stress in ten Philippine stress languages[21]

	1	2	3	4	5	6	7	8
Ilokano (C)	O	O	O	O	O	U	P	P
Itawis(C)	U	U	U	U	?	N	N	N
Pangasinan (C)	U	U	U	N	?	U	?	N
Kapampangan (CL)	O	U	U	N	U	N	N	N
Tagalog (CP)	O	O	O	N	O	O	U	U
Bikol (CP)	O	O	O	N	O	U	?	P
Hanunóo (SM)	U	U	U	N	U	U	?	N
Palawan Batak (P)	U	U	U	N	?	U	?	N
Cebuano (CP)	P	P	P	N	P	U	?	?
Aklanon (CP)	P	P	P	N	P	U	?	?

Condition 1: The data sample used here makes no allowance for Galton's Problem (the requirement that units of comparison be historically independent), but about half of the languages agree in showing an oxytone stress pattern in reduplicated monosyllables. As noted above, Aklanon and some other Bisayan languages have reversed this pattern. The situation in Itawis is more complex. Original CVCCVC reduplications which contained an obstruent as the first member of the consonant cluster lost or completely assimilated this segment, and stress in such forms is always paroxytone: *bábak* 'that which peels off' < *bakbak, *ŋáŋat* 'chewing tobacco' < *ŋatŋat* 'gnaw', *kúkud* 'ankle of an animal' < *kudkud* 'hoof', *sússuk* 'concealed place' < *suksuk* 'insert'. In original CVCCVC reduplications a preconsonantal sonorant assimilated in place to the next segment and stress cannot be predicted: *dandám* 'sentiment' < *demdem* 'think, mull over', *zínziŋ* < *diŋdiŋ* 'wall'.

Condition 2: This condition generally conforms to Condition 1. The only known exception is Kapampangan, which appears to show an oxytone pattern in reduplicated monosyllables, but an unpredictable pattern in non-reduplicated bases that contain a medial consonant cluster. However, the available data is limited, as the only dictionary of Kapampangan that marks stress (Forman 1971) does not mark it on all forms.

Condition 3: Although all languages sampled evidently treat stress assignment identically under conditions 2 and 3, there is some uncertainty for Ilokano words with a homorganically prenasalised medial stop. Rubino (2000:xxx) reports that an oxytone pattern prevails if the penultimate syllable is closed ([bas.nót] 'whip', [tak.kí] 'excrement', etc.), but a paroxytone pattern is used in loanwords or in native words that contain -*ŋk*- before the last vowel, as with *láŋka* 'jackfruit', *bibíŋka* 'rice cake', or *súŋka* 'kind of native game'. However, there are many exceptions to this statement, as *aŋkít* 'asthma', *baŋkág* 'upland field', or *tiŋkáb* 'pry open', and it appears more likely that oxytone stress is found

---

[21]   Jason Lobel (p.c., August 3, 2007) notes that condition 6 is true only for Manila Tagalog (the opposite generally holds for southern Tagalog, as well as written sources for sixteenth-century Tagalog). Condition 8 is said to be generally true of central Philippine languages.

in all native words with a closed penultimate syllable, and paroxytone stress in loanwords, whether from Spanish or from other AN languages.

Condition 4: Among the languages considered here this condition applies only to Ilokano and Itawis. The relationship of stress to consonant gemination will be discussed at greater length below.

Condition 5: This condition is stated for Ilokano by Rubino (2000:xxx), who notes that it applies only to native vocabulary, but not to loanwords from Chinese or Spanish. Bikol appears to stress a vowel that immediately follows such a derived consonant cluster, whether the vowel is final or penultimate, as in *banwá* 'town, country' (*banua), *sadyáʔ* 'make something on special order' (Malay *sədia* 'ready'), *patyának* 'supernatural association between a naturally aborted fetus and an animal double' (Malay *puntianak*). Unlike Ilokano, where loanwords are exempt, in Bikol the Spanish palatal nasal was interpreted as a cluster of *n + y*, thereby requiring oxytone stress where the Spanish original is paroxytone: *banyó* (Species *báño*) 'bathroom'.

Conditions 6-8: The requirement that identical vowels separated by a glottal stop take oxytone stress appears to be unique to Tagalog, as this condition is absent even from such closely related languages as Bikol, Cebuano and Aklanon. The last two conditions are either inapplicable or unknown in most of the languages sampled.

Conditions for the predictability of stress in particular environments have been stated in terms of absolutes, but this formulation may conceal tendencies that a more fine-grained analysis would reveal. In general the matter is complicated by the often contrary tendencies in stress placement introduced by loanwords, especially, but not exclusively from Spanish. In addition, it should be noted that in trisyllabic bases the stressed vowel generally follows a closed syllable even if it is penultimate, as in Ilokano *biŋkúlog* 'clod of earth', *taŋkíran* 'sturdy old age', *taŋkúloŋ* 'kind of child's stroller'. Rather than stating stress patterns in terms of an oxytone : paroxytone contrast, then, it may be more accurate to state them in terms of position immediately following the conditioning factor.

Trisyllables are far less common than disyllables, and stress rarely falls on the first syllable of a trisyllabic base. Stress placement thus normally selects either of two positions, penultimate or final. As a result of the above conditions, which disfavor stress in a nonfinal syllable that is closed, there is a tendency for final stress to be more frequent than penultimate stress in most Philippine languages. Table 4.7 shows the distribution of penultimate versus final stress in the ten languages sampled above. Calculations are based on the first twenty and last twenty words from the *a-*, *b-*, *k-*, *l-* and *s-* sections of the relevant dictionaries, for a total database of 200 words. Unambiguous loanwords and different derivatives of the same base have been discarded, and the languages are arranged from the highest to the lowest percentages of penultimate-stressed forms.

Table 4.7 Relative frequency of penultimate vs. final stress in Philippine languages

Language	Penult	%	Final	%
Hanunóo (SM)	128	64	72	36
Cebuano (CP)	124	62	76	38
Aklanon (CP)	121	60.5	79	39.5
Itawis (C)	107	53.5	93	46.5
Ilokano (C)	97	48.5	103	51.5
Kapampangan (CL)	96	48	104	52
Palawan Batak (P)	93	46.5	107	53.5
Pangasinan (C)	81	40.5	119	59.5
Tagalog (CP)	81	40.5	119	59.5
Bikol (CP)	80	40	120	60

Although preliminary, these patterns suggest that Central Philippine languages tend to have more oxytone forms, a tendency that follows from the number of environments in which final stress is automatic. Bisayan languages such as Aklanon or Cebuano are artificial exceptions, since they have reversed an earlier pattern of oxytonality on the vowel following a closed syllable, thus inflating the number of paroxytone forms.

The most deviant language in Table 4.7 is Hanunóo, which is heavily paroxytone without the artificial inflation caused by automatic penultimate stress in closed nonfinal syllables seen in the Bisayan languages. Since nearly all non-stress languages in the Philippines are paroxytone, it is possible that Hanunóo is on its way to losing phonemic stress completely. This assumes that when stress contrasts are lost they do not disappear instantly, but rather one position is increasingly favored to the gradual exclusion of the other. While it may be too early to determine whether this is the case in Hanunóo, the tendency to lose contrastive stress becomes increasingly clearer south of the Bisayas. In almost all languages of Mindanao stress is penultimate (Subanen, Manobo, Danaw languages), or final (Bilic languages). This is also true of most of southern Bikol (Jason Lobel, p.c.), and of Tausug. Since Tausug has been in intensive contact with non-stress Sama-Bajaw languages for some seven centuries (Pallesen 1985), contrastive stress may have been lost in this case as a result of prolonged contact with non-stress languages.

One language that seems to be losing contrastive stress is Mansaka of southeastern Mindanao, which is described with contrastive vowel *shortness*, marked by an acute accent, as in *bagaʔ* 'lung' : *bágaʔ* 'swelling' (Svelmoe and Svelmoe 1990). Short vowels are rare in Mansaka, and correspond to unstressed vowels in languages such as Tagalog (cf. Tagalog *bágaʔ* 'lung' : *bagáʔ* 'abscess', where the acute accent on the penult has a value opposite to that in Mansaka). To rephrase this statement by interpreting a short penultimate vowel as oxytone stress, a count of oxytone vs. paroxytone patterns in Mansaka, yields 184 paroxytone vs. 16 oxytone bases in a data sample equivalent to that collected for the ten languages in Table 4.7. Although Mansaka still has contrastive stress, then, its functional load has been drastically reduced, since some 92 percent of all bases carry penultimate stress, and it appears that Mansaka may well be on its way to eliminating oxytone bases entirely.

These observations raise questions for which we still do not have complete answers. If most Philippine stress languages have a preponderance of oxytone bases, why is the direction of stress reassignment from oxytone to paroxytone in nearly all non-stress languages? One possibility is that Proto Philippines was a stress language dominated by paroxytone bases, and that the preponderance of oxytone bases in most attested languages

is a product of subsequent change. However, PPH oxytone bases appear to have been at least 50 percent more common than paroxytone bases, based on agreements of phonemic stress in cognates shared by Cordilleran and Central Philippine languages. This raises the further question as to how such a preponderance of oxytone bases could have arisen in the first place. It is likely that penultimate schwa played a key role in this process.

As will be seen in greater detail below, mid-central reflexes of PAN *e (schwa) in open penultimate syllables are shorter than other vowels, and many languages with regular penultimate stress substitute an oxytone pattern under this condition. Unlike PAN, Proto Philippines had many heterorganic consonant clusters in medial position, most of which were products of schwa syncope. At some stage in the formation of Proto Philippines, then, two innovations apparently occurred: 1) stress automatically shifted rightward when the penultimate vowel was schwa, and 2) in the environment VC__CV the schwa then elided, giving rise to a subset of uniformly oxytone bases that contained a heterorganic consonant cluster in medial position, as with PAN *qapedu > Ilokano *apró*, Tagalog *apdú* 'bile, gall'. Given this correlation of consonant clusters and final stress many languages then extended the pattern to reduplicated monosyllables, and to bases with a homorganically prenasalised obstruent in medial position. However, these changes could not have given rise to a preponderance of oxytone bases if the language in which they began had regular penultimate stress, since the great majority of PAN bases were disyllabic and lacked a medial consonant cluster.

The current distribution of stress languages in the Philippines, includes a more-or-less solid block from northern Luzon to the southern Bisayas, but none in the islands north of Luzon, and almost none in Mindanao and the Sulu Archipelago. This distribution can be interpreted historically in either of two ways: 1) stress may have been innovated at some point in the central or northern Philippines, and then spread through contact, or 2) stress languages may originally have been present throughout the Philippines, with a tendency for fixed stress to develop at the geographical extremes of north and south. Comparative evidence strongly favors the latter alternative. Despite differences of detail, Philippine languages generally agree in the stress pattern of cognate morphemes, and lexical stress is therefore reconstructed for Proto Philippines, as in PPH *matá 'eye' (Ilokano, Isneg, Tagalog, Bikol *matá* 'eye'), or PPH *láŋit 'sky' (Ilokano, Isneg, Tagalog, Bikol *láŋit* 'sky'). It follows that Proto Philippines almost certainly had phonemic stress, and discrepancies in the correspondence pattern are due to language-specific differences in the conditions under which stress has been made predictable, as with Ilokano, Isneg, Bikol *dálan*, but Tagalog *daʔán* 'path, road' (with shift to required oxytone stress after loss of *l and development of automatic glottal stop between like vowels). It may also be noted in passing that phonemic stress was lost and reacquired through contact in Pangasinan (Zorc 1979) and perhaps in some Central Philippine languages.

Philippine languages are also unusual with respect to the behaviour of stem-initial glottal stop. Although vowel-initial bases are not preceded by an automatic glottal stop in most AN languages, this is common in the Philippines. Since there is no surface contrast for V- and ʔV-, the glottal stop is not normally written in initial position, even though it is phonemic in other environments. When a prefix is added to a vowel-initial base, however, the glottal stop usually remains, as in Tagalog *áral* [ʔáɾal] 'admonition, counsel' : *paŋ-áral* [paŋʔáɾal] 'preaching, counsel' (cf. Malay *ajar* [ʔádʒaɾ] 'instruction' : *pəŋ-ajar* [pəŋádʒar] (or, less commonly, [pəŋʔádʒaɾ]) 'instructor', from which the Tagalog forms are borrowed). In non-Philippine languages that do have an automatic glottal onset to initial vowels, as Malay, the phonetic glottal stop in the base disappears in affixed forms, whereas

in Philippine languages it remains. For this reason some scholars believe that word-initial glottal stop is phonemic in Tagalog and some other Philippine languages.

Most recently Lobel and Riwarung (2009, 2011) have made the remarkable 'discovery' that Maranao, a language which has been the subject of considerable linguistic attention at least since McKaughan (1958), has four previously unrecognized consonants (called 'heavy consonants' pending further phonetic specification) that have arisen from derived clusters of voiced + voiceless obstruents. As noted by the authors of these two studies, native Maranao writers have --- not surprisingly --- been aware of this distinction for many years, and at least some (but not all) indigenous writers have marked these consonants by distinctive orthographic devices in the fairly extensive literature (much of it religious) intended only for the Maranao community. In addition to showing that not all Philippine languages have 'simple' consonant systems, this paper is important in revealing the role of cultural presupposition in putting blinders on the investigator, since the Maranao are Muslims, and have been studied primarily by Christian missionaries working for the Summer Institute of Linguistics, who could have corrected this error much earlier had they considered the available Islamic textual material in this language.

In a second important contribution to revising the phonological study of Philippine languages, Lobel and Hall (2010) have shown that Southern Subanen has developed a contrastive set of aspirated consonants each of which can be traced to "a Proto-Subanen cluster of *kp, *kt, *gk (via *kk), or *ks, which evolved into geminates in pre-Southern Subanen" (Lobel and Hall 2010:329). Again, earlier material on this language, as the wordlists in Reid (1971), completely overlooked this distinction.

4.1.3 Borneo (and Madagascar)

Phonological typology in Borneo is quite varied, with certain features recurring in widely separated areas. Apart from Malagasy, which will be treated separately at the end of this section, most phoneme inventories probably fall within the general range of 19-25 segments, but on average they are somewhat larger than those typical of the Philippines, since many languages have a separate palatal series that includes *j, ñ,* and in some cases *c,* and others have developed six-vowel systems by monophthongisation of the diphthongs *-ay* and *-aw* > *-e* and *-o*. Many languages of central and western Borneo have segment inventories similar to that of Kayan, with 19 consonants and six vowels /p t k ʔ b d j g m n ñ ŋ s h v l r w y; i u e ə o a/, or of Highland Kenyah, with 18 consonants and six vowels /p t c k ʔ b d j g m n ñ ŋ s l r w y; i u e ə o a/. The largest Bornean segment inventory known probably is that of Sa'ban (headwaters of the Baram River, northern Sarawak), with 22 consonants (many of which may occur long or short) and ten vowels. Ignoring length, Sa'ban can be analyzed as having 32 phonemes. The problem with counting segments in highly innovative systems such as this, is that some segments can be interpreted phonemically either as units or as clusters. This is true both for the voiceless nasals and liquids of Sa'ban (*hm, hn, hŋ, hl, hr*), and for the postploded medial nasals of Narum (m^b, n^d, $ñ^j$, $ŋ^g$), which has only a slightly smaller inventory (24 consonants, 6 vowels). Because of descriptive gaps the smallest inventory is unclear. However, most languages of Sabah have fairly small segment inventories, since they have merged the palatal and dental series, have not expanded the vowel inventory, and have merged some voiced stops with sonorants. The languages of Sabah for which data is available have 15-18 consonants and four vowels. Okolod Murut, with 14 consonants and four vowels /p t k ʔ b g m n ŋ s l r w y; i u o a/, is a good candidate for the smallest phoneme inventory in Borneo. In general,

then, most Bornean segment inventories fall within the same 19-25 phoneme range as those of Philippine languages, but Philippine languages tend to cluster around the lower end of this range while Bornean languages, especially those south of Sabah, tend to cluster around the upper end.

Far more striking than variations in size are variations in the content of Bornean segment inventories, some of which are typologically very unusual. For historical reasons that will be explored in a later chapter, much of the diversity in the phonological typology of Bornean languages has arisen within a single subgroup (North Sarawak) which occupies a relatively restricted region. If this one small area did not exist the island of Borneo would be typologically much more uniform than it is. Table 4.8 shows phoneme inventories for Sa'ban, Bario Kelabit, Narum, and Bintulu of northern Sarawak:

Table 4.8 Phoneme inventories for four languages of northern Sarawak

Sa'ban (Blust 2001e)					**Bario Kelabit** (Blust 1993a)				
p	t	c	k	ʔ	p	t		k	ʔ
b	d	j			b	d	(j)	g	
m	n		ŋ		bʰ	dʰ		gʰ	
hm	hn		hŋ		m	n		ŋ	
	s			h		s			h
	1					l			
	hl					r			
	r				w		y		
	hr								
w		y							
vowels: i, ɪ, e, ɛ, ə, a, u, ʊ, o, ɔ					vowels: i, u, e, ə, o, a				
22 + 10 = 32					20 + 6 = 26				

Narum (Blust n.d.(a))					**Bintulu** (Blust n.d.(a))				
p	t	c	k	ʔ	p	t		k	ʔ
b	d	j	g		b	d	j	g	
m	n	ñ	ŋ		ɓ	ɗ			
mᵇ	nᵈ	ñʲ	ŋᵍ		m	n	ñ	ŋ	
f	s			h		s			
	l				v	z			
w	r	y				l			
						r			
					w		y		
vowels: i, u, e, ə, o, a					vowels: i, u, ə, a				
24 + 6 = 30					21 + 4 = 25				

Sa'ban is a highly innovative dialect of Kelabit which has developed a number of phonemes that have not been reported elsewhere in Borneo. Among the consonants both nasals and liquids have voiceless counterparts. Since Sa'ban has initial and medial consonant clusters these segments can be interpreted as clusters of *h* + sonorant, and if this analysis is adopted the Sa'ban inventory will drop to 27 phonemes. Equally surprising

from the standpoint of areal typology is the large vowel inventory of Sa'ban in comparison with languages like Bario Kelabit, since lexicostatistically the two are dialects of the same language. The principal basis for the addition of new vowels in Sa'ban has been the development of laxed/lowered counterparts of the high and mid vowels. Although these changes may have begun as conditioned, they have been so obscured by subsequent innovations that there is no longer any question that the new phones are phonemes.

While it is bizarre from the standpoint of areal typology, the phoneme system of Sa'ban is at least compatible with the general expectations of universal typology. The segment inventory of Bario Kelabit, on the other hand, contains a series of stops that is among the rarest known in any language (Blust 1969, 1993a, 2006a). Next to the plain voiced obstruents *b, d, g* (and a voiced palatal obstruent with low functional load, found mostly in loans), Bario Kelabit has a second series of voiced obstruents b^h, d^h, g^h that begin voiced and end voiceless, with variable voiceless onset to the following vowel, as in $təb^huh$ [tə́bpʰuh] 'sugarcane', $id^huŋ$ [ʔídtʰʊŋ] 'nose', or $ug^həŋ$ [ʔúgkʰəŋ] 'spin without wobbling, of a top'.[22] These segments thus differ from the fully voiced murmured stops of Hindi or other Indo Aryan languages that are still sometimes called 'voiced aspirates'. Phonetically the Kelabit voiced aspirates occur only word-medially following a stressed vowel. As the onset to the final syllable of a base they are phonemic, as just shown, but as the onset to the final syllable of a suffixed base they are almost entirely predictable, as in $kətəd$ 'back (anat.)' : $kətəd^h$-$ən$ 'be left behind, as in walking single-file and not keeping up', or $arəg$ 'fragment, small piece of something broken' : $rəg^hən$ 'break something into small pieces'. A similar set of segments is also found in the Long Lellang dialect of Kelabit, in the distinct but closely related language known variously as Lun Dayeh, Lun Bawang, or Southern Murut, and in Ida'an Begak of eastern Sabah. Unlike Kelabit or Lun Dayeh, in Ida'an Begak a cluster analysis appears to be preferable (Goudswaard 2005). As will be seen in greater detail later, the voiced aspirates of Bario Kelabit correspond to single voiced obstruents in PMP and to various types of single-consonant phonemes in other Kelabit dialects, as Pa' Dalih (b^h : p, d^h : s, g^h : k), or in other languages, such as Bintulu (b^h : ɓ, d^h : ɗ, d^h : j, g^h : g).

Bintulu ɓ and ɗ are strongly imploded. Unlike the coronal implosive of some other languages, which is retracted to post-alveolar or even cerebral position, Bintulu ɗ appears to be alveolar. As in many other languages globally, Bintulu has only labial and alveolar implosives. Implosives also occur in some Lowland Kenyah dialects such as Long San, Long Sela'an, and Long Ikang, where they are found at labial, alveolar, palatal and velar positions. In these languages, which have final stress in citation forms, the phoneme inventory includes at least 17 consonants and six vowels: /p t k ʔ b d j g m n ñ ŋ s l r w y; i u e ə o a/. What is unclear is the phonemic relationship between plain obstruents and their implosive counterparts. In Long Ikang the voiced obstruents have strongly implosive allophones before a stressed vowel, but plain allophones elsewhere: $baʔ$ [ɓaʔ] 'voice' (but $baaʔ$ [baáʔ] 'swollen'), $pədəw$ [pəɗə̃w] 'gall', jo [ɗʸoː] 'one' : $kələja$ [kələɗʸáː] 'work' (but $jalan$ [dʸalán] 'path, road'), $gəm$ [ɠəm] 'foot, leg', $sagaʔ$ [saɠẵʔ] 'dance' (but $gatən$ [gatə́n] 'itchy'). A palatal implosive was once considered phonetically impossible, and such segments must in any case be rare. The palatal implosives of Long Ikang, however, are actually palatalised alveolars (hence [ɗʸ]). A similar distribution of plain and implosive voiced obstruents is found in the dialects of Long San and Long Sela'an. In the early 1970s

---

[22]   In Kelabit and many other languages of western Indonesia /e/ and /ə/ are in complementary distribution, but this complementation is regarded as accidental. As a result the practical orthographies use /e/ for both phonemes.

these dialects were undergoing reduction of nasal-stop clusters by loss of the nasal, as with Long Sela'an *təŋgan ~ təgan* 'floor', or *ŋgaŋ ~ gaŋ* 'handspan'. This change introduced new instances of plain voiced obstruents as the onset of a stressed syllable, and the allophonic status of the implosive stops was undergoing transformation into one of phonemic contrast. However, since *mb and *nd have become *v* and *r* in Long Sela'an, implosion appears to be contrastive only for the palatal and velar stops. The result appears to be an emerging typological anomaly: a system of voiced obstruents /b, d, j, g, ɗ^y, ɠ/, in which the phonemic implosives favor the back of the oral cavity, contrary to universal tendencies in which such segments favor the front of the oral cavity (Maddieson 1984:111ff). However, phonetic implosives still occur at all four points of articulation, and only before a stressed vowel.

The Narum segment inventory touches on another point of some typological interest. In addition to plain nasals which occur initially, medially, and finally, Narum has a set of postploded nasals that occur only medially. These segments reflect earlier nasal + voiced obstruent, and as a result the oral release in this type of segment is always voiced. Like clusters of nasal + voiced obstruent, postploded nasals begin with an oral closure and lowered velum which is raised prior to releasing the oral closure. The difference between these articulations is determined by the time interval between velic closure and oral release. In a cluster such as -mb- the velum is lowered for about half the duration of the total sequence, but in -m^b- the velum is lowered for most of the duration, giving the impression of a relatively long primary nasal articulation with an extremely brief oral release. This is consistent with the observation of Cohn (1990:199) that in languages generally the nasal of an -NC- cluster takes up about half the duration of the cluster where the obstruent is voiceless, but more than half where it is voiced. In other words, nasal articulations have a universal tendency to spread into following voiced obstruents but not voiceless obstruents, and languages with postploded medial nasals have simply carried this tendency further than is normal for allophonic variation. Because of the oral release of postploded nasals vowels following them are never nasalised.

Postploded nasals are known only from western Indonesia. In Borneo they are found in Narum and other Lower Baram languages of northern Sarawak, in Land Dayak and Iban of southern Sarawak-west Kalimantan, and in Tunjung of southeast Kalimantan. Beyond Borneo they are either attested, or inferred for earlier periods in the Chamic languages of Vietnam and southern China, in Urak Lawoi' of peninsular Thailand, in some non-standard dialects of peninsular Malay (such as Ulu Muar), in Acehnese and Rejang of Sumatra, in Lom of Bangka Island between Borneo and Sumatra, and in Sundanese of west Java. This distribution is puzzling, since it is sufficiently confined to suggest that it might be due to areal contact or subgrouping, but it is geographically discontinuous and not aligned with established subgroup boundaries.

A number of the languages of Borneo also have preploded final nasals, which have arisen from word-final simple nasals in syllables that have a non-nasal onset. These are perhaps best known from the Land Dayak-Kendayan ([kən^{d}ájaᵗn]) Dayak area of southern Sarawak and adjacent parts of the Indonesian province of West Kalimantan, but they have a geographical distribution which is very similar to that of postploded medial nasals. The term 'nasal preplosion' was coined by Court (1967) in a brief description of the phonology of Mentu Land Dayak. In this language nearly all final nasals are preceded by a brief voiceless oral onset, as in *əsɨ^{p}m* 'sour', *burə^{t}n* 'moon', or *turaᵏŋ* 'bone'. The exceptions fall into two classes: personal names, and bases in which the final syllable begins with a nasal consonant. While the first class of exceptions evidently is motivated by conscious

manipulation of the phonetic substance for purposes of marking semantic fields, the second is due to purely phonetic factors. Phonetic information on vowel nasality is poorly reported for most AN languages, but relatively good data is available for the languages of Borneo, and in these it is clear that vowels are most strongly nasalised by a preceding primary nasal consonant in a process that can be described as 'onset-driven' nasal harmony (Blust 1997c). Coda-driven nasal harmony is generally absent, but some minor leakage of nasality into vowels that precede a final nasal must take place in most languages, or final nasals would be universally preploded. The preplosion of final nasals can therefore be seen as a strategy for blocking nasal spreading in the 'wrong' direction. When a final syllable begins with a nasal consonant this is impossible, as in Mentu Land Dayak *inəm* 'six'. The oral element in a preploded nasal is voiceless in some languages, such as Mentu Land Dayak, voiced in others (as Bau/Senggi Land Dayak), and mixed in Bonggi, spoken on Banggi Island north of Sabah, where Boutin (1993:111) reports -^{b}*m*, -^{d}*n,* but -^{k}*ŋ.*

One other feature of nasal consonants is noteworthy. In Highland Kenyah dialects such as Lepo' Sawa (spoken at Long Anap), syllabic nasals occur both prevocalically and preconsonantally. These segments occur only in initial position before a vowel or voiced obstruent in bases that would otherwise be monosyllabic. It thus appears that they are allophones of the non-syllabic nasals, an interpretation borne out by the further observation that in the dialect of Long Anap variation occurs between non-syllabic and syllabic nasals before a vowel, as in *nəm* 'six', or in words that optionally lose CV- before a prenasalised stop, as *məndəm ~ ndəm* 'darkness'.

Other details of the consonant inventories of Bornean languages which merit some notice are the occurrence of geminate consonants in several languages of northern Sarawak, as Berawan, Kiput, and Narum, the occurrence of a voiced labiovelar stop *g*^{w} in Kejaman and Bekatan of the upper Rejang (also spelled 'Rajang') River in south-central Sarawak, as well as in Bintulu as recorded around 1900 (Ray 1913), and the occurrence of a rounded labio-dental fricative *f*^{w} in Kiput of northern Sarawak.

Vowel inventories are generally small in the languages of Borneo. Many languages retain the inherited four-vowel system of Proto Austronesian (*i, *u, *a, *e, where *e = schwa), or have modified this system by shifting the schwa to a back rounded or rounded mid-central vowel (*i u a o*), and Banjarese of southeast Kalimantan has merged PAN *e and *a in all positions, reducing the system to the basic vowel triangle (*i u a*). However, a number of languages south of Sabah that have retained the schwa in most or all positions have developed *e* and *o*, either through lowering of high vowels under conditions that were subsequently removed by further change, or by monophthongisation of *-ay and *-aw.

As noted above, vowel nasalisation in Borneo is generally onset-driven, and coda-driven nasal spreading is negligible. Very little information on allophonic vowel nasality is available for Formosan or Philippine languages, but what there is suggests that the pattern is the same. Certain consonants are transparent to the rightward spread of vowel nasality. These always include glottal stop and *h*, usually include semivowels *w* and *y*, and in some languages also include *l*, but not necessarily *r*.

Phonemic vowel nasality is rare in AN languages, and when it appears it may have a very restricted distribution. Usually restrictions on the distribution of nasalised vowels are statable in terms of phonetic environment, but in two languages of northern Sarawak phonemically nasalised vowels are found in a single known morpheme. In Bintulu the general marker of negation is *ã* [ʔã]. Since other words that begin with *a-* invariably have an oral vowel, the nasality of the negative marker cannot be explained as conditioned. However, in a field corpus of nearly 800 available lexical items no other examples of

unconditioned vowel nasality are known. It might be argued that many languages (English among them) have colloquial negative markers with deviant phonetic properties such as morpheme-specific vowel nasality, but a second example from northern Sarawak cannot be explained in this way. Miri, spoken near the mouth of the Baram River, has the minimal pair *haaw* 'rafter' : *hããw* '2sg'. Here vowel nasality may have arisen through the process of 'rhinoglottophilia' (Matisoff 1975), in which glottal and pharyngeal consonants lower the velum, with concomitant vowel nasality. But in this case the conditioning would have to be morpheme-specific, since *h* reflects *k in both words (PMP *kasaw 'rafter', *kahu '2sg'). Again, vowel nasality in this form appears to be unique, and the question of how to represent it phonemically presents problems.[23]

In addition to vowels, a number of the languages of central and northern Sarawak have rich systems of diphthongs, and even triphthongs. Kiput, spoken at the junction of the Baram and Tutoh Rivers, has ten level or falling diphthongs with a high vocoid as coda (*-iw, -ew, -uy, -oy, -əy, -əw, -ay, -aay, -aw, -aaw*), at least four diphthongs with a mid vocoid as coda (*-iə, -eə, -uə, -oə*), and two triphthongs (*-iəy, -iəw*). As these four diphthongs show, Kiput is also typologically unusual in having a mid-central glide which is essentially a non-syllabic schwa (Ladefoged and Maddieson 1996:323).

Unlike the Philippines, stress is not phonemic in any language of Borneo. However, unlike much of the AN world, where stress falls predictably on the penult, many of the languages of coastal Sarawak, and some interior languages have final stress in citation forms, but penultimate stress in phrasal context. Final stress thus appears to serve as a marker of decontextualised lexical data. Information on syllabification is not available in the descriptions of most AN languages, but fieldnotes collected by the writer show the interesting fact that in Highland Kenyah words such as *ləkoʔ* 'bracelet', which contain a single phonetically geminated medial consonant, are syllabified *lək.koʔ* , while words such as *lundoʔ* 'sleep', which have a medial consonant cluster, are syllabified *lu.ndoʔ*.

Malagasy must be treated with the languages of Borneo on subgrouping grounds, but because of its physical separation from the other languages of Borneo it shows a sharp typological divergence. In general, Malagasy phonology is highly innovative, perhaps in part because of Bantu substratum (Dahl 1954). As will be seen in Chapter 7, however, the system of voice marking is very conservative. Typological divergence between Malagasy and the Barito languages with which it subgroups has thus occurred primarily because of innovative phonological developments in Malagasy, but innovative syntactic developments in the languages of southern Borneo. Table 4.9 gives the phoneme inventory for standard Malagasy as described by Dahl (1951:32ff; note that /u/ is conventionally written *o*).

The phoneme inventory of Malagasy is relatively rich in fricatives, but diverges most markedly from that of Bornean languages in having four affricates written *ts, tr, dz*, and *dr*. Dahl (1951:33) describes all of these as alveolar, and further notes that the rhotic release of *tr* is voiceless. Dyen (1971b:213), however, distinguishes the place features of *ts* and *tr* and their voiced counterparts as alveolar and alveolar retroflex respectively.

Whether prenasalised obstruents are unit phonemes in Malagasy—an issue on which different writers have taken different stands—is taken up below in a general discussion of this problem in AN languages, and will not receive further attention here. Malagasy is prosodically interesting because it shows historically secondary stress contrasts that have arisen as a result of the addition of a word-final supporting vowel *–a*. Most writers on

---

23   Adelaar (2005a:20) lists /õ/ and /ũ/ as phonemes of Salako in southern Sarawak, but all examples of /õ/ that he cites are products of nasal spreading. The 2sg possessive suffix, however, is /ũ/ (< *-mu), as in *beber* 'lip' : *beber-ũ* 'your lip', or *nasiʔ* 'cooked rice' : *nasiʔ-ũ* 'your rice'.

Malagasy at least since Richardson (1885) agree that, although stress generally falls on the penult, lexical stress contrasts can be found. One often-cited pair included in Richardson's dictionary, and cited by Dyen (1971b:214), is *tánana* 'hand' : *tanána* 'village'. Similarly, Beaujard (1998) marks stress throughout his extensive dictionary of Tañala, although this generally falls on the penult.

Table 4.9 Phoneme inventory for standard Malagasy (Merina dialect)

p	t	k	
b	d	g	
	ts, tr		
	dz, dr		
m		n	
f	s		h
v	z		
	l		
	r		

vowels: i, u, e, a
19 + 4 = 23

4.1.4 Mainland Southeast Asia

In mainland Southeast Asia the phonological typology of AN languages has been determined in many cases by contact-induced change. This has been described in greatest detail for the Chamic languages of Vietnam and southern China by Thurgood (1999).

Most segment inventories of mainland Southeast Asian AN languages are fairly large, the record probably being held by Phan Rang Cham, spoken in coastal Vietnam, for which Thurgood (1999) gives 31 consonants and eleven basic vowels. Malay, typical of many languages in western Indonesia, and of the immediate ancestor of the Chamic languages before Mon-Khmer contact influence, has eighteen consonants and six vowels in native vocabulary, but additional consonants in loanwords from Arabic and European languages. Although this is a large phoneme inventory for island Southeast Asia, it may be the smallest for any AN language on mainland Southeast Asia (in Table 4.10 below numbers following a slash include phonemes in borrowed vocabulary).

Most Chamic languages distinguish clear and breathy registers. Among consonants the breathy register is seen in Phan Rang Cham *bh, dh, jh, gh*. In other Chamic languages these segments are written with voiceless stops, as their distinctive feature is the breathy quality of the following vowel, rather than voicing. The series *ph, th, ch, kh*, on the other hand, is aspirated, although some writers qualify this by noting that Chamic languages allow a number of word-initial consonant clusters, and the sequences *ph*, etc. could therefore be treated as clusters rather than units. This is the case in many of the Mon-Khmer languages, in which an infix may be inserted between an initial voiceless stop and a following *h*. In Jarai the sequences *ph, th, ch, kh* may be pronounced in careful speech with a short, schwa-like release preceding the aspiration, again suggesting that these are phoneme sequences rather than unitary segments. The orthographic parallelism in *ph, th, ch, kh vs. bh, dh, jh, gh* is thus misleading, since the aspirate represents a consonant in the first series, but a laryngeal setting which produces breathy voice on the following vowel in the second series. In addition to this vocalic distinction which is marked on the consonants in the

traditional Brahmi-based Cham syllabary, Phan Rang Cham and various other Chamic languages have a series of voiced glottalic consonants. Thurgood (1999:268), partly following earlier sources, writes the first two of these segments as implosives, but the third as preglottalised. In Jarai all three are preglottalised, but not noticeably imploded. In this respect they are rather like the preglottalised stops of south-central Formosan languages such as Thao or Bunun, and differ from the implosives of such Bornean languages as Bintulu or Lowland Kenyah, where the similar segments are produced with a noticeable lowering of the larynx and a strong ingressive airstream.

Table 4.10 Phoneme inventories for Western Cham and Malay

Phan Rang Cham (Thurgood 1999:268ff)					Malay (Wilkinson 1959)				
p	t	c	k	ʔ	p	t	c	k	
ph	th	ch	kh		b	d	j	g	
b	d	j	g		(dz)				
bh	dh	jh	gh		m	n	ñ	ŋ	
ɓ	ɗ	ʔj			(f)	s	(ʃ)	(x)	h
m	n	ñ	ŋ		(z)				
s	ś		h		l				
	l				r				
	r				w		y		
w	y								
vowels: i, ɨ, u, e, ɛ̃, ə, o, ɔ̃, ɛ, a, ɔ					vowels: i, u, e, ə, o, a				
31 + 11 = 42					18/23 + 6 = 24/29				

Chamic vowel inventories show considerable expansion of the four-vowel system of Proto Austronesian. This must be due at least in part to Mon-Khmer contact influence, given the extremely rich vowel systems of many of these languages. Thurgood (1999) suggests that the typological reformation of the Chamic languages was due primarily to contact with the Bahnaric branch of Mon-Khmer, although Sidwell (2005) has drawn attention to problems with this explanation, and prefers an unattested Mon-Khmer substrate language as the source of many Chamic innovations. As Thurgood (1999) has shown in detail, some vocalic innovations in Chamic are products of register differences, since these affect vocalic quality in complex ways. Although most Chamic languages form a more-or-less continuous block, three are geographically separated from the rest. The first of these is Western Cham, located around the large Tonle Sap lake in central Cambodia, which became separated when the southern Cham capital of Vijaya fell to the Vietnamese in 1471 AD. Western Cham has breathy and clear vowel registers, like many Mon-Khmer languages. According to some writers, such as Thurgood (1993b, 1999) Eastern Cham has incipient tone under Vietnamese contact influence, although this claim is disputed by Brunelle (2005, 2009). The second geographical isolate is Acehnese, spoken in northern Sumatra. It will be treated in the following section. The third geographically separated Chamic language is Tsat, spoken by about 4,500 people near the southern tip of Hainan Island in southern China. Thurgood (1999) has shown that Tsat subgroups with Northern Roglai of Vietnam, and that both Acehnese and Tsat probably reached their present locations as refugee populations after the Vietnamese conquest of the northern Cham capital of Indrapura in 982 AD. Since Tsat is surrounded by tone languages, both Tai-

Kadai, and Chinese, it is perhaps not surprising that it has developed phonemic tone as well. Thurgood (1999:215) has carefully documented the diachronic steps by which five contrasting tones have developed in this language.

In addition to Tsat, tonogenesis has also been reported for one dialect of Moken, spoken in the Mergui Archipelago, a chain of small islands reaching from southern Burma to peninsular Thailand. Although earlier reports suggested that nearly all Moken speakers were sea nomads who generally avoided contact with outsiders and had only begun to settle some coastal areas, Larish (1999) found that the Moken in the region of Phuket Island, Thailand, use two clearly distinguished dialects, one (Moken) spoken by boat people, and the other (Moklen) spoken by closely related land-based populations. The sedentary Moklen are in contact with southern Thai, and as a result of contact this dialect has begun to develop tonal distinctions that are not present in Moken. A similar process of tonogenesis through contact with southern Thai has been reported for Pattani Malay, spoken just north of the Thai-Malaysian border (Tadmor 1995).

Other features of the phoneme inventories of AN languages in this area include a palatal series in most languages, initial geminates in some non-standard dialects of peninsular Malay, word-final pharyngeal fricatives or uvular stops in other local Malay dialects, and phonemic vowel length in Moken/Moklen. Malay has an inherited palatal series including voiceless and voiced affricates *c* and *j*, and the palatal nasal *ñ*. Formal standard Malay also has an unusually large number of fricatives (six), all but two of which are confined to loanwords from Arabic, European languages, or both. In colloquial speech most of these are replaced by native stops or affricates, as in the Arabic loanwords *fikir* (formal) : *pikir* (colloquial) 'to think', *xabar* (formal, written *khabar*) : *kabar* (colloquial) 'news', and *zaman* (formal) : *jaman* (colloquial) 'era, period of time'. Word-initial geminates are found in most eastern peninsular dialects of Malay. Other Malay dialects have phonological traits that are quite foreign to the standard language. Dialects of the upper Trengganu River in northeast Malaya, for instance, have developed an epenthetic final affricate *-kx* following earlier high vowels (Collins 1983b:48), the Besemah and Seraway dialects of southeast Sumatra contrast alveolar and uvular rhotics, and Pattani Malay has twelve phonemic vowels, including four that are nasalised.

4.1.5 Sumatra, Java, Madura, Bali, and Lombok

The area represented here is not unitary in the sense of the Philippines (a single, fairly close-knit archipelago), Borneo (a single island), or even the AN portion of mainland Southeast Asia, where most of the languages have experienced similar areal adaptations in typology. It is rather chosen as a convenient catch-all category for a number of islands in western Indonesia.

The largest phoneme inventory in this area apparently is that of Madurese, for which Stevens (1968:16ff) gives 26 consonants and nine vowels. Javanese, with 26 consonants and eight vowels (Nothofer 1975:8), and Acehnese, with 20 consonants and ten oral vowels (Durie 1985:10ff) are not far behind. However, the Javanese total includes five prenasalised obstruents which should be regarded as clusters, and the Acehnese total excludes seven nasal vowels and ten diphthongs which, depending upon the assumptions made in counting segments, could very well put Acehnese in the lead. The smallest phoneme inventory certainly is that of Enggano, off the west coast of southern Sumatra, with ten consonants and six vowels, all of which have nasalised counterparts.

Table 4.11 Phoneme inventories of Madurese and Enggano

Madurese (Stevens 1968:16ff)						Enggano (Kähler 1987)					
p	t	ṭ	c	k	ʔ	p	(t)	(c)	k		ʔ
b	d	ḍ	z	g		b	(j)				
bh	dh	ḍh	zh	gh		m	n	ñ			
m	n		ñ	ŋ		(f)			x		h
	s				h						
	l						(l)				
	r						(r)				
w		y						y			

vowels: i, ə, u, e, ʌ, o, ɛ, a, ɔ   vowels: i, ə, u, e, o, a, plus nasality
26 + 9 = 35   10/16 + 6 = 16/22

The languages of this area show several phonological traits that are fairly rare in AN languages. To begin at the northern end of Sumatra, since the end of the nineteenth century it has been recognised that although Acehnese seems closely related to Malay, it has many typological features that are distinctively Mon-Khmer, and most of these are shared with languages of the Chamic group. By the middle of the twentieth century the close relationship of Acehnese with Chamic had become widely accepted. Thurgood (1999) carried this recognition one step further, by suggesting that Acehnese is not a sister language of the Chamic group, but rather a geographically displaced member of the Chamic group, although this claim is challenged by Sidwell (2005). For centuries after their arrival in Sumatra, the Acehnese were engulfed in Malay cultural and linguistic influence, further increasing the lexical similarities between two languages that are in any case fairly closely related. Despite a good deal of lexical borrowing, the phonological and grammatical differences between Acehnese and Malay are substantial, even if much of the phonological divergence is on the level of phonetics. Durie (1985:10) describes two series of stop phonemes in Acehnese: /p t c k/ and /b d j g/. However, positional allophones include aspirated voiceless stops and 'murmured' voiced stops that are distinguished from their modal variants primarily by having "a whispery voice onset lead." Durie (1985:10) notes that the term 'murmur' is used inconsistently in language descriptions, but suggests on the basis of spectrographic data that the murmured stops of Acehnese are acoustically similar to the breathy consonants of Gujarati. In addition to these, Acehnese has a number of other consonant segments that are unusual from a general AN standpoint, including postploded medial nasals (called 'funny nasals'), automatically glottalised labial and dental stops word-finally, voiced allophones of the glottal fricative *h*, and a laminal alveodental fricative with a wide channel area for which no appropriate IPA symbol is available.

Javanese also has several phonological traits that are typologically unusual. The first of these is the partial devoicing of the voiced obstruents /b, d, ḍ, j, g/, and the presence of what Ladefoged and Maddieson (1996:63) call 'slack voice' in the following vowel (a setting of the vocal folds which is more open than that of modal voice). Javanese speakers of Indonesian sometimes 'import' this phonetic feature of their native language in speaking the national language, much to the amusement (and sometimes annoyance) of Malay/Indonesian speakers from Sumatra. Although Ladefoged and Maddieson (1996) describe the Javanese stops as produced with a different laryngeal setting than the breathy register of Mon-Khmer and Chamic languages, the similarities are striking. Nonetheless,

contact as an explanation for the appearance of breathy voice consonants in Javanese seems unlikely, given the absence of anything similar in any known dialect of Malay.

Javanese also differs from most other AN languages in having two retroflex stops (a retroflex nasal occurs preceding these segments, but this is a positional variant of *n*). Although the term 'retroflex' is used in the literature on Javanese, this is a misnomer, as Javanese *ṭ-*, and *ḍ-* are not retracted nearly as much as the apico-domal consonants of Dravidian or Indo-Aryan languages. In AN languages *t* is often a voiceless unaspirated postdental stop, and *d* is alveolar. Javanese has retained this contrast, but added voiceless alveolar and voiced postdental stops to the inventory. The contrast today is thus between voiced and voiceless stops at postdental and alveolar or slightly postalveolar positions. Dahl (1981a:23) argued that since retroflex consonants occur only syllable-initially in both Indic languages and Javanese, the Javanese retroflex stops must have arisen through Indic contact influence. But retroflex stops probably are universally associated with syllable-initial position. Given the quite different phonetics of 'retroflexion' in the two languages and the presence of 'retroflex' stops in such Javanese words such as *ṭɛk* 'sound produced by striking a bamboo tube', the hypothesis of an Indic origin for Javanese *ṭ* and *ḍ* is not convincing. Another argument against the Indic origin of Javanese retroflexion derives from the observation that the voiced retroflex consonant is far more common than its voiceless counterpart. No statistical data is available to make this point, but a sample can easily be constructed. The *Javanese-English dictionary* of Horne (1974) does not segregate postdental and retroflex stops alphabetically. Table 4.12 shows the relative proportions of retroflex to dental stops for the first 100 words in both the *d* and *t* sections of the dictionary. This provides a picture of relative frequency in word-initial position. To provide a similar picture of relative frequency in word-medial position the first 100 examples of dental or retroflex stops have been culled from the dictionary beginning with the *a* section. Duplications which arise through cross-referencing in the dictionary have been eliminated:

Table 4.12 Relative frequency of dental and retroflex stops in Javanese

d-	ḍ-	t-	ṭ-	-d-	-ḍ-	-t-	-ṭ-
33	67	93	7	40	60	67	33

Although preliminary, these numbers show a clear asymmetry between voiced and voiceless stops in relation to retroflexion. For the voiced stops the retroflex segment is roughly twice as frequent as the dental segment word-initially, and 1.5 times as frequent word-medially. For the voiceless stops this pattern is reversed, most markedly in word-initial position, where only about 7% of voiceless coronal stops are retroflex. If the 'retroflex' consonants of Javanese were a product of borrowing from literary Sanskrit or colloquial Prakrit it is difficult to see why retroflexion was not also extended to the nasals, as it is in Indic languages, or why retroflexion would greatly favor voiced over voiceless stops, since a parallel asymmetry apparently was not found in Sanskrit.[24]

Javanese has had a complex history of contact, and it is likely that borrowing from other AN languages has played a key role in the development of retroflexion as a contrastive feature in the stop system of this language. Dempwolff (1937) noted that Malay loanwords

---

[24]   In his comparative dictionary of Indo-Aryan languages Turner (1966) lists 98 forms with initial voiceless retroflex stops, and 106 with voiced retroflex stops. A random sampling in intervocalic position following *a-, u-, ka-, ki-, ku-* and *ga-* shows 95 voiceless and 51 voiced retroflex stops.

with *d* are borrowed into Javanese with the retroflex stop, as seen in doublets such as *ḍalam* 'internal' (< Malay *dalam*), but *daləm* 'inside of, within; inner rooms of a house' (native). A similar pattern is seen in loanwords from European languages, as with *ḍefinisi* 'definition', *ḍinamit* 'dynamite', or *ḍobəl* 'double(d), duplicated', borrowed from Dutch. The reason for this phonological adaptation is straightforward: as in most AN languages for which adequate phonetic descriptions are available, *t* in Malay is postdental, but *d* is alveolar, and hence phonetically closer to Javanese *ḍ-* than to Javanese *d*. For the same reason we would expect Malay loanwords that contain *t* to be borrowed into Javanese with the dental *t*, not its retroflex equivalent.

In Javanese has borrowed fairly heavily from Malay at various times in its history, and we would thus expect the number of retroflex stops to be inflated relative to dental stops for the voiced series, but the reverse for the voiceless series. During the Dutch colonial period Javanese also borrowed fairly heavily from Dutch, since Java was the preeminent post to which colonial officials were assigned, and this would further tend to inflate the percentages of retroflex to dental stops for the voiced series. Curiously, even though both *t* and *d* are alveolar in Dutch and English, *t* is always borrowed as a postdental stop, as in *tabɛl* 'table (of figures, facts)', *taksi* 'taxi', or *tɛkstil* 'textile'. Given these observations the origin of the Javanese dental : retroflex contrast seems almost certainly rooted in a history of borrowing, but borrowing from other AN languages (and later, European languages) rather than borrowing from Indic sources.

In addition to borrowing from other languages Javanese has also been a source of loanwords. Because Javanese has been a language of considerable local influence and prestige at various times, retroflex stops have been borrowed into both Madurese and Balinese from a Javanese source. It is also likely that the so-called 'voiced aspirates' of Madurese, which include segments at labial, dental, alveolar, palatal and velar positions (Stevens 1968:16) are a product of Javanese contact. Cohn (1994) maintains that these segments are voiceless aspirates in contemporary Madurese, but the fact that they derive historically from plain voiced stops strongly suggests that they passed through a stage in which some other laryngeal mechanism was superimposed on ordinary voicing.

Both Madurese and Javanese have considerably expanded vowel inventories. In at least Madurese some vowel splitting appears to have been conditioned by voiced and aspirated stops (Cohn 1993b). In addition, some dialects of Rejang in southern Sumatra have as many as ten diphthongs (McGinn 2005), a total matched elsewhere in insular Southeast Asia only by some of the Lower Baram languages of northern Sarawak such as Kiput (Blust 2002c). Other typologically noteworthy traits in the languages of this area include consonant gemination, bilabial trills, preploded final nasals, and the relatively common occurrence of voiceless velar fricatives. Consonant gemination is phonemic in Toba Batak of northern Sumatra, and in Madurese. In the former language it derives from earlier sequences of nasal plus voiceless stop, and it is still represented in the Indic-derived Batak syllabary in this way. A voiceless velar fricative *x* is fairly common in the languages of Sumatra, being reported from Simalur, Sichule, Nias, Enggano, Lampung and 'Middle Malay' (Besemah and Seraway). Because it derives from several different etymological sources (*p and *k in Simalur, *j in Sichule and Nias, *R in Lampung and Middle Malay) this makes the common occurrence of this segment all the more surprising. Catford (1988) reports the presence of three trills in Nias: bilabial, alveolar and post-alveolar (slightly retroflex). The middle member of this set is the ordinary trilled *r* of many languages, but the other two are very rare in global perspective. Both are sometimes prenasalised, most commonly in medial position, and the bilabial trill is described as *endolabial*, that is, "the

lips are somewhat everted and held loosely together so that their inner surfaces are in contact" (Catford 1988:153). According to Catford, all three trills are short, involving normally only one to three taps. In addition, Sichule has what Kähler (1959) describes as a length contrast in a single vowel. Surprisingly, this is the schwa, a vowel that in almost all other AN languages of insular Southeast Asia is inherently shorter than any other.

4.1.6 Sulawesi

Sneddon (1993) divided the languages of Sulawesi into ten microgroups: 1) Sangiric, 2) Minahasan, 3) Gorontalo-Mongondow, 4) Tomini, 5) Saluan, 6) Banggai, 7) Kaili-Pamona, 8) Bungku-Tolaki, 9) Muna-Buton, and 10) South Sulawesi. Mead (2003a,b) recognises an additional Sulawesian microgroup, Wotu-Wolio (formed by joining the previously unclassified Wotu language with Wolio, extracted from Sneddon's Muna-Buton group). At the same time he has shown that Banggai should be included in the Saluan group, and that there is some evidence for a Celebic supergroup which encompasses all of the microgroups he recognises except 1-3, and 10.

The languages in Sulawesian microgroups 1-3 belong to the Philippine subgroup, and are in many respects typologically similar to them. Those further south show a divergence from Philippine typologies which in general increases with distance. On the whole the segment inventories of Sulawesian languages vary little in size within microgroups, but considerable divergence may occur across microgroup boundaries. Almost all languages of central and northern Sulawesi have segment inventories ranging between 18 and 23 segments. Sangil, a Sulawesian language in southern Mindanao which apparently split off from Sangir several centuries ago, and some of the languages of central and southwest Sulawesi have slightly larger inventories (Sangil 18 + 6 = 24, Buginese 18 + 6 = 24, Barang-Barang 20 + 5 = 25, Bada/Besoa 21 + 6 = 27), but the largest phoneme inventories are found in southeastern Sulawesi, as with Muna (25 + 5 = 30), Tukang Besi (27 + 5 = 32), Kulisusu (29 + 5 = 34), or Wolio, spoken on Buton Island, which apparently has the record, with 31 consonants and five vowels, for a total of 36 segments. The smallest inventory found so far in any of the languages of Sulawesi probably is shared by several languages, including Sa'dan Toraja, spoken in the central mountain massif of the island, and by some of the Tomini languages, of which Lauje can be taken as representative. Both Sa'dan Toraja and Lauje have 13 consonants and five vowels in native vocabulary, for a total of 18 segments.

At first sight the segment inventories of most languages in Sulawesi seem ordinary. In relation to most other languages of insular Southeast Asia, however, there are several differences, both positive and negative, which are noteworthy. One positive feature that marks this region as different from others surveyed so far is the widespread presence of word-initial prenasalised obstruents. These are rare in the north of the island (attested so far only in Ratahan, or Toratán of the Sangiric group), and apparently are absent in the southwest, but are relatively common in the languages of central and southeast Sulawesi, and are the major factor accounting for the greater size of phoneme inventories in this part of the island. The analysis of this problematic set will be taken up below.

Other distinctive typological traits in segment inventories include the presence of a retroflexed lateral flap in most Sangiric languages, in several Gorontalo-Mongondic languages, and Tonsawang; the occurrence of a voiced retroflex fricative in Talaud; the occurrence of a dental vs. alveolar contrast for voiced (but not voiceless) stops in Muna; and the absence of a palatal series almost everywhere except in the South Sulawesi

languages. Perhaps the most striking negative typological feature of the languages of Sulawesi, however, is the rarity of diphthongs and of schwa, particularly in languages of the 'Celebic supergroup'. Throughout Sulawesi vowel inventories also show little variation. In every language for which descriptive data has been examined the vowel system is either *i, u, e, o, a*, or this set of five vowels plus schwa. No language in Sulawesi is known to have fewer than five vowels, or more than six.

In addition to these distinguishing traits, implosive consonants are fairly common in the southeastern languages (Wolio, Muna, Tukang Besi). Anceaux (1952:4ff) describes implosive stops at labial and alveolar places for Wolio. He notes that the alveolar implosive is slightly retroflexed, and stresses that the corresponding non-implosive labial and alveolar stops are produced with 'high muscular tension.' Muna has a single labial implosive, but Tukang Besi has ɓ and ɗ. A few descriptions note that *t* is postdental, but *d* alveolar; as with Selayarese (Mithun and Basri 1986:214), and Barang-Barang (Laidig and Maingak 1999:80). Finally, as will be seen in greater detail below, geminate consonants are found in Talaud, and in a number of the South Sulawesi languages.

4.1.7 The Lesser Sundas

The Lesser Sundas as defined here include all islands from Sumbawa through Timor (but not Wetar). The languages of Timor form a reasonably clearcut subgroup, to which Rotinese also belongs, and this is reflected in a relatively uniform typology. West of Timor, however, there appears to be considerable genetic diversity, and the typology is equally variable.

The largest segment inventory known in the Lesser Sundas is that of Waima'a (Timor), with 31 consonants and 5 vowels. Other relatively large phoneme inventories are concentrated near the western end of the Lesser Sunda chain, as with Komodo (29C, 6V), Manggarai of west Flores (26C, 6V), and Bimanese of eastern Sumbawa (24C, 5V). In each of these languages, as with Komodo, the higher numbers result in part from a series of prenasalised obstruents, and the arguments for including these as unit phonemes will be examined below. The smallest inventory is that of Dawan, also called Timorese or Atoni, spoken at the western tip of Timor, with 12 consonants and seven vowels (the voiced palatal affricate *j* is rare, and its phonemic status is unclear). Most languages of Timor have inventories containing 19-21 phonemes. Some, known only from short wordlists, appear to have five vowels, but this may reflect underanalysis. If it does not, then some of the languages of Timor may have just 17 phonemes.

Table 4.13 Phoneme inventories of Waima'a and Dawan

Waima'a (Belo, Bowden, Hajek and Himmelmann 2005)	**Dawan** (Steinhauer 1993)

(p) t k ʔ
(pʰ) tʰ kʰ
p' t' k'
b d g
(f) s   h
  s'
m n
mʰ nʰ
m' n'
  l
  lʰ
  l'
  r
  r'
w
wʰ
w'
	p t k ʔ
	b (j)
	m n
	f s h
	l

vowels: i, u, e, o, a 28/31 + 5 = 33/36	vowels: i, u, e, ɛ, o, ɔ, a 11/12 + 7 = 18/19

Although Dawan lacks *d*, Steinhauer (1993:131) notes that *t* is dental, while *n* and *l* are alveolar in all positions. Steinhauer posits seven vowels for Atoni, but my (limited) fieldnotes on the Molo dialect suggest that the e/ɛ and o/ɔ difference is conditioned, the tenser vowels occurring before surface vowels or consonants that can metathesise to produce a prevocalic mid-vowel, as in *mese* [mɛ́sɛʔ] 'one' vs. *tenu* [ténu] (but rapid speech [*téun*]) 'three'. Most typological features found in the languages of the Lesser Sundas have already been encountered in other parts of insular Southeast Asia. The two outstanding exceptions are found in the same language. Belo, Bowden, Hajek and Himmelmann (2005) report that Waima'a, spoken on the northeast coast of East Timor, has four series of obstruents: voiceless unaspirated /p t k ʔ/, voiceless aspirated /pʰ tʰ kʰ/, voiceless ejective /p' t' k'/, and voiced plain /b d g/. These are illustrated by the four-way minimal set *kama* 'bed' (from Portuguese) : *kʰama* 'eat already' : *k'ama* 'scratch' : *gama* 'shark'. Contrastive aspiration is rare in AN languages, and contrastive glottalisation has previously been reported only in Yapese of western Micronesia, so the recent report of an ejective series in one of the languages of Timor is surprising. Belo, Bowden, Hajek and Himmelmann describe the ejectives of Waima'a as produced with complete closure of the oral cavity and simultaneous closure and raising of the larynx. They reportedly occur only word-initially before vowels. During work on Dawan or Atoni of west Timor in the early 1970s the writer also recorded an apparent contrast between plain and glottalised consonants, in words such as *mak* 'say' vs. *m'iʔ* 'urine', *neʔ* 'six' vs. *n'eʔu* 'right side', *neke* 'wild kapok tree' vs. *n'uku-n* 'whole arm', and most strikingly in the minimal pair *l'iʔ* 'left side' vs. *liʔ* 'fold; bend the knee'. As in Waima'a, these segments occur only word-initially before a vowel. However, they also appear to be restricted to sonorants. In [ʔnanan] ~ [tnanan] 'inside', variant forms were recorded with a glottalised nasal and a nasal preceded by *t*, and

it is possible that all instances of glottalised consonants observed in Atoni are products of low-level phonetic processes of crasis between a proclitic grammatical marker and a following noun. Hajek and Bowden (2002:223, fn. 3) are also exceptional in noting that in most of the languages of Timor *t* is dental, but *d* is alveolar.

Implosives are fairly common in the western Lesser Sundas, being found e.g. in Bimanese, Komodo, Ngadha, and Hawu. They thus appear to be an areal feature here (and in southeast Sulawesi just north of this area). Verheijen (1967-70) did not record them for the principal dialect of Manggarai in west Flores, but noted (xii) that in West Manggarai and a few other dialects labial and alveolar implosives occur in addition to the other consonants of the standard language. If so, these dialects presumably would have 28 consonants and six vowels, making their segment inventories nearly as large as that of Komodo. In both Bimanese and Ngadha the implosive corresponding to the alveolar series is retracted. For Ngadha Djawanai (1977:10) describes it as retroflex, and in Bima it was recorded by the writer as apico-domal. Hawu, like Lowland Kenyah dialects in northern Sarawak, is typologically unusual in having implosives at labial, alveolar, palatal, and velar places of articulation. In Bimanese implosives never occur after a nasal, a distributional trait that undoubtedly is shared with other languages.

Among the non-implosive stops several languages near the western end of the Lesser Sunda chain as defined here, including Bimanese, Komodo, and Manggarai have a palatal series that includes both voiced and voiceless affricates. Ngadha *t* and *d* are described by Djawanai (1977:10) as differing in place, but surprisingly *d* is dental, and *t* alveolar, an atypical situation also reported for Rotinese (Fox and Grimes 1995:613).

Most languages of the Lesser Sundas appear to have three nasals (labial, alveolar, velar). However, several languages of Timor, including at least Atoni, Tetun, and Kemak, have just *m* and *n*, and Hawu has *m*, *n*, *ñ*, and *ŋ*. Kambera of eastern Sumba is described by Klamer (1998:10) as having nasals *m*, *n*, *ŋ*, and a 'prenasalised semivowel' *ny*. If so, however, this feature would be unique among AN languages, as prenasalisation is otherwise found only with stops, fricatives, and affricates. Since no argument is given for this analysis, and since a phonetic palatal nasal forms part of a series with *y* and *j* (both given as 'alveolar') it seems best to include all three segments in a palatal series.

In general the languages of this area have a five vowel system like that of Dawan (my analysis), or a six vowel system like that of Komodo. Diphthongs are rare, but some languages have developed suprasegmental vowel features such as nasality, reported in Lamaholot where, according to Pampus (1999:25) all six vowels may occur either oral or nasal, length, reported for the non-mid vowels in Kambera (/i u e o a i: u: a:/), and possibly stress, reported for Sawu, where it is normally penultimate, but sometimes final (*méla* 'trace' : *melá:* 'gold, silver'). For the latter Walker (1982:7) suggests underlying forms *mela* and *melaa*, thus allowing a statement of penultimate stress without exception.

Finally, Tetun Dili, the dialect of Tetun spoken in the capitol of East Timor, and long used as the language of the colonial elite, has a far larger consonant inventory than any other language of Timor, although the vowels are unchanged. This is because Portuguese loanwords have been incorporated to such an extent that they have altered the phonology of this sociolinguistically very distinctive dialect.

4.1.8 The Moluccas

Most languages of the Moluccas have few speakers, and are poorly known. Any statement about the size of segment inventories in this region must therefore be understood

as provisional. Keeping this qualification in mind, segment inventories tend to be small, and to show limited variation in size.

The largest segment inventory for any Moluccan language noted to date is that of Buli in south Halmahera (19C, 5V). Maan notes that Buli *j* is found only in loanwords. However, there are a number of loans, and this consonant appears to be fully integrated into the phonological system of the language, filling a gap in the system of consonants that occur in native vocabulary. Several other languages in the Moluccas have 23 phonemes (North Babar of the Babar Archipelago, Larike of Ambon Island, Taba of south Halmahera), and the figures for most of these will vary slightly depending on whether or not loan phonemes are regarded as fully integrated into the phonological system. The smallest inventory apparently is that of Nuaulu in Seram (11C, 5V). In general the languages of southern Halmahera have larger phoneme inventories than those in the central and southern Moluccas, in part because they have a palatal series and a larger inventory of fricatives than most other languages of the region.

Table 4.14 Phoneme inventories of Buli and Nuaulu

Buli (Maan 1951)					**Nuaulu** (Bolton 1989)			
p	t	c	k		p	t	k	
b	d	(j)	g		m	n		
m	n	ñ	ŋ			s		h
f	s			h	l			
	l				r			
	r				w	y		
w		y						
vowels: i, u, e, o, a					vowels: i, u, e, o, a			
19 + 5 = 24					11 + 5 = 16			

The following observations are noteworthy regarding the content of segment inventories in the Moluccas. First, several of the languages of southern Halmahera have a palatal series, although it may be defective (Taba has *c* and *j,* but no palatal nasal; Buli has *c* and *ñ* in native words, and has added *j* in loans). Second, several languages of the Moluccas have a single labiovelar consonant, k^w. This is true of Dobel, and perhaps other languages of the Aru Islands (southern Moluccas), and of Alune, and perhaps other languages of Seram (central Moluccas). Yamdena, of the southern Moluccas has a typologically odd stop system, with /p t k b d/, and a voiced velar stop. Instead of the expected /g/, however, it has /gb/, described as a 'double stop'—evidently a co-articulated labiovelar (Mettler and Mettler 1990). A second asymmetry in the Yamdena consonant system is seen with the 'nasalised stops' /mp/ and /nd/. This proposed series is odd because it is defective in two ways: it contains no velar member, and it selects one voiced and one voiceless member from the pre-velar stops. Finally, several of the languages of the Aru Islands in the southern Moluccas have a voiceless bilabial fricative [ɸ], a segment that is otherwise rather uncommon in AN languages.

Nearly all languages of the Moluccas have five vowels. The known exceptions are found mostly in the southwestern Moluccas, as with Letinese /i u e o ɛ ɔ a/, or North Babar /i u e ə o ɛ a/. Contrastive stress, length, and nasality are rare, but at least two languages are tonal. According to Remijsen (2001) Maˈya and Matbat, spoken on the island of Misool in the Raja Ampat group have true lexical tone, not pitch-accent, as in a number of the

Papuan languages of neighbouring areas in New Guinea. He describes Ma'ya as having three contrastive tones (rising, falling, level), and Matbat as having five (extra high falling, high level, low rising, low level, falling). In Matbat most lexical bases are monosyllabic. The suprasegmental phonology of Ma'ya is highly unusual in having both lexical stress and three tonemes.

4.1.9 New Guinea

It is difficult to generalise about the AN languages of New Guinea, as large areas of the island are still poorly known. Thanks primarily to Ross (1988), many previous gaps have been filled, and basic typological and historical information is now available for most languages of the eastern (Papua New Guinea) half of the island. The AN languages of Irian Jaya remain seriously underdescribed.

For many languages phoneme inventories must be extracted from comparative word-lists, or from statements of historical development. This is a tedious process, and oversights are to be expected. Nonetheless, the largest phoneme inventory found so far for any AN language of New Guinea is that of Bukawa, spoken in the Huon Gulf on the northeast coast of the island, and described by Ross (1993) as having 30 consonants, seven vowels and two contrastive tones. Its close relative Yabem (also tonal) has a segment inventory that is only slightly smaller. The smallest inventory is very difficult to determine. The best candidate for this distinction probably is Mekeo (eastern shores of the Gulf of Papua, southeast New Guinea), but the phonological analysis of this language remains unclear. Like virtually all AN languages of the New Guinea area, Mekeo has a five vowel system /i u e o a/. Published information, however, shows disagreements in the number of consonants. Pawley (1975) is concerned primarily with comparative issues, and provides no explicit description of the synchronic phonologies of the languages he compares, but an examination of his data suggests that Mekeo has the ten consonants /p k ʔ g m n ŋ f v l/ for a total inventory of 15 segments. Ross (1988) is likewise concerned with comparative issues. As a result, all information about the synchronic phonologies of the languages he compares must be extracted from the tables of reflexes of Proto Oceanic phonemes. These contain seven consonant phonemes for Mekeo: /p k ʔ f m n l/. The principal difference between the Mekeo consonant inventory implied by Pawley (1975) and that implied by Ross (1988) seems to be that *f/v*, *k/g*, and *n/ŋ* are treated as allophones of the same phoneme in the latter work.

Jones (1998) provides a detailed account of the four recognised dialects of Mekeo. He maintains (1998:14) that "Their consonantal inventories consist of not more than eight phonemes …and, in the case of two varieties, only seven." In his list of consonant phonemes, however, he gives nine segments in Northwest Mekeo (/p (t) k β g m ŋ w y/), ten in West Mekeo (/p (t) k b d̲ g m ŋ l w/), nine in East Mekeo (/p (t) k ʔ/Ø f s̲ m ŋ l/), and eight in North Mekeo (/b (t) g m ŋ v z̲ l/). He tries to reconcile these numbers with his statement by treating the parenthesised voiceless dental stop as extrasystematic, although this still leaves nine consonants in West Mekeo, and leaves a number of attested examples of *t* ([ts] before *i*) in North Mekeo unexplained. Underlined segments, which result from the fortition of transitional glides by some speakers and so constitute a change in progress, are described as 'intrusive', but counted as full phonemes. Jones provides a 200-word vocabulary of all four Mekeo dialects in an appendix to his book, and the orthography used there, unfortunately, is inconsistent with the earlier description that he gives of the phonemic analysis of this language. Given these various uncertainties the Mekeo material

should be treated with caution, but it seems safe to infer that the North dialect of Mekeo has just eight consonant and five vowel phonemes. This makes it the same size as the better known Hawaiian inventory of eight consonants and five vowels, with the difference that Hawaiian has phonemic vowel length, while Mekeo does not. If Jones is justified in excluding /t/ from the North Mekeo phoneme inventory this dialect will usurp the distinction, long held by Hawaiian and several other Eastern Polynesian languages (cf. Table 4.1), of having the smallest segment inventory known for any AN language, and will have one of the three smallest inventories reported for any language. The segment inventories of Bukawa and North Mekeo are given in Table 4.15. Jones (1998:14) gives *p, (t), k* as *b, (t), g*, but elsewhere (559) he writes the last of these as *k*. His *b* is described as "an unaspirated bilabial stop with late voice onset," and *k* as "an unaspirated velar stop with early voice onset" (Jones 1998:559). I follow Ross (1988) in writing only voiceless stops:[25]

Table 4.15 Phoneme inventories for Bukawa and North Mekeo

Bukawa (Ross 1993)				**North Mekeo** (Jones 1998, with changes)		
p	t	k	ʔ	p	(t)	k
p^w		k^w		m		ŋ
b	d	g		v	z	
b^w		g^w			l	
^mb	nd	^ŋg				
$^mb^w$		$^ŋg^w$				
m	n	ñ	ŋ			
m^w						
	s		h			
	^ns					
	l					
	l^h					
w	y					
w^h		y^h				
vowels: i, u, e, o, ɛ, ɔ, a				vowels: i, u, e, o, a		
30 + 7 = 37				8 + 5 = 13		

New Guinea is a large island with an extraordinary number of both Papuan and AN languages, and there is thus considerable typological variation in this region. The genetic boundary between South Halmahera-West New Guinea languages and Oceanic languages falls around the mouth of the Mamberamo River in Irian Jaya, and as one moves east from that point certain typological features begin to appear, or to become more frequent in the phonology. The most obvious of these are the labiovelar consonants, which are quite common in the Oceanic languages of western Melanesia, but are rather rare elsewhere. As seen already, *g^w* occasionally turns up in the languages of Borneo, and *k^w* is found in

---

[25]   As Blevins (2009) points out, the Northwest dialect of Mekeo, which Jones reports with nine consonants can be seen as typologically anomalous, since it lacks unambiguous coronal phonemes in the mainstream lexicon (although it has coronal allophones). This interpretation requires that /t/, described by Jones (1996:16) as part of "the special register of baby talk and the 'exotic' stop, found in certain words, not all loanwords" be treated as extra-systematic, and what Jones (1996:14) describes as y/e be treated as non-coronal.

several languages of the central and southern Moluccas. However, all of these non-Oceanic languages have a single labiovelar phoneme. Many of the Oceanic languages of western Melanesia, by contrast, have three labiovelars, some have four, and a few have five, as with Kilivila, spoken in the Trobriand Islands (Senft 1986). As seen above, Yabem and Bukawa have expanded the inventory of labiovelar consonants even further through the addition of prenasalised members of the series:

Table 4.16 The Phoneme inventory of Kilivila, with five labiovelar consonants

p	t	k
p^w		k^w
b	d	g
b^w		g^w
m	n	
m^w		
	s	
v		
	l	
	r	
w		y

vowels: i, u, e, o, a

Vowel systems in most AN languages are remarkably stable in comparison with even closely related Indo-European languages such as English, Dutch, and German. In the AN languages of New Guinea this stability is extreme: with few exceptions the system of vowels is /i u e o a/. Suprasegmental features are rarely phonemic. The only noteworthy exception to this statement is that tone has developed in two discontinuous areas. Mor, a South Halmahera-West New Guinea language of Irian Jaya, is reported as having a two-tone system (high vs. low), and possible stress contrasts (Laycock 1978:290). If so it parallels phonological developments in Maʼya of the Raja Ampat Islands as reported by Remijsen (2001). As already noted in passing, a two-way tonal contrast is also found in Yabem and Bukawa, spoken on the northeast coast of New Guinea. In both Moor and Yabem-Bukawa tonogenesis is associated with the development of a monosyllabic base canon.

4.1.10 The Bismarck Archipelago

The languages of the Bismarck Archipelago are spread over several islands or island groups that are not widely separated, but which nonetheless reflect a history of considerable isolation in their linguistic development. The most notable exception to this statement is seen in southern New Ireland-northeast New Britain, where there has been clear linguistic migration and diffusion between adjacent parts of these two islands. As a result of the general isolation between the three main island groups (the Admiralties, New Ireland-St. Matthias, New Britain-French Islands) certain aspects of the phonological typology varies strikingly from one island cluster to the next. The largest phoneme inventory known for any of the languages of this area is that of Lindrou, spoken at the western end of Manus Island in the Admiralty group, with 21 consonants and five vowels.

The smallest inventory is shared by Kilenge and Amara of west New Britain, which have identical sets of 11 consonants and five vowels. It is represented here by Kilenge.

Table 4.17 Phoneme inventories of Lindrou and Kilenge/Amara

Lindrou (Blust n.d. (b))					**Kilenge** (Goulden 1996)		
p	t		k	ʔ	p	t	k
b	d	j	g		m	n	ŋ
bʷ			gʷ			s	
	dr				v		ɣ
m	n	ñ			l		
mʷ					r		
	s			h			
	l						
	r						
w		y					
vowels: i, u, e, o, a					vowels: i, u, e, o, a		
21 + 5 = 26					11 + 5 = 16		

Although Maddieson (1984) records a number of languages with labial, alveolar and palatal nasals but no velar nasal (e.g. Spanish), the Lindrou system is highly unusual for AN, where /m n ñ/ almost invariably imply the presence of a velar nasal.

The languages of the Bismarck Archipelago present several typological features that are rare in languages generally. Many languages of the island of Manus, especially those in the center and east of the island, have prenasalised bilabial and alveolar trills, written here as *br* and *dr*. Maddieson (1989a) has described the mechanism through which these sounds arise. The alveolar trill may precede any vowel, but the bilabial trill precedes only rounded vowels, particularly *u*, as in Nali (eastern Manus) *draye-n* 'his/her blood', *dret* 'green frog', *drihow* 'first-born child', *drow* 'hardwood tree: *Intsia bijuga*', *drui-n* 'its bone (fish)', but *brupa-n* 'his/her thigh', *brusas* 'foam', *kabrow* 'spider, spiderweb', Ahus *bro-n* 'his penis/her vulva', *brokop* 'hermit crab'. In some languages both trills occur in the same morpheme, as in Leipon *brudr* 'banana', and in others a prenasalised trill may follow another nasal consonant, as in Nali *dromdriw* 'sword grass, *Imperata cylindrica*'. Rarely, the bilabial trill is found before front vowels, as in Leipon *bri-n* 'her vulva', or *brekop* 'hermit crab,' but where etymologies are available these derive from *u (Proto Admiralty *bui- 'vulva, vagina', *bua-kobV 'hermit crab'). This observation suggests that bilabial trills arise only in environments in which they are phonetically motivated, but once they exist they can be maintained in other environments. Bilabial trills are also phonemic in some of the languages of Vanuatu (Blust 2007a, Crowley 2006a), and occur subphonemically in languages such as Galeya (Ferguson Island), which shows optional allophonic trilling of *b* before a high back rounded vowel.

Four other rare typological features are found in a single language of the Admiralty Islands. The first of these is phonemic vowel nasality, which appears in Seimat, spoken in the extensive Ninigo Lagoon to the west of the main island of Manus. All five Seimat vowels /i u e o a/ may be nasalised, but nasal vowels occur only after *h* and *w*, and there are some indications that the locus of nasality is in the consonants rather than the vowels that follow them (Blust 1998a). The second unusual feature that appears to be confined to one language in the Admiralties is phonemic aspiration. Drehet, spoken in western Manus,

has the unusual obstruent system /p t c k kʰ pʷ dr/, with a single aspirated stop. This segment developed from *dr*, and *dr* was then borrowed back into the language, probably from neighbouring and closely related Levei. The third atypical feature in the phonologies of Admiralty languages is a retroflexed and partly affricated voiceless stop which occurs in Lenkau, spoken on the island of Rambutjo in the southeastern Admiralty Islands. This segment (here written *tr*) occurs both syllable-initially and syllable-finally, and corresponds either to *dr* or to *t* in the languages of Manus: Nali *dria-n*, Lenkau *tria-n* 'his/her abdomen', Nali *draye-n*, Lenkau *troh-i* 'his/her blood', Nali *dras*, Lenkau *tres* 'sea, saltwater', Nali *drow*, Lenkau *troh* 'hardwood tree: *Instia bijuga*', Nali *nayat*, Lenkau *lalatr* 'stinging nettle', Nali *ma-saŋat*, Lenkau *soŋotr* 'hundred', Nali *kut*, Lenkau *kutr* 'louse'. Fourth, Wuvulu, spoken on an atoll well to the west of the main island of Manus, has an interdental lateral, written in German sources from the early twentieth century as *dl*. This segment is similar to the voiced interdental fricative of English, but is laterally released. In addition, citing earlier literature that had been largely overlooked in more recent discussions, Hajek (1995) has drawn attention to the presence of two-tone systems in several languages of New Ireland, including Kara, Barok and perhaps some dialects of Patpatar. Nothing is known to date about the history of this development.

Other noteworthy phonological features of this area are the presence of palatal stops and nasals in a number of the languages of Manus and some southeastern Admiralty Islands, such as Nauna, and the complete absence of labiovelars in New Ireland and New Britain, the western islands of the Bismarck Archipelago (Wuvulu-Aua, Kaniet, Seimat), and the languages of the southeastern Admiralties. In many of the languages of Manus consonant gemination is phonemic, but no adequate published description is yet available for any of these. Some of the languages of the Admiralty Islands, such as Loniu (Hamel 1994) have seven vowel phonemes, and some of the languages of New Ireland, as Madak, may have as many as eight. Finally, stress appears *a priori* to be contrastive in Wuvulu and Lindrou, but in both cases it is better explained in other ways. Wuvulu shows phonetic contrasts such as [gúfu] 'island' : [gufú] 'kinsman', but closer attention to morphology shows that the latter is /kufu-u/ 'my kinsman', and a rule of rightward stress shift together with the contraction of morphologically-derived sequences of identical vowels accounts for the surface contrasts. In Lindrou some words were recorded with penultimate, and others with final stress, as [kápak] 'my cheek' vs. [madák] 'my eye'. In many cases, however, a stressed penultimate vowel appears to be followed by a geminated consonant, and differences of stress may be epiphenomena of consonant length.

4.1.11 The Solomon Islands

The AN languages of the Solomon Islands present few typologically noteworthy traits in phonology. A linguistic survey of the nation of Solomon Islands, which excludes Bougainville-Buka at the western end of the Solomons chain, was conducted by Tryon and Hackman (1983), who provide a vocabulary of 324 items and reflexes of Proto Oceanic phonemes for 111 languages. This material can be mined for information about largest and smallest segment inventories, although the inventories themselves can more reliably be extracted from descriptive sources.

The largest phoneme inventory found for any language in the Solomons, including the Santa Cruz Archipelago to the southeast of the main Solomons chain, is that of Cheke Holo (Santa Isabel, central Solomons), with 32 consonants and five vowels. The smallest

inventory is more difficult to determine, but a good candidate is 'Āre'āre (Malaita, southeast Solomons), with 10 consonants and five vowels.

Table 4.18 Phoneme inventories of Cheke Holo and 'Āre'āre

Cheke Holo (White et al 1988)	**'Āre'āre** (Geerts 1970)
p t c k ʔ	p t k ʔ
pʰ tʰ kʰ	m n
b d j g	s h
m n ñ ŋ	r
hm hn hñ hŋ	w
f s h	
v z ɣ	
ɣʰ	
l	
lʰ	
r	
rʰ	
w	
vowels: i, u, e, o, a	vowels: i, u, e, o, a
32 + 5 = 37	10 + 5 = 15

The phoneme inventory of Cheke Holo (= Maringe) is not only large, it also contains some of the more unusual segments found in languages of the Solomons. The great majority of Solomons languages have 15-17 consonants and the stable five vowel system seen in both of the above languages. Most of the larger consonant inventories, and those with the most unusual segments are found in languages of Santa Isabel, and perhaps Choiseul (for which little reliable data is available). As will be seen below, the most elaborate vowel inventories are found in languages of the Santa Cruz Islands.

Most languages of the Solomons have matching series of voiced and voiceless stops at labial, alveolar and velar places. In some there is a gap for the voiceless labial, and an unmatched glottal stop (as in Lau, Kwaio /t k ʔ b d g/. Cheke Holo, Bugotu, and a few other languages of the central and western Solomons have a palatal series, but Cheke Holo is unusual in having further developed a plain/aspirated contrast for voiceless stops, and a voiced/voiceless contrast for nasals and liquids. It is, moreover, highly unusual in having an aspirated voiced velar fricative. Whether these phonetically complex segments should be analyzed as unit phonemes, as White, Kokhonigita and Pulomana (1988) propose, or as consonant clusters is an open question, as many clusters are permitted in the language in both initial and medial positions.

Several languages in the central and western Solomons have more fricatives than is usual for an Oceanic language. Roviana, of the New Georgia Archipelago, for example, has /s h v z ɣ/ while Bugotu of Santa Isabel Island has /f s h v ɣ/, with [ð] as an allophone of /l/ before non-high vowels. Sa'a, which has no other voiced stops, has a phoneme *d* that Ivens (1929:43) described as an apparent trill, 'equivalent to *dr*'. In the 1920s it had a voiceless allophone [ʧ] before *i*, but half a century later Tryon and Hackman (1983) recorded it as [ʧ] in all environments. It thus apparently has followed the same phonetic trajectory as *dr* in languages of the Admiralty Islands, such as Levei, Likum, and Pelipowai, which also show the change *dr > [ʧ]. Finally, while labiovelars are common in

the southeast Solomons they tend to favor primary velar consonants in the languages of northern Malaita, and primary labial consonants in the languages of southern Malaita, as seen in the north Malaita language Kwaio, with /k^w g^w ŋ^w/, as against the south Malaita languages Sa'a, with /p^w m^w/ or Arosi of San Cristobal (Makira) Island, with /p^w b^w m^w/.

In view of the small phoneme inventories of most Polynesian languages, it is surprising that Pileni, a Polynesian Outlier language in the Santa Cruz Islands, has one of the larger phoneme inventories in the Solomons. Elbert (1965) provided preliminary data on this language, but more detailed information is now available. Hovdhaugen, Næss and Hoëm (2002) describe Pileni with the usual five Polynesian vowels, but a consonant inventory that has been expanded through the introduction of an aspirated/non-aspirated contrast in the voiceless stops, and a voiced/voiceless contrast in the sonorants. These phonemic expansions almost certainly are due to contact with the typologically aberrant non-Polynesian languages of the Santa Cruz Islands. Since Pileni also has a few words with consonant clusters, however, there is some question as to whether the aspirated stops and voiceless sonorants are best analyzed as clusters or as unit phonemes:

Table 4.19   The consonant inventory of Pileni (Santa Cruz Islands)

p	t	k		
p^h		t^h	k^h	
(b)		(d)	g	
m	n			
m^h		n^h		
f	(s)			h
	l			
	l^h			
v				

As already noted, most languages of the Solomons retain the Proto Oceanic system of five vowels unchanged. This is true even for languages such as Cheke Holo, which have greatly expanded the inventory of consonants. By contrast, Lödai of the Santa Cruz Islands is described by Lincoln (1978) as having ten vowels that occur either oral or nasalised. However, this language is highly aberrant in other respects, with little lexical evidence of AN affinity, and there has consequently been disagreement regarding its classification as AN or Papuan (Wurm 1978).

4.1.12 Vanuatu

The languages of Vanuatu have traditionally been divided into those of the far south (Erromango, Tanna, Anejom), which were regarded as 'aberrant', vs. the rest. The bulk of information on these languages has been published since 1980, largely through the efforts of the Australian linguists John Lynch and Terry Crowley. For information on most languages, however, it is still necessary to consult the survey data in Tryon (1976), which provides a vocabulary of 292 words for 179 language communities.

Most languages in Vanuatu have between 17 and 21 phonemes. The largest inventory identified to date is that of Lonwolwol, spoken on Ambrym Island, with 25 consonants and 13 vowels. The smallest inventory apparently is that of Matae/Navut, spoken on the island of Espiritu Santo, with 11 consonants and five vowels:

Table 4.20 Phoneme inventories for Lonwolwol and Matae

Lonwolwol (Paton 1973)					**Matae** (Tryon 1976)		
p	t		k	ʔ	p	t	k
	tˢ	c				ts	
b	d		g		m	n	
bᵛ	dᶻ					s	
bʷ					v		
m	n		ŋ		l		
f	s	ʃ		h	r		
v					w		
		l					
		ɾ					
		r					
w			y				
vowels: i, ɪ, e, ɛ, a, ɑ, u, ʊ, o, ɔ, ø, ə, ʌ					vowels: i, u, e, o, a		
25 + 13 = 38					11 + 5 = 16		

Paton, who was a missionary for fifteen years among the people of Ambrym and who only received formal training in linguistics after his departure from the island, reports far more vowels for this language than is typical for the region. Those that might need special explanation are: ɪ = 'lower-high front unrounded vowel', ɛ = 'open, mid-front unrounded vowel', a = 'low-front unrounded vowel', ɑ = 'low-back unrounded vowel' (written as /a/ with a raised dot to the right), ʊ = 'lower-high back rounded vowel', ɔ = 'lower mid-back rounded vowel', ø = 'higher mid-central unrounded vowel', ə = 'mid-central unrouded vowel', ʌ = 'lower-mid back unrounded vowel.' He gives two vowel symbols, ʊ and ɔː with the same definition, and I have combined them here. Despite the questions that inevitably arise in connection with an analysis such as this, it is only fair to point out that Guy (1974) reported a twelve vowel system for Sakao, spoken on a small island of the same name off the southeastern coast of Espiritu Santo. It thus does appear that some languages of central Vanuatu have exceptionally large vowel inventories. A number of other languages in central Vanuatu have phoneme inventories that are nearly as small as that of Matae. In almost all of these *n and *ŋ have merged as *n*, and *m and *mʷ have merged as *m*, leaving just the two nasals *m* and *n*.

Many phonological traits in Vanuatu languages are familiar from a general Oceanic perspective, but some are highly distinctive. Without question the most remarkable phonological trait in Vanuatu is the presence of a series of apicolabial, or linguo-labial consonants in a number of the languages of north-central Vanuatu. Maximally these include a stop, a nasal and a voiced fricative, which Tryon (1976) writes /p̼ m̼ v̼/, as in the Mafea words /ap̼a/- 'wing', /m̼em̼e/- 'tongue', or /v̼itu/ 'moon'. In nearly all languages these appear to be allophones of /p m v/, the labials occurring before rounded vowels and the linguo-labials elsewhere. François (2002:15) describes the linguo-labial consonants of Araki as 'pronounced with the tip of the tongue touching the middle of the upper lip,' an articulation that is easy to learn, but surprisingly rare in the world's languages. These unusual articulations in the languages of north-central Vanuatu have been described in detail by Maddieson (1989b), who uses the label 'linguo-labial' rather than 'apicolabial.' A similar series is reported for the Macro-Ge language Umotina, spoken in the Mato Grosso region of west-central Brazil, but is otherwise unknown outside north-central

Vanuatu (Ladefoged and Maddieson 1996:18). Linguo-labials are reported in eight languages of Vanuatu which are distributed along the southern coast of the island of Espiritu Santo and the northern coast of the island of Malakula, and it is clear that this is an areal feature (Lynch and Brotchie 2010). Attention to the historical phonology also makes it clear that linguo-labial consonants once had a wider geographical distribution. At least thirteen languages, including some in the north of Espiritu Santo (Tolomako and Sakao) reflect *m as *n* before non-round vowels, but as *m* elsewhere, and *p as *ð* or *θ* before non-round vowels, but as *v* elsewhere, showing that the linguo-labial nasal and fricative, which arose from labials, have further evolved into alveolar or interdental consonants.

The preceding developments reflect another typological feature of a number of languages in Vanuatu: the presence of voiced and voiceless interdental fricatives. A third feature of typology that is unusual for AN languages is a contrast between two rhotics, one flapped and the other trilled. This contrast is found in Sakao and Araki of Espiritu Santo, and in Lonwolwol of central Ambrym. In many of the languages of western Melanesia the labiovelars /mʷ/, /pʷ/, /kʷ/ appear to be primary labial or velar articulations that are labialised—that is, they are pronounced with distinct lip rounding. In some, perhaps most of the languages of Vanuatu the corresponding segments are co-articulated nasals or stops /ŋmʷ/, /kpʷ/ that are velarised—that is, pronounced with lip spreading. Schütz (1969b:15) describes Nguna *p̃* as "a bilabial lenis implosive stop with varying degrees of labialisation." Ladefoged and Maddieson (1996:82ff) indicate that implosive stops may be either voiced or voiceless, but that voiceless implosives are so rare that they were long considered to be nonexistent. The probability that Nguna *p̃* is a true implosive stop is thus fairly low. Co-articulated consonants such as *ŋmʷ* or *kpʷ* are sometimes described as 'suction stops' since they appear to involve a mildly ingressive airstream, but there apparently is no significant lowering of the larynx in producing them, which is sharply different from true implosives.

Other noteworthy phonetic or phonological elements are a preploded velar lateral in Hiw of the Torres Islands (François 2010), a retroflex consonant *ʈ* that developed from *d ([nd]) in Mafea, spoken on a small island of the same name just "offshore from southeastern Espiritu Santo to the north of Tutuba" (Lynch and Crowley 2001:47), bilabial and velar fricatives in a number of languages, and a palatal series in at least Lonwolwol and Anejom, the former language having a voiceless palatal fricative, a segment that is common in many parts of the world, but not in AN languages. Finally, although he provides no phonetic descriptions, Tryon (1976:11ff) lists *dr*, *nr*, and *ndr* as reflexes of POC *nt and *d in more than 30 languages of Vanuatu, all of which evidently represent a prenasalised alveolar trill. This is confirmed by Crowley (2006a:25), who notes the presence of prenasalised trills at both bilabial and alveolar positions in Avava of central Malakula.

4.1.13 New Caledonia and the Loyalty Islands

Nemi, with 43 consonants and five vowels (which occur both oral and nasalised) has the largest segment inventory in New Caledonia or the Loyalty Islands. As noted earlier, this is also the largest segment inventory known for any AN language. The smallest segment inventory known for any language of this region is that of Cèmuhî, with 19 consonants and seven vowels which occur both oral and nasalised, and may carry any of three contrasting tones. These two languages are spoken in nearly adjoining territories on the east coast of the north-central New Caledonian mainland, separated only by speakers of Pije:

Table 4.21 Phoneme inventories for Nemi and Cèmuhî

Nemi (Haudricourt 1971)				**Cèmuhî** (Rivierre 1994)			
p	t	c	k	p	t	c	k
p^h	t^h		k^h	p^w			
p^w				b	d	j	g
p^{wh}				b^w			
b	d	j	g	m	n	ñ	ŋ
b^w				m^w			
m	n	ñ	ŋ		l		
hm	hn	hñ	hŋ				h̃
hm^w				h̃			
pm	tn	cñ	kŋ	hw̃			
pm^w							
m^w							
f	s		x	h			
v			ɣ				
	l						
	hl						
	r						
	rh						
w		y					
hw	hy						
vowels: i, u, e, o, a + nasality				vowels: i, e, ɛ, u, o, ɔ, a + tone			
43 + 5 = 48				19 + 7 = 26			

Languages of this area are typologically unusual in having contrastive aspiration, which Nemi applies even to the voiceless labiovelar stop, and in having voiceless sonorants and postnasalised stops (*pm, tn*, etc.; for the latter see Haudricourt 1964, and Ozanne-Rivierre 1975, 1995).[26] As in most other Oceanic languages, voiced obstruents are automatically prenasalised. This is true, for example, of all of the voiced stops of Nemi (/b d j g b^w/). Since the smallest New Caledonian inventory noted so far is larger than the average for AN languages as a whole (which typically fall in the range of 19-25 segments), it is clear that the languages of New Caledonia and the Loyalty Islands have exceptionally large segment inventories. Table 4.22 provides an overview to make this point (phonemes found only in recent loanwords are not counted):

---

[26]   Despite their superficial similarity to the preploded final nasals of western Indonesia, the postnasalised stops of northern New Caledonian languages have a completely different origin, arising from the loss of an unstressed vowel between a word-initial stop and the nasal onset of a following syllable, resulting in what appears to be a consonant cluster, rather than a unit phoneme (or allophone of a unit phoneme).

Table 4.22 size of phoneme inventories in New Caledonia and the Loyalty Islands

Language	Consonants	Vowels	Total
Nemi	43	5 (+ nasal)	48
Gomen	38	7	45
Nêlêmwa	35	7 (+ 5 nasal)	42
Iaai	32	10	42
Jawe	34	7 (+ long + 5 nasal)	41
Pije	36	5 (+ 5 nasal)	41
Nyelâyu	35	5 (+ long + nasal)	40
Tinrin	30	8 (+ 6 nasal)	38
Xârâcùù	26	10 (+ 7 nasal)	36
Nengone	29	7 (+ long)	36
Goro	25	10 (+ short + nasal + 2 tones)	35
Dehu	27	7 (+ long)	34
Unya	25	7 (+ long + 5 nasal + 2 tones)	32
Paicî	18	10 (+ 7 nasal)	28
Cèmuhî	19	7 (+ nasal + 3 tones)	26

The foregoing listing is by no means complete, but neither is it selective. Rather, it represents the languages for which adequate information is available, and this clearly has a random association with number of phonemes. For both New Caledonia and the Loyalties, then, segment inventories are exceptionally large by AN standards. In this sample of 15 languages there are 564 segmental phonemes, or an average of 37.6 per language. However, many of these languages also have phonemic nasality, phonemic length, or both in the vowel system, allowing for even greater possibilities of contrast. Goro, in the extreme south of New Caledonia, as described by Rivierre (1973:89) is typologically unusual in that the system of long oral vowels (10) is larger than the system of short oral vowels (7), and the system of long nasal vowels (7) is larger than the system of short nasal vowels (5). Despite a predominance of long vowels, however, the vowel system of Goro does not violate the language universal first proposed by Ferguson (1963) that the number of nasal vowels will never exceed the number of oral vowels.

Most strikingly, several of the languages of New Caledonia are tonal. Both Unya and Goro of the far south have contrasting high and low tones (these are also found in the language of the Isle of Pines, which reportedly is a dialect of Goro). Cèmuhî, spoken on the northeast coast of New Caledonia, is described as having a three-way contrast between high, mid, and low tones. Unlike other AN tone languages, which arguably acquired pitch contrasts through contact with tone languages belonging to other families, the evolution of tone in New Caledonia appears to have been purely system-internal.

4.1.14 Micronesia

Unlike Polynesia, which has a high degree of linguistic and cultural homogeneity, Melanesia and Micronesia are residual categories. However, these regions differ in fundamental respects. Whereas Melanesia contains both AN and non-AN ('Papuan') languages, all languages in Micronesia are AN. In this sense they are more uniform than those of Melanesia. However, if we consider only the AN languages the situation is reversed. All AN languages of Melanesia are either Oceanic or South Halmahera-West

New Guinea, hence Eastern Malayo-Polynesian and so part of a single, continuous eastward and southward expansion of AN speakers into the Pacific. By contrast, the languages of Micronesia are a heterogeneous collection that have reached their historical locations as the result of several quite distinct migrations. The largest category, sometimes called 'Nuclear Micronesian' (NMC), contains a number of Oceanic languages that clearly subgroup together, and probably expanded into eastern Micronesia from some location in Melanesia, most likely the southeast Solomons. Nauruan is the most divergent and least well-described of these, but almost certainly belongs to the NMC group. Yapese, spoken in western Micronesia, also is an Oceanic language, but one that has no close subgroup relatives (Ross 1996c). Chamorro and Palauan are non-Oceanic languages that are not closely related to one another, or to any other language. Given the heterogeneous AN origins of the languages of Micronesia we would not expect the typology of this area to be uniform, and it is not.

The Micronesian language with the largest phoneme inventory is Kosraean (35 consonants, 12 vowels. The smallest inventory appears to belong to Gilbertese (I-Kiribati), with 10 consonants and five vowels. (Kosraean *sr* is a palatal fricative; superscript *w* represents velarisation, and superscript *o* marks rounding):

Table 4.23 Phoneme inventories for Kosraean and Gilbertese

Kosraean (Lee 1975)				**Gilbertese** (Sabatier 1971)		
p	t		k		t	k
p^w	t^w		k^w	b		
	t^o		k^o	b^w		
m	n		ŋ	m	n	ŋ
m^w	n^w		$ŋ^w$	m^w		
	n^o		$ŋ^o$	w	r	
f	s	sr				
f^w	s^w	sr^w				
	s^o	sr^o				
	l	r				
	l^w	r^w				
	l^o	r^o				
	y					
w^w	y^w					
w^o	y^o					
vowels: i, ɨ, u, e, ə, o, ɛ, ʌ, ɔ, æ, ɑ, oa				vowels: i, u, e, o, a		
35 + 12 = 47				10 + 5 = 15		

Kosraean provides a useful starting point for discussing Nuclear Micronesian phoneme inventories. Consonant places are bilabial, labiodental, dental, retroflex, and velar, and stops, nasals, fricatives and liquids all come in plain, velarised (lip spread) and labialised (lip rounded) varieties. All consonants may be velarised, but only non-labial consonants may be labialised. This permits a two-way contrast between plain and velarised labials, but a three-way contrast between plain, velarised, and labialised consonants for the rest. A similar set of distinctions was proposed by Bender (1968) for Marshallese, where consonants are assigned to plain, palatalised, velarised, and labialised types. Another feature of Kosraean that is found elsewhere in Micronesia but is rare in other AN

languages, is a retroflex series of consonants (/sr sr^w sr^o r r^w r^o/). Other NMC languages that have at least one retroflex consonant include Pohnpeian, Puluwat, and Woleaian. Most NMC languages allow consonants to be geminated both word-initially and word-medially. Some languages, such as Pohnpeian, Puluwat, or Woleaian also have underlying word-final geminates. In at least the Chuukic languages these surface as single consonants in final position, but as geminates before a vowel-initial suffix, as with Puluwat *kacc* 'good' ([kats]), but *kacc-ún ŕáán* 'good day' ([*kats:-ɨn ræ:n*]).

One of the most distinctive traits of the phonological typology of NMC languages is the pervasive manner in which the features of vowels have become interdependent with those of adjacent consonants. Historically, vowel features have 'smeared' onto adjacent consonants, as in POC *Rumaq > PMC *im^{w}a/um^{w}a 'house', where the rounding of the vowel has been transferred to the labial consonant in most attested languages. While transfers of rounding from vowels to consonants occur in other Oceanic languages (as in Vanuatu), this generally is an isolated phenomenon. In Micronesia the interpenetration of vowel and consonant features is far more pervasive, and has undergone reanalysis so that synchronically it is sometimes simpler to argue that a transfer of phonological features is occurring not from vowels to consonants, but rather from consonants to vowels.

The vowel systems of Nuclear Micronesian languages are typically larger than the Oceanic five-vowel norm, but surface vowels must be distinguished from underlying vowels. Marshallese, for instances, has twelve surface vowels, but Bender (1968) has shown how these reduce to four underlying vowels distinguished only by height, with palatal or velar colouration introduced by the qualities of adjacent consonants. In other NMC languages relatively large numbers of surface vowels appear to correspond to an equal number of phonemic vowels, as with the nine-vowel systems of Chuukese or Puluwat. Lee and Wang (1984) state that the 12 surface vowels of Kosraean almost certainly are not phonemic, but they are unable to provide the vowel phonemes, and so fall back on the surface vowels as a default phonemic analysis. Given the size of its consonant inventory, however, Kosraean probably would still have the largest segment inventory of any Micronesian language, even if the vowel contrasts were reduced to four. The largest number of phonetic vowels claimed for any Nuclear Micronesian language probably is the 14 vowels of Sonsorol. Capell (1969) implies, but does not defend the proposal that these are phonemes.

Other noteworthy features in the phonology of NMC languages include the presence of a palatal series (Mokilese, Woleaian), and the presence of two rhotics (Puluwat, where the contrast is between a double-tap trill and a retroflex continuant, and Nauruan, where the contrast is between an alveolar flap or trill conditioned by stress, and a partially devoiced fortis rhotic). In addition, NMC languages have some articulations that are not widely reported elsewhere, such as the voiceless velarised bilabial fricative of Woleaian, described by Sohn and Tawerilmang (1976:xiii) as a 'candle blowing sound' made by retracting the tongue while producing bilabial friction.

Outside Nuclear Micronesian Yapese has the largest phoneme inventory for any Micronesian language, with 31 consonants and eight vowels. Yapese is almost unique among AN languages in using phonemic glottalisation for voiceless stops, nasals, fricatives, laterals, and glides, to yield eleven glottalised consonant phonemes /p' t' k' m' n' ŋ' f' θ' l' w' y'/. Yapese vowels occur long or short, and fall into two sets, described by Jensen (1977a) as 'light' or 'palatalising', and 'plain'. Chamorro has a palatal series which consists of a palatal nasal plus alveolar affricates written *y* [dz] and *ch* [ts]; it also has a single labiovelar stop *g^w*. Palauan, as it is usually analyzed, has just ten consonants /t k ʔ b

d [ð] m ŋ s l r/ and six vowels. However, this analysis arbitrarily eliminates glides *w* and *y*, which are conventionally written as vowels, as in *e-uíd* 'seven', *eánged* 'sky' (phonemically *e-wid*, *yaŋd*). With the addition of two glides the Palauan inventory increases to 12 consonants and six vowels. Finally, Palauan is a typological rarity in having no alveolar nasal, due to the historical change *n > *l*.

The phoneme inventory of Nauruan is still imperfectly known, and its position vis-à-vis other Oceanic languages of Micronesia remains unclear. Lynch, Ross and Crowley (2002:890) treat Nauruan and the Nuclear Micronesian Family as coordinate branches of the Micronesian Family, thus suggesting that Nauruan is not a Nuclear Micronesian language, but is the closest external relative of this group. Nathan (1973) posits a system of 24 consonants and six vowels which may occur either short or long, but he stresses that this analysis is preliminary. Kayser (1993), proposes a practical orthography for the language with 13 consonants and five vowels, but provides no phonemic analysis, and gives only superficial and cursory information on phonetics.

4.1.15 Polynesia, Fiji, and Rotuma

The Polynesian languages form a well-defined subgroup that is generally believed to have its closest relationships with Fijian and Rotuman. Within this region of the central and eastern Pacific average segment inventories are considerably smaller than the AN norm. Segment inventories in the languages of Triangle Polynesia do not vary widely, the principal pattern being a progressive reduction in the number of phonemes as one moves eastward across the Pacific. However, as shown earlier with Pileni of the Santa Cruz Archipelago, contact-induced change has caused some of the Polynesian Outlier languages to diverge rather sharply from those of Triangle Polynesia both in size of segment inventory and in the types of segments allowed. A Polynesian Outlier language that evidently has developed an unusual typology, purely through system-internal change, is Luangiua, or Ontong Java, in the Solomon Islands, which has a consonant system /p k ʔ m ŋ v s h l/, and so, along with Palauan, is one of the two AN languages known to lack an alveolar nasal. Whereas Palauan reduced its system of nasals to *m* and *ŋ* through a change *n > *l*, Luangiua has developed an identical system of nasal consonants through a change *n > *ŋ*.

The largest segment inventory within Triangle Polynesia-Fiji-Rotuma is probably that of Wayan, or Western Fijian, with 19 consonants and five vowels (plus length). Rotuman has 14 consonants and 10 surface vowels, but the latter manifest an underlying system of five vowels, with allophones derived through complex synchronic patterns of metathesis and vowel fusion. As shown earlier, with the possible exception of Mekeo in southeast New Guinea, the smallest AN segment inventory is shared by five Eastern Polynesian languages (South Island Maori, Rurutu, South Marquesan, North Marquesan, Hawaiian), all of which have eight consonants, and five vowels plus length. As the best-known of these, the Hawaiian inventory is reproduced below:

Table 4.24 Phoneme inventories for Wayan and Hawaiian

Wayan (Pawley and Sayaba 2003)[27]	**Hawaiian** (Elbert and Pukui 1979)
t k	p k ʔ
kʷ	m n
b d g	h
gʷ	l
dr	w
m n ŋ	
m̄	
ŋʷ	
s	
v ð	
l	
r	
w	
vowels: i, u, e, o, a (+ length)	vowels: i, u, e, o, a (+ length)
19 + 5 = 24	8 + 5 = 13

Western Fijian (and a few Eastern Fijian language communities) differ from all other languages in this region in having a series of labiovelar consonants. It is primarily this feature that makes the Wayan phoneme inventory larger than that of standard Fijian, which has *k* for *k*^{w}, *g* for *g*^{w}, and *m* or *ŋ* for *ŋ*^{w}. A second feature that distinguishes these phoneme inventories is the presence of a segment that Pawley and Sayaba (2003) write *m̄*, and describe as a long bilabial nasal. Since consonant length is otherwise not attested this phonetically long segment is treated as part of the phoneme inventory.

Virtually all Fijian dialects have a prenasalised alveolar trill, that is highly distinctive in global or even pan-AN perspective, but which is also found in languages of the Admiralty Islands and Vanuatu. There is little else to add, except perhaps to note that Rotuman alone among the languages of this region has a palatal 'series' represented by a single voiceless affricate, conventionally written *j*.

4.2   Morpheme structure (phonotactics)

The inherited morpheme structure of AN languages can be expressed by the canonical formulas CVCVC and CVCCVC. In the second of these the abutting consonants are either homorganically prenasalised obstruents (*tumbuq 'grow', *punti 'banana'), or heterorganic segments in a fossilised reduplication (*butbut 'pluck, pull out', *pakpak 'clap, flap the wings'). More than 90% of PAN or PMP base morphemes were disyllabic, and all but a few of the others were trisyllables. Many daughter languages, particularly those in Taiwan, the Philippines and Indonesia, have preserved this canonical shape with little change (although Formosan witnesses show virtually no trace of homorganically prenasalised obstruents).

---

[27]   Fijian orthography represents the prenasalised voiced velar stop by *q*, the prenasalised alveolar trill by *dr*, the velar nasal by *g*, and the voiced interdental fricative by *c*. In the interest of broad comparabilty I have changed *q* and *qw* to *g* and *g*^{w}, *g* and *gw* to *ŋ* and *ŋ*^{w}, and *c* to *ð*.

4.2.1 Limitations on the distribution of consonants

Limitations on the distribution of consonants in AN languages can be divided into those that are inherited from PAN or other early proto languages, and those that arose later. The former type is naturally more widely shared among contemporary languages, although some historically secondary constraints—such as the absence of CVC syllables—are characteristic of large areas.

Chrétien (1965) presented a comprehensive statistical overview of the phonological structure of morphemes in Dempwolff (1938). Although Dempwolff's comparative dictionary is now dated, most generalisations that Chrétien proposed can be maintained in a somewhat reworked form. Where these constraints have been passed on to large numbers of daughter languages they will be mentioned here.

The first distributional limitation on PAN phonemes affects the palatals. Dempwolff reconstructed six palatals, that are now commonly written *s, *c, *z, *j, *ñ, and *y. *c, *z, and *ñ never occur word-finally, and it is these that are often reflected as palatals [ʧ], [ʤ] and [ɲ] in the modern languages when they do not merge with another phoneme. In most daughter languages with one or more palatal stops or nasals, then, these segments do not occur word-finally, a constraint inherited from PAN. Exceptions to this statement fall into two categories: 1) languages that have developed final palatals from earlier non-palatal segments, and 2) retained palatals that have come to be word-final through loss of final vowels or -VC. An example of the first type is seen in the Long Jegan dialect of Berawan in northern Sarawak, where alveolar and velar stops and nasals have palatalised word-finally following *i (reconstructions are pre-Berawan): *kulit > *kolaic* 'skin', *tumid > *toməñ* 'heel', *lamin > *lamaiñ* 'house', *betik > *bətiəic* 'tattoo', *kabiŋ > *kakəiñ* 'left side'. An example of the second type is seen in the languages of Manus, where *-VC has been lost, making some medial palatal stops or nasals word-final: *laje > Loniu *lac* 'coral limestone', *poñu > Loniu *poñ* 'green turtle'. Where this has happened it is clear that final -*ñ* is unstable, as closely related languages with an *ñ* : *ñ* correspondence differ in -*ñ* vs. -*y*: Loniu, Bipi *ñaman* 'fat, grease', Loniu, Bipi *ñamon* 'mosquito', but Loniu *moñ*, Bipi *moy* 'pandanus species', Loniu *poñ*, Bipi *puy* 'green turtle').

A second inherited constraint on phoneme distribution that is widespread in attested languages is the absence of prevocalic and word-final schwa. All other vowel sequences are allowed, and all other vowels may occur in every possible position except that the contrast of schwa and *a* is often neutralised as *a* before glottal stop and *h*.

Somewhat more subtle is a constraint against dissimilar labials in successive syllables. Chrétien (1965) provides information both on phoneme frequencies by position, and on combinatorial frequencies. He notes the following labial-V-labial frequencies for initial and medial consonants (positional frequencies in relation to a CVCVC base are: *p = 185-159-56, *b = 237-180-10, *m = 40-60-50):

Table 4.25 Labial-vowel-labial frequencies in Dempwolff (1938)

	-p-	-b-	-m-	-p	-b	-m
p-	8	0	0	0	0	5
b-	1	4	0	0	0	0
m-	3	1	0	0	1	3
-p	1		0	0		
-b	0		0	0		
-m		0	0	0		

The associative and dissociative tendencies reflected in Table 4.25 are coloured by the positional frequency of particular phonemes. What stands out clearly is that there is no dissociative tendency with identical labials, but that dissimilar labials in successive syllables are strongly disfavored (the constraint appears to be weaker with regard to initial *p and final *m). The same constraint is found in many modern languages, and as will be seen, plays a critical role in the phenomenon of 'pseudo nasal substitution.'

Additional conclusions based on modifications of the Chrétien counts are that final *b and *g are rare, that nasals are rare word-initially (several instances of 'morpheme-initial' *m- in Dempwolff (1938) have proven to be prefixes that have fossilised in most daughter languages), and *k and *ŋ are extremely frequent in final position. These structural properties of morphemes are reflected rather faithfully in many AN languages of insular Southeast Asia, but not in the Oceanic languages, which on the whole show much more phonological erosion. Historically secondary constraints have arisen in the evolution of individual daughter languages. In some cases these are shared among languages as a result of common inheritance from an exclusive proto language, while in others they are shared as a result of convergent development.

4.2.1.1   The phonotactics of liquids

Where segments are sensitive to one another in adjacent syllables we can speak of 'interference effects.' Most AN languages have both lateral and rhotic liquids, and as in other language families, these segments are subject to interference effects in some languages. While inherited constraints such as the avoidance of *bVp, etc. have no reconstructable history, interference effects with liquids are clearly products of historical assimilation or dissimilation that has given rise to restrictions on permitted liquid sequences in successive syllables. Undoubtedly the most perplexing feature of interference effects among liquids is that they work in opposite directions in different languages, sometimes eliminating, and sometimes introducing dissimilar liquids in successive syllables. Equally surprising is the observation that liquid interference effects are easy to find in the AN languages of insular Southeast Asia, but are rare or absent in Oceanic languages. Given a single lateral and a single rhotic liquid there are four logically possible avoidance patterns in a liquid-vowel-liquid sequence: 1) no lVl, 2) no rVr, 3) no lVr, 4) no rVl.

No lVl. No AN language is known which has *l* in its inventory of phonemes, but disallows the sequence *lVl*.

No rVr. Javanese allows *lVl* (*lalər* 'housefly'), *lVr* (*larik* 'line, row'), and *rVl* (*rila* 'wholeheartedly willing'). It also allows *rVr*, but this is disfavored, and words with such sequences often have variant pronunciations with *lVr* (Horne 1974). In all such cases where etymological information is available it shows that an earlier sequence *rVr* has dissimilated to *lVr*: PMP *duha > dua > rua > Old Javanese *ro/roro*, modern Javanese *loro* 'two', *daRa > da > ra > Old Javanese *rara* > modern Javanese *lara* 'virgin'.

No lVr. Languages that avoid *lVr* include Maranao of the southern Philippines, Kelabit of northern Sarawak, and the Batak languages of northern Sumatra. In all three languages *l > l when interference effects are not present. The first of these languages permits *lVl*, *rVl* and *rVr*, and the latter two permit only *lVl* and *rVr*. However, *rVl* sequences in Maranao are rare, and occur mainly in loanwords. Where a sequence of dissimilar liquids arose through regular change it was eliminated by assimilation: in Maranao, *alud 'kind of boat' > alor > *aror* 'raft', *laja > lara > *rara* 'weave a mat or basket', *lujan > loran > *roran*

'cargo', *zalan > dalan > ralan > *lalan* 'road', *zuluŋ > duluŋ > ruluŋ > *loloŋ* 'prow of a boat', in Kelabit, *salaR > alar > *arar* 'nest', *qaluR > alur > *arur* 'small stream', *qateluR > telur > *terur* 'egg', in Toba Batak, *saluR 'flow' > *sarur* 'diarrhoea', *raraŋ* 'forbidden' (Malay *laraŋ*), *rura* 'dale, valley' (Malay *lurah*). In Maranao earlier *lVr* became *rVr* and earlier *rVl* became *lVl*, eliminating sequences of dissimilar liquids by regressive assimilation. Examples of earlier *rVl* are lacking for the other languages and it therefore is unclear whether they would become *rVr* or *lVl*.

No rVl. Although some languages that have phonemes *l* and *r* rule out the sequence *rVl*, this appears to occur only when the sequence *lVr* is also ruled out, as with Maranao, Kelabit and the Batak languages. In this sense few if any languages seem to distinguish (or better, dissociate) patterns 3 and 4 from one another.

4.2.1.2   The phonotactics of sibilants

In the general phonetics literature the term 'sibilant' is commonly reserved for fricatives in the dental and alveolar regions of the hard palate (Ladefoged and Maddieson 1996:145-64). Several well-attested phenomena in AN languages target a class of segments that includes standard sibilants plus some other sounds that resemble sibilants in either articulatory or acoustic terms. No ready-made term is available for this larger class of consonants, and in the interest of simplicity the term 'sibilant' will therefore be extended to encompass it.

Interference effects with sibilants are rarer than those with liquids, but a band of languages across southern Borneo which includes at least Iban in the west and Ngaju Dayak in the east rules out the sequences *sVs, cVc, sVc*, and *cVs*. Where such sequences existed earlier they were eliminated by dissimilation of the first segment to *t*, as in Iban, *tacat* 'incomplete, blemished' (Malay *cacat* 'defect; shortcoming'), Iban *tasak* 'plait' (Malay *sasak* 'wattle; plaitwork'), Iban *təsal* (Malay *səsal*) 'regret', Iban *ticak* (Malay *cicak*) 'gecko', Iban *tisir* (Malay *sisir*) 'comb', Iban *tucok* (Malay *cucok*) 'pierce, stab', or *seseD > Ngaju Dayak *təsər* 'dive, submerge', *sisik > Ngaju Dayak *tisik* 'fish scale', *susu > Ngaju Dayak *tuso* 'breast'. The evidence for a similar restriction on voiced sibilants in successive syllables is less clear-cut, but Iban shows some examples of dissimilated jVj: *dajaʔ* (Malay *jaja*) 'to hawk, peddle', *dəjal* (Malay *jəjal*) 'stop up, plug', *dujul* (Malay *jujul*) 'stick out, project'.

A somewhat different type of interference effect between sibilants in adjacent syllables is found in three Formosan languages. Since this remains productive in at least one of these languages it will be treated below under 'Phonological processes.'

4.2.1.3   Consonant clusters

Many AN languages allow no consonant clusters, and most languages that do permit them only medially. Table 4.26 lists types of medial clusters in 39 languages arranged geographically. RM = cluster in reduplicated monosyllable, PO = prenasalised obstruent, G = geminate, HC = heterorganic cluster in a non-reduplicated base. 'Prenasalised obstruent' excludes examples of homorganic nasal + obstruent sequences that arise from what is essentially heterorganic consonant clustering. Prenasalisation normally refers to homorganic nasal-obstruent sequences, but this includes -ŋs- in some languages. All statements exclude transparent loanwords:

Table 4.26 Patterns of medial consonant clustering in Austronesian languages

	RM	PO	G	HC
Taiwan				
Thao	+	–	–	+
Paiwan	+	–	–	–
Puyuma	+	–	–	+
Bunun	+	–	–	+
Kavalan	+	–	+	–
Amis	+	–	–	+
Philippines				
Ilokano	+	+	+	+
Bontok	+	–	+	+
Tagalog	+	+	–	+
Borneo				
Kadazan Dusun	–	+	–	–
Kelabit	–	–	–	–
Kiput	–	–	+	–
Ma'anyan	–	+	–	–
Sumatra-Java-Bali				
Toba Batak	+	+	+	+
Rejang	–	+	–	–
Balinese	+	+	–	–
Sulawesi				
Sangir	–	+	–	–
Banggai	–	+	–	–
Makasarese	–	+	+	–
Eastern Indonesia				
Manggarai	–	+	–	–
Ngadha	–	–	–	–
Wetan	–	–	–	+
Soboyo	–	+	–	–
New Guinea-Bismarcks				
Motu	–	–	–	–
Kuruti	–	–	–	+
Tigak	+	–	–	+
Solomons				
Roviana	–	–	–	–
Nggela	–	–	–	–
Arosi	–	–	–	–
Vanuatu - New Caledonia				
Nguna	–	–	–	–
Paamese	+	–	–	+
Lenakel	–	–	–	+
Xârâcùù	–	–	–	–
Micronesia				
Chamorro	+	+	+	+
Mokilese	+	+	+	+

	RM	**PO**	**G**	**HC**
Kosraean	–	–	+	+
Polynesia-Fiji				
Fijian	–	–	–	–
Tongan	–	–	–	–
Hawaiian	–	–	–	–

In a few cases the line between prenasalised obstruents and heterorganic clusters is not easy to draw. Ma'anyan of southeast Borneo and Makasarese of southwest Sulawesi, for example, permit only homorganically prenasalised stops and *-nr-*, the latter reflecting *nd. Since no other heterorganic clusters occur in either language it seems best to treat this sequence as part of the well-attested pattern of homorganically prenasalised medial obstruents. Some languages, such as Kelabit, allow no intramorphemic clusters, but permit morphologically derived clusters within a word: *tə-bakaŋ* 'spread wide, as the legs': *simbakaŋ* < /t<in>ə-bakaŋ/ 'was spread wide by someone, of the legs.' Others, as Rejang, allow prenasalised voiced obstruents, but not prenasalised voiceless obstruents. Finally, any statement about consonant clusters in AN languages must deal with the phonemic interpretation of prenasalised obstruents, a topic that is given separate treatment below.

Although the data in Table 4.26 may not be fully representative, some patterns emerge from it. First, languages that permit no consonant clusters tend to be found in eastern Indonesia and the Pacific. Second, medial consonant clusters in reduplicated monosyllables are virtually universal in Formosan and Philippine languages, where they are inherited from PAN, but have a thin and scattered distribution elsewhere. Third, homorganic nasal-stop clusters are rare or absent in Formosan languages, but are common in the Philippines and western Indonesia, and turn up in some languages of eastern Indonesia. Most Oceanic languages lack true clusters of nasal + obstruent, since prenasalisation is an automatic feature of voiced obstruents. Fourth, the distribution of geminate consonants in AN languages is wide and scattered, reflecting the derivation of these segments through numerous independent historical changes. Finally, heterorganic medial consonant clusters occur in Taiwan, Indonesia and the Pacific, but are rare. By contrast, such clusters are common in Philippine languages, where they are generally products of vowel syncope in the environment VC__CV. Table 4.27 shows the types of medial clusters found in native Tagalog words (reduplicated monosyllables are excluded, and unique attestations are underlined):

Table 4.27 Medial consonant clusters in nonreduplicated native Tagalog words

	p	t	k	?	b	d	g	m	n	ŋ	s	h	l	w	y
p						pd			pn		ps	p<u>h</u>	pl	pw	py
t					tb	td			tn			th	tl		ty
k	<u>kp</u>	kt			kb	kd		km	kn		ks	kh	kl	kw	ky
?															
b		bt							bn		bs		bl	bw	by
d	<u>dp</u>											dh	dl	dw	dy
g	gp	gt	gk		gb	gd		gm	gn		gs	gh	gl	gw	gy
m	mp				mb						ms		ml		my
n		nt				nd					ns	nh	nl	nw	ny
ŋ		ŋt	ŋk				ŋg	ŋm			ŋs	ŋh	ŋl	ŋw	ŋy
s	sp	st	sk		sb	sd	sg	sm		sŋ			sl	sw	sy
h															
l	lp	lt			lb	ld					ls	lh		lw	ly
w															
y															

Table 4.27 must be qualified in certain ways. First, it shows attested clusters, but apart from unique instances it gives no idea of their frequency, which can vary greatly (e.g. -pC- is rare, while -Cl- is common). Second, this table shows only intramorphemic consonant clusters, although other surface clusters occur across a morpheme boundary as a result of vowel syncope (*atíp* : *apt-án* 'roof'). Third, it omits common clusters such as *-ts-* which are largely confined to loanwords. Having said this, certain generalisations do emerge from this table. It is apparent that there are no geminates, and no clusters with glottal stop. With regard to the latter feature Tagalog stands apart from most Central Philippine languages, which allow glottal stop to cluster with other consonants, but often only in one order (-ʔC- or -Cʔ-, depending upon the language). Like many Philippine languages, Tagalog does not permit *h* or a glide as the first member of a cluster. Other gaps that appear to be non-accidental are the absence of -Cŋ-, and the rarity of velar stops as the second member of a cluster.

Between languages such as Tagalog, which allow a wide range of medial consonant clusters, and languages such as Kadazan Dusun, in which only homorganically prenasalised obstruents or homorganic voiced-voiceless stop clusters occur, are a few languages that allow prenasalised obstruents and preconsonantal liquids in medial position, as with standard Malay, which has *-mb-*, *-mp-*, *-nd-*, *-nt-*, *-nj-*, *-nc-*, *-ŋs-*, *-ŋg-*, *-ŋk-*, *-rC-*, and Toba Batak, which has *-mb-*, *-nd-*, *-nj-*, *-ŋg-*, *-ts-*, *-lC-*, *-rC-*, and voiceless geminates *-pp-*, *-tt-* and *-kk-* that reflect *-mp-, *-nt- and *-ŋk-, and are still written as prenasalised stops in the Indic-derived Batak syllabary.

While medial consonant clusters are by far the most common type, and for that reason the most difficult to describe in generalising terms, some languages have consonant clusters in initial position, and a few have final clusters. Most AN languages that permit initial consonant clusters also permit clusters in medial position. Examples include Thao of central Taiwan, Javanese of western Indonesia, Alune of the central Moluccas, Cheke Holo of the Solomon Islands, and Lonwolwol of central Vanuatu. Both initially and medially, consonant clusters are almost always limited to two segments. Exceptions include Thao, where clusters of voiceless lateral plus two consonants may occur initially but never medially, as in *lhckis-in* 'be finished up, consumed entirely' ([ɬθk]), or *lhqnis-i* 'bear down,

as in childbirth (imper.)' ([ɫqn]), Javanese, where clusters of homorganically prenasalised obstruent plus liquid or glide occur medially but never initially, as in *luŋkrah* 'enervated, weak', *sampyuk* 'splash, splatter', or *timblis* 'wooden mallet', and Palauan, where obstruent-liquid-obstruent clusters occur initially, although phonetically these are broken up by excrescent schwas, as in *b<l>deŋ* 'stagnant state (of liquid)', *k<l>dols* 'fatness; thickness', or *klsuk* 'Palauan money'. Most triliteral initial clusters result from infixation with the nominaliser -*l*-. A few languages such as Rotinese, Tetun, and Yapese, have only initial consonant clusters. In this connection it is important to note that the orthography of some sources can be misleading. Fey (1986), for example, writes initial clusters in Amis words such as *cka* 'thorn', *croh* 'to add water', *dfak* 'bed of dirt for gardening', or *hci* 'fruit', but she has reached this result by arbitrarily choosing to write the penultimate schwa of this language as zero.

Final clusters are the rarest type in AN languages. Homorganically prenasalised stops occur word-finally in languages of the Markham Family, as in Wagau *mbɛyomp* 'cloud', *palêŋk* 'vine', *muŋg* 'old', or Mapos *ŋgʷimb* 'cassowary', *βund* 'banana' (Hooley 1971). Similar clusters are reported for Saluan of eastern Sulawesi, although no data is given (Busenitz and Busenitz 1991:fn. 8). Most clusters in Markham languages are intramorphemic, but in forms such as Zenag *miaŋk* 'mouth', *nuluŋk* 'nose', *kʷaŋk* 'neck', or *nəmaŋk* 'hand', the final cluster almost certainly is the first person singular possessive pronoun -*ŋk*. Given the rarity of final consonant clusters in AN languages generally, these final consonant sequences probably reflect the diffusion of phonotactic patterns from neighbouring Papuan languages. Among the few other examples of final consonant clusters that can be found in AN languages are the numerous heterorganic clusters of Palauan, which—like non-final clusters—are normally broken up by excrescent schwas: *bsibs* [bsíbəs] ~ [bsíbəsə] 'termite', *yaŋd* (written *eánged*) [jáŋəð] ~ [jáŋəðə] 'sky', *durs* [ðúrəs] ~ [ðúrəsə] 'sleep'. With the possible exception of Saluan, for which little information is available, Palauan appears to be unique among AN languages in permitting consonant clusters in initial, medial and final positions.

4.2.1.3.1 The sonority hierarchy in consonant clustering

Hajek and Bowden (2002) note that many phonologists recognise a universal sonority hierarchy of the form vowels > glides > liquids > nasals > obstruents 'which has been frequently observed as governing the permitted organisation of consonant clusters.' The expectation is that consonants in sequence will increase in sonority as they approach the syllable nucleus, and decrease in sonority as they move away from it. Greenberg (1978a:260) incorporated this principle in a proposed universal that further distinguishes voiced from voiceless obstruents: 'Except for voiced nasal followed by homorganic unvoiced obstruent, an unvoiced consonant or sequence of unvoiced consonants in initial systems immediately preceding a vowel is not itself preceded by one or more voiced consonants.' Hajek and Bowden have shown that some languages of the Moluccas violate this principle: *mt-, ms-, np-, nk-, lp-, rp-, rt-, rk-, rs-* are found in Roma, *mt-, mk-, rs-* and *ms-* in Leti, and *mt-, mk-, ms-, np-, nk-, ns-, lp-, lk-, lt-,* and *ls-* in Taba. They note that the Roma clusters and the first two Leti clusters are phonetically heterosyllabic, and that many of these clusters contain a morpheme boundary. Several other AN languages also violate this universal ordering principle, as Dorig of northern Vanuatu (François 2010). All of the initial clusters cited by Hajek and Bowden contain a nasal or liquid followed by an obstruent. Palauan has clusters of this type, but also clusters of voiced stop followed by voiceless stop or fricative, hence *bk-, bs-, bt-, lk-, ls-, lt-, mk-, ms-, mt-, rs-, rt-*. All such

clusters occur within lexical bases, and many appear to be tautosyllabic: *bkáu* 'large tree in rose family', *bsibs* 'termite', *btuch* 'star', *lkes* 'sandbar', *lsal* 'very', *ltúkel* '(someone) is to be remembered', *mkar* 'gave medicine to', *msur* 'bent down', *mtab* 'found (landmark)', *rsáol* 'border between deep and shallow sea', *rtáŋel* 'is to be pounded'.

Yapese, a language that has borrowed heavily from Palauan, but which is only distantly related to it, also has several word-initial consonant clusters that fail to conform to the expectations generated by Greenberg's typology. Among problematic syllable onsets Jensen (1977a:48) gives *rch-* (*rchaq* 'blood'), and *bp-* (*bpiin* 'woman'), and in his Yapese-English dictionary Jensen (1977b) also lists clusters of *b* plus glottalised *p*, as in *bp'aŋiin/p'aŋiin* 'its top, cutting, of a plant topped to plant.' He notes that the word for 'blood' may also be pronounced *rachaq*, but implies that *rchaq* is more commonly heard.

Among the languages cited by Hajek and Bowden, Taba is a member of the South Halmahera-West New Guinea group, and many other languages within this group also have one or more word-initial consonant clusters that decrease in sonority in approaching the syllable nucleus. An example is Numfor-Biak of west New Guinea, in which a number of words begin with *ms-*, and several others with *rm-*: *msāf* 'bifurcated', *msirən* 'fish species', *msor* 'angry', *msun* 'to fall, of a ripe fruit', *rmonen* 'lust after, seek revenge.' Most examples of typologically aberrant consonant clusters in Palauan, Yapese, and Numfor are morpheme-internal. A search for aberrant clusters that straddle a morpheme boundary undoubtedly would turn up further examples in other languages.

4.2.1.4 *Final consonants*

One of the most striking phonotactic differences among AN languages is seen in the range of permissible final consonants. Table 4.28 illustrates differences in the proportion of final consonants (F) to total consonants (T) in six languages. X = consonant present in final position, 0 = consonant present only in non-final position. Blanks show total absence of a consonant in the language in native forms. MAR = Maranao (southern Philippines), MAL = Malay, MIN = Minangkabau (southwest Sumatra), LgT = Long Terawan Berawan (northern Sarawak), BGS = Buginese (southwest Sulawesi), BIM = Bimanese (western Lesser Sundas).

Table 4.28 Proportion of final to total consonants in six Austronesian languages

	MAR	MAL	MIN	LgT	BGS	BIM
p	X	X	0	0	0	0
t	X	X	0	0	0	0
c		0	0	0	0	0
k	X	X	0	0	0	0
ʔ	X		X	X	X	0
b	X	0	0	0	0	0
d	X	0	0	0	0	0
j		0	0	0	0	0
g	X	0	0	0	0	0
ɓ						0
ɗ						0
m	X	X	X	X	0	0
n	X	X	X	X	0	0
ñ		0	0	0	0	
ŋ	X	X	X	X	X	0
f						0
s	X	X	0	0	0	0
h		X	X	X		0
v						0
l	X	X	0	0	0	0
r	X	X	0	0	0	0
w	X	X	X	X	0	0
y	X	X	X	X	0	
F/T	15/15	12/18	7/19	7/19	2/18	0/20
%	100	66.7	36.8	36.8	11.1	zero

The six languages in Table 4.28 have been selected to illustrate progressive stages of collapse in the system of final consonants. Maranao shows no tendency in this direction, as all consonant phonemes that occur in the language are permitted in final position. In this respect it is similar to most AN languages of Taiwan and the Philippines. A number of languages in western Indonesia show a comparable distributional freedom, but have several palatal consonants which, as already noted, are almost invariably barred from final position. Malay exemplifies the first step in the eventual eradication of closed syllables: voiced stops have merged with voiceless stops word-finally. A further stage in this drift is seen in Minangkabau, where all final stops have merged as -ʔ, earlier *s* and *h* have become -*h* and liquids have disappeared, leaving only three nasals, two laryngeals and two glides in final position. Long Terawan Berawan shows heavy reduction of final contrasts, but via a distinctly different diachronic route. Although final voiceless stops have merged as -ʔ and final *s has weakened to -*h*, final voiced stops and liquids have merged with the corresponding nasals. Moreover, counteracting this general weakening of syllable endings is another diachronic development, the addition of -*h* after original final vowels: *mata > mattəh* 'eye', *laki > lakkeh* 'male', *batu > bittoh* 'stone'. Buginese shows the system of final consonants at its last stand, where all obstruents and liquids have merged as -ʔ, all nasals have merged as -ŋ, and final glides have disappeared through monophthongisation: *-ay > -e*, *-aw > -o*.

All AN languages in Taiwan, the Philippines and Borneo allow closed syllables, and elsewhere the absence of closed syllables is rare and areally concentrated. In western Indonesia the only languages that exclude closed syllables appear to be Nias and Enggano, both spoken in the Barrier Islands west of Sumatra, but otherwise not closely related. In both cases the absence of closed syllables is a historical product of final consonant loss. In Malagasy, by contrast, the absence of closed syllables has resulted from the addition of a supporting vowel *-a* to words that originally ended in a consonant. Both final consonant loss and vowel addition are responsible for the disappearance of closed syllables in many Oceanic languages (PMP *zalan > Samoan *ala*, Mussau *salana* 'path, road', PMP *ma-takut > Samoan *mataʔu*, Mussau *matautu* 'afraid'). A number of Oceanic languages that have eliminated closed syllables through consonant loss or vowel addition have subsequently lost final vowels, and so returned to their earlier syllabic condition with a reduced canonical shape of CVC for lexical bases. This is true of all languages of the eastern Admiralties, most languages of Vanuatu, and nearly all Nuclear Micronesian languages. Finally, a number of Oceanic languages have lost word-final *u only after *m, producing a small number of *m*-final forms in languages that otherwise are consistently vowel-final. In Kilivila syllabic *m* is the only consonant that clusters with another in initial or medial position, and non-syllabic *m* is the only consonant that occurs in syllable final position, as in *kabitam* 'wisdom', or *kukʷam* 'you eat' Senft (1986:18). A similar distributional anomaly is seen in Wuvulu, where all morphemes end either in a vowel or in *-m*, the latter seen most commonly, if not exclusively, in nouns suffixed with *-m* '2sg', as in *pani-u* 'my hand', *pani-m* 'your hand', *pani-na* 'his/her hand'. Manam allows slightly greater latitude to final consonants. According to Lichtenberk (1983:21) 'At the phonemic level, the syllable structure of Manam is (C)V(N).' The only nasals that can occur in this position are *-n* and *-ŋ*, and historically these derive from forms that earlier ended in a high vowel: POC *pani > *an* 'give', *kani > *ʔan* 'eat', *punuq > *un* 'beat', *danum > *daŋ* 'water', *manuk > *maŋ* 'chicken', *poñu > *poŋ* 'green turtle'.

Less apparent than the presence or absence of final consonants are phonotactic differences in penultimate syllables. PAN permitted heterorganic consonant clusters in reduplicated monosyllables such as *buCbuC 'pluck, pull out', *kiskis 'scrape', or *tuktuk 'knock, pound, beat'. These have been retained in many languages, and in several of the Formosan languages they are the only clusters allowed. All Central-Eastern Malayo-Polynesian languages and some languages in western Indonesia have reduced such clusters, with the exception that nasal-obstruent sequences are sometimes retained: *seksek > Manggarai *cəcək* 'stuff, cram in', *bejbej > Kambera *búburu* 'bundle', *bukbuk > Tetun *fuhuk* 'wood weevil', *sepsep > Javanese *səsəp* 'suck', *buŋbuŋ > *wuwuŋ* 'ridge of the roof', *bukbuk > Malay *bubok* 'wood weevil', but *diŋdiŋ > *dindiŋ* 'wall'. As these examples illustrate, the loss of syllable-final consonants in medial position is independent of the loss of word-final consonants.

4.2.2 Limitations on the distribution of vowels

Restrictions on the distribution of vowels are uncommon in AN languages. The one constraint that stands out most clearly is inherited from PAN: the schwa may not occur prevocalically or word-finally. Some languages, such as Bilaan of southern Mindanao, Balinese, or non-standard dialects of peninsular Malay, have developed final schwa from unstressed *a, but no language is known to have developed prevocalic schwa from any earlier vowel.

A second constraint that involves schwa (often written /e/) developed in many of the languages of western Indonesia, and parts of the southern Philippines. In some languages just the low vowel *a*, but in others all vowels, neutralised as schwa in prepenultimate syllables. This constraint affects the vowels of polysyllabic bases, and of many affixes. Philippine languages such as Western Bukidnon Manobo preserve the shape of affixes before vowel-initial bases by separating affixal and base vowels with a glottal stop: WBM *amur* 'gather' : *kə-ʔəmur-an* 'a celebration of any kind', *itəm* : *mə-ʔitəm* 'black', *upiya* 'to do good' : *mə-ʔupiya* 'good'. In all languages of Borneo south of Sabah for which relevant information is available, and in a few other languages in western Indonesia, the prevocalic schwa of such prefixes drops: Lun Dayeh *mə-lauʔ* 'hot', *mə-bərat* 'heavy', *mə-ditaʔ* 'tall', but *m-abuh* 'dusty', *m-itəm* 'black', *m-ulaʔ* 'many, numerous', Javanese *banjir* 'flood' : *kə-banjir-an* 'be caught in a flood', *kilap* 'lightning' : *kə-kilap-an* 'forget, overlook', *turu* 'sleep' : *kə-turu-n* 'fall asleep accidentally', but *amuk* 'fury, violence' : *k-amuk* 'be the object of an irrational or frenzied attack', *ijɛn* 'alone' : *k-ijɛn-an/k-ijɛn-ən* 'lonely', *undur* 'backward motion' : *k-undur-an* 'get backed into'.

In most AN languages vowel sequences are limited to two vowels, but in some of the Polynesian languages sequences of three vowels are not uncommon, and sequences of four or possibly more exist: Samoan *sōia* 'dehortative particle', *taeao* 'morning, early morning', Hawaiian *pueo* 'owl', *ʔaiea* 'shrubs of the genus *Nothocestrum'*, etc. A similar, but somewhat less extreme freedom for vowels to cluster is seen in Enggano of the Barrier Islands west of Sumatra: *búai* 'disperse', *eʔiau* 'a bird: *Garrulus javanicus'*.

4.2.3 The problem of prenasalised obstruents

Most AN languages outside Taiwan allow phonetic sequences of nasal + homorganic obstruent either within a morpheme or across a morpheme boundary. In the languages of the Philippines and most of western Indonesia, of which Tagalog and Malay can be taken as paradigm cases, these sequences are invariably interpreted as consonant clusters, and the question never arises whether they might be viewed as unit phonemes. In Oceanic languages such as Fijian, on the other hand, there has never been any doubt that [mb], [nd] or [ŋg] are phonemically voiced stops with automatic nasal onsets. These two poles in the continuum are thus uncontroversial. However, the question whether prenasalised obstruents are unit phonemes or consonant clusters has arisen repeatedly in the discussion of languages which share some phonological traits of languages such as Tagalog or Malay, and some traits of Oceanic languages such as Fijian.

How phonetic nasal-obstruent sequences should be interpreted phonologically has been a recurrent issue in Sulawesian linguistics, but one that is clouded at times by an apparent confusion of criteria for determining membership in the same syllable vs. membership in the same phoneme. As Quick (n.d.) has pointed out, scholars working with similar data in different languages have reached opposite conclusions, sometimes opting for a unit phoneme analysis, and sometimes for a cluster analysis, with no consistent theoretical guidelines. Since the same problem has arisen, either implicitly or explicitly, in the analysis of widely separated AN languages without the benefit of mutual scholarly awareness, the result has been a patchwork of dissimilar analyses of similar data, or of similar analyses of dissimilar data, and in a survey volume such as this the problem needs to be addressed as a unified whole. Table 4.29 provides a bird's-eye view of phonological analyses that have been proposed for nasal-obstruent sequences in AN languages, and some of the criteria that have been considered relevant to the analysis: PNO = prenasalised

obstruent, I = PNO occurs word-initially, M = PNO occurs word-medially, CC- = language allows other initial clusters, UO = prenasalisation affects unvoiced (voiceless) obstruents, VO = prenasalisation affects voiced obstruents, MB = PNO occurs across morpheme boundary, FC = language has final consonants, UP = PNO regarded (in at least some analyses) as a unit phoneme:

Table 4.29 Distributional traits for prenasalised obstruents in Austronesian languages

Language	I	M	CC-	UO	VO	MB	FC	UP
Tagalog	--	+	--	+	+	+	+	no
Malagasy	(+)	+	--	+	+	+	--	yes
Karo Batak	+	+	--	+	+	+	+	no
Javanese	+	+	+	+	+	+	+	yes
Ratahan	+	+	--	+	+	+	+	no
Pendau	+	+	--	+	+	+	+	no
Uma	+	+	--	+	--	--	+	yes
Balantak	+	+	--	+	+	+	+	no
Muna	+	+	--	+	+	+	--	yes
Manggarai	+	+	--	+	+	--	+	yes
Bimanese	+	+	--	+	+	+	--	yes
Rotinese	+	--	--	--	+	--	+	yes
Kambera	+	+	--	--	+	--	--	yes
Yamdena	+	+	+	--	+	+	+	yes
Fijian	+	+	--	--	+	--	--	yes

In a lucid discussion of the problem, Van den Berg (1989:19) gives the following reasons for regarding nasal-obstruent sequences in Muna as unit phonemes: 1) there are no unambiguous consonant sequences in the language, 2) there are no word-final consonants, 3) the prenasalised consonants function as units in the morphological process of full reduplication, thus *lambu* 'house' : *ka-lambu-lambu* 'small house', but *pulaŋku* 'staircase' : *ka-pula-pulaŋku* 'small staircase', where the diminutive is formed with *ka-* plus a copy of the first two syllables of the base, showing a syllabification *pu.la.ŋku*, 4) native speakers of this previously unwritten language agree in the syllable divisions *la.mbu* and *pu.la.ŋku*. These arguments are cogent, but are they evidence for assigning nasal-obstruent sequences to single phonemes or to single syllables?

Logically, there are eight possible relationships that phonetic prenasalised obstruents (NC) can have to morphemes, syllables, and phonemes. Probably the simplest way to represent these relationships is through a numerical code in which 2 = belongs to two (morphemes, syllables, phonemes) and 1 = belongs to only one of these units, as shown in Table 4.30:

Table 4.30 Relationships of NC to morphemes, syllables, and phonemes

Type	1	2	3	4	5	6	7	8
morpheme	2	2	2	2	1	1	1	1
syllable	2	2	1	1	2	2	1	1
phoneme	2	1	2	1	2	1	2	1

In Types 1-4 an NC sequence belongs to two morphemes, and in Types 1-2 and 5-6 it belongs to two syllables. If a morpheme boundary intervenes between the nasal and obstruent it is clear that this sequence cannot be a single phoneme.[28] It is also clear that if a syllable boundary intervenes between the nasal and obstruent this phonetic sequence cannot be a single phoneme. This leaves only 7 and 8 as cases in which NC sequences could be interpreted as unit phonemes. As Table 4.30 shows, if an NC sequence is both tautomorphemic and tautosyllabic it may or may not be a single phoneme, and the choice of which analysis to adopt depends on additional criteria. The error that appears repeatedly in the AN literature on prenasalised consonants is to assume that necessary conditions for determining unitary phonemes are automatically sufficient.

In Fijian and most other Oceanic languages, the phonemic interpretation of NC sequences is non-controversial because of bilateral predictability: voiced obstruents are always prenasalised, and prenasalised obstruents are always voiced. A cluster analysis would not be optimal for Fijian since: 1) if [nd] were written /nd/ it would include redundant phonetic information, 2) if [nd] were written /nt/ it would allow voicing to be predicted from prenasalisation, but it could not be applied to the prenasalised alveolar trill [ndr], which lacks a voiceless counterpart, or to [mb], since Fijian /p/ occurs only in a few loanwords, 3) in these cases syllable structure would be modified to accommodate phonetic clusters that on independent grounds must be treated as unit phonemes.

Objections 1) and 2) do not hold in Muna, where voicing and prenasalisation are independent. As as result, although Fijian voiced obstruents can be written with single symbols *b* = [mb], *d* = [nd] and *q* = [ŋg], Muna voiced obstruents cannot, since [p] contrasts with [mp] and [b] with [mb]. The arguments assembled by van den Berg are relevant to determining syllabicity, but they cannot be used to distinguish unit phonemes from consonant clusters. Since these arguments are similar to those used in analyses of prenasalised obstruents in other languages of central and southern Sulawesi, the same reservations must be expressed with regard to all such cases. In the case of Manggarai, no argument is given for the analysis chosen (Verheijen 1967:xii, Verheijen and Grimes 1995), and as seen in Table 4.30, the reservations regarding Muna and other languages of Sulawesi apply here as well. The rejection of a unit phoneme analysis means that CCV-syllables must be allowed even in languages that lack syllable codas. Since prenasalised obstruents are perhaps the least marked consonant clusters, however, it should not be surprising that some languages allow these sequences and no others syllable-initially.

Both Dempwolff (1937:72) and Dahl (1951:33) include /mp nt nts ntr ŋk mb nd ndz ndr ŋg/ as part of the phoneme inventory of Malagasy, and Beaujard (1998:9) adopts a similar analysis for the Tañala (= Tangala) dialect of southeast Madagascar. Dyen (1971b:213), on the other hand, includes prenasalised counterparts for only the voiced obstruents (hence five additional phonemes rather than ten), but does not explain why his analysis departs from the others. Although Malagasy syllables normally contain only single consonant onsets, all major dictionaries of Malagasy, including Richardson (1885), Abinal and Malzac (1970), and Beaujard (1998) give some bases with initial clusters of *mb-, mp-, nd-,*

---

[28]   The widespread process of nasal substitution creates affixed forms in which a morpheme boundary is in a sense internal to the replacing nasal, as with Malay *pukul* 'hit' : *mə-mukul* 'to hit' (where the second *m* contains manner features of a prefixal nasal and place features of a base-initial obstruent). Similarly, in Javanese prefixation with *ke-* (< *ka-) causes high-low or low-high vowel sequences to fuse as a mid vowel, as in *udan* 'rain' : *k-odan-an* 'be caught in the rain', leaving the morpheme boundary 'stranded' within the first-syllable vowel. Such cases probably are best described as showing ambiguity in the location of a morpheme boundary rather than as instances of a single phoneme straddling two morphemes.

ndr-, nt-, and the like, and this evidently is the basis for the claim that Malagasy has a separate series of prenasalised obstruent phonemes. However, words that begin with such sequences are rare. In a dictionary with well over 10,000 base forms, Richardson (1885) gives eleven examples of *mb-,* six of *mp-,* three of *nd-,* six of *ndr-,* and two of *nt-.* The frequency of prenasalised word-initial obstruents in the Tañala dialect as described by Beaujard (1998) is equally limited. Moreover, some of these examples clearly have arisen from medial prenasalised obstruents by loss of the base-initial vowel. Richardson (1885), for example, gives *mba* 'too, as well as, in order that', but cross-references this to *omba* 'follow, accompany, associate with' (seen also in *mbamy < (o)mba amy* 'with, together with, including'), and a similar relationship holds for a number of other words, such as *mbávy/ambávy* 'tree or shrub which supplies a kind of torch', *mbazáha/ambazáha* 'manioc, cassava', *mbe/ombe* 'ox, cow', *mby/omby* 'arrived', *njía/jía* 'a path', etc. In some cases it is possible that forms which occur only with a nasal + obstruent onset are Bantu loans. All-in-all, then, the basis for recognising a series of prenasalised obstruent phonemes in Malagasy seems precariously weak. Dyen's decision to recognise such a series only for the voiced obstruents of Malagasy evidently is modeled on the relationship between voiced and voiceless obstruents in languages such as Fijian, but this view is clearly misguided, since prenasalisation and obstruent voicing are mutually predictive in Fijian, but not in Malagasy.

Writers such as Horne (1974) and Nothofer (1975) have also recognised a similar series of prenasalised obstruent phonemes in Javanese, but on equally insubstantial grounds. The Javanese data differ from Malagasy in two crucial respects. First, unlike Malagasy, Javanese does allow unambiguous word-initial clusters of nasal + obstruent. The initial clusters of Javanese thus cannot be interpreted as medial clusters which have optionally come to be initial through aphaeresis. Second, however, unlike the great majority of AN languages Javanese allows some additional consonant clusters in word-initial position. Horne (1974) lists several hundred bases which begin with a stop or nasal followed by *l* or *r,* and she often distinguishes these from similar bases in which these consonants are separated by schwa, as in *bələt* 'mud, muck' vs. *bləs* 'entering', *jələh* 'bored, tired of' vs. *jləg* 'a quick downward motion', or *mərəŋ* 'dried rice straw' vs. *mrəŋgi* 'kris-maker'. Given this relaxation of canonical form there is no reason why initial sequences of nasal + stop could not be analyzed as consonant clusters. To make matters worse for the unit phoneme analysis, virtually all examples of word-initial prenasalised obstruents appear to incorporate a morpheme boundary, as with *m-bayar* 'to pay', *n-dələŋ* 'to see', *n-jəmblaŋ* 'be bloated, of the belly', or *ŋ-gawa* 'to carry'.

The most difficult problems for the phonemic interpretation of nasal-obstruent sequences are encountered in languages of the Lesser Sundas. Kambera of eastern Sumba allows only open syllables, and only voiced obstruents may be prenasalised:

Table 4.31 Consonant inventory of Kambera (eastern Sumba)

p	t		k	
ɓ	ɗ	j		
m	n	ñ	ŋ	
mb	nd	nj	ŋg	
				h
	l			
	r			
w		y		

Klamer (1998:10ff) proposes a unit phoneme analysis for the prenasalised obstruents of Kambera for the following reasons: 1) they appear as syllable onsets both word-initially and word-medially, 2) the nasals *n* and *m* cannot be codas, 3) the language lacks unambiguously complex onsets such as *tr-*, *pl-* or *st-*, and loanwords with such sequences are reparsed into two open syllables, 4) loanwords with initial voiced obstruents are always prenasalised (thus Malay *baca* is borrowed as *mbaca* 'to read'). The first three of these arguments, like those given by van den Berg for Muna, are cogent but misplaced, since they provide evidence only for membership in the same syllable, not evidence for a unit phoneme. The fourth observation is more interesting, and raises questions about how Kambera speakers perceive the phonetic properties of *b-* or *d-* when these do not match either of the voiced obstruent types found in the native vocabulary. Since only labial and alveolar stops are imploded we would not expect the same phonetic mismatch with loanwords that have *j-*, and such forms should be borrowed without prenasalisation. This is, indeed, the case, as with Kambera *jala* 'casting net' (< Malay *jala*, ultimately from Sanskrit), *jara* 'drill' (< Malay *jara*), or *jàriku* 'citrus fruit' (< Malay *jəruk*).

Since only voiced obstruents may be prenasalised in Kambera the relationship between plain and prenasalised obstruents in this language resembles that in Oceanic languages more closely than it resembles any of the cases from Sulawesi or the Lesser Sundas that have been examined so far. However, even in Kambera it cannot be argued that prenasalisation is predictable, since there is a series of voiced obstruents which is not prenasalised. The series of simple voiced obstruents differs from the prenasalised series in two respects. First, it is defective, lacking a velar member. This allows the inference that [ŋg] is /g/, but predictable prenasalisation would then hold for only one member of the voiced obstruent series, leaving the other three members as phonemic clusters. Since the labial and alveolar voiced obstruents are implosives, it might be inferred that these stops are distinct from the plain labial and alveolar stops which are prenasalised. But it clearly would be simpler to treat implosion as predictable in Kambera. Again, the proposal that nasal-obstruent sequences must be treated as unit phonemes is defeated by the evidence: although obstruent voicing in Kambera is predictable from prenasalisation, prenasalisation is not predictable from obstruent voicing.

Rotinese presents a similar set of relationships. This language has voiceless stops *p, t, k, ʔ*, plain voiced stops *b, d* and prenasalised stops *nd, ŋg* (Fox and Grimes 1995). As in Kambera, prenasalisation predicts voicing, but voicing cannot predict prenasalisation. The structural role of prenasalisation in Rotinese is thus more like that in Kambera, or various of the languages of Sulawesi than it is like that in Fijian, or other Oceanic languages. Details differ in other languages, as in Uma of central Sulawesi, where all prenasalised stops are *voiceless* (Martens 1995), or in Yamdena of the southern Moluccas, which has [mp], [nd], [ndʒ] next to plain voiced and voiceless stops (Mettler and Mettler 1990), but no known AN language outside the Oceanic group shows the bilateral predictability of prenasalisation from obstruent voicing and of obstruent voicing from prenasalisation that marks prenasalised obstruents as unambiguous unit phonemes in languages such as Fijian.

Finally, Mithun and Basri (1986:218ff) argue that Selayarese of southwest Sulawesi contrasts prenasalised stop phonemes with N + S clusters in intervocalic position, as in [bóːᵐbaŋ] 'wave' (noun) vs. [bómːbaŋ] 'bamboo skin used for binding'. Since both [b] and [mp] also occur in the language it is clear that voicing and prenasalisation are not bilaterally predictive. Given the different syllabification of [ᵐb] as against [mːb] the question naturally arises whether the 'prenasalised voiced stops' described for Selayarese

might not be postploded nasals (usually written [m^b], etc. as in 4.1.3.), since in other languages these segments 1) arise from clusters of nasal plus voiced stop, 2) are restricted to medial position, and 3) are invariably tautosyllabic.

4.2.4 The special case of geminate consonants

Geminates occupy an ambivalent position between individual segments and consonant clusters. Unlike the distinction between long vowels and sequences of like vowels, which can often be made on the basis of rearticulation, long consonants and sequences of like consonants are often phonetically indistinguishable, although they may be distinguished phonologically. Given their wide distribution and sometimes problematic status they clearly merit a special discussion. Many languages in insular Southeast Asia automatically geminate most consonants after schwa. Historically this gave rise to phonemic geminates in widely separated languages where the schwa has merged with some other vowel. The discussion here is concerned only with phonemic gemination.

Just as consonant clusters tend to be more common and varied in medial position, so geminates are most common intervocalically (Thurgood 1993a). However, geminates also occur word-initially, and rarely in word-final position. Word-initial geminates are found in a number of Oceanic languages, where they have arisen from CV- reduplication. Included in this category are a number of Nuclear Micronesian languages, in which the historical process of geminate formation seems to have been completed, some Polynesian Outlier languages, such as Takuu of the Solomon Islands, and Mussau of northwest Melanesia, where the process of geminate formation appears to be in progress, with generational differences in longer vs. shorter forms (Blust 2001a). Most of these languages have geminates both initially and medially, as with Mussau *kikiau* (older generation) : *kkiau* (younger generation) 'megapode', or *kukuku* : *kukku* 'dove species'. However, Sa'ban of northern Sarawak and Iban of southwest Sarawak are typologically unusual in having geminate consonants only word-initially (Blust 2007b).

All known cases of word-final geminates are found in Micronesia. Rehg (1981:36) notes that Pohnpeian *-l* and *-m^w* are doubled in forms such as *mall* 'clearing in the forest', *m^well* 'shellfish species', *kull* 'cockroach', *lemm^w* 'afraid of ghosts', and *romm^w* 'calm'. Although Pohnpeian limits word-final geminates to sonorants, Puluwat geminates both sonorants and obstruents word-finally: *haakk* 'halved coconut shell', *m^wónn* 'appetite or fondness for', *rall* 'smooth, as sea or cloth', *rapp* 'capsised', *wutt* 'boathouse', *wúttútt* 'choose, select' (cp . *wútúút* 'rub sticks to make fire'), and *yááhh* 'to fly' (Elbert 1972). Elsewhere Elbert (1974:2) notes that word-final geminates occur only in utterance-medial position, when followed by a vowel; utterance-finally they merge with single consonants. Sohn (1975) gives examples of both voiced and voiceless final geminates in Woleaian, but notes that these are followed by voiceless vowels which generally are not written: *tapp* [tap:a̦] 'clan, clan member', *tapp* [tap:i̦] 'kind, sort', *yaff* [jaf:i̦] 'coconut crab', *faŋŋ* [faŋ:i̦] 'itchy'. To summarise, in Chuukic languages such as Woleaian or Puluwat word-final geminates are underlying, and surface as geminates only when followed by a vowel, either in alternations or in the phonetics of careful speech. In Pohnpeian, on the other hand, utterance-final geminates surface as phonetic geminates because they are invariably sonorants, and so have the inherent phonetic capability of extension without the support of a following vowel. Finally, as a result of the loss of unstressed vowels under certain conditions Palauan permits many consonant clusters in initial, medial and final positions.

Among final clusters is the sequence -*ll*, as in *biáll* [bijál:ə] 'k.o. shark', *chull* [ʔúl:ə] 'rain', or *ketit-all* [kətitál:ə] 'is to be picked out'.

Table 4.32 indicates the approximate number of languages with phonemic geminates in major geographical zones. A number in parentheses indicates initial geminates that occur only across a morpheme boundary:

Table 4.32 Austronesian languages known to have geminate consonants

Zone	Total	Initial	Medial	Final
Taiwan	4	1	4	1
Philippines	21	1	21	
Borneo	6	2	4	
Mainland SE Asia	1	1		
Sumatra-Lombok	4		4	
Sulawesi	14	(2)	14	
Lesser Sundas	5		5	
Moluccas	13		13	
New Guinea	0			
Bismarcks	10	1	10	
Solomons	4	1	4	
Vanuatu	7	4	7	
New Caledonia	0			
Micronesia	15	10	15	4
Central Pacific	0			
Total	93	18	90	4

Taiwan: Li and Tsuchida (2006:5) note that "Kavalan is one of two Formosan languages having geminate consonants." They claim that geminates may occur initially, medially or finally in Kavalan, as in *llut* 'beak of a bird' (cf. *luq* 'big eyes'), *babbar* 'hit with fist' (cf. *babuy* 'pig'), or *qan-, qann-* 'eat'. However, final geminates occur only in suffixed forms, thus *qan pa iku* (eat **fut 1sg.nom**) 'I am going to eat', but *qann-i ka zau* (eat-imp imp this) 'Eat this!'. In addition they note that gemination appears to have a morphological function in marking intensity, as in *sukaw* ' bad' : *sukkaw* 'very bad', or *kikia* 'a little, a moment' : *kikkia* 'a very brief moment'. The second Formosan language for which these authors report geminates is the now extinct Favorlang (Ogawa 2003). These are written only word-medially. Tsuchida, Yamada and Moriguchi (1991) have republished vocabularies that were collected during the Japanese colonial period for some of the now extinct aboriginal languages of Taiwan. Among these the lists for Basai and Trobiawan, also show orthographic geminates. Since the relevant data was collected by Erin Asai, who was both a well-known linguist and a native speaker of a language with phonemic geminates, it seems very likely that consonant gemination (noted only word-medially) was phonemic in both Basai and Trobiawan.

Philippines: Common in northern Luzon (Agta, Atta, Bontok, Gaddang, Ibanag, Ifugaw, Ilokano, Isinay, Isneg, Itawis, Itneg, Kalinga, Kallahan, Kankanaey, Yogad). Rare in the central Philippines, reported only for Rinconada Bikol (Jason Lobel, p.c.), Sama Abaknon, and Kagayanen, a geographically separated Manobo language spoken on Cagayancillo Island between Palawan and Negros Islands. Found in the Samalan languages (Central Samal, Mapun, Yakan) and in Tausug of the southern Philippines. Most Mansakan languages (except Kamayo) allow a small subset of consonants to be geminated,

but these are rare. Initial geminates are reported for Mapun (Collins, Collins and Hashim 2001), but these may also be analyzed as beginning with schwa.

Borneo: Found in all four dialects of Berawan (Long Terawan, Long Teru, Long Jegan, Batu Belah), in two of the three Lower Baram languages for which phonetically adequate data is available (Kiput, Narum), and in Sa'ban of northern Sarawak. Although they evidently do not occur in Miri, it is likely that geminates also occur in other Lower Baram languages such as Belait of Brunei, which is a close relative of Kiput. The geminates of Berawan-Lower Baram languages may reflect a single innovation, but those of Sa'ban, which are found only in initial position, are historically independent. Medial geminates are also found in some of the non-standard Malay dialects of Borneo, as in Berau Malay (Collins 1992). Apart from this, Scott (1956:vii) reported variable pronunciations in Iban, such that bases of the shape $C_1 \partial C_1 VCVC$ vary with $C_1 C_1 VCVC$ pronunciations, the result of deleting a schwa between identical consonants in rapid speech, as in *tətawak ~ ttawak* 'large gong' or *gəgudi ~ ggudi* 'a kite'.

Mainland Southeast Asia: The only known geminates are found in nonstandard dialects of Malay where, as in Sa'ban and Iban, they are typologically unusual in being restricted to initial position. Collins (1983b:53) notes that word-initial geminates derived from partial reduplication and syncope are found in most eastern peninsular dialects of Malay. In addition, medial geminates are reported from Sri Lankan Malay (Adelaar 1991), where they arose from automatic consonant lengthening after schwa, followed by regressive vowel assimilation: SLM *kiccil* 'small' (SM *kəcil*), SLM *punnu* 'full' (SM *pənuh*), SLM *tubbu* 'sugarcane' (SM *təbu*).

Sumatra-Java-Bali-Lombok: Geminates are found in Karo, Toba and Angkola Batak of northern Sumatra, and in Madurese. The geminates of Toba and Angkola Batak, which are restricted to *-pp-*, *-tt-* and *-kk-*, are still written as prenasalised stops in the traditional Indic-based syllabary. By contrast, Madurese permits all consonants except the glottal stop to be geminated (Cohn, Ham and Podesva 1999).

Sulawesi: Found in the far north (Talaud), in Totoli and Boano of the Tomini Gulf (Himmelmann 2001:71), and in the region of the southwestern peninsula (most South Sulawesi languages, including Barang-Barang, Buginese, Duri, Makasarese, Mamuju, Mandar, Massenrempulu, Pitu Ulunna Salo, Sa'dan, Seko, and Selayarese). In some of these languages, such as Talaud (Sneddon 1984:22) or Massenrempulu (Mills 1975:1:112) phonological geminates are phonetically preglottalised singleton consonants. This is especially true of the voiced stops, less so of the voiceless stops, and apparently never true of sonorants. In some other languages, such as Konjo, geminates occur after schwa or a vowel intermediate between schwa and *a*, but are not phonemic (Friberg and Friberg 1991). Selayarese (Mithun and Basri 1986), and Barang-Barang (Laidig and Maingak 1999:57) have initial and medial geminates, but initial geminates occur only across a morpheme boundary. Most other South Sulawesi languages appear to permit geminates only intervocalically, although the available descriptions are not always explicit.

Lesser Sundas: There are very few examples in this area. Likely cases include Kodi, Lamboya, and Weyewa of western Sumba, Dhao, spoken on a tiny island of the same name just west of Roti, and the Alor Kalabahi dialect of Lamaholot. Little reliable descriptive data is available for any of these languages, and inferences about gemination must be extracted from short preliminary wordlists. All examples of apparent geminates noted so far in these languages are word-medial.

Moluccas: Geminates are found in several languages of the southwest Moluccas. The 193-word survey vocabulary in Taber (1993) shows medial geminates in Kisar, Luang

(= Letinese), West Damar, Dawera-Daweloor, Dai, Iliun, Roma, North Babar, Central Masela, Emplawas, Tela-Masbuar, and Imroing. Of these, reasonably good descriptive data is available only for Kisar and Luang. van Engelenhoven (2004:49) describes true geminates in Leti both initially and intervocalically, as well as 'pseudo geminates' that alternate with heterorganic consonant clusters in intervocalic position.

Outside this tightly circumscribed area geminates appear to be rare in the Moluccas. Coward (1989:20) reports -*tt*-, -*mm*ᵂ- and -*kk*- in Selaru of the Tanimbar Archipelago, but only one of the examples that she gives is not across a morpheme boundary.

Bismarck Archipelago: Phonetic data is generally inadequate, but it is clear that several of the languages of Manus in the Admiralty Islands have contrastive medial geminates. These include at least Nali, Pelipowai, Mondropolon, Kele, Drehet, Levei, Likum, Lindrou, and Penchal. In addition, as noted already, speakers of Mussau born after about 1930 have geminates in both initial and medial position; older speakers pronounce these with a vowel between the identical consonants, as in *kkiau*/*kikiau* 'megapode', *mmuko*/*mumuko* 'sea cucumber', *rrana*/*rarana* 'mangrove', *gorru*/*goruru* 'seaweed species', *mamma*/*mamama* 'yawn', or *makkile*/*makikile* 'sour' (Blust 2001a). Geminates are unknown from any of the AN languages of New Ireland, New Britain, or New Guinea and its immediate satellites.

Solomons: Rare, but Tryon and Hackman (1983) recorded geminates in several languages of Choiseul Island, including Sengga, Lömaumbi and Avasö. At least some of these are due to automatic lengthening after a historically secondary schwa (not the reflex of PAN *e), as with Avasö *pəssa* 'three', *ləmma* 'five', *təkka* 'taro', *tənne* 'to say', *rərri* 'to fly', but others are contrastive, as *vittu* 'seven' : *vati* 'four', or *ummi-na* 'his beard' : *ulumu* 'blunt'. Initial geminates also are found in the Polynesian Outliers of Takuu, and Bellona, where they have arisen through partial reduplication and syncope. On the basis of a preliminary contact Elbert (1965:439) reported the use of emphatic gemination in Pileni, a Polynesian Outlier spoken in the Santa Cruz Islands. If correct this usage presumably would be morphological. However, nothing similar is reported in the more complete description by Hovdhaugen, Næss and Hoëm (2002).

Micronesia: All languages of Micronesia for which descriptions are available have contrastive gemination. Most have both initial and medial geminates (Gilbertese, Marshallese, Pohnpeian, Mokilese, Chuukese, Puluwat, Woleaian, Ulithian, Sonsorol-Tobi, Carolinian). Chamorro, Palauan, Yapese, Nauruan, and Kosraean appear to lack initial geminates. As already noted, Pohnpeian, Puluwat, and Woleaian also have word-final geminates. Geminates are rare in Yapese and Kosraean, and in Palauan the status of these consonants as units vs. clusters remains in doubt.

Some languages allow only certain consonants to be geminated. Although voiceless stops are geminated in many languages, for example, the glottal stop rarely is part of this set. In some cases historical information may help to explain these limitations. Table 4.33 lists consonants by natural class and indicates which of these natural classes may occur geminated (+) in a given language. If only some members of a natural class may be geminated this is marked +/-. For many languages the available data is insufficient for generalisation, and for this reason the following table is limited to languages represented either by dictionaries or by grammars which explicitly state the range of phonemes that may occur geminated: (1) = voiceless stops except glottal stop, (2) = glottal stop, (3) = voiced stops, (4) = fricatives, (5) = nasals, (6) = liquids, (7) = glides. Blanks indicate the absence of a segment from the phoneme inventory of the language in question:

Table 4.33 Phonemic patterns of gemination in Austronesian languages

Language	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bontok	+	--	+	+	+	+	+
Ifugaw	+	+	+	+	+	+	+
Ilokano	+		+	+/-	+	+	+
Isneg	+	--	+	+	+	+	+
Itawis	+	--	+	+	+	+	+
Kankanaey	+	--	+	+	+	+	+
Mapun	+/-	--	+/-	+/-	+	+/-	+
Yakan	+/-	--	+/-	+/-	+	+/-	+
Tausug	+/-	--	+/-	+/-	+	+	+
Mansaka	--	--	--	--	+	+/-	--
Berawan	+	--	+	--	+	+	--
Kiput	+	--	--	+/-	+	+	--
Sa'ban	+	--	+	+/-	+	+	--

All Ilokano consonants except *h* occur geminated (Rubino 2000:xxxiv). In Ifugaw *h* may be geminated (*bahhó* 'do, achieve, perform'), possibly because *h* < *s, and an older pattern of gemination may have survived the process of lenition. In Itawis *h* and *r* may not be geminated; no examples of -*ww*- appear in Tharp and Natividad (1976), but this may be an accidental gap, since *y* is doubled in *arayyú* 'far, distant.' Similar constraints are found in Mapun, where *j*, *r*, *h*, and *ʔ* do not geminate, and only one example of a geminate glide is known (*awwal,* a borrowing of Malay *awal* 'early'). In Yakan nearly all geminates follow the reflex of *e (formerly schwa, now [ɛ]), although non-geminated consonants also occur in this environment. Other than this, limitations on phonemic content are identical to Mapun. The same is true for Tausug, except that *r* can be geminated.

Mansaka has unusual limitations on gemination. The phonemes of this language are /p t k ʔ b d g m n ŋ s l r w y; a i ə u/, but only *l* and nasals occur long. Moreover, there is a striking imbalance in the frequency of geminated sonorants. Svelmoe and Svelmoe (1990) list 99 examples of -*ll*-, 3 of -*mm*-, 13 of -*nn*-, and 3 of -*ŋŋ*-. In addition, -*ll*- shows very different frequencies in relation to the preceding vowel: 50 examples of -*all*-, 8 of –*ill*-, 31 of -*ull*-, and 10 of -*əll*-.[29] Unlike many other AN languages, in which consonants geminate following schwa, most of if not all examples of geminate *l* in Mansakan languages appear to result from the total assimilation of *d* to *l* in medial clusters derived by syncope: *allaw* 'day' (earlier *adlaw), *tulluʔ* 'finger' (earlier *tudluʔ).

Several languages allow voiceless stops, but not voiced stops to geminate. In Kiput, for example, voiceless stops, nasals, *l* and *r*, and *s* may be geminated, but examples of geminate voiced stops, *f*, or glides are unknown (Blust 2003b). Similarly, in Toba Batak *p*, *t*, *k* occur geminated, but *b*, *d*, *j*, *g* do not. Kiput acquired geminated stops and *s* by the full assimilation of nasals in nasal-obstruent sequences, although other processes must have been operative to produce geminate sonorants. In Toba Batak *mp > *pp*, *nt > *tt*, *ŋk > *kk* and *ns > *ts*. In both of these languages, then, nasals show a stronger tendency to assimilate in manner to following voiceless obstruents than to following voiced obstruents. Berawan dialects, which have developed geminates independently of Kiput, lack this restriction: all consonants were geminated if they were the onset of an open final syllable

---

[29]   Somewhat confusingly, Svelmoe and Svelmoe (1990) write schwa as *u,* and the high back vowel /u/ as *o.*

(Blust 1995a). In Sa'ban, where geminates are found only word-initially there appear to be few restrictions, as attested geminates include *pp-, tt-, kk-, ss-, bb-, dd-, jj-, mm-, nn-, ll-* and *rr-*. The only non-accidental gap is seen with glides, voiceless nasals and voiceless liquids, which are never long.

Other languages that show significant restrictions on gemination include Luang and Kisar of the southwest Moluccas, where only *n* and *l* may occur long. The first of these often occurs across a morpheme boundary with the 3sg possessive suffix *-ne,* but both geminates also occur morpheme-internally. In at least some cases these result historically from the assimilation of an adjacent *d or *n to *l.

4.2.5 Morpheme and word size

Chrétien (1965) showed that some 2,081 of the 2,216 lexical reconstructions in Dempwolff (1938), nearly 94%, are disyllabic. This figure may be slightly inflated, since some of Dempwolff's disyllables have since proved to be trisyllabic. However, any adjustments that need to be made are minor, and it is clear that the great majority of unaffixed word bases consisted of two syllables. Many daughter languages retain this pattern with little change. Table 4.34 provides a syllable count for unaffixed bases that occur in a variant of the Swadesh 200-word list in ten AN languages which have been selected to provide broad geographical coverage, and to illustrate the known range of variation: Paiwan (Taiwan), Tagalog, Tboli (Philippines), Malay (western Indonesia), Tetun (Lesser Sundas), Hitu (Moluccas), Tigak (Bismarck Archipelago), Dehu (Loyalty Islands), Chuukese (Micronesia), and Hawaiian (Polynesia). Because of multiple equivalents for some entries and gaps in others the length of this list varies from one language to the next. Known loanwords are excluded, and reduplicated bases that do not occur in any other form are counted as unaffixed, as with Paiwan *kuma-kuma* 'spider' (but not *qaya-qayam* 'bird', next to *qayam* 'any omen bird'. The category of quadrisyllables also includes words of more than four syllables, although these are rare, and almost always contain two free morphemes. For some less well-described languages, such as Hitu of Ambon, the figure for trisyllables may be slightly inflated by undetected morpheme boundaries. Percentages are given in parentheses:

Table 4.34 Number of syllables for ten Austronesian languages

Language	1-syllable	2-syllables	3-syllables	4-syllables
Paiwan	13 (6.4)	156 (76.5)	31 (15.2)	4 (2.0)
Tagalog (45.5)	3 (1.5)	184 (92.5)	12 (6.0)	
Tboli (37.4)	55 (26.7)	148 (71.8)	3 (1.5)	
Malay (58.0)	2 (1.0)	186 (93.0)	10 (5.0)	2 (1.0)
Tetun (45.5)	17 (8.1)	177 (84.7)	13 (6.2)	2 (1.0)
Hitu (39.7)	5 (2.5)	153 (77.7)	30 (15.2)	9 (4.6)
Tigak (19.3)	38 (20.8)	108 (59.0)	31 (16.9)	6 (3.3)
Dehu (9.8)	72 (27.5)	121 (46.2)	56 (21.4)	13 (5.0)
Chuukese (37.8)	122 (51.3)	100 (42.2)	12 (5.0)	4 (1.7)
Hawaiian (32.5)	12 (5.7)	131 (61.8)	58 (27.4)	11 (5.2)

Although the determination of lexical bases is not always straightforward, the data sampled for Table 4.34 shows an interesting pattern. In addition to the percentages given for the frequency of each syllable type in the corpus, a number appears after all language

names except Paiwan. This figure represents the percentage of basic vocabulary that has been retained from Proto Malayo-Polynesian (Blust n.d. (d)). Since Paiwan is not a MP language a similar percentage cannot be directly computed for it, but the percentage of Paiwan forms that are cognate with the reconstructed basic vocabulary of PMP is 40.2. Languages that are lexically more conservative, then, tend to show higher percentages of disyllables. The most notable exception is Chuukese, which has undergone extensive sound changes, including loss of *-VC in unsuffixed bases. We can say that lexical bases in most AN languages are predominantly disyllabic, and that this pattern tends to be lost in languages that are highly innovative in lexicon, in phonology, or in both.

This observation is important in looking at historical change, as many innovations evidently have been influenced by the predominant disyllabism of the most frequent words. Mean morpheme length in number of syllables for the languages in Table 4.34 is: Paiwan (2.13), Tagalog (2.05), Tboli (1.75), Malay (2.06), Tetun (2.0), Hitu (2.22), Tigak (2.03), Dehu (2.04), Chuukese (1.57), Hawaiian (2.32). This measure clearly is less important than the predominance of a particular type, since it is the frequency of an actual canonical type rather than an abstract representation of norms that is ultimately responsible for psychological salience. Needless to say, while canonical pressures may have existed in PAN, PMP or even POC as a consequence of the psychological salience of disyllabic word form, these pressures have ceased to operate in languages such as Chuukese or Dehu, where fewer than half of basic word bases are disyllables. Although it cannot be fully explored here, the effect of complex affixation obviously alters the frequencies of disyllabic *words* (as opposed to morphemes), but preliminary word counts in sample texts still show a mean frequency closer to 2.0 than to any other whole number.

Languages that have developed a high degree of monosyllabism in relation to the AN norm include: 1) Sa'ban of northern Sarawak, 2) Modang of northeast Kalimantan, 3) the Chamic languages of mainland Southeast Asia, 4) some languages of the Aru Islands in the southern Moluccas, 5) languages of the Raja Ampat Islands in the northern Moluccas, 6) most languages of the Admiralty Islands, 7) Yabem and Bukawa on the north coast of New Guinea, 8) many languages of Vanuatu, 9) some of the languages of New Caledonia, and 10) most Nuclear Micronesian languages. In Oceanic languages monosyllabism tends to be superficial, since surface monosyllables are often underlyingly disyllabic both for nouns and for verbs, as with Seimat *min* 'hand' : *mina-k* 'my hand', *put* 'navel' : *puto-m* 'your navel', *sus* 'breast' : *susu-n* 'her breast', or *hoŋ* 'hear' : *hoŋo-hoŋ* 'hearing', *taŋ* 'cry' : *taŋi-taŋ* 'crying', *un* 'drink' : *unu-un* 'drinking'. In the Chamic languages in particular, but also to a large extent in Sa'ban and Modang of Borneo, and in languages of the Aru and Raja Ampat Islands, a monosyllabic canonical form has penetrated more deeply into the structure of the language.

Languages that have developed an exceptionally large percentage of polysyllabic bases in relation to the AN norm include: 1) Malagasy, 2) Enggano, 3) a number of Oceanic languages reaching from Mussau in the St. Matthias Archipelago, through New Ireland to the western Solomons, and 4) many languages of the D'Entrecasteaux and Louisiade Archipelagos southeast of New Guinea, including Sudest, Dobuan, Molima, and Kilivila. Just as the monosyllabism of Oceanic languages tends to be superficial, so does the polysyllabism of Malagasy, as seen in word pairs such as *láhatra* 'order, row, rank' : *lahàr-ana* 'be arranged, set in order', *lalóna* 'trees whose wood is largely used in house-building' : *lalom-bavy* 'a shrub or tree', *mi-tànika* 'to boil, cook', but *tanèh-ina* 'boiled, cooked', or *póitra* 'apparition, vision' : *poìr-ina* 'be made to appear, be brought to sight',

where surface trisyllables with final *-a* are seen to be underlying disyllables to which a low vowel has been added after a final consonant.

4.2.6 Syllabification

In most AN languages syllable boundaries appear to conform to universal patterns: a CVCVC canonical shape is syllabified CV.CVC, and a CVCCVC canonical shape is syllabified CVC.CVC. As in languages generally syllable boundaries are not isomorphic with morpheme boundaries. The Thao phrase *i-zay a azazak* (that LIG child) 'that child', for example, is syllabified *i.za.ya.a.za.zak*, with syllable breaks splitting a morpheme, or joining part of one morpheme with another (i.-za.y-a). In addition, under suffixation syllable boundaries shift to ensure that consonants are onsets in prevocalic position, as in Tagalog *sulat* [sú.lat] 'writing' : *ka-sulat-an* [ka.su.lá.tan] 'document, piece of writing', *saksák* [sak.sák] 'stab' : *saksak-in* [sak.sa.kín] 'be stabbed'. There are very few exceptions to this pattern. Even in exceptional languages such as Thao which permit a range of word-initial consonant clusters, syllabification appears to split medial clusters that are allowed word-initially: *qtilha* [qtí.ła] 'salt', *aqtalha* [aq.tá.ła] 'pork'. Little data is available to determine how syllabification works in a language where shift of syllable boundaries under suffixation would place a morpheme-final consonant in non-permitted syllable-initial position.

4.3   Phonological processes

The preceding section examined phoneme inventories and distributional constraints, as well as prominent problems in determining what constitutes a segment (prenasalised obstruents, geminates). This section is concerned instead with phonological processes. I will use the term 'process', disconnected from its use in any theory of phonology, to describe both allophony and alternation. Although allophony is synchronically a 'static' set of relationships based on distribution, and alternation a 'dynamic' set of relationships based on substitution, both are products of conditioned sound change, and may result from the same phonological innovation (as with the *h/k* relationship in Toba Batak, described below). I treat most aspects of historical phonology in a later chapter, but since many synchronic processes are residues of historical change it will sometimes be difficult to draw a clear line between synchrony and diachrony here. For reasons of length many processes that are found in individual languages but are not characteristic of large areas are omitted. Exceptions to this practice are made in some cases where a phonological characteristic of a language is particularly well-known. In cases where allophony and alternations reflect the same historical change the two are treated together, and where it appears relevant the absence of phonological processes that are common in other language families will be noted in passing. The general division is into processes that are primarily concerned with consonants as against those that primarily affect vowels.

4.3.1 Processes affecting consonants

Because vowels tend to be more stable than consonants in AN languages, most synchronic residues of sound change affect consonants. This seems to be true more of alternation than of complementation, which presumably is due to the intersection of two factors: 1) most phonological alternations occur at morpheme boundaries, and 2) the

standard CVCVC canonical form of most AN languages rarely is found without at least one of the outside consonants. Vowel alternations appear to be less common than consonant alternations, then, although this may not be true of vocalic allophones, since allophony is not boundary-sensitive. From a diachronic perspective synchronic alternations are of two types: 1) the innovative form of a morpheme occurs under affixation, as in Western Bukidnon Manobo *baləy* 'house' : *bə-valəy* 'build a house' (< *balay), 2) the retained form of a morpheme occurs under affixation, as in Sumbawanese *ratis* 'hundred' : *ratus-an* 'hundreds of' (< *Ratus).

4.3.1.1 *Palatalisation and assibilation*

Palatalisation is a common process in many of the world's languages. Bhat (1978) distinguishes tongue-fronting and tongue-raising as distinct gestures in palatalisations, and points out that velar consonants are most commonly fronted before front vowels, while apical consonants are most commonly raised (hence palatalised) before palatal glides. More basically, this seems to imply that *kiCV or *keCV sequences commonly undergo palatalisation of the initial consonant, while *tiCV or *teCV sequences do not. For palatalisation to be favored with apical stops the targeted segment must be followed by a vowel sequence in which the first vowel is front (*tiVC or *teVC) so that, dependent upon stress, the front vowel loses its syllabicity and becomes a palatal glide.

In the Indo-European family velars have been recurrent targets of palatalisation for millennia in languages as diverse as Sanskrit, Germanic, and Romance. The Slavic languages have undergone historical cycles of palatalisation that have varied in the inclusiveness of the class of consonants affected. However, where the class is restricted it has targeted velars. In many other languages, such as Japanese or Korean, *s* is palatalised before a high front vowel. AN languages depart from these global tendencies, and are remarkably consistent in failing to undergo the most common types of palatalisation. Although many examples of reconstructed *ki- and *si- exist, *k* and *s* in AN languages rarely palatalise, even in phonologically very innovative languages. The same may be said for *d and *n except in the sequences *diV or *niV. As seen already, voiceless palatal fricatives are uncommon in the segment inventories of AN languages, and one reason for this situation is that no palatalisation process has operated to produce them.[30]

Like Finnic languages, many AN languages show assibilation (/t/ > [s]) rather than palatalisation before a high front vowel, before any front vowel (Motu) or before any high vowel (Arosi and other languages of the southeast Solomons). In Tongan *t* and *s* contrast as a result of English loans which have introduced *s* before vowels other than *i* (*sāmani* 'salmon', *same/sāme* 'psalm') or *t* or *d* before a high front vowel (*tī* 'tea; golf tee', *tia* 'deer', *tikite* 'ticket', etc.), but in the native vocabulary [s] is an allophone of *t* before high front vowels. Arosi and some other languages of the southeast Solomons lost *t, and *s became *t. This new *t* then developed a new allophone *s* before high vowels. Subsequent change and borrowing produced *t* in the same environment, but without obscuring the fact that these segments were once in complementary distribution, with [s] before high vowels, and [t] elsewhere. While *t > s/_i, (u), is a historical change, it has left a synchronic residue in many languages in the form of allophony, alternation or neutralisation of the *t/s* contrast before a high front vowel. Languages that show this change include Agta, Atta,

---

[30]   In West Futunan , a Polynesian Outlier in southern Vanuatu, *t > ʃ/_i. While this development may have been *t > s/_i followed by palatalisation of the sibilant, it is more likely that the path was *t > č/_i, an outcome that is preserved in the Aniwa dialect, followed by lenition of the affricate to a palatal fricative.

and Isneg of the northern Philippines, Kelabit of northern Sarawak, Bolaang Mongondow of northern Sulawesi, Motu, Suau and Dobuan of southeastern New Guinea, Bulu, Uvol and Kilenge of New Britain, Lihir, Sursurunga and Tanga of New Ireland, Varisi and Ririo of the western Solomons, the Polynesian Outlier Pileni of the Santa Cruz Archipelago, Mota, Raga, Litzlitz and Anejom of Vanuatu, and Tongan of western Polynesia. A detailed subgrouping picture for these languages is needed to determine how many of the reported cases are historically independent, and how many are retentions of a single change in an immediate common ancestor. In several cases, however, it is clear that the change is relatively recent, since it occurs in Kelabit, but not in the closely related Lun Dayeh, in Bolaang Mongondow, but in no other Gorontalo-Mongondic language, and in Pileni and Tongan-Niue which are only distantly related members of the Polynesian language group.

In at least Isneg, Kelabit, and Bolaang Mongondow the change *t > s/_*i* produced alternations. In all three languages the reason is the same: word bases may be infixed with *-in-*, marking perfectivity in verbs and forming some deverbal nouns. When this happens in a base that begins with *t-* the conditions for assibilation are met, as in Isneg *tabbāg* 'answer' : *t<um>bāg* 'to answer' : *s<in><um>bāg* 'answered', *totón* 'carrying on the head' : *mag-totón*, *toton-án* 'to carry on the head' : *s<in>otón* 'was carried on the head', *tupáʔ* 'cutting up' : *mag-tupáʔ* 'to cut up' : *s<in>upáʔ* 'meat cut into pieces', or Bario Kelabit *tabun* 'heap, pile' : *nabun* 'to heap or pile up' : *s<in>abun* 'was heaped or piled up', *tudo* 'to sit' : *nudo* 'to set something down' : *s<in>udo* 'was set down by someone', and Bolaang Mongondow *taboy* 'smoking of fish, meat, etc.' : *mo-taboy* 'to smoke fish or meat' : *s<in>aboy* 'smoked fish or meat', or *tobut* 'unloading' : *mo-tobut* 'to unload; buy off' : *s<in>obut* 'unloaded; what was unloaded'.

Where palatalisation occurs it almost always affects *t*, not *k* or *s*, and mimics the process of assibilation in other languages, as in Halia (western Solomons), Aniwa (Polynesian Outlier in Vanuatu) *t* → [ʧ]/_*i* but *t* → [t] elsewhere, or Wuvulu (western Admiralty Islands), Banoni (western Solomons) *t* → [ʧ]/_*i,u*, but *t* → [t] elsewhere. It is possible that all instances of *t > s* before high vowels began as *t > ʧ*, with subsequent merger of *ʧ* and *s*, since assibilation is unknown in any language that has a palatal series. While this hypothesis may help reconcile universal tendencies with the way consonants react to adjacent high vowels in AN languages, the failure of *k and *s to undergo palatalisation remains unexplained. If palatalisation is motivated *only* by universal phonetic forces, why should it target velars in some language families but apicals in others?[31] As will be seen, the question why universally motivated changes favor some language groups more than others can also be raised in connection with different subgroups of the same language family.

Finally, a synchronic process that palatalises *s* after *n* is found in a number of AN languages. Adelaar (1981:8) notes that in Mandailing Batak *s* is realised as [ʧ] in the cluster *ns*. Virtually identical allophony is found in Pamona of central Sulawesi, where [ʧ] occurs within a morpheme only following *n*, and where base-initial *s* alternates with [ʧ] (written *c*) when a nasal-final prefix is added, as in *siku* 'elbow' : *san-ciku* 'unit of measure from tip of middle finger to elbow joint'. Bimanese shows further examples, as in the reduplicated base *sinci* 'finger ring', or the alternation in *saŋa* 'branch; bifurcation' : *n-caŋa* 'to branch, bifurcate'. As noted by Ross (1988:71ff) a similar process probably determined the transition from PMP *ns to POC *j ([ndʒ] or [nʧ]). What appears to have operated recurrently in the history of the AN languages is a sequence of two processes.

---

[31]   Among the rare exceptions in AN are the palatalisation of *k*, *ŋ*, and *l* following *i* in Moronene of southeast Sulawesi (Mead 1998:113), and of *ts* to [č] and *s* to [ʃ] before *i* in Amis of eastern Taiwan.

First, the timing of velic opening and oral closure is imperfect in nasal-sibilant sequences, giving rise to pronunciations of the type [nts]. These are heard as sequences of nasal + alveolar affricate, and the affricate [ts] is then reinterpreted as a palatal, presumably because this articulation is less highly marked (Maddieson 1984:38).

4.3.1.2 Sibilant assimilation in Formosan languages

As noted earlier, in order to account for the class behaviour of certain groups of consonants in various AN languages, the term 'sibilant'must be extended to cover a wider range of segments than is usual.

Three Formosan languages show historical evidence of 'sibilant assimilation'. In this process sibilants in adjacent syllables show interference effects resembling those in the English tongue-twister 'she sells sea shells by the sea shore'. The details of this process differ, however, in interesting ways. In Saisiyat PAN *s and *C normally became *h* and *s* respectively, and *S became *ʃ*. However, *s and *C are reflected as *ʃ* if a palatal fricative occurs later in the word: *liseqeS > *liʔʃiʃ* 'nit, egg of a louse', *CiŋaS > *ʃiŋaʃ* 'food particles caught between the teeth.' On the other hand, *j became *z*, and if *S preceded or followed this segment it became *s* rather than *ʃ*: *Sajek > *sazək* 'smell', *qajiS > *œzis* 'boundary of a field'. Normal changes in Paiwan include *s > *t*, *S > *s*, and *C > *ts*. However, interference effects are observed in etymologies such as *Sasaq > *tataq* 'whet, sharpen' (*S assimilates to *s before *s > *t), *Caŋis > *tsaŋits* 'weep, cry' (*s assimilates to *C), or *liseqeS > *lisəqəs* 'nit, egg of a louse' (*s assimilates to *S).

The most extensive system of sibilant assimilation in Formosan languages is found in Thao, where normal changes include *C > θ, *d and *z > *s*, *S > *ʃ*, and *R > *ɬ*. Under assimilative influence, however, the following changes are attested (Blust 1995b):

Table 4.35 Patterns of sibilant assimilation in Thao

PAN segment	Unconditioned reflex	Sibilant assimilation
C	θ	ʃ before ʃ
		ɬ after ɬ
d	s	ʃ before ʃ
		ɬ before ɬ
z	s	ɬ before or after ɬ
b	f	ɬ before ɬ
S	ʃ	s before s

Sibilant assimilation continues to operate in the contemporary language, leading to variant pronunciations such as *fiɬaq ~ ɬiɬaq* 'to spit', *ɬaθkað* or *ɬaɬkað* 'bumblebee', or *ʃmauɬin ~ ʃmauʃin ~ ɬmauɬin* 'to swing'.

4.3.1.3 Nasal spreading

Most descriptions of AN languages say little about allophonic nasality. However, where information is available nasality usually does not spread from a syllable coda as in many Indo-European languages, but rather from a syllable onset. Field data supporting this claim are available for languages in Taiwan, the Philippines, Borneo, Sumatra, and Java.[32] In

---

[32]   Schachter and Otanes (1972:8, 21) report variously that "As in American English, nasalised vowels frequently occur in Tagalog before or after the nasal consonants…', and 'vowels that precede nasal

addition to nasalising the vowel that immediately follows a primary nasal consonant, onset-driven vowel nasalisation typically penetrates certain consonants to affect vowels that are further removed from the source of nasality. The class of consonants that is transparent to nasal spreading varies, but often includes *ʔ*, *h*, *w*, *y*, and sometimes *l*, as seen in Table 4.36 (+ = transparent, - = not transparent, Ø = consonant not found in this language):

Table 4.36 Consonants that are transparent to nasal spreading

Language	**w**	**y**	**ʔ**	**h**	**l**	**r**
Kiput	–	–	–	–	–	–
Rejang (Sumatra)	+	+	–	–	–	–
Long Anap Kenyah	+	+	+	Ø	–	–
Mukah Melanau	+	+	+	+	–	–
Sundanese	–	–	+	+	–	–
Uma Juman Kayan	+	+	+	+	+	–

The Uma Juman dialect of Kayan (central Sarawak) can serve to illustrate: *ñaʔuy* [ɲãʔũj] 'to scream', *ñuhuʔ* [ɲũhõʔ] 'to rise, of the river', *ñiwaŋ* [ɲĩwãŋ] 'thin, of people and animals', *ŋuyuʔ* [ŋũjõʔ] 'provisions, food taken on a journey', *m-alit* [mãlĩt] 'heal'. In some languages nasal spreading through an oral consonant is so complete that in a nasalising environment the underlying oral consonant can no longer be distinguished from a nasal consonant with similar place features. In Ngaju Dayak of southeast Borneo, for example, the addition of a prefixal nasal before the first vowel of a word base triggers rightward nasal spreading. Among the consonants that are transparent to nasal spreading are *w* and *y*, and although *w* remains a glide in nasalising environments, *y* alternates with *ñ*, as in *kayu* 'wood; firewood' : *ma-ŋañu* 'to gather firewood', *uyah* 'salt' : *m-uñah* 'to salt something', or *payoŋ* 'umbrella' : *ma-mañoŋ* 'to shelter with an umbrella'. In Narum of northern Sarawak a similar process affects *l*, as in *alaut* 'boat' : *ŋ-anaut* 'paddle a boat', *pilaiʔ* 'choice' : *minaiʔ* 'choose', or *hulet* 'skin' : *m-unet* 'to skin'. It is notable that although *y* alternates with *ñ* in Ngaju Dayak, this does not happen in Narum, which also has a palatal nasal and allows nasality to spread through glides. Also, *l* alternates with *n* in Narum but not in Ngaju Dayak, although it is unclear whether *l* is transparent to nasal spreading in this language. Second, the kinds of alternations seen in Ngaju Dayak and Narum must be independent (the languages are widely separated on the island of Borneo, and have had no historical contacts), yet they are curiously circumscribed in geographical terms, as nothing similar has ever been reported elsewhere in the AN language family.

In languages such as French final nasals that have disappeared have left a trace of their former presence through nasalisation of the preceding vowel. Many AN languages have lost original word-final consonants, but in only one known case has this resulted in contrastively nasalised final vowels. In the Lamaholot dialects of the eastern Lesser Sundas (eastern Flores, Adonara, Solor, Lembata) word-final consonants have been preserved in some dialects, but lost in others. Where they have been lost, vowels that preceded an earlier final nasal are now contrastively nasalised, as in PMP *zalan > Ile Ape *laran*, Lamalera *larã* 'path, road', PMP *hikan > IA *ikan*, L *ikã* 'fish', PMP *quzan > IA *uran*, L *urã* 'rain' (cf. PMP *anak > IA, L *ana* 'child', *taneq > IA, L *tana* 'earth, land', *deŋeR >

---

consonants are normally nasalised." This statement appears to be impressionistic, and is open to question.

IA, L *dəŋa* 'to hear'). It is unclear from the available data whether regressive nasalisation remains as part of the synchronic grammar. These developments imply a pattern of leftward nasal spreading that is rare in AN languages for which adequate phonetic information is available. The data given by Keraf (1978) show some inconsistencies, as with PMP *wanan > IA *wanan*, L *fana* 'right side' (no nasalisation), or PMP *mata > IA *mata-k*, L *matã* 'eye' (unexpected nasalisation), but in most cases there are plausible explanations for these. Lamalera *fana* for expected *fanã*, for example, may have a final nasal vowel which is 'masked' by the preceding nasal for a recorder who was writing in Indonesian, a language with rightward nasal spreading. Likewise, IA *mata-k* shows an obligatory possessive pronoun ('my eye') where the default possessive marker in most languages is the 3sg. *-n*, as in PMP *buaq > IA *wua-n*, L *fuã* 'fruit' (both < *fua-n 'its fruit'). Given this larger pattern L *matã* 'eye' probably reflects earlier *mata-n 'his/her/its eye'. For the Lamaholot dialect of Lewolema Pampus (1999:29ff) reports what he describes as 'morphophonemic' processes nasalising final vowels, but the examples he gives suggest that nasalisation actually is a morpheme, since in at least some cases it appears to be the sole carrier of semantic changes in bases that differ in oral vs. nasal vowels: *dikəʔ* 'real, true' : *dikə̃ʔ* 'truth', *goʔe* 'I' : *goʔẽ* 'my', *pəlate* 'warm' : *pəlatĩ* 'something that is warm', *gike* 'spicy' : *k<ən>ikĩ* 'spicy food', *laŋoʔ* 'house' : *laŋũʔ* 'his/her house', *aho* 'dog' + *anaʔ* 'offspring' > *aho anãʔ* 'puppy'.

As in some other language families, when they are adjacent to laryngeal segments vowels may also be nasalised by 'rhinoglottophilia' (Matisoff 1975). In the Polynesian Outlier language of Rennell Island in the southern Solomons, for example, *hahine* 'female, woman' is [hãhĩne], with clear nasalisation of the low vowel [ã], and less obvious nasalisation of the high vowel [ĩ]. In the Malay dialect of Kedah non-final *r* is realised as [ʁ], and final /r/ as [q]. Before [q] high vowels are pronounced with a mid-central offglide (as the tongue is pulled down to produce the uvular stop), and nasalised, as in [ʔaljə̃q] 'to flow', but [ʔaliʁán] 'current'. In a few languages rhinoglottophilia has produced contrastive nasality for reasons that remain unclear, as with Narum (northern Sarawak) *hããw* '2sg.' (< *kahu), but *haaw* 'rafter' (< *kasaw).

Finally, Robins (1957) reported a puzzling pattern of nasal spreading in Sundanese that initiated a decades-long controversy among phonologists trying to make sense of the data he presented. Based on kymograph recordings, Robins maintained that Sundanese vowels are nasalised after a nasal consonant and that nasality spreads rightward until interrupted by any segment other than glottal stop or *h*. This in itself is unexpected, since if glottal stop and *h* are transparent to nasal spreading, glides normally are as well (Table 4.36). Most unexpectedly, however, nasality is said to skip over the plural infixes *-al-* and *-ar-* to nasalise the *second* vowel after the infix, as in *niʔir* [nĩʔĩɾ] 'pierce' : *n<al>iʔir* [nãliʔĩɾ] 'pierce (pl.)', *miak* [mĩãk] 'stand aside' : *m<ar>iak* [mãriãk] 'stand aside (pl.)', but *marios* [mãrios] 'to examine' (where [o] is not nasalised). This reported behaviour gave rise to several proposed explanations, beginning with Anderson (1972), who argued that *-ar-* is an underlying prefix that is infixed by metathesis after nasal spreading. Since the infixes *-um-* and *-in-* show normal nasalising behaviour, however, it became necessary to maintain that infixes with a nasal consonant are inserted prior to nasalisation, but infixes with an oral consonant are inserted after nasalisation. It is troubling that there is no historical support for such a scenario, since PAN *-ar- must be reconstructed as an infix. Some subsequent researchers have questioned Robins' interpretation of his data. According to Latta (1977) the original kymographs show nasality spreading through the infixes *-al-* and *-ar-* in some cases, and so do not consistently support Robins' claims. Moreover, there is a plain but

unstated inconsistency in the implication that glides block nasal spreading, when the transcription [mĩãk] clearly should be [mĩỹãk], with a nasalized automatic transitional glide. Cohn (1993a), who has made independent acoustic recordings of the language, supports the general accuracy of Robins' kymographs, differing primarily in her claim that *l* is transparent to nasal spreading, while *r* is not.

4.3.1.4 Nasal preplosion and postplosion

A number of AN languages in western Indonesia and mainland Southeast Asia have preploded nasals. These segments, which reflect simple nasal consonants, are found only word-finally, as in Selako [tadʒaᵖm] (Malay *tajam*) 'sharp', Selako [ba-dʒaaᵗn] (Malay *bər-jalan*) 'to walk', or Selako [tuwaᵏŋ] (Malay *tulaŋ)* 'bone'. In most languages plain and preploded nasals are in complementary distribution, with plain allophones occurring non-finally, or word-finally in syllables that begin with a nasal consonant, as with Selako *aŋan* (Malay *taŋan*) 'hand', or Selako *aŋin* (Malay *aŋin*) 'wind'. Preploded nasals are found in Bonggi, spoken on Banggi and Balembangan Islands north of Sabah, most of the Land Dayak languages of southwest Borneo, Tunjung of southeast Borneo, Lom of Bangka and Belitung Islands between Borneo and Sumatra, in some speech registers of Rejang in southern Sumatra, and in some dialects of Mentawai in the Barrier Islands west of Sumatra. In other languages the merger of nasal consonants with homorganic voiceless stops under the same conditions that govern nasal preplosion (word-finally, except in syllables that begin with a nasal) shows that preploded final nasals were once found in languages that no longer have them. Languages that have no preploded nasals now, but must have had them at an earlier time include Urak Lawoi', a Malay dialect spoken in southwest peninsular Thailand, Northern Roglai of Vietnam, Tsat, of Hainan Island in southern China, most dialects of Mentawai, and some Land Dayak languages. This can be seen in Urak Lawoi' *asap* 'bitter; tamarind tree' (Malay *asam*), *hujat* 'rain' (Malay *hujan*), or *tulak* 'bone' (Malay *tulaŋ*), with change of nasal to homorganic voiceless stop only word-finally, next to *tanam* 'to plant, bury' (Malay *tanam*), *aŋɛn* 'wind, air' (Malay *aŋin*), or *kuniŋ* 'yellow' (Malay *kuniŋ*), with no preplosion because the final syllable has a nasal onset.

This innovation is both widespread and independent in many languages, yet it is heavily concentrated in Borneo, Sumatra, and mainland Southeast Asia, and is found nowhere else in the AN language family.[33] Neither contact nor subgrouping provides a plausible explanation for this distribution, which is maximally discontinuous even on the island of Borneo (Bonggi in the far north, Selako and Land Dayak in the far southwest, Tunjung in the far southeast), and includes individual members of different well-defined subgroups, but excludes other languages within the same subgroup. Nasal preplosion thus appears to be a drift, but one that raises questions both of motivation and of geographical circumscription: if this change is motivated by some more fundamental property of WMP languages, why is it limited to a teritorially-defined subset of them? As will be seen, this problem recurs in various guises throughout the AN family—certain changes that

---

[33]   Despite a superficial similarity, the postnasalized consonants of languages in northern New Caledonia (Haudricourt 1964, Ozanne-Rivierre 1975, 1995) are structurally and historically quite distinct from the preploded nasals of insular Southeast Asia. Whereas the latter occur only word-finally, and never in a syllable with nasal consonant onset, the postnasalized consonants of New Caledonia typically are found in syllable-initial position. Likewise, postploded final nasals have arisen historically from simple nasals, while the postnasaled consonants arose from loss of a vowel between a stop and a following nasal, giving rise to initial consonant clusters that have been reanalyzed as unit phonemes.

presumably are motivated by some generally shared property are far more common in one geographical region than in others.

The motivation for nasal preplosion seems to be associated with the patterning of allophonic vowel nasality in these languages. Final nasals cannot be preploded in syllables that begin with a nasal consonant because the rightward spread of nasality ensures that vowels preceding final nasals are fully nasalised. Final nasals in syllables that begin with an oral consonant are preceded by oral vowels, but in languages that lack preploded nasals some contragrade nasality probably colours the vowel preceding a final nasal, just as some contragrade nasality colours vowels following nasal consonants in languages with leftward-spreading nasality such as English. In this context nasal preplosion appears to be a reaction to contragrade nasality, an attempt to prevent any nasality at all from spreading leftward from a word-final nasal.

As noted in 4.1.3, a number of languages in western Indonesia also have postploded nasals word-medially, as in Narum of northern Sarawak (stress is final) [tum^{b}áŋ] 'to fell trees', [mən^{d}áwʔ] 'to bathe', [miñ^{ʤ}ám] 'borrow', or [əŋ^{g}áp] 'scorpion'. These segments reflect prenasalised voiced obstruents in which the duration between velic closure and oral release is shortened until the obstruent is barely perceptible, if at all. For some languages postploded medial nasals are described as simple nasals in which the following vowel is fully oral, just as it would be following *-mb-*, *-nd-*, *-nj-* or *-ŋg-*, but such nasals are longer than simple nasal consonants preceding nasalised vowels, and in Acehnese they are said to show reduced nasal airflow (Durie 1985:15). Because some languages have both types of segments, and their geographical distribution is nearly coterminous, it is tempting to see these articulations as inherently related. However, the mechanisms of nasal preplosion and nasal postplosion appear to be fundamentally different.

4.3.1.5   *Nasal substitution and pseudo nasal substitution*

Nasal substitution is a process found in nearly all WMP languages, and in a few other AN languages. It is usually triggered by reflexes of PMP *maŋ- 'active verb', and *paŋ- 'agent/instrument', which have several phonologically conditioned allomorphs. In Malay, for example, the prefixal coda is unchanged before vowel-initial bases, as in *məŋ-ukur* 'to measure'(base: *ukur*), but assimilates in place to voiced obstruents, as in *məm-bantu* 'to help' (base: *bantu*), and is deleted before nasals, liquids or glides, as in *mə-makan* 'to eat' (base: *makan*), *mə-lukis* 'to draw' (base: *lukis*), or *mə-waris-i* 'to inherit, as property' (base: *waris*). With voiceless obstruents the prefixal coda assimilates in place and then replaces the base-initial consonant, as in *mə-milih* 'to choose' (base: *pilih*), *mə-nulis* 'to write' (base: *tulis*), *mə-ŋulit-i* 'to skin, excoriate' (base: *kulit*), or *mə-ñurat-i* 'to write to someone', base: *surat*. Dempwolff (1934) used the terms 'nasal accretion' and 'nasal substitution' for patterns of the type seen in Malay *məm-bantu* vs. *mə-milih*. As can be seen from these examples, replacement nasals are normally homorganic with base-initial obstruents, but bases with *s-* take the palatal nasal in languages that have an *n* : *ñ* distinction.

Although the basic groundplan of nasal substitution is similar cross-linguistically, differences of detail in individual languages make this a challenging phenomenon from the standpoint both of linguistic reconstruction and of general phonological theory. To mention only a few of the more salient points, the class of consonants that undergoes nasal substitution differs across languages. Some languages (such as Malay) permit nasal substitution only for voiceless obstruents, while others, as Tausug, include *p-*, *t-*, *k-*, *s-* and *b-*, but not *d-* or *g-*. Bikol allows nasal substitution for bases that begin with *p-*, *t-*, *k-*, *s-*, *b-*

and *d-*, but not *g-*, while Bontok nasal substitution affects all base-initial obstruents. In still other languages, such as Tagalog, nasal substitution affects base-initial voiceless obstruents, and *some* bases that begin with *b-*, but not others. A second variable feature is the size of the prefix which triggers nasal substitution. In languages such as Malay the verbal prefix that triggers nasal substitution has the shape *məŋ-*. In other languages, such as Kelabit of northern Sarawak, or Javanese the 'prefix' is just nasal substitution, as in Kelabit *mətad* 'to separate' (base: *pətad*), *munuʔ* 'go to war' (base: *bunuʔ*), *nələn* 'to swallow' (base: *tələn*), *niŋər* 'to hear' (base: *diŋər*), *ŋulit* 'to skin, excoriate' (base: *kulit*), *ŋətəp* 'to bite' (base: *gətəp*) or *niri* 'to straighten' (base: *siri*). Most Kenyah dialects of central Borneo, and some other languages show both long and short forms of prefixes with nasal substitution, as they are evidently in transition from long forms to short forms as a result of the general tendency to favor a disyllabic word shape. A third variable feature of nasal substitution is less common, but surfaces in several widely separated languages. If a lexical base has a medial prenasalised consonant this will affect prefixal allomorphy. In Ngaju Dayak of southeast Borneo, for example, nasal substitution normally affects base-initial voiceless obstruents, but not base-initial voiced obstruents (*batal* 'boil, abscess' : *mam-batal* 'to swell up'). Base-initial voiced obstruents undergo nasal substitution, however, if there is a prenasalised obstruent later in the word, as in *buŋkus* 'package' : *ma-muŋkus* 'to wrap, make into a package.' In Timugon Murut of Sabah as described by Prentice (1971) the same affixed base may show syntactically contrasting patterns of nasal substitution and nasal accretion, as with *ma-nutu* 'subject will pound object' vs. *man-tutu* 'subject will pound [object]' (base: *tutu*). Bases that contain a medial prenasalised obstruent show dissimilation of the first cluster, hence *ma-numbuk* 'subject will thump [object]' vs. *ma-tumbuk* 'subjects will thump each other' (base: *tumbuk*). Other variations are more exotic, and need not concern us here (for a fuller discussion cf. Blust 2004a, 2012b).

In addition to these cross-linguistic variations, some examples of nasal substitution that were phonetically transparent in their inception have become opaque as a result of sound change. Examples of this kind fall into two classes: 1) cases where sound change has altered the place features of a base-initial obstruent, 2) cases where sound change has transformed a nasal to an oral consonant. Most base-initial consonants are extremely stable in those languages that make active use of nasal substitution, but there are a few exceptions. Several languages in the Philippines and western Indonesia have undergone the sound change *s > h, and these languages consequently have a synchronic alternation of *n* or *ñ* with *h*, as in Ifugao (northern Luzon) *ma-náaŋ-ka* 'begin cooking' (base: *háaŋ*), *ma-nígid* 'person who sweeps something' (base: *hígid*), *ma-nukáp* 'cover a basket' (base: *hukáp*), or Kayan (central Borneo) *ñatuŋ* 'swim' (base: *hatuŋ*), *ñərut* 'sip' (base: *hərut*), *ñigəm* 'hold' (base: *higəm*). Malagasy also shows an *h/n* alternation in active verbs, but one that has arisen from different sources, *k > h, and *ŋ > n. Together these changes have converted an earlier *k/ŋ* alternation into an *h/n* alternation, as in *huditra* 'skin, bark' : *ma-nuditra* 'to flay, remove the skin' (< *kulit : *ma-ŋulit), or *havitra* 'hook' : *ma-navitra* 'to hook something' (< *kawit : *ma-ŋawit). Other sound changes have introduced different patterns of phonetic opacity into the process of nasal substitution. In Gorontalo of northern Sulawesi, for example, *b became *h* before *u*, and an earlier pattern in which *b* alternated with *m* under conditions of nasal substitution has been transformed to one in which *h* alternates with *m*, but only before a high back rounded vowel, as in *huluto* : *mo-muluto* 'to husk a fruit' (<*bunut : *ma-munut), *huwato* : *mo-muwato* 'to lift' (< *buhat : *ma-muhat), or *huwoŋo* : *mo-muwoŋo* 'to split'. In addition, Gorontalo has some bases that

begin with *bu-* (presumably products of borrowing), and these also alternate with *m-* under the same conditions of prefixation. Synchronically, the basis of the *h/m* alternation in Gorontalo is thus thoroughly disguised.

Although the preceding cases involve phonological alternations that are phonetically opaque they may still be called examples of nasal substitution, since an oral consonant is replaced by a nasal consonant under prefixation. In Palauan, sound change has caused a rearrangement of phonological relationships that in some respects is even more drastic. Here, the sound change *n > l has produced a typologically aberrant consonant system in which the only nasals are *m* and *ŋ*, and in which a phonetically transparent pattern of *t/n* nasal substitution has been replaced by a phonetically opaque alternation in which *d* ([ð]) alternates with *l* under the same conditions in which nasal substitution is otherwise found: *dáləm* : *mə-láləm* 'to plant' (< *tanem : *ma-nanem), *dúb* 'dynamite tree (used to obtain fish poison)' : *mə-lúb* 'to bomb, dynamite' (< *tuba : *ma-nuba), as against *káud* 'dam' : *mə-ŋáud* 'to dam' (< *kapet : *ma-ŋapet), or *kárd* : *mə-ŋárd* 'to nibble, bite' (< *karat : *ma-ŋaRat). Clearly the *d/l* alternation in Palauan can no longer be described as 'nasal substitution', yet it patterns exactly like nasal substitution with base-initial velar stops.

A second type of nasal substitution apparently is restricted to the Ponapeic languages of Micronesia, of which Pohnpeian and Mokilese are the best known. The type of nasal substitution found in these languages can be described as dissimilatory: oral geminates that arise through reduplication dissimilate by altering the first half of the geminate to a homorganic nasal, as in *pap* 'swim', *pampap* (< *pap* + *pap*), *sas* 'stagger', *sansas* (< *sas* + *sas*), or *did* 'build a wall', *dindid* (< *did* + *did*). A second nasal substitution process in Pohnpeian affects identical non-coronal consonants that come together in the flow of speech without regard to reduplication. Ponapeic nasal substitution is thus fundamentally distinct from the more widespread type of nasal substitution characteristic of the AN languages of the Philippines and western Indonesia.

Finally, there also is a phenomenon that can be called 'pseudo nasal substitution' (PNS). Unlike nasal substitution, which is triggered by prefixation with *maŋ-, or *paŋ-, PNS is triggered by infixation with *-um-. As noted by Chrétien (1965), PAN strongly disfavored non-identical labials in successive syllables. Bases that reflect *p<um>VCVC, *b-um-VCVC or both therefore lose CV- in many AN languages, as with Thao *patash* 'write' : (*p<um>atash >*) : *matash* 'to write', or *pulhbuz* 'make something sink' : (*p<um>ulhbuz >*) : *mulhbuz* 'to sink'. Nasal substitution and PNS can be distinguished by distribution: PNS operates only with bases that begin with a labial consonant, while true nasal substitution is not constrained in this way. In addition, unlike nasal substitution, which is unknown in Formosan languages, PNS is found in both Formosan and Malayo-Polynesian languages.

4.3.1.6   *Other alternations of initial consonants*

Other base-initial consonant alternations can be classified as lenitions or fortitions. Lenitive alternations of initial consonants occur in languages that have lenited intervocalic consonants, and built some words with vowel-final prefixes. Western Bukidnon Manobo has undergone the historical changes *b > -v-, *d > -z- and *g > -ɣ-. Morpheme-internally this has had no consequences for the synchronic phonology, but at morpheme boundaries it has when CV- prefixes or –VC suffixes are added, or a base is reduplicated: *basuk* 'till the soil' : *mə-vasuk* 'industrious in farming', *bulawan* 'gold' : *kə-vuləwan-an* 'wealth, splendor', *dukiləm* 'become night' : *mə-zukiləm* 'night', *duwa* 'two' : *ikə-zuwa* 'second', *gəvuʔ* 'weakness, fragility' : *mə-ɣəvuʔ* 'weak, fragile', *guraŋ* 'old' : *mə-ɣuraŋ* 'old person; old', *bunsud* 'to set some long object on its end' : *bunsuz-an* 'area at the foot of a ladder',

buwad 'to reproduce' : *ke-vuwaz-an* 'one's posterity', *basa* 'to read' : *basa-vasa* 'a proverb framed in archaic language', *duwa* 'two' : *duwa-zuwa* 'to doubt', etc.

Nasal accretion before a base-initial obstruent usually affects the place features of the nasal, but has no effect on the non-nasal consonant with which it comes in contact. However, in a few languages base-initial obstruents alternate under prefixation, as with Pamona *salili* 'side' : *mon-calili* 'carry on one side', *sindu* 'ladle' : *man-cindu* 'scoop up with a ladle', *ma-repe* 'flat' : *mon-depe-gi* 'pound flat' or Bimanese *saŋa* 'fork of a branch' : *n-caŋa* 'to fork, of a branch', *raʔa* 'blood' : *n-daʔa* 'to bleed'. Synchronically these alternations are best described as fortitions, since base-initial consonants have greater constriction after a nasal-final prefix than in the unaffixed base. Historically, however, these alternating bases began with *s or *d, and the s/c ([s]/[ʧ]) alternation is thus a product of fortition, but the *r/d* alternation is a product of lenition. Malagasy shows a similarly complex history in the alternations associated with nasal accretion: *v-* : *mb-*, *r-* : *ndr-*, *z-* : *ndz-*, and *h-* : *ŋg-* have arisen historically from the lenition of PMP *b, *d, *z, and *g (where *z was a voiced palatal affricate) when not preceded by a nasal consonant, but *l-* : *d-* has arisen both from the lenition of *d and the fortition of *l under specific conditions, as in *lalina* 'deep, profound' : *man-dalina* 'to deepen' (PMP *dalem* 'deep'), but *lua* 'vomit' : *man-dua* 'to vomit' (PMP *luaq* 'spit out food').

Many languages of eastern Indonesia show voicing alternations of base-initial consonants that correlate with person-marking in the verb. Stresemann (1927:119-125) described several patterns of 'verbal conjugation' (*Flexion*) in languages of the central Moluccas and Lesser Sundas. Historically these have arisen from the fusion of the base-initial consonant with a preposed person marker (sometimes called 'agreement markers') *ku- '1sg', *mu- '2sg', *na- '3sg', *ta- '1pl in', *ma- '1pl ex', *mi- '2pl', and *da- '3pl'. Sika of eastern Flores illustrates a pattern typical of much of this region:

Table 4.37 Person and number marking by initial consonant alternation in Sika

PMP		*butbut 'pluck'	*taŋis 'weep, cry'	*kita 'see'
1sg	aʔu	pupu	tani	ita
2sg	ʔau	bupu	dani	gita
3sg	nimu	bupu	dani	gita
1pl/i	ita	pupu	tani	ita
1pl/e	ami	bupu	dani	gita
2pl	miu	bupu	dani	gita
3pl	rimu	pupu	tani	ita

In effect, stops that came to be prenasalised by the addition of proclitic pronouns and syncope of the prefixal vowel yield voiced reflexes, as in pre-Sika *kau mu-pupu (> *kau m-pupu > *ʔau bupu*), and those that were not prenasalized in this way yield voiceless reflexes, as in pre-Sika *aku ku-pupu (> *ʔau k-pupu) > *ʔau pupu*.

A pattern of initial consonant 'mutation' superficially resembling this is also widely attested in the languages of central Vanuatu. Crowley (1991:180) notes that many of these languages have two series of verb-initial consonants which, following Schütz (1968) he calls 'primary' and 'secondary'. In the Nakanamanga language of north Efate the primary set consists of *v-*, *w-*, *k-* and *r-* and their secondary counterparts of *p-*, *pʷ-*, *ŋ-* and *t-*. The primary consonants reportedly occur after the preverbal particles *pʷa* 'imperative', *ŋa* 'future, subjunctive', *pe* 'conditional', and in some other grammatical contexts, thus *e pe vano* 'if s/he goes', but *e pano* 's/he goes/went'. The details of this alternation differ from

language to language, but in every case the choice of verb-initial consonant is grammatically conditioned, generally by a difference of tense or mood. Historically, these grammatically conditioned differences arose from phonological conditions involving the contrast of oral grade and nasal grade consonants.

4.3.1.7 Subphonemic alternations

The examples of *s*/*c* alternation mentioned above raise another point: use of the expression 'morphophonemic alternation' for such processes is misleading, since the segments that alternate are not phonemically distinct. In both Pamona and Bimanese [ʧ] is an allophone of /s/ after nasals, yet this alternation is formally identical to that of *r* and *d*, which are distinct phonemes. To maintain terminological consistency alternations of allophones should be called 'morphophonetic alternations', and this is good reason for abandoning the older terminology and calling all such segment substitutions under affixation 'phonological alternations.' A similar set of relationships is seen in Toba Batak of northern Sumatra, where [h] occurs before vowels and [k] elsewhere, but under suffixation with *-an* or *-on* base-final [k] alternates with [h], as in *anak* 'child' : *par-anah-on* 'relationship of father and child', or *lapuk* 'mildew' : *lapuh-on* 'mildewed.' In this case both complementation and alternation are the outcome of a single sound change *k > h/_V.

4.3.1.8 Alternations of final consonants

With rare exceptions, as where medial liquids or glides alternate with nasals in Narum or Ngaju Dayak, segmental alternations, particularly where they affect consonants rather than vowels, normally occur at the edges of morphemes. Most alternations considered so far have been concerned with base-initial consonants. Many AN languages also have alternations that affect final consonants, final vowels, or both. Most languages that show alternations of final consonants can be divided into two types: those in which a limited class of word-final consonants alternates with a larger class of morpheme-final consonants, and those in which zero alternates with an essentially open class of thematic consonants. A few languages show final alternations that do not fit comfortably into this general schema.

To begin with the first type, most South Sulawesi languages as reported by Mills (1975) show extreme reduction of word-final contrasts. Mandar allows only /ʔ n~ŋ r l s/, Sa'dan Toraja only /ʔ k n ŋ/, and Buginese and Makasarese only -*ʔ*, and -*ŋ*, although historically –*r*, -*l*, and -*s* have merged as -*ʔ* in Buginese, but have been preserved by the addition of an echo vowel in Makasarese. Since all of these languages have active suffixes the final consonants of bases may appear either word-finally, or word-medially. According to Mills (1975:451) Buginese -*ʔ* may alternate with -*r*, -*s*, or -*k* in suffixed forms, although this appears to be sporadic, and -*ŋ*, which historically reflects final *m, *n, and *ŋ, shows no alternation at all under suffixation: *inuŋ* 'drink' : *aŋ-inuŋ-ən* 'place to drink' (< *inum). Sangir, of northern Sulawesi, shows somewhat clearer patterns of alternation between a reduced set of word-final consonants and the fuller set from which they derive historically, and which continue to function as underlying segments in the synchronic grammar. In Sangir as shown by Sneddon (1984) word-final *p, *t, and *k merged as glottal stop, and word-final *m, *n, and *ŋ merged as -*ŋ*. Under suffixation each underlying consonant that is neutralised as -*ʔ* or -*ŋ* surfaces in its distinctive form, except that *-t and *-n surface as -*k* and -*ŋ*: *ma-nəda*ʔ 'to set, of the sun' : *sədap-əŋ* 'west' (PMP *sejep), *m-awi*ʔ 'to climb' : *la-awik-aŋ* 'place where one climbs up' (PMP *abit), *balu*ʔ 'to sell' : *palahəm-baluk-aŋ* 'market place' (Proto Sangiric *baluk), *maŋ-inuŋ* 'to drink' : *inum-aŋ* 'a drink' (PMP

*inum), *maŋ-ambuŋ* 'cover with earth, sand, or leaves' : *ambuŋ-aŋ* 'filled up, heaped up'
(PMP *ambun), *ma-niruŋ* 'to shelter' : *siruŋ-aŋ* (in poetry *sirum-aŋ*) 'place where one
seeks shelter' (Proto Sangiric *siduŋ). These relationships are summarised in Table 4.38:

Table 4.38 Alternations of *-ʔ* and *-ŋ* with other segments under suffixation in Sangir

*-p	>	-ʔ ~ p	*-m	>	-ŋ ~ m
*-t	>	-ʔ ~ k	*-n	>	-ŋ ~ ŋ
*-k	>	-ʔ ~ k	*-ŋ	>	-ŋ ~ ŋ ~ m (poetic)

Sangir shows some apparent back-formations in which original final velar nasals
alternate with word-medial bilabials in what appears to be a figurative or poetic speech
register, as with *sirum-aŋ* (poetic for *siruŋ-aŋ*) 'place where one seeks shelter', or *mə-tuluŋ*
'to help' : *tuluŋ-aŋ* (poetic: *tulum-aŋ*) 'to help' (PMP *tuluŋ). The origin of this alternation
is unclear. Primary historical change produced three underlying sources for both *-ʔ* and *-ŋ*,
but two alternations for the former and only one for the latter. The unique alternation of
word-final *-ŋ* with word-medial *-m-* would have provided an unambiguous model for back-
formations which could have undergone secondary semantic reshaping. Although the data
sample is too small to establish statistical significance, if back-formation took place while
the reflex of *-n was still alveolar it would explain why the only examples of back-formed
-m- found so far come from *ŋ, and not *n.

A second area in which a reduced set of final consonants alternates with a fuller set of
non-final consonants is found in some of the Malayic languages of southwest Sumatra. In
Minangkabau, as described by Adelaar (1992:13) bases that end in *-a*, *-iə* or *-uə* show a
thematic liquid when suffixed with the transitivity marker *-i* or the nominaliser *-an*: *kapuə*
'chalk' : *ma-ŋapuər-i* 'to plaster, whitewash', *badiə* 'gun' : *sa-pam-badiəl-an/sa-pam-
badiər-an* 'the distance of a gunshot'. If the base ends in a glottal stop this is either
followed or replaced by *p* or *t*: *rambuyʔ* 'hair of head' : *rambuyʔt-an* 'a fruit with hairy
skin, the rambutan', *sakiʔ* 'ill' : *pa-sakit-an* 'difficulty, impediment', *tutuyʔ* 'closed' :
tutuyʔp-an 'prison'. If the base ends in *-h*, this is sometimes replaced by *-s*: *manih* 'sweet' :
manis-an labah/manih-an labah 'honey' (*labah* = 'bee').

The best known final consonant alternations in AN languages undoubtedly are those
connected with what is sometimes called the -Cia suffix. The great majority of Oceanic
languages have lost original word codas, but in suffixed forms these often reappear as
'thematic' consonants. The suffix to which these consonants most commonly attach is the
reflex of the Proto Oceanic passive/imperative marker *-ia. Table 4.39 gives examples of
the -Cia suffix from Wuvulu of the Admiralty Islands, and from Samoan:

Table 4.39 Thematic consonants in Wuvulu and Samoan suffixed verbs

POC	Wuvulu	Samoan	English
*qutup	uʔu/uʔu-f-ia	utu/utu-f-ia	submerge to fill
*inum	inu/inu-m-ia	inu/inu-m-ia	drink
*ranum		faʔa-lanu-m-ia	fresh water
*tanom		tanu/tanu-m-ia	bury
*tasim	ʔati/ʔati-m-ia		whet, sharpen
*apaRat		afã/afã-t-ia	storm
*kabit	api-ʔ-ia		pinch, squeeze
*kulit	uli/uli-ʔ-ia		peel (yams, etc.)
*puput	fufu-ʔ-ia	futi	pluck, pull out
*aŋin		aŋi/aŋi-na	wind
*paŋan		fafaŋa/faŋa-ina	feed
*paŋun		faŋu/fa-faŋu-ina	wake up
*qusan		ua/ua-ina	rain
*salan		faʔa-ala-ina	path, road
*taŋis	ʔai/fa-ʔai-k-ia	taŋi/taŋi-s-ia	weep, cry

As shown by the POC reconstructions, the thematic consonants that occur before the suffix *-ia* in widely separated Oceanic languages such as Wuvulu and Samoan often reflect historical word codas. In Samoan this is somewhat obscured by the fact that the expected sequence *-nia* regularly metathesised to *-ina*. This variant (seen also in *aŋi-na* < *aŋi-ina) then acquired distinctive syntactic properties that allowed it to be generalised to word bases which did not originally contain *-n. However, not all thematic consonants in suffixed verbs are etymologically faithful. Arms (1973) argued that in Fijian, where they are often etymologically deviant, the thematic consonants before the verbal suffixes *-i* and *-aki* have acquired a generalised semantic value. The basis for his semantic characterisations is rather impressionistic, and no similar claim has since been made for any other Oceanic language. As will be seen, it is striking that the thematic consonant of the -Cia suffix in Oceanic languages never shows the lenition that one would expect for a stage-by-stage reduction of word-final contrasts, as in the languages of Sulawesi.

In some languages the class of thematic consonants is severely limited. In Thao of central Taiwan, for example, a few bases show a thematic *-h* before a suffix that begins with a vowel, as in *bisu* 'beard' : *tan-bisu-h-an* 'bearded', or *t<m>ala* 'cut underbrush or firewood' : *tala-h-an* 'be cut with a sweeping motion'. In Tagalog a thematic glottal stop is regularly inserted between like vowels and a thematic *-h* between unlike vowels in suffixed bases, as in *abó* 'ash' : *abu-h-án* 'ashpit', *tubó* 'sugarcane' : *tubu-h-án* 'sugarcane plantation'. Some bases allow both possibilities in different affixed forms, as with *matá* 'eye' : *ma-mata-ʔ-án* 'see by chance' : *mata-h-án, mata-h-ín* 'big-eyed'.

Other alternations of final consonants that are striking, but difficult to assign to general categories, include several noted by Li (1977a, 1980a) for Formosan languages. In the Paran dialect of Seediq as described by Yang (1976) word-final *k* alternates with both *p* and *b* before the imperative suffix: *kayak* : *kiyap-i* 'cut meat', *atak* : *tap-i* 'cut with scissors', *cəhak* : *cəhəp-i* 'lick', *rubəruk* : *ruburub-i* 'broil' or *əluk* : *ləb-i* 'close'. The underlying forms of these words must be assigned labial stops for two reasons: 1) if velar : labial alternations were derived from underlying velars there would be no way to predict the voicing contrast, and 2. in base : imperative pairs such as *piyuk* : *puyuk-i* 'blow with the breath', and *gəmuk* : *gumuk-i* 'cover', the underlying form must contain -*k*. Li (1980a:379)

reports similar alternations together with *-m ~ -ŋ* in the Skikun dialect of Atayal, and notes that word-final labials were changing to velars among speakers born around 1950, were completely unchanged in a speaker born around 1900, and were changed in some words but not in others among two speakers born around 1930.

Following Egerod (1966), Li (1980a) reported that Squliq Atayal *t* becomes [ʧ] before *i*. Paran Seediq, however, shows a reversal of this natural process, as seen in the following base : imperative pairs: *qiyuc* : *quyut-i* 'bite', *rəŋac* : *ruŋat-i* 'chirp, growl', *haŋuc* : *huŋəd-i* 'cook', *tugakac* : *tugukad-i* 'kneel'. Whether palatalisation in Paran Seediq has developed from an earlier process like that in Squliq Atayal, or is historically independent, is unclear. What is clear is that synchronically the process is unexpected, since palatalisation is natural before front vowels, but not before word boundaries.

Finally, historical devoicing of word-final obstruents is a source of voicing alternations in some languages (as in German). Since final devoicing has occurred in a number of AN languages, we would expect similar alternations to be fairly common under suffixation. Surprisingly, however, they are not. Li (1977a:387) reports two cases of this type in Formosan languages, one in Atayal-Seediq, the other in Pazeh. However, Blust (1999a:326) has argued that the Pazeh alternations are better described as instances of intervocalic voicing. This leaves only cases such as Atayal *hgup* : *hbg-an* 'do magic', *hop* : *hab-an* 'stab', and *m-gop* : *gob-un* 'share one cup' as likely examples of voiced : voiceless alternations that result from final devoicing. According to Li (1980a:357ff) the voicing alternation in Atayal is limited to labial stops, and is rare (*-g-* alternates with *-w*, and there is no *d*). In other languages that have undergone historical final devoicing, such as Malay, no such alternation is observed before a suffix: Malay *uŋkap* 'open' : *məŋ-uŋkap-i* 'open things up to make them clear' (PMP *uŋkab), *laut* 'sea' : *laut-an* 'ocean' (PMP *lahud), *surut* 'to ebb' : *mə-ñurut-i* 'reduce or decrease' (earlier *surud), *məŋ-udut* 'smoke tobacco' : *udut-an* 'tobacco pipe' (earlier *udud).

4.3.1.9   Consonantal sandhi

All phonological processes discussed so far are triggered, at least historically, by affixation. However, a few languages have processes that occur only at word boundaries, and hence can be called 'sandhi.' One of the most notable examples of such phonological behaviour is seen in Toba Batak of northern Sumatra, where word-final nasals and stops assimilate either completely or partially to a following word-initial consonant in normal speech.[34]   Table 4.40 sketches the patterns of assimilation to a following consonant for word-final nasals and for a single final stop, *t*. Thus -m + p- > pp, -m + b- > bb, -t + h- > tt, etc.:

---

[34]   Despite the general thoroughness of his classic grammar of Toba Batak, originally published in 1864-67, van der Tuuk (1971) says remarkably little about this phenomenon. Percival (1981:28ff) discusses these kinds of alternations under the general heading of 'automatic morphophonemics', and Nababan (1981:57ff) refers to them somewhat more appropriately as 'external sandhi.' The material that I cite here represents the dialect of North Tapanuli, and was collected in 1968 from Mangantar Simanjuntak.

Table 4.40 Patterns of word sandhi in Toba Batak

	-m	-n	-ŋ	-t
p-	pp	pp	kp	pp
b-	bb	bb	ŋb	bb
m-	mm	mm	ŋm	bm
t-	mt	tt	ŋt	tt
d-	md	dd	ŋd	dd
n-	mn	nn	ŋn	dn
s-	ms	ss	ks	ts
l-	ml	ll	ŋl	dl/ʔl
r-	mr	rr	ŋr	tr/ʔr
j-	mj	jj	ŋj	jj
k-	pp	ŋk	kk	kk
g-	mg	ŋg	ŋg	gg/ʔg
ŋ-	mm	ŋŋ	ŋŋ	kŋ
h-	pp	kk	kk	tt

The following examples of usage show the sandhi forms of the verb *m-inum* 'to drink' with various objects. Some of these were chosen from English or made up by the informant in order to achieve a full set of combinatory possibilities for a single verb: p: *purik* 'rice water' = *minup purik* 'drink rice water', b: *bir* 'beer' = *minub bir* 'drink beer', m : *milk* = *minum milk* 'drink milk', t : *tuak* 'rice wine' = *minum tuak* 'drink rice wine', d : *Diet Cola* = *minum diet cola* 'drink Diet Cola', n : *noodle* = *minum noodle* 'drink (soup) noodles', s: *susu* 'milk' = *minum susu* 'drink milk', l : *Lipton tea* = minum Lipton tea 'drink Lipton tea', r : *rujak* 'k.o. soft fruit salad' = *minum rujak* 'drink *rujak*', j : *jamu* 'a drink used to keep slim' = *minum jamu* 'drink *jamu*', k : *Cola* = *minup pola* 'drink Cola', g : *gula* 'sugar (water)' = *minum gula* 'drink sugar water', ŋ : *ŋudik* (nonceword) = *minum mudik* 'drink *ŋudik*', h : *hopi* 'coffee' = *minup popi* 'drink coffee'. Before words that begin with a vowel assimilation is impossible (hence *minum aek* 'drink water', etc.).

The system underlying these and similar sandhi alternations is theoretically challenging. Despite the standard orthography, it should be noted that [h] is a prevocalic allophone of /k/, and behaves phonologically like a velar stop. Alternating words that end with a bilabial nasal show only full assimilation, and this occurs before words that begin with a labial consonant or *k/h* (but not *g*). Words that end with a velar nasal, on the other hand, show assimilation in manner and voicing (becoming *k* only before voiceless consonants), but never in place. The sandhi forms of word-final labials and velars thus appear to exhibit a complementarity based on place (+ place for labials, -place for velars). As might be expected on general theoretical grounds, word-final coronals are more catholic in their assimilatory behaviour. The coronal nasal shows complete assimilation to all following consonants except velar stops, and before these it assimilates only in place. The voiceless coronal stop shows complete assimilation to following stops (including *h*), assimilates in both place and voicing (but not manner) to *l*, *m* and *n*, and in place (but not voicing or manner) to *ŋ*. The clusters *dl*, *tr* and *gg* were also recorded as *ʔl*, *ʔr* and *ʔg*. In general the sandhi alternations described here agree with those reported by Nababan (1981), but differ somewhat from those of Percival (1981), whose description was based on the urbanised dialect of Medan, a multiethnic major population center.

A language which shows final consonant alternations that do not fit neatly into either of the types already described is Kerinci of south Sumatra. According to Steinhauer (2002)

Kerinci shows distinctive phrase-final and phrase-medial forms of many words, and these involve the simultaneous alternation of both vowels and consonants. Examples of -*ŋ/n* and *e/ɔ* alternation are illustrated by *bateŋ* 'tree', but *batɔn ño* 'his/her/its tree', *batɔn pinaŋ* 'areca nut tree', *batɔn pinan licayn* 'smooth areca nut tree', *batɔn pinan* (*licen*) *itoh* 'that (smooth) areca nut tree'. The conditions governing these alternations in Kerinci are complex, and will not be described in detail here. The essential point to note is that, like the initial consonant alternations of languages in eastern Indonesia or in central Vanuatu these phonological processes are now grammatically conditioned, but in every known case they began under statable phonological conditions.

4.3.2 Processes affecting vowels and suprasegmentals

Until now the discussion of phonological processes has focused on consonants. Although vowel allophony is fairly rich, as seen with vowel nasalisation and its effects on consonants, in general vocalic alternations in AN languages are less spectacular than consonantal alternations. This difference may be due at least in part to the relatively small vowel inventories of most languages, and to the infrequency of vowels at morpheme boundaries. Where vowels do occur often at morpheme boundaries, as in Oceanic languages that have a CV syllable canon, phonological alternations generally involve the addition of thematic consonants rather than processes that affect the vowels themselves. In general, vowel alternations are few and simple in languages of insular Southeast Asia, and in some of the better-described parts of the Pacific, such as the Southeast Solomons, Polynesia and Fiji. Complex vowel alternations occur in some Formosan languages, Palauan, many Nuclear Micronesian languages, and in a number of the languages of the Admiralty Islands.

Unlike consonants, vowel alternations are sensitive to stress and intonation, which remain among the least studied aspects of AN languages. This is particularly true of intonation. Some characteristics of stress are nonetheless widely shared, and merit a brief discussion in connection with processes affecting vowels.

4.3.2.1 Stress rules

As noted earlier, most languages of the Philippines have phonemic stress. This is particularly true of the northern and central Philippines, but is not true throughout much of Palawan, Mindanao and the Sulu Archipelago. In these areas, and over much of the rest of the AN world, most languages place primary stress on the penultimate vowel of the word, whether it is monomorphemic or morphologically complex. However, close inspection of the literature suggests that there are many distinguishing features in the stress systems of AN languages. The following survey is not meant to be exhaustive, nor does it treat secondary stresses.

Malay type: Most Malay words are stressed on the penultimate vowel. However, if this vowel is schwa and is followed by a single consonant stress shifts rightward to the final syllable, hence *barat* [bárat] 'west', but *bərat* [bərát] 'heavy'. In native words Malay permits two types of medial consonant clusters: -NC- and -rC-, where -NC- is any homorganically prenasalised obstruent (/mp/, /mb/, /nt/, /nd/, /nc/, /nj/, /ŋk/, /ŋg/, /ŋs/), and -rC- is a cluster consisting of a tapped or trilled /r/ plus a consonant. If a penultimate schwa is followed by -NC- stress does not shift, as in *əmpat* [ə́mpat] 'four', *bənci* [bə́nʧi] 'hate', or *gəŋgam* [gə́ŋgam] 'grasp', all with penultimate stress. If a penultimate schwa is

followed by -rC- stress appears to be variable: *kərbaw* 'carabao' is generally heard with final stress, whereas *pərnah* 'ever' is often heard with penultimate stress, although stress is never contrastive in this environment.

Some writers, as Odé (1994), and Tadmor (2003:30) have claimed that Malay has no word accent, but this is contrary to the position of virtually all other writers, as Kähler (1965:39), Macdonald and Darjowidjojo (1967:31), Halim (1974:70ff), Adelaar (1992:9), Moeliono and Grimes (1995:449), Mintz (1998:30-31), or Sneddon, Adelaar, Djenar and Ewing 2010).

Tiruray type: Post (1966) describes stress in Tiruray of the southern Philippines as falling mainly 'on the penult or antepenult of polysyllabic stems except when the vowels of those syllables are shortened, in which case it falls on the ultima. Primary stress falls on both syllables of unsuffixed bisyllabic stems having identical closed syllables.' Her examples suggest that stress actually falls on the first syllable of a lexical base unless the nucleus of that syllable is schwa followed by a single consonant, in which case stress shifts rightward. In trisyllables such as *darabay* [dárabaj] 'help along' stress thus falls on the antepenult, while in disyllables such as *dogot* [dógot] 'sea', it falls on the penult. Like Malay, a schwa (or high central vowel) can bear stress immediately before -NC-, as in *bayɨŋkig* 'mumps', which has penultimate stress. Contrary to this pattern, bases of the shape CVVC are final-stressed, as with *siuk* [sijúk] 'fish trap', *uit* [ʔuwít] 'take', or *ŋiaw* [ŋijáw] 'meow' (cf. *duyuy* [dújuj] 'type of song', where two vowels separated by an underlying glide maintain the dominant pattern of initial stress). The impression of two primary stresses in CVCCVC reduplications such as *gisgis* [gísgís] 'chafe' applies to many of the AN languages of insular Southeast Asia, but instrumental testing may show that the first syllable carries stronger stress in normal speech rhythm.

Hawaiian type: Following earlier work by Albert J. Schütz, Elbert and Pukui (1979) describe Hawaiian stress in terms of what they call 'stress groups' (adapted from Schütz's 'accent groups'). They maintain (1979:16) that in Hawaiian "The stress in a stress group is always on the next to the last syllable or on a long vowel marked by a macron." They note further that stress groups in Hawaiian "consist most commonly of two syllables, often of one or three syllables, but never of more than three except in proper names." They illustrate words containing a single stress group with *akaaka* 'clear', *hale* 'house', *Hanauma* 'a place name', *kanaka* 'person' and *malama* 'light'. Although the first example appears to be inconsistent with their definition, it seems clear that the stresses in these forms are [aká:ka], [hále], [hanáwma], [kanáka] and [maláma]. It follows, then, that before stress is assigned sequences of identical vowels collapse to a single long vowel, and sequences of non-identical vowels resyllabify as vowel + glide if the first is non-high and the second is a high vowel that is penultimate within the stress group. Words containing two stress groups include *Hana.lei* 'a place name', *hei.au* 'ancient temple', and *kā.naka* 'people' (where a period marks stress group boundary). Words containing three stress groups can be illustrated with *hoʔo.lau.leʔa* 'celebration', where again each stress group carries primary stress on the penultimate vowel. No mention is made of secondary stresses, and so the (perhaps questionable) implication is that there can be as many primary stresses as there are stress groups.

Maori type: Unlike most Polynesian languages, Maori places primary stress on the first syllable of a lexical base, as in *manawa* [mánawa] 'belly', or *taŋata* [táŋata] 'person, human being' (cf. e.g. the Samoan cognate *taŋata* [taŋáta]). In at least place names, where morpheme boundaries have been weakened, primary stress may also fall on the first

syllable of polymorphemic words, as in the well-known place name Waikato [wájkato] (= *wai* 'water' + *kato* 'flowing, flood').

Acehnese type: Durie (1985:30ff) distinguishes phrase stress from word stress in Acehnese, and notes that "Phrase stress falls on the stressed word in the phrase (usually the final or penultimate word). Word stress falls on the final syllable of a word." He notes that this rule holds for citation forms and for stressed words in a phrase, but that unstressed words in a phrase are cliticised, as none of the syllables are stressed.

Uma Juman type   For the Uma Juman dialect of Kayan in south-central Sarawak Blust (1977c) recorded a variable stress pattern that is sensitive to syntactic context: in citation forms stress is final, but in phrasal context it is penultimate, as in /mataʔ/ [matáʔ] 'eye', but /mata-n do/ [mátan do:] 'sun' ('eye of the day'), or /udik/ [udíjək] 'headwaters of a river', but /haʔ udik/ [haʔ údijək] ~ [húdijək] 'upriver' (= loc. + headwaters). A similar pattern, whereby final stress appears to serve as a marker of citation forms, is widespread in coastal and downriver areas of Sarawak.

Kokota type   Palmer (2009) has described the stress pattern in Kokota, spoken on Santa Isabel Island in the central Solomons, in terms of trochaic feet. He notes (2009:31) that 'Feet are aligned with the left margin of the word. Stress is then assigned to the trochee, or leftmost syllable or mora in each foot. This means stress is assigned to the first syllable or mora of each word, then to every odd syllable or mora after that.' This results in initial stress for disyllables such as *kame* [káme] 'hand, arm' and trisyllables such as *makasi* [mákasi] 'bonito' (since the first two syllables of each of these words forms a trochaic foot, with the last syllable of *makasi* left unassigned), but in secondary stress on the first syllable and primary stress on the penult in *dihunare* [dìhunáre] 'rough, of sea' (since *dihu* forms one trochaic foot, and *nare* another).

Many other patterns of stress assignment also occur, some of them quite quirky, as with the Sawai (South Halmahera) rule that stress falls on the penult unless that vowel is [ɛ] and the final vowel is anything other than [ɛ]: *baŋa* [báŋa] 'forest', *yɛgɛt* [jégɛt] 'oil', *lɛgaɛ* [lɛgáɛ] 'man', but *musɛla* [musɛlá] 'woven mat', *lɛlit* [lɛlít] 'mango', or *dɛlut* [dɛlút] 'parents' (Whisler and Whisler 1995).

4.3.2.2 Stress-dependent alternations

With a few outstanding exceptions (e.g. Rehg 1993) data on allophonic stress, length, pitch, intonation and other suprasegmental phenomena are inadequate or entirely lacking for most AN languages. Certain problems in characterising prosodic phenomena in these languages are nonetheless reasonably clear. To cite a particularly salient example, it appears difficult to characterise many AN languages as either syllable-timed or stress-timed, as these terms are commonly used in relation to Indo-European languages. In syllable-timed languages such as Spanish all syllables reportedly have roughly equal durations. In stress-timed languages such as English, on the other hand, two or more syllables between stress peaks can be compressed into a time-slot equivalent to that of a single stressed syllable. Many languages in the southern Philippines and western Indonesia show a pattern of vowel reduction and deletion in prepenultimate position, whereby vowels are centralised, shortened, and then often lost if they come to be word-initial. Examples are seen in Western Bukidnon Manobo (WBM) of Mindanao, where the contrast of *a* and *ə* is neutralised in prepenultimate syllables, producing alternations under suffixation, as in *apuʔ* 'grandparent; grandchild' : *əpuʔ-an* 'line of descent', *panuŋ* 'keep in captivity, as a fish, eel, or crab' : *pənuŋ-an* 'woven container in which a fish, eel or crab is kept in the water', or *mə-tazəy* 'straight, of a trail' : *pəkə-təzay-ən* 'make something clear or plain,' and in

Bario Kelabit of northern Sarawak, where the penultimate vowel of free morphemes alternates with schwa or zero under suffixation, as in *dalan* 'path, road' : *dəlan-an* 'path made by repeated walking over the same course', *guta* 'wade across a river' : *gəta-an* 'fording place', *taban* 'kidnapping, elopement' : *təban-ən* 'be kidnapped or taken in elopement', *aduŋ* 'adoption', *ŋ-aduŋ* 'adopt' : *duŋ-ən* 'be adopted', *irup* 'what is drunk' : *m-irup* 'to drink' : *rup-an* 'watering hole in the jungle where animals drink', or *itun* 'a question' : *ŋ-itun* 'ask a question' : *tun-ən* 'be asked'. Whereas WBM shows alternations only between *a* and schwa when the former vowel is placed in prepenultimate position, Bario Kelabit shows alternations between *a, i, u* and schwa or zero under the same conditions.[35]   Since these languages also show vocalic neutralisations of the same type within unaffixed bases of three or more syllables, it is clear that this difference in patterns of stress-dependent alternation is due to a hierarchy of vowel lenition in historical change: *a first neutralises with schwa in prepenultimate syllables, followed by the high vowels. Even closely related languages differ in this detail, as Lun Dayeh vs. Bario Kelabit, or Minangkabau vs. Malay, where the first language in each pair has weakened only *a, while the second language has weakened all prepenultimate vowels. Li (1980a) reports similar alternations of full vowels with schwa, which he writes as phonemic zero, in various dialects of Atayal (northern Taiwan). All five vowels of Atayal are affected, as seen in the following stems and suffixed passives: *kihuy* : *khoy-un* 'dig', *suliŋ* : *sliŋ-un* 'burn', *leliq* : *lliq-un* 'hold up', *hobiŋ* : *hbeŋ-un* 'cut meat', *paqut* : *pqut-un* 'ask'.

Stress-dependent vocalic alternations are unknown in any of the languages of the Philippines that have phonemic stress. Since the languages that show vocalic alternations under suffixation generally have a fixed penultimate stress, prepenultimate position corresponds to pretonic position. The behaviour of unstressed vowels in such languages raises an interesting question about general typology. If all languages must be either syllable-timed or stress-timed, to which category do we assign languages with stress-dependent vowel reductions of the kind illustrated here?   In Philippine languages with phonemic stress such as Tagalog stressed vowels in the penult are longer than unstressed vowels, and in some analyses length is considered primary (Schachter and Otanes 1972:8). Languages such as Tagalog therefore cannot be syllable-timed. But neither are they stress-timed, since they show no reductions of unstressed vowels. Rather, all unstressed syllables are approximately equal in length, as in syllable-timed languages, but stressed syllables are extra long. Languages with stress-dependent vocalic alternations show almost the mirror image of this pattern: with qualifications to be noted below, all vowels following the antepenult appear to be about equal in duration, whether they are stressed or not, but pretonic syllables are extra short. In effect, Tagalog words are syllable-timed with penultimate stressed vowels that are extra long, while Western Bukidnon Manobo or Kelabit words are syllable-timed with a single deviant vowel that is extra short. This characterisation oversimplifies the facts, but it does so to highlight the difficulty of applying asserted universal criteria derived from the behaviour of languages in one language family (Indo-European) to the behaviour of languages in others.

---

[35]   In at least WBM this process is suspended in reduplications, as in *alaŋʔalaŋ* 'deceive by hypocrisy', *basavasa* 'a proverb phrased in archaic language', or *kalaŋkalaŋ* 'be overtaken by night on a journey.' It is often claimed that reduplications are exempt from general phonological processes that would make the two copies of a base dissimilar. However, since general phonological processes affect the intervocalic stops of words such as *basavasa*, or *duwazuwa* 'to doubt' (from *duwa* 'two'), it is more likely that stress-dependent vocalic alternations are absent in WBM reduplications because these words carry two primary stresses.

The development of historically secondary schwa in pretonic syllables leads us to the role played by reflexes of PAN *e in the stress patterns of many AN languages of insular Southeast Asia. The special problems connected with the behaviour of this vowel will be taken up below in connection with stress shift and consonant gemination. We need only mention here that the schwa is extra-short, and cannot hold stress without compensatory lengthening of a following prevocalic consonant. The alternation of pretonic vowels in AN languages shows that length, stress, and centralisation are interconnected, but the arrow of causality is reversed: whereas the historically primary schwa is extra-short and so cannot carry stress, pretonic vowels are unstressed and so become extra-short, or drop.

A few AN languages appear to fit the proposed universal schema of syllable-timed and stress-timed languages. Palauan of western Micronesia, for example, is similar to American English in having many words of four or more syllables that carry a single primary stress, and in which most or all unstressed vowels are reduced to schwa, as in *məŋ-chəsóls* [məŋʔəsóləsə] 'to chant', *bləkərádəl* [bləgəráðələ] 'mannerisms, ways of behaving', or *kləŋəréŋər* [kləŋəréŋərə] 'hunger'. Unstressed syllables in Palauan, as in American English, are much shorter than stressed syllables, and it is not uncommon for words to contain five syllables with what are at least impressionistically approximately equal durations for 1) the two pretonic syllables as a unit, 2) the stressed syllable, and 3) the two posttonic syllables as a unit. As might be expected from such a prosodic system, Palauan has many stress-dependent vocalic alternations: *búsəʔ* 'feathers; fur; body hair' : *bsəʔé-l* 'his/her pubic hair', *kar* 'medicine' : *kərú-l a sokəl* 'medicine for ringworm', *mə-lík* 'line bottom of pot or basket with leaves' : *lkə-l* 'its lining'.

Table 4.41 Rightward stress shift under suffixation

	Base	**Suffixed base**	
Thao	paru [páɾo]	paru-an [paɾówan]	'hammer'
	furaz [fóɾað]	furaz-in [foɾáðin]	'moon'
	pilhnac [píɬnaθ]	pilhnac-an [piɬnáθan]	'thunderclap'
	saran [sáɾan]	s\<in>aran-an [sinaɾánan]	'road'
	iup [íjup]	iyup-i [júpi]	'to blow'
Malay	məŋ-aŋkat [məŋáŋkat]	aŋkat-an [aŋkátan]	'to raise'
	batu [bátu]	məm-batu-i [məmbatúwi]	'stone'
	məŋ-gigit [məŋgígit]	gigit-an [gigítan]	'bite'
	surat [súɾat]	mə-ñurat-i [məñuɾáti]	'write'
	məŋ-ukur [məŋúkuɾ]	ukur-an [ukúɾan]	'measure'
Wuvulu	ake [áxe]	ake-u [axéw]	'chin, jaw'
	uko [úgo]	uko-u [ugów]	'head hair'
	uli [úli]	uli-na [ulína]	'tree bark'
	pepea [pɛpéja]	pepea-u [pɛpejáw]	'intestines'
	upu [úpu]	upu-u [upú]	'grandfather'

4.3.2.3 *Rightward stress shift*

Although it often goes unreported, many AN languages exhibit a process of stress shift in suffixed bases. Most AN languages have penultimate word stress. While stress falls on the penult in an unaffixed base, then, in a suffixed base it shifts to the last syllable of the

base. Table 4.41 shows examples of rightward stress shift in three widely separated languages, Thao of central Taiwan, Malay of western Indonesia, and Wuvulu of the Admiralty Islands in western Melanesia.

In these languages, and in many others which behave in much the same way, stress is penultimate in the word, whether this is monomorphemic or polymorphemic. Prefixation and infixation have no effect on stress placement, but suffixation does, since it changes the relationship of the stressed syllable to the larger word of which it is a part. In languages with phonemic stress the same process generally applies, so that oxytone bases shift stress to the suffix, a pattern that is common to many languages of the northern and central Philippines, as with Bikol *hákot* 'transport' : *hakót-on* 'to transport', but *apód* 'call' : *apod-ón* 'to call'. In some Central Philippine languages, however, the stress pattern of suffixed bases is unpredictable, as with Tagalog *apúy* [apój] 'fire', but *ápuy-an* [ápojan] 'fireplace' (expected final stress). Similarly, in a few cases individual lexical items in languages that have a fixed penultimate stress appear to be exceptions to this process, as with Malay *minum* [mínum] 'to drink', but *minum-an* 'a drink', normally pronounced with stress on the first syllable: [mínuman].

4.3.2.4   *The mora*

Stress is penultimate in most AN languages, but in insular Southeast Asia this often involves certain complications. Malay has six native vowels written *i, u,* e (written *é*), *o,* ə (written *e*), and *a*. Of these, *e* and *o* are infrequent and distributionally restricted and the schwa, although common, never occurs before a vowel, *h*, or word boundary. Stress falls on any penultimate vowel except schwa followed by a single consonant, in which case it is deflected rightward, as in *barat* [bárat] 'west', but *bərat* [bərát] 'heavy'. This can be called 'lexical stress shift' to distinguish it from the morphological stress shift that is equally automatic in many languages under conditions of suffixation. If schwa is followed by a consonant cluster stress does not shift, as in *əmpat* [ə́mpat] 'four', or *pərlu* [pə́rlu] 'necessary'. This behaviour derives from subphonemic differences of length in the vowels of PAN and their reflexes in many daughter languages, in which the schwa appears to be extra short. The mora is thought to be the minimal unit of length. In most AN languages that have a schwa, however, this interpretation is difficult to apply, since most vowels in lengthening environments are long (hence two moras), most vowels in unstressed syllables are short (hence one mora), while the schwa is extra-short, hence less than one mora. If the schwa is monomoraic, short vowels must have two moras and long vowels three, but this is not true of languages that lack an extra-short vowel, and to adopt this interpretation would make cross-linguistic phonetic comparison inconsistent.

Whatever interpretation is adopted, it clearly is the short duration of the schwa that makes it stress-resistant. However, the ability of schwa to hold a stress when followed by a consonant cluster leads to another observation: in many languages consonants are geminated after schwa, and there is no stress shift (Blust 1995a:127). In Kelabit of northern Sarawak, for example, stress always falls on the penultimate vowel, but if this is schwa it automatically geminates most following consonants, as in *bəkən* [bə́k:ən] 'other, different', *bəŋəl* [bə́ŋ:əl] 'deaf', *əluŋ* [ʔə́l:ʊŋ] 'mouth of a river', *əpit* [ʔə́p:ɪt] 'bamboo tongs', *kətəd* [kə́t:əd] 'back', *pəman* [pə́m:an] 'feed', or *tənəb* [tə́n:əb] 'cool, cold'. There are two types of exceptions to gemination after schwa in Kelabit. First, the single Kelabit rhotic [ɾ] remains a flap after schwa rather than becoming a trill, and stress is deflected rightward, as in *bəra* [bəɾáː] 'husked rice', *ərət* [ʔəɾə́t] 'belt', or *tərur* [təɾúɾ] 'egg'. Second, since the voiced aspirates of Kelabit are phonetically longer than other voiced stops they are not

lengthened further by a preceding schwa. Thus, the aspirated stops in *təbʰuh* [təbpʰuh] 'sugarcane' or *bədʰuk* [bədtʰʊk] 'short-tailed monkey' do not appear to be appreciably longer than those in *ubʰo* [ʔúbpʰoː] 'stop, rest, take a break', or *idʰuŋ* [ʔídtʰʊŋ] 'nose'. The behaviour of stress in relation to schwa in both Malay and Kelabit is thus most simply explained by assuming that syllables can hold stress only if they contain at least one full mora, and a Cə- syllable falls short of this minimal requirement.

Some other phonological processes in AN languages are also mora-sensitive. In most AN languages monosyllabic content morphemes must be bimoraic. Alternations between monomoraic and bimoraic pronunciations of the same morpheme, as in Thao *ma-raʔin nak a taun* 'my house is big' ([nak]), vs. *i-zay a taun nak* 'that house is mine' ([naːk]) suggest that the bimoraic condition on monosyllables applies to phonological words rather than content morphemes, since in *nak a taun* the morpheme *nak* 'my, mine' is part of the larger phonological word [náka], with an incorporated ligature.

In Kiput of northern Sarawak both medial consonants and last-syllable vowels occur contrastively long: *mataay* 'die, dead' : *mattay* 'kingfisher' : *lattaay* 'chain', *lay* 'dry season' : *laay* 'male', *tot* 'fart' : *toot* 'needle' (Blust 2003b). In addition, the onset of monosyllables such as *lay* or *laay* is automatically long if the nucleus is short, and vice versa. This suggests that a constant unit of length must be maintained over a stressed CV-sequence, a phenomenon also reported for Madurese (Cohn and Ham 1999), and one that is reminiscent of the behaviour of schwa in relation to stress in other languages. In Kiput, however, the requirement that a stressed CV- sequence be equally long regardless of the length of the vowel affects only monosyllables, since the stressed first syllable of words such as *pana* [pánaː] 'to cook' contains a short consonant and a short vowel. If phonetic consonant gemination in forms such as *lay* [lːaj] is motivated by the same bimoraic length constraint as the phonetic vowel gemination in e.g. Thao *nak* [naːk], then it follows that Kiput is one of the few reported languages in which onsets have moraic value and so contribute to syllable weight.

While the behaviour of schwa in the AN languages of island Southeast Asia suggests that this vowel may be less than one mora and hence extra-short, Blevins and Harrison (1999) have argued that some prosodic constituents in Gilbertese of eastern Micronesia are extra-long. According to them, 'the typical foot in Gilbertese contains three moras.' Trimoraic constituents are reportedly units of stress, and also define minimal prosodic word size. Nothing similar has been reported for any other AN language.

4.3.2.5 Harmonic alternations

Vowel harmony is rare in AN languages, although some harmonic features play a role in phonotactics or in alternations. Examples of harmonic alternations that involve complete or nearly complete feature agreement are found in some of the languages of eastern Sulawesi. For Banggai of northeast Sulawesi, Van den Bergh (1953:126) describes a causative construction marked by *a-*, *e-* or *o-*, which reportedly results from the reduction of *pa-*, *pe-* and *po-* (PAN *pa- 'causative'). This variation is assimilatory: *a-* occurs if the first base vowel is central, *e-* if it is front, and *o-* if it is back: *a-lakit* 'load cargo', *e-teleŋen* 'bend something backward', *o-kolo* 'lay in a cradle', *o-sukup* 'make sufficient or complete'. Prefixal vowel assimilation occurs after other alternations are triggered by suffixation. Next to *a-lakit*, for example, van den Bergh cites the frequentative form *o-lokit-i*: addition of the suffix *-i* changes the *a* of *lakit* to *o* (from earlier schwa), and the prefixal vowel then assimilates to this derived base vowel. Probably the most far-reaching system of affixal vowel harmony reported for any AN language is found in Balantak, one

of the Saluan languages spoken in the northeastern peninsula of Sulawesi. Busenitz and Busenitz (1991) note that the nuclei of the Balantak verbal prefixes mVŋ-, nVŋ- and pVŋ- show full feature harmony with the first vowel of the following stem, as in *maŋa-wawau* 'to do', *meŋe-memeli* 'to cool', *miŋi-limbaʔ* 'to move', *moŋo-roŋor* 'to hear', or *muŋu-yuŋot* 'to shake'. Since nasal-sonorant consonant sequences are disallowed, an excrescent vowel is inserted between prefix and base, and both prefixal vowels are fully assimilated to the first base vowel. In addition to the regressive assimilation seen in these prefixal alternants Balantak has a system of progressive assimilation which affects the surface form of the 2sg possessive suffix, as in *tama-am* 'your father', *tambue-em* 'your green beans', *kopi-im* 'your coffee', *tigo-om* 'your tobacco', or *apu-um* 'your fire.'

Outside the Saluan languages of eastern Sulawesi such thoroughgoing systems of affixal vowel harmony are rare. Busenitz and Busenitz (1991:46, fn. 17) report that even within Balantak vowel harmony is not complete in all dialects, some of which appear to limit the alternations of prefixal vowels to just *a* and *o*. Two of the clearest cases of affixal vowel harmony are found in Kampampangan of central Luzon, and in Seediq of northern Taiwan. In Kapampangan the nucleus of the verb prefix *mVm-* harmonises with the first base vowel: *áŋin* 'wind' : *mam-áŋin* 'to blow, of the wind', *íkat* 'a braid' : *mim-íkat* 'to braid', *urán* 'rain' : *mum-urán* 'to rain'. Kapampangan has a five-vowel system, but since penultimate *e* and *o* are generally restricted to loanwords, *mVm-* has only three allomorphs. Li (1977a:402), describing earlier work by Yang (1976), notes that Seediq has a five-vowel system, and the vowel of various prefixes including *mu-*, *pu-* and *ku-*, assimilates completely to the first (stressed) vowel of a base. However, harmonic assimilation is restricted in two ways. First, assimilation occurs only if the base begins with a vowel or *h*. All other base-initial consonants block prefixal vowel harmony. Second, since *h + o* does not occur in Seediq, and *o* never occurs in a vowel sequence, prefixal vowel harmony cannot apply to bases that contain a penultimate mid-back vowel. The result is a four-way harmonic alternation: *ma-adis* 'bring', *me-eyah* 'come', *mi-imah* 'drink', *mu-uyas* 'sing', *mi-hido* 'expose to the sun', etc.[36]

The preceding examples illustrate affixal vowel harmony in which the harmonising vowel either assimilates completely to the nearest base vowel (Balantak, Kapampangan, Seediq), or assimilates in frontness (Banggai). In distinguishing subtypes these might be called instances of global harmonic alternations. Other examples of harmonic processes that occur across a morpheme boundary or a word boundary are more restricted in their application. In Bolaang Mongondow of north-central Sulawesi, for example, the actor voice infix *-um-* appears as *-im-* in bases that contain a high front vowel in the penult: *dompaʔ* : *d<um>ompaʔ* 'swoop down, of birds', *kutad* : *k<um>utad* 'swell up, as a corpse', *takoy* : *t<um>akoy* 'to ride (horse, vehicle, etc.)', but *kilat* 'lightning' : *k<im>ilat* 'to flash, of lightning', *siup* 'space under house' : *s<im>iup* 'enter the space under a house'. Since assimilation does not occur in e.g. *kuliliŋ* 'surroundings' : *k<um>uliliŋ* 'circle around an area', *kulit* 'skin' : *kulit-an* (**kilit-an) 'excoriated', or *kumi* 'moustache' : *kumi-an* (**kimi-an) 'have a moustache' it appears that frontness harmony in this language is a morphologically restricted phonological process. An apparently even more tightly constricted harmonic process occurs in Central Tagbanwa, spoken on the island of Palawan in the central Philippines, where Scebold (2003:35) reports that "Inflectional prefixes with high back vowels influence derivational prefixes that follow them in the following manner: /a/ → /u/ within the derivational prefix." In illustration he gives two

---

[36]   Li marks stress in his transcriptions, but stress is not phonemic in Seediq, and it is difficult to see from the data that he gives what role it plays in the harmonic alternations of prefixal vowels.

examples: *pa-tabas* 'to have pruned' + *pug-* → *pug-pu-tabas* 'having pruned', and *paŋ-aral* 'to lecture' + *pu-* → *pu-puŋ-aral* 'lecturing someone'. Here vowel harmony evidently is triggered only by a preceding affixal vowel, and affects only the vowel of a following affix, excluding all stem vowels.

Chamorro has a harmonic process of vowel fronting that is triggered by particles or affixes that contain any of the front vowels *i, e* or *æ*. When a base that contains a back vowel in the first syllable follows an affix or particle that contains a front vowel, the first vowel of the base is fronted, with no change of height: *gumaʔ* 'house' : *i gimaʔ* 'the house', *foggon* 'stove' : *ni feggon* 'the stove', *lagu* 'north' : *sæn-lægu* 'toward the north.' Three things are distinctive about Chamorro vowel harmony. First, harmonic changes are unidirectional—all harmonisation involves fronting. Second, unlike the previous examples, in which feature values spread from a base to an affix, in Chamorro feature values spread from an affix or particle to the first base vowel. Third, as *ni feggon* shows, Chamorro vowel harmony is partial rather than complete—unlike languages such as Turkish, it does not require that all vowels in a word agree in the harmonic feature.

Chamorro vowel harmony can be called 'fronting harmony.' Other AN languages exhibit types of unidirectional harmonisation that can be called 'backing harmony' or 'rounding harmony'. In Loniu, spoken on the island of Manus in the Admiralty Islands, for example, rounded vowels trigger rounding in both following vowels and consonants under certain conditions. A striking example is seen in *kaman* 'bachelor's house', but *lo kaman* > [lo komʷan] 'in the bachelor's house', where the vowel of the locative preposition *lo* has no apparent influence on the following velar stop, but rounds the first vowel of the base and the labial nasal immediately after it. The rarest type of unidirectional vowel harmony is one in which a colourless mid- or high-central vowel is the activator of an assimilation process. This is attested in two Formosan languages. Li (1977a:403), reporting on earlier work of Ting (1976), notes that in Saaroa of south-central Taiwan, the affix *(-)um-* alternates with *(-)ʉm-* when the first base vowel is *ʉ* (a high central vowel): *um-ala* 'take', *um-usal(ʉ)* 'to rain', *t<um>aŋi* 'weep', but *ʉm-ʉmʉcʉ* 'to touch', *l-ʉm-ʉmʉkʉ* 'to plant', etc. Similarly, in Pazeh of central Taiwan the active verb prefix *mu-* (< *-um-) surfaces as *mə-* before bases in which the first vowel is *ə*: *mu-laŋuy* 'to swim', *mu-bisu* 'to write', *mu-bulax* 'to split', but *mə-bəxəs* 'to spray water from the mouth', *mə-dəsək* 'to hiccup/to sob' (Blust 1999a). These cases are exceptional, since in most AN languages schwa is the neutral vowel *par excellence*, and so tends to be a follower rather than a controller of phonological processes.

Some morpheme structure constraints resemble vowel harmony, but are distinct from it. Kroeger (1992) describes what he calls 'vowel harmony' in three languages of Sabah. The details of this process vary from one community to the next, but the central observation relevant to several Dusunic and Murutic communities is that the sequence *oCa* is avoided, usually through a change of *o* to *a* iteratively within the same phonological word. Historically, non-final *o* in Dusun derives from schwa, and it is not clear whether the avoidance of *oCa* arose before or after this sound change. To illustrate, Tindal Dusun, spoken in the Kota Belud area, has the four-vowel system *i, u, a, o* (plus an emergent fifth vowel *e* in a small number of forms). In this dialect the attributive prefix *o-* may occur before any CV- base unless the first base vowel is *a*: *o-nibaʔ* 'short in length', *o-kuguy* 'narrow', *o-somok* 'near'. However, where the first base vowel is *a* the prefixal vowel is also *a*: *a-naru* 'long', *a-laab* 'wide', *a-lasuʔ* 'hot', *a-ralom* 'deep', *a-taraŋ* 'bright'. The same type of alternation affects all affixes and affixed bases in the language. Thus the vocalism of the prefix in *mog-onsok* 'to cook, fry' differs from that in *pag-ansak-an*

'hearth' as a result of changes initiated by the locative suffix -*an*. In Kimaragang Dusun as described by Kroeger (1992) the past affix *noko-* becomes *naka-* if the first vowel of the base is *a*, but otherwise does not change: *noko-dagaŋ > naka-dagaŋ* 'sold'. The recursive conversion of *o* to *a* is carried to even greater lengths in forms such as *mog-ogom* 'set something down', but *pa-agam-an* 'place where something is set' (< *po-ogom-an* >), or *ma-nanam* 'to plant', *pa-nanam-an* 'time/place of planting' (< *poŋ-tanom-an*). What is noteworthy is that this assimilatory process does not apply to the sequence *aCo*: *dagaŋ* : *dagaŋ-on* 'buy', *lapak* : > *lapak-on* 'to split', *tanay* 'termite' : *tanay-on* 'termite-infested', *rataʔ* 'flat' : *pa-rataʔ-on* 'flatten', *suaŋ* 'enter' : *po-suaŋ -on* 'be allowed to enter'.

In true vowel harmony some feature value agrees throughout a given phonological domain. Often this is the value for the feature [back], the feature [high], or the feature [round]. In typical Turkish or Mongolian languages, for example, all vowels within a phonological word must agree in frontness. In Chamorro, on the other hand, adjacent vowels across a morpheme boundary or close word boundary agree in frontness, although non-adjacent vowels are not subject to this constraint. Here there is still true vowel harmony, but over a smaller domain. Unlike these or other known examples, the *a/o* alternation in Dusun does not show a harmonic pattern, since it is sequence-sensitive, targeting *oCa*, but not *aCo*. In a few specific affix-base combinations the prohibited sequence is avoided by *a > o* rather than *o > a*: *a-gayo* 'big, large' : *mama-gayo* 'enlarge' : *po-goyo-on* 'be enlarged' (**pa-gayo-on), *a-maluʔ* 'shy, embarrassed' : *po-moluʔ-on* 'embarrass someone' (**pa-maluʔ-on), *panaw* 'walk, go' : *ka-panaw* 'able to walk' : *po-ponow* 'leave' (**pa-panaw). The latter mechanism is less well-understood than the former, but together they achieve the same effect, namely avoidance of the sequence *oCa*. The *a/o* alternation in Dusunic languages thus appears to fall between two relatively well-defined categories: vowel harmony, and language-specific conspiracies. Because oCa avoidance is sequence-sensitive it lacks criterial traits of true vowel harmony. Because both *o > a* and *a > o* are assimilations, *oCa* avoidance also lacks a criterial trait of true conspiracies, namely, that structurally variable strategies serve a common functional end.

Finally, since discussions of vowel harmony in AN languages usually address harmonic alternations, it is easy to overlook the few reported examples of structural vowel harmony (harmonic relations within a base). Winstedt (1927:48-49) noted that in Malay a mid vowel cannot be followed by a high vowel in the next syllable. However, high vowels can be followed by mid vowels, and Wilkinson (1959) lists some forms of both types. At best then, height harmony in Malay vowels appears to be a tendency, and is perhaps best expressed, like the *a/o* alternation in Dusunic languages, as a sequence constraint. However, true height harmony within a base does appear in some of the languages of northern Sarawak. Long Labid, Long Lamai and Long Merigam are three closely related Penan dialects spoken on the Tinjar and Tutoh tributaries of the Baram River. All three language communities have six contrastive vowels, *i*, *u*, *e*, *ə*, *o* and *a*, as well as a height constraint that operates only with vowels of the same frontness. The vowels *e* and *o* occur in the penult only if a mid front or mid back vowel occurs in the final syllable. Within a word, any combination of vowels may appear in successive syllables except *uCo*, *iCe*, or *uCe*, and where these would be expected on the basis of comparative-historical information the high vowel is lowered. Long Labid can be used to illustrate. In this dialect *u lowered to *o* before word-final glottal stop (from the PAN uvular stop *q), *h* (< *s), and occasionally *k and *ŋ. At a later time high vowels in the penult lowered if they were followed by a mid vowel of the same value for frontness: *puluq > *poloʔ* 'ten', *tuzuq > *tojoʔ* 'index finger', *pusuq > *posoʔ* 'heart', *nipis > nipih > nipeh > nepeh > *nepe* 'thin,

of materials', *titis > titih > titeh > teteh > *tete* 'drop of water', *itiq > itiʔ > iteʔ > *eteʔ* 'breast'. High and mid vowels that differ in frontness are treated differently, depending upon the combination. The sequence *uCe* evidently changed to *oCe*, as in Long Labid *ŋoreh* 'to scratch, as a cat' (*kuris), *mosen* 'rat', *loʔen* 'frog', or *ŋe-romek* 'to crush'. However, the sequence *iCo* is common in all three languages. Although the data sample is too small to generalise safely, it is noteworthy that all known examples of structural vowel harmony in AN languages are forms of height harmony, while this is not true of any known example of harmonic alternation.

Since consonant harmony is far less common than vowel harmony it is best treated here rather than in a separate section. As noted above, some PAN phonotactic constraints have been transmitted to many daughter languages. In general these constraints disfavor segment sequences or segments in some positions, but their effect is negative: in disfavoring dissimilar labials in successive syllables there is no positive requirement about the resulting sequence. Systems of consonant harmony go further in requiring all consonants within a given phonological domain to agree in some feature value. Several scholars, including Dempwolff (1939), Bradshaw (1979), and Ross (2002b) have noted a development of this kind in Yabem (Jabêm), spoken on the Huon Peninsula of New Guinea. As in many languages, Yabem vowels carry lower pitch following voiced obstruents and higher pitch following voiceless obstruents. These phonetic differences remain largely predictable, but have become contrastive in some environments, thus making Yabem one of the few AN tone languages. Within a morpheme voiced and voiceless obstruents or high and low tones may not co-occur. With certain qualifications that can be ignored here sonorants and *s* are treated as neutral. Apart from s, then, all obstruents within a stated domain must either be voiced or voiceless. It follows that Yabem has a system of obstruent harmony based on the feature [voice], and its close association with pitch height. This harmonic relation holds within a morpheme, but also triggers voicing alternations, as in the realis verb forms *ka-puŋ* 'I plant', *ka-taŋ* 'I make a sound', or *ka-ko* 'I stand' vs. *ga-bu* 'I insult', *ga-duc* 'I bow', or *ga-guŋ* 'I spear'.

Following earlier work by Fokker (1895), Adelaar (1983:57) has argued that in the history of Malay 'there was a constraint in disyllabic lexemes against the occurrence of a medial stop or nasal, homorganic with, but not with the same voicing as, the word initial stop.' He calls this 'consonant harmony'. Although examples such as *dataŋ* < *dateŋ 'come, arrive', *tidur* < *tiduR 'sleep', *tanah* < *taneq 'land', *tanam* < *tanem 'plant, bury', *tuna* < *tuna 'freshwater eel', *tunu* < *tunu 'burn', or *natar* 'background; level' < *nataD 'clear area around house' initially appear to violate this claim they can be reconciled with it through the observation that while /d/ and /n/ are alveolar, /t/ is dental in Malay, a phonetic feature that is widespread in AN languages, and probably inherited from at least PMP. However, as seen above, for the term 'consonant harmony' to be used in a way that strictly parallels the general use of 'vowel harmony', true harmonic systems for consonants must be limited to cases in which a dynamic process ensures agreement between certain feature values (height, frontness, voicing, etc.). Since the morpheme structure constraints that Adelaar identified in Malay, and that are implicit (although unstated) in Chrétien (1965:262-266) are based only on a pattern of avoidance they meet the necessary, but not the sufficient conditions defining harmonic systems, and we are forced to conclude that the only example of consonant harmony reported to date for any AN language which stands up to close scrutiny is that of Yabem.

4.3.2.6 Syncope

Syncope is an active process in some AN languages that is often overlooked. Blake (1925:300), for example, devotes an eight page appendix to Tagalog syncope, but this is rarely mentioned in more recent descriptions, such as Ramos (1971), or Schachter and Otanes (1972). Any unstressed vowel in Tagalog can drop between consonants flanked by vowels (VC_CV), but historically this was particularly true of *e (schwa): *atíp* 'roof' : *apt-án* 'roofed' (*qatep), *asín* 'salt' : *asn-ín* 'salted' (*qasin), *bukás* : *buks-án* 'to open' (*bukas), *tahíp* 'up and down motion of grains being winnowed' : *taph-án* 'winnowing basket' (*tahep). Vowel syncope in the environment VC_CV is a historical change in several languages of central Manus, and will be addressed in the discussion of historical phonology. A number of Oceanic languages also syncopate vowels between identical consonants in partially reduplicated bases, giving rise to initial geminates. This is mostly a historical process, but may be a change in progress in some languages. Milner (1958) reported that in the Polynesian language of the Ellice Islands (now Tuvalu) and in the Polynesian Outlier of Kapingamarangi atoll in Micronesia unstressed vowels delete between identical consonants, giving rise to surface aspirates p^h, t^h, k^h. The conditions for deletion in these languages typically arise in either of two ways: 1) through CV-reduplication, in which case any of the base-initial consonants *p, t, k* may surface as an aspirate, or 2) through syncope and fusion of the common noun article *te* with a following noun that begins with *t-*, in which case only the dental stop is aspirated as in *eai te puaka teenei* 'whose pig is this?' vs. *eai t^hao teenei* 'whose spear is this?' (*te tao* > t^h*ao*). Given the conditions that produce it, aspiration is not surprisingly generally confined to the initial syllable. In considering how to represent aspiration in Tuvaluan and Kapingamarangi the line between synchrony and diachrony is somewhat blurred, but paradigmatic considerations suggest that syncope and aspiration are products of the synchronic grammars of these languages.[37]   Another Oceanic language that syncopates vowels between identical consonants exhibits two interesting differences from the Polynesian cases just described: 1) a vowel syncopates between identical consonants whether the first consonant is word-initial or not, and 2) syncope affects stressed as well as unstressed vowels. Once again, the change appears to be in progress. In Mussau of the St. Matthias Archipelago north of New Ireland, speakers born c. 1930 or later drop the first vowel that occurs between like consonants, counting from the right margin of the word, while older speakers do not: *ai gagali* : *ai ggali* 'razor', *gigima* : *ggima* 'tree species', *ai kakala* : *ai kkala* 'broom', *kikiau* : *kkiau* 'megapode', *mamaa* : *mmaa* 'gecko', *mumuko* : *mmuko* 'sea cucumber', *papasa* : *ppasa* 'connecting sticks for outrigger', *rarana* : *rrana* 'mangrove', *raraŋa* : *rraŋa* 'sea urchin', *tutulu* : *ttulu* 'housepost', *gagaga* : *gagga* 'tidal wave', *kabitoto* : *kabitto* 'nit', *katoto* : *katto* 'star', *miroro* : *mirro* 'fish species', *mumumu* : *mummu* 'to suck'. Stress is regularly penultimate in Mussau, and rather surprisingly syncope thus deletes stressed vowels to produce medial geminates. Blevins (2008) has proposed a historical scenario which attempts to account for how this theoretically unexpected situation may have come about. However, her explanation depends upon published accounts of stress-marking which differ from those I explicitly recorded for the language, and the whole matter of whether Mussau deletes stressed vowels or not thus remains in limbo.

---

[37]   In his more recent description Besnier (2000:618) describes syncope as part of the synchronic phonology, but as producing geminates rather than aspirated consonants in Tuvaluan.

4.3.2.7 *Alternations of final vowels*

The appearance of thematic consonants before vowel-initial suffixes in many Oceanic languages clearly is a historical product of final consonant loss, even if subsequent changes altered the etymological integrity of the alternating segment. In much the same way, thematic final vowels have arisen from the loss of final vowels that were not 'protected' by a suffix, as in Seimat (Admiralty Islands) *min* 'hand' : *mina-k* 'my hand', *kaw* 'forehead' : *kawã-k* 'my forehead', *kinaw* 'neck' : *kinawe-k* 'my neck', *waku* 'testicle' : *wakue-k* 'my testicle', *leh* 'tooth' : *leho-k* 'my tooth', *ut* 'penis' : *uti-k* 'my penis', *sus* 'breast' : *susu-n* 'her breast', or *um* 'house; nest' : *um^wo-n* 'his/her house; its nest'. Few AN languages have lost final vowels without also losing final consonants, and consequently in languages that have lost final vowels phonetic 'erosion from the right' typically has led to the loss of -VC. This has happened in most languages of the Admiralty Islands, in many of the languages of Vanuatu, and in nearly all Nuclear Micronesian languages. As a result, suffixed forms of verbs or nouns sometimes differ from the corresponding unaffixed base in having not just a thematic consonant, but in having a thematic -VC sequence, or some subsequently modified form of this sequence. It is difficult to find examples of such alternations, possibly because they are historically unstable, but a few are given in Table 4.42 for Mota (Banks Islands, northern Vanuatu), and Mokilese (central Micronesia):

Table 4.42 Thematic -VC sequences in Mota and Mokilese

POC	Mota	Mokilese	English
*taŋis	taŋ (i)	jɔŋ	weep
*taŋis-i	taŋis	jaŋid	weep for someone
*qumun	um	um^w	earth oven
*qumun-i	—	um^wun	bake in earth oven
*tutuk	tut		beat with fist
*tutuk-aki	tutg-ag		thump upon

Other processes affect the quality of final vowels. In many dialects of Javanese, for example, -/a/ is rounded to [ɔ], and this allophonic change triggers regressive rounding assimilation: /lara/ > [lɔrɔ] 'ill, painful', /mata/ > [mɔtɔ] 'eye', *maŋsa* > [mɔŋsɔ] 'time, season', *sanja* > [sɔndʒɔ] 'visit, drop in' (/a/ does not occur in antepenultimate syllables). Under suffixation low vowel rounding is blocked, resulting in alternating forms: *kənɔ* 'be subjected to' : *kə-kəna-n* 'subjected to something unpleasant', *lɔrɔ* 'ill, painful' : *lara-n* 'prone to illness', *mɔtɔ* 'eye' : *kə-mata-n* 'having excessively large eyes'.

4.3.2.8 *Vowel lowering/laxing*

Many AN languages show phonetic differences of vowel quality that is associated with syllable type. The general pattern is for vowels in closed syllables to be lower or more lax than those in open syllables. In some languages the category 'closed syllable' may include syllables that end with any consonant, while in others this category may be more restricted. Topping (1973:19ff) describes a fairly complex set of conditions for automatic vowel lowering in Chamorro. The high vowels are said to have high allophones [i] and [u] only if the vowel is stressed and in an open syllable. Lowered allophones are said to occur in closed syllables, or in open syllables that are unstressed: [hí.hʊt] 'near', [pú.gas] 'husked rice', [mú.mʊ] 'fight', [lá.hɪ] 'male'. The mid vowels have three allophones that are distinguished by height, and by somewhat different conditions than those governing high

vowels. For /e/ these conditions divide the environment in which high vowels are lowered by distinguishing closed syllables from open unstressed syllables. In particular, /e/ is realised as [ɪ] (overlapping with [ɪ] < /i/) in an open syllable that is unstressed, as [e] in an open syllable that is stressed, and as [ɛ] in a closed syllable: [óp.pɪ] 'respond', [pé.ga] 'attach', [mɛ́g.gai] 'many'. For /o/ the conditions are more specific. This phoneme is realised as [ʊ] (overlapping with [ʊ] < /u/) in any unstressed syllable, as [o] in a stressed syllable that does not end with *k* or *ŋ*, and as [ɔ] in a stressed syllable that ends with *k* or *ŋ*: /mapput/ [máp.pʊt] 'difficult', /oppi/ [óp.pɪ] 'respond', /toktok/ [tɔ́ktʊk] 'hug'.

A number of languages in central and western Borneo also show allophonic vowel lowering or laxing in certain environments. In general these languages have a canonical shape CVCVC, and lowering occurs only before certain final consonants. Preceding a final consonant only the vowels *i*, *u*, *ə* and *a* occur in most native words. In Bario Kelabit the high vowels have lowered allophones [ɪ] and [ʊ] preceding any final consonant other than glottal stop or *h*. Moreover, where closed syllables are derived in non-final position as a result of vowel syncope high vowels also lower: *b<in>adaʔ* [binádaʔ] 'was advised by someone', but *p<in>ə-taʔut* [pɪntáʔʊt] 'was frightened by someone'. In the Uma Juman dialect of Kayan high vowels are lowered before word-final glottal stop, *h*, *l* and *r*: *lakiʔ* [lákeʔ] 'male', *uruʔ* [ʔúroʔ] 'grass', *hivih* [híveh] 'lower lip', *duh* [doh] 'female', *uil* [wel] 'lever', *bakul* [bákol] 'basket', *tumir* [túmeɾ] 'heel', *atur* [ʔátoɾ] 'arrange, put in order'. Mukah Melanau lowers high vowels only before -*ʔ* and -*h*, and many Kenyah dialects show a similar process only before final glottal stop, although phonemic mid vowels have arisen before -*h*, which then disappeared. Some languages in this area, such as Kiput, have no vowel lowering. It is perhaps worth observing that the final consonants which trigger vowel lowering in these languages correspond closely to those that are transparent to nasal spreading in the same languages, although there is no obvious connection between the two phenomena. The role of glottal stop and *h* is particularly noteworthy, as these occur in the environment both for vowel lowering (Kayan, Melanau and Kenyah), and for resistance to vowel lowering (Kelabit).

Among Formosan languages Thao lowers high vowels adjacent to *r* (an alveolar flap), and laxes them in closed syllables: *rima* [ɾéma] 'five', *rusaw* [ɾósaw] 'fish', *irush* [éɾoʃ] 'saliva', *lhmir* [ɬmeɾ] 'grass, weeds', *turu* [tóɾo] 'three', *pish-tiŋtiŋ* [pɪʃtéŋtɛŋ] 'crabby, irritable', *hibur-in* [hiʔbóɾɛn] 'be mixed together', *duruk-ik* [ʔdoɾókɪk] 'I stabbed it'. Whereas lowering and laxing are inseparable properties of high vowel allophones in most of the languages of Borneo for which we have information, they are separate in Thao. In addition, like several other Formosan languages, Thao lowers high vowels adjacent to the uvular stop *q*. More specifically, /u/ is lowered before or after /q/, as in *tuqris* [tóqɾes] 'noose trap' or *qusaz* [qósað] 'rain', but /i/ is lowered with a mid-central offglide preceding /q/, and lowered with a mid-central onglide following /q/: *mish-tiqur* [mɪʃteəqoɾ] 'to stumble', *qilha* [qəíɬa] 'rice wine'. Similar effects of *q* on high vowels (due to lowering and retraction of the body of the tongue in producing the stop) have been observed in other Formosan languages, including Atayal, Paiwan, Amis, and Bunun. In Bunun the body of the tongue appears to be pulled down even further by a uvular stop, producing ongliding and offgliding phonetic sequences closer to [ijaq] and [qai]: *ciqay* [tsíaqaj] 'branch', *daqis* [dáqaiʃ] 'face'. The conditions governing vowel lowering and centralisation in Amis differ somewhat from those in other Formosan languages. Amis has two glottal phonemes /ʔ/ and /h/; word-finally /ʔ/ is realised as an epiglotto-pharyngeal stop, and /h/ as an epiglotto-pharyngeal fricative (Edmonson, Esling, Harris and Huang 2005). Both consonants lower

adjacent high vowels: *nuliq* [nolé?ˤ] 'hundred', *puluq* [polóə?ˤ] 'ten', *upih* [upéəнħ] 'feather', *fanuh* [fanóəнħ] 'body hair'.

All of the examples of vowel lowering or laxing given above precede tautosyllabic consonants. Tagalog shows a pattern of lowering that is conditioned instead by position in the word. In native Tagalog words /u/ (but not /i/) lowers to a mid-vowel in the last syllable, independent of syllable type and of stress. The lowering of last-syllable /u/ in Tagalog is allophonic in native words, but has become phonemic through the introduction of loanwords from Spanish and English. What was earlier a subphonemic phonological alternation has thus become a morphophonemic alternation of /o/ and /u/ under suffixation, as in *apóy* [apój] 'fire' : *ápuy-an* [ápujan] 'fireplace', *buhók* [buhók] 'head hair' : *buhuk-án* [buhukán] 'hairy', *límot* [límot] 'oblivion' : *limut-in* [limútin] 'try to forget', or *súso* [súso] 'breast' : *susúh-in* [susúhin] 'suck at the breast'. Schachter and Otanes (1972:8ff) describe Tagalog as having lax varieties of /i/, /e/, /u/, and /o/ under certain conditions. They suggest that tense and lax allophones for all four vowels are completely interchangeable for some speakers, but that for other speakers the tense allophone is more common in stressed open syllables.

4.3.2.9 *Vowel lengthening*

As noted already under 4.3.2.4., some AN languages automatically lengthen vowels in certain environments. In several languages of northern Sarawak word-final vowels are predictably long, at least in citation forms, as in Bintulu *ba* [ba:] 'two', *lima* [limá:] 'five', *bivi* [biví:] 'mouth', or *rədu* [ʁədǔ:] 'female, woman'. In the Uma Bawang dialect of Kayan vowels lengthened before final glottal stop. Glottal stop was then added after final vowels, producing length contrasts before final glottal stop (*-a? > -[a:?], *-a > -[a?]) . Finally, short high vowels were lowered before final glottal stop, and final glottal stop was lost after high vowels. The result is that vowel length is contrastive in this dialect today, but only for the low vowel *a* before a final glottal stop, as in *dua?* 'two' (Pre-Kayan *dua), but *buaa?* 'fruit' (pre-Kayan *bua?).

The foregoing examples illustrate allophonic vowel lengthening in the last syllable of polysyllables. In many AN languages vowels are also automatically lengthened in monosyllables. This is the familiar 'minimal word constraint' of general phonological theory. In Hawaiian, for example, vowel length is phonemic, but non-clitic monosyllabic bases other than interjections are bimoraic: *ā* 'jaw, cheekbone', *?ā* 'fiery, burning', *hē* 'grave', *kī* 'the *ti* plant: *Cordyline terminalis*', *nō* 'to leak, ooze, seep', *pū* 'conch shell'. The bimoraic requirement for monosyllabic bases tends to be noticed more in content morphemes than in functors, because the latter are commonly cliticised to adjacent content morphemes to form part of a larger phonological word.

4.3.2.10 *Vowel breaking*

Vowel breaking is well-known in the Romance languages, where it generally affected stressed mid-vowels. A number of the languages of Borneo, Sumatra and mainland Southeast Asia exhibit historical breaking phenomena, some of which led to restructuring, and others to active processes that are preserved in the synchronic grammars. In Borneo, most languages of the Melanau dialect chain, which extends along the coast of Sarawak from Balingian in the north to at least the mouth of the Rejang River in the south, have allophones of *i* and *u* that are pronounced with a mid-central offglide before final *k* and *ŋ* (but not *g*). Some languages, as Mukah Melanau, have extended this even to /a/ (> [eə]),

but in general it is restricted to the high vowels. This phonetic detail also appears in some Lower Baram languages, and some dialects of Kenyah and Kayan that are spoken in the more accessible reaches of the major river systems of Sarawak, but not in dialects that are located in more remote interior areas, suggesting that vowel breaking is an areal feature that spread from Melanau.

In the Kampung Teh dialect of Dalat Melanau (KT) *i and *u preceding final *k* or *ŋ* developed a mid-central offglide, while high vowels that were final or that preceded final *ʔ* or *h* developed a mid-central onglide:

Table 4.43 Vowel breaking in Dalat Melanau

Pre-Melanau	Phonemic form	Phonetic form	English
*titik	titik	[titíjək]	drop of liquid
*buk	buk	[búwək]	head hair
*kuniŋ	kuniŋ	[kuníjəŋ]	yellow
*nibuŋ	nibuŋ	[nibúwəŋ]	palm species
*kami	kaməy	[kamə́j]	1pl ex
*qulu	uləw	[ulə́w]	head
*putiʔ	putiʔ	[putə́iʔ]	white
*lasuʔ	lasuʔ	[lasə́uʔ]	hot
*paqis	paʔih	[paʔə́ih]	roasted fish or meat
*bibiR	bibih	[bibə́ih]	lip
*qateluR	təluh	[tələ́uh]	egg

Before final consonants this historical process has left a synchronic residue in the form of complementation. The development of final diphthongs, however, is better treated as phonemic restructuring, since the borrowing of some words with -*i* and -*u* has led to contrast between final high vowels and diphthongs with a mid-central onglide.

Breaking with a mid-central offglide does not occur before any other final consonant in KT. Surprisingly, this includes *g*: *lilig* ([lilíg]) 'tree resin', *ribig* ([ribíg]) 'to pinch', *muug* ([mũũg]) 'to rub, as in scrubbing a floor', *tug* ([tug]) 'heel', *utug* ([utúg]) 'a piece of something'. Phonemic sequences of high vowel plus schwa may precede a voiced velar stop in word-final position, as with *pieg* ([pijə́g]) 'shivering', and here the placement of stress shows that the phonemic form contains two vowels rather than a single high vowel that undergoes phonetic breaking (cf. e.g. [titíjək] 'drop of liquid'). A similar limitation on breaking with a mid-central offglide before final *k* and *ŋ* but not before final *g* is found in all known languages of Sarawak that preserve a voicing distinction in final stops.

Mukah Melanau shows an exceptionally complex and innovative system of vowel breaking, probably because it is near the center from which this innovation spread along the coast of Sarawak and up the major river systems (Blust 1988c) Languages that have acquired vowel breaking through contact, such as the Uma Juman dialect of Kayan (Baluy branch of the Rejang River), Murik (main branch of the Baram River), or the Long Wat dialect of Kenyah (Tutoh branch of the Baram), show simpler forms of vowel breaking, or breaking in completely unrelated environments. In Murik, for example, breaking occurs before final velars (there is no -*g*), but only for underlying *i*: -*uk* and -*uŋ* are realised as [ʊk] and [ʊŋ], with lax allophones of the high back vowel that fluctuate between lower high and lower mid, but -*ik* and -*iŋ* are realised as [ejək] and [ejəŋ] (Blust 1974c). A similar set of allophonic relationships is found in Uma Juman Kayan, where *u* is lowered to [o] before final velars, but -*ik* and -*iŋ* are offglided to [ijək] and [ijəŋ]. In the Kenyah

language of Long Wat, on the other hand, both high vowels are pronounced with a mid-central offglide before word-final *k* and *ŋ*, but in conducting fieldwork on the language breaking was also heard inconsistently with the sequences *-un* and *-ut*: *avun* [avúən] 'cloud', *məñun* [məɲúən] 'to sit', *ujun* [udʒúwən] 'mouth', *məñut* [məɲúwət] 'to wash clothes'. Breaking was not recorded for *i* in this environment, or in the sequence *-ud*.

Finally, since vowel breaking in the better-known Romance languages was limited to stressed vowels, the relation of stress and vowel length to breaking in the languages of Sarawak is of interest. Historically, these languages almost certainly had a pattern of primary stress on the penult unless this vowel was schwa. The contemporary languages, however, tend to be oxytone in citation forms, but paroxytone in phrasal context. Stress in Mukah Melanau generally was recorded as penultimate, implying that vowel breaking affected unstressed vowels, although it is possible that the language was oxytone when vowel breaking was innovated, and has since shifted under the influence of paroxytone Malay. A consideration of Malay, however, raises an interesting point of comparison. Although no reported Malay dialect has breaking rules like those of Melanau, in standard peninsular Malay high vowels in native vocabulary are lowered under conditions that closely parallel those governing vowel breaking in coastal Sarawak. Adelaar (1992:10), summarising earlier literature, notes that *-ik*, *-uk*, *-ih*, *-uh*, and *-uŋ* are realised with allophonic vowel lowering, but that *-iŋ* is not. While this does not correlate perfectly with the set of environments in which vowel breaking takes place in coastal Sarawak, it corresponds rather closely. A number of languages in southern Sumatra, including the Malayic languages Minangkabau and Kerinci, and the strongly Malayicised Rejang also have complex histories of vowel breaking. Since these produced phonemic restructuring in most cases, however, they will be treated in the discussion of historical change.

4.3.2.11 Vowel nasality

As noted under 4.3.1.3, where adequate phonetic information is available, it appears that allophonic vowel nasalisation in AN languages usually proceeds from a syllable onset to a following vowel rather than from a syllable coda to a preceeding vowel. This is clear in a number of the languages of Borneo, where progressive vowel nasality blocks the otherwise general preplosion of final nasals, and where nasal spreading has produced consonant alternations of the type [l] (following oral vowels) ~ [n] (following allophonically nasalised vowels) in Narum, or [j] (following oral vowels) ~ [ɲ] (following allophonically nasalised vowels) in Ngaju Dayak. It is also true in Oceanic languages such as Seimat, where phonemic nasalisation has resulted from the nasalising effects of earlier syllable onsets (*mʷV > wṼ, but *wV > wV; and *rV (> hṼ, but *pV > hV). The nasalisation phenomena noted above for Bornean languages generally remain part of the synchronic grammars while in Seimat they have led to phonemic change, but in both cases there is clear evidence of directionality, with nasalisation spreading from a consonant onset to a following vowel rather than from a consonant coda to a preceding vowel. A number of the languages of New Caledonia exhibit similar patterns of historical vowel nasalisation in which the feature [nasal] has been transferred from a syllable-initial consonant to a following vowel, although "Bilateral assimilation originating from intervocalic consonants has also had a part in some languages … and there has been spontaneous nasalisation in a few languages of the Far North" (Ozanne-Rivierre and Rivierre 1989).

The one area that is known to be a clear exception to this pattern is the Lamaholot-speaking region of the Lesser Sunda Islands (eastern Flores through Lembata/Lomblem). Keraf (1978) provides unmodified Swadesh 200-word lists for 35 language communities

within this extensive dialect continuum, and many of these show what appears to be phonemic vowel nasality which has arisen historically through regressive nasalisation of a last syllable vowel followed by loss of the final consonant, as seen in PMP *zalan > Lamalera *larã* 'path, road', *ipen > *ipã* 'tooth', *kaen > *kã* 'eat', or *pafã* 'fallen down', next to forms with a final nasal in other dialects (Mulan, Belang *pawaŋ*, Kalikasa *pawan*, etc.). Although these forms are products of historical change which apparently has not left a synchronic residue, other etymologies suggest that regressive vowel nasalisation remains part of the synchronic grammar. Several terms which did not originally end with a nasal, for example, now have nasalised final vowels in Lamalera, as with *beqbaq > *fəfã* 'mouth', *mata > *matã* 'eye', or *buaq > *fuã* 'fruit'. These are words that in many AN languages are obligatorily possessed ('mouth', 'eye'), or conceived in a part to whole relationship ('fruit'), and which therefore may earlier have ended with *-n* '3sg possessor/3sg genitive'. If this is the case then forms of these nouns that did not carry this suffix may appear with an oral vowel in the final syllable. For the dialect of Lewolema Pampus (1999:29ff) clearly states that vowel nasality plays a role not only in the phonology, but also in the morphology or syntax of the language, as shown in *beləʔ* 'big', which appears with a nasalised schwa when following *laŋoʔ* 'house' in the expression 'the big house'. However, it is not yet clear how this process arose, and whether vowel nasality in Lamaholot dialects has a single source or more than one.

4.3.2.12 *Other types of vowel allophony*

One other type of vowel allophony should be mentioned here not because it is widespread, but because it is distinctive not just for AN, but for languages generally. Bender (1968) showed that Marshallese has twelve surface vowels defined by the intersection of front, back unrounded and back rounded qualities with four heights (high, high-mid, mid, low). Word-finally, where vowels usually are short, minimal sets that illustrate all twelve contrasting values can easily be found, and in initial position evidence of contrast also exists. Under these circumstances a traditional phonemic analysis would posit twelve contrasting vowels. However, a traditional analysis is awkward, since the distribution of vowels is restricted by the class of adjacent consonants: in CVC syllables front vowels are found between palatalised labial or dental consonants (called 'J consonants'), back unrounded vowels between velarised labials or dentals, or phonetically non-labialised velars or liquids (called 'K consonants'), and back rounded vowels between phonetically labialised velars or liquids (called 'Q consonants'). In mixed environments vowel quality changes in transition from onsets of one class to codas of another. These relationships are schematised in Figure 4.2, where F = front vowel, BU = back unrounded vowel, and BR = back rounded vowel:

	F	BU	BR
J	+	−	−
K	−	+	−
Q	−	−	+

Figure 4.2 Relationship of Marshallese vowel and consonant quality

Where one initially sees evidence of vowel contrast, closer attention to the consonant environment shows complementation. In short, vowel quality (not height) is predictable from the class of adjacent consonants. Despite its richness in allophones, then, Marshallese appears to have only four vowel phonemes, each representing a distinct height. Although

this set of structural relationships seems to be unique to Marshallese, the interpenetration of vowel and consonant features in this language is widely shared in the historical phonology of other Nuclear Micronesian languages, and historically it was the vowels that conditioned the qualities of adjacent consonants rather than the reverse.

4.3.2.13 Rich vowel alternations

As noted earlier, although most AN languages show few vowel alternations, there are some notable exceptions. Table 4.44 illustrates singular possessive paradigms in two languages of western Manus which show innovative vowel alternations. Where they are available the POC antecedents are given for comparison (all POC common nouns and their continuations in Proto Admiralty, are understood to require the preposed common noun marker *na):

Table 4.44 Innovative vowel alternations in Levei and Pelipowai of western Manus

	Base	1sg	2sg	3sg	
POC	mata	mata-gu	mata-mu	mata-ña	'eye'
Levei		moto-k	muto-ŋ	mʷato-ŋ	
Pelipowai		ndʷə-k	ndiə-m	ndʷa-n	
POC	qate	qate-gu	qate-mu	qate-ña	'liver'
Levei		ete-k	itie-ŋ	etæ-ŋ	
Pelipowai		ati-k	ate-m	ati-n	
POC	natu	natu-gu	natu-mu	natu-ña	'child'
Levei		nesu-k	nisie-ŋ	nesu-ŋ	
Pelipowai		nacu-k	nacu-m	nacu-n	
POC	taliŋa	taliŋa-gu	taliŋa-mu	taliŋa-ña	'ear'
Levei		cinu-k	cinie-ŋ	cinu-ŋ	
Pelipowai		pʷiki-caniə-k	pʷiki-caniə-m	pʷiki-cani-n	

Although base forms are given for POC, in most Oceanic languages base forms of obligatorily possessed nouns cannot occur alone. Base forms thus appear to be more homogeneous in Pelipowai than Levei, showing real complexity in the possessive paradigm only in items such as 'eye'. Other languages in western Manus have few or none of these alternations, suggesting that they are relatively recent innovationss in Levei and Pelipowai (cf. e.g. Lindrou *mada-k, mada-m, mada-n* 'eye', *ade-k, ade-m, ade-n* 'liver', *nadu-k, nadu-m, nadu-n* 'child', *drañe-k, drañe-m, drañe-n* 'ear').

Another language with rich vowel alternations is Palauan, which has acquired them through extensive conditioned sound changes. Many of these alternations are seen in comparisons of a lexical base with its 3sg possessed form. Where the base is CVCVC the first base vowel alternates with schwa under suffixation: *dáŋəb : dəŋəb-él* 'cover', *ródəʔ : rədəʔ-él* 'fruit', *díŋəs : dəŋəs-él* 'satiation'. In most CVC bases no alternation occurs: *ʔur : ʔur-ál* 'tongue', *diŋ : diŋ-ál* 'ear' (but *ʔur : ʔər-íl* 'laughter'), suggesting that Palauan once underwent a process of prepenultimate vowel neutralisation, like many of the languages of Borneo. Most of the complexity in Palauan vowel alternations arises from the suffixation of bases that contain a vowel sequence, or glide + vowel. If the first of two vowels is *a* it

usually deletes: *dáit* : *dit-él* 'taro', *káud* : *kud-él* 'dam', *táod* : *tod-él* 'fork'. In certain cases, however, the result is unpredictable, as with *dáob* : *dǝb-él* 'sea'. Where the first vowel is not *a* the form of the suffixed base appears to be generally unpredictable: *buík* : *bik-él* 'boy', *tóuʔ* : *tuʔ-él* 'foot disease' (loss of the first vowel), *bóes* : *bos-él* 'gun; blowgun', *tuáŋǝl* : *tuŋǝl-él* 'entrance, door' (loss of the second vowel). Other bases show patterns of vowel alternation that appear to be idiosyncratic, as *ʔúi* : *ʔiú-l* 'hair of the head'.

4.4 Metathesis

Metathesis is commonly regarded as marginal and sporadic. Although this is true of many AN languages, metathesis holds a special interest in this language family because it is sometimes regular, and may even play a role in the grammatical system. A complete treatment of the phenomenon must include historical change, and this will be discussed in a later chapter.

The scope of regular metathesis varies widely in AN languages. Many languages of the central Philippines permit medial consonant clusters that contain a glottal stop, but these are often restricted to -ʔC- or -Cʔ-, and derived clusters with glottal stop in the wrong order undergo metathesis. In Hanunóo, spoken on Mindoro Island, the glottal stop follows, but never precedes another consonant in medial clusters: *bagʔú* 'tree species', *báŋʔit* 'bite', *káwʔit* 'picking something up with the toes', *kírʔum* 'fear, worry', etc. Clusters derived by syncope that would otherwise contain a preconsonantal glottal stop undergo metathesis, as in *ʔusá* 'one' : *ká-sʔa* 'once', *ʔúpat* 'four' : *ká-pʔat* 'four times', or *ʔúnum* 'six' : *ká-nʔum* 'six times'. Two things are noteworthy about metathesis in these forms. First, other derived clusters do not metathesise, even if they are highly marked, as with *túlu* 'three' : *ká-tlu* 'three times'. Second, other languages of the central Philippines, as standard Bikol, allow medial clusters with glottal stop only in preconsonantal position. Over much of the central Philippines, then, there is a restriction on the order of glottal stop in consonant clusters, but the preferred order is language specific, or even dialect-specific. What is general to the entire area is a condition that glottal stop may not occur *both* preconsonantally and postconsonantally in medial clusters in the same language.

Steinhauer (1993) has shown that in Atoni or Dawanese of west Timor (C)VCV verb stems before the suffix *-ku* metathesise the final CV if the vowel is non-low (low vowels delete), hence *lomi* : *loim-ku* 'want', *sapu* : *saup-ku* 'sweep', *inu* : *iun-ku* 'drink', *tepo* : *teop-ku* 'hit'. However, his statement (1993:130) that Dawanese uses 'metathesis as an inflectional morphological process' is not supported by such examples, which suggest that it is a purely phonological concomitant of affixation. Steinhauer (1996a:224) makes a renewed case for the use of metathesis as a grammatical device in Dawanese, since here underlying and metathesised forms of a base are illustrated with the same affix, as in *noni mnatu qina n-lomi* (money gold **3sg 3sg**-like) 'Gold money he likes' vs. *qina n-loim noni mnatu* (**3sg 3sg**-like money gold) 'He likes gold money'. Again, however, it is difficult to see from these examples how metathesis alone carries syntactic or semantic information independently of word order, and it must be questioned whether a convincing case has yet been made for metathesis as a grammatical device in this language.

My fieldnotes on the Molo dialect show another dimension of Dawanese metathesis. Metathesis is sometimes regarded as incompatible with the notion of gradual sound change. But this is true only in theories of sound change that fail to recognise the key role of variation. As seen in Table 4.45, the metathesis of Dawan CV syllables in words used in isolation appears to be a change in progress:

Table 4.45 Metathesis as a change in progress in Dawan of western Timor

PMP	Careful speech	Rapid speech	English
*ma-isa	mɛsɛʔ	mɛs	one
*duha	nua	nua	two
*telu	tenu	teun	three
*epat	ha	ha	four
*lima	nim	nim	five
*enem	nɛʔ	nɛʔ	six
*pitu	hitu	hiut	seven
*walu	fanu	faun	eight
*siwa	seɔʔ	seɔʔ	nine
*sa-ŋa-puluq	boʔɛs	boʔɛs	ten

Other examples include PCMP *Ratus > *natun mɛsɛʔ* (careful), *naut nɛs* (rapid) '100', *malip > *ʔat mani* (careful), *ʔat main* (rapid) 'to laugh', *m-atay > *ʔat mate* (careful), *ʔat maɛt* (rapid) 'to die', *kali > *ʔat hani* (careful), *ʔat hain* (rapid) 'to dig', and *ma-qasu > *na masu* (careful), *na maus* (rapid) 'smoky'. The terms 'careful speech' and 'rapid speech' are used here to distinguish the style in which metathesised forms appear in Dawan, but it is possible that this is not the primary parameter determining variation. In citation form the first eight numerals maintain the historical order of consonants and vowels, but in serial counting or phrasal context metathesis of the final -CV occurs. Other words evidently occur only in a form that shows historical metathesis, as with 'nine' above, and *deŋeR > *ʔat nɛɛn* 'to hear'. Although the Dawan word for 'ten' does not reflect PMP *sa-ŋa-puluq, *boʔ nua* '20', *boʔ tenu* '30' and the like show that *boʔɛs* consists of *boʔ* 'unit of ten', plus a metathesised variant of the last syllable of 'one'.

The form of metathesis in these examples agrees with Steinhauer's description: only a final CV is affected, and low vowels apparently delete rather than metathesise. What is noteworthy is that these forms generally involve no affixation or syntactic differences: metathesis apparently is governed entirely by speech tempo/register. Metathesis as a change in progress has interesting implications for general linguistic theory. Although it is often assumed that segmental transpositions could not fail to intrude on a speaker's awareness, the single Dawan speaker with whom I worked appeared to be no more conscious of the difference between metathesised and unmetathesised forms of the same word than speakers of American English are of unreleased vs. released final stops. When confronted with a metathesised pronunciation he would sometimes insist emphatically that only the unmetathesised form is correct. If other matters intervened so that stylistic differences were backgrounded, however, the metathesised variants again begin to surface. In Dawanese metathesis thus appears to be both a marker of speech tempo or register, and a phonological concomitant of certain morphological or syntactic processes. This is surprising, and since the form of metathesis is identical in the two cases it would be analytically preferable if they could be united under a single condition. Since the metatheses observed in the numerals involve no variation in morphological or syntactic environment this condition would have to be tempo/register, but the data presented by Steinhauer do not appear to support this. Deck (1933-1934:38) described a similar condition for metathesis in Kwara'ae of the southeast Solomons, where the final CV (regardless of vowel height) reportedly metathesises only in rapid speech, as in *leka > leak*, *ʔaemu > ʔaeum*, *aliali > ailail*, or *likotai > lioktai*.

Metathesis has also been reported as a grammatical device in Letinese of the southern Moluccas (van Engelenhoven 1996, 1997), and most famously in Rotuman of the central Pacific, a language that has received considerable attention from general linguists. This phenomenon was first described in 1846 by Horatio Hale, who cited examples such as *hula* 'moon; month', but *hual rua* 'two months', or *uhi* 'yam', but *uh rua* 'two yams'. These two forms of a word in Rotuman have been variously labeled the 'absolute' and 'construct' forms (Hocart 1919), the 'complete' and 'incomplete' phases (Churchward 1940), and 'indefinite' vs. 'definite' (Laycock 1981).

Churchward, who provides what might be called the 'classical' account, describes the incomplete phase as a product of three processes: 1) elision of the final vowel, as in *haŋa* : *haŋ* 'to feed' or *hoto* : *hot* 'to jump', 2) elision of the final vowel with alteration of the vowel that remains, as in *tɔfi* : *tɛf* 'to sweep', *mose* : *mös* 'to sleep', or *futi* : *füt* 'to pull', 3) metathesis of -CV, as in *seseva* : *seseav* 'erroneous', *pure* : *puer* 'rule, decide', or *tiko* : *tiok* 'flesh'. Most subsequent analyses have sought to find a unity underlying Churchward's fragmented description by showing that the incomplete phase is uniformly derived by -CV metathesis. These analyses are conveniently summarised by Schmidt (2003). Where the vowels brought together by metathesis are identical they coalesce into a single vowel of the same quality. Where the vowels brought together by metathesis are non-identical they either remain distinct, as in */tiko/* > [tijok], or they fuse into a single novel surface vowel, as in */futi/* > [füt]. Churchward describes Rotuman metathesis as 'frequently responsible for differences of meaning,' and he illustrates this statement with *famori feʔen* 'the people are zealous' vs. *famör feʔeni* 'the zealous people.' This and other examples show that one function of metathesis in Rotuman is to distinguish predicate adjectives from attributive adjectives, but the phases are also used to contrast definiteness in nouns, as in *famori ʔea* 'the people say' vs. *famör ʔea* 'some people say.' Unlike Dawanese, where metathesis evidently conveys differences of meaning only together with affixation or changes of word order, in Rotuman metathesis is often the only marker of grammatical distinctions, and its role as a grammatical device is therefore beyond question.

Blevins and Garrett (1998) argue that CV metathesis arises from the phonological reinterpretation of long durational cues. What they call 'compensatory metathesis' "is a gradual development whereby the articulation of a vowel shifts temporally from a weak peripheral position to an adjacent tonic position. We have distinguished three discrete parameters of this gradual sound change: extreme coarticulation or vowel copy; peripheral vowel reduction; and peripheral vowel loss. But these are interconnected changes under our account, all aspects of articulatory migration" (1998:549). They consider several AN cases, including Rotuman, Kwara'ae, Dawan and Letinese, and conclude that compensatory metathesis is a phonetically motivated process, like most other types of sound change. This article provides a useful survey of AN examples in terms of a unified theory of regular metathesis. One of its consequences is to imply that the superficially similar phenomenon of V…V or C…C metathesis seen historically in e.g. PMP *qudip > Malay *hidup* 'living, alive' or POC *laŋo > Hawaiian *nalo* 'housefly' is unrelated to 'compensatory metathesis', since cases of this type could not conceivably result from a reinterpretation of long durational cues. In effect, then, it is implied that the traditional term 'metathesis' is a cover term for more than one type of unrelated change. To the extent that 'compensatory metathesis' correlates with regular metathesis, and other types of metathesis with sporadic metathesis, this may not be an undesirable result. However, the fact that Dawan speakers have lento/allegro speech variants such as *tenu ~ teun* 'three' with no evidence of coarticulation does seem difficult to reconcile with this interpretation.

One might argue that the stages leading to metathesis were historical scaffolding that has since been lost while leaving both variant forms, but this is easier to maintain where there is phonological or grammatical conditioning that is independent of speech tempo. Whether the Blevins and Garrett theory can be extended to account for ʔC or Cʔ metathesis in languages of the central Philippines is also unclear.

Finally, Cook (2004) has presented a novel argument that metathesis in Hawaiian may be motivated by semantic iconicity. While some of the etymologies he uses to illustrate his thesis are open to question, the thesis itself may be worth exploring further.

4.5   Conspiracies

Place assimilation and metathesis have no structural property in common, yet these processes can co-conspire to insure that certain outputs are avoided. In Tagalog *asín* 'salt' : *asn-ín* 'salted' syncope triggers no other change, but in *atíp* 'roof' : *apt-án* 'roofed' the expected cluster *-tp-* metathesises to *-tp-*. This is not an isolated example, but part of a larger process in which derived clusters of stops or nasals in the order coronal-noncoronal metathesise if they have the same value for nasality, but otherwise undergo assimilation of the nasal to the stop, whether the nasal precedes or follows the consonant to which it assimilates (Blust 1971):

Table 4.46 Coronal-noncoronal consonant cluster avoidance in Tagalog

Base	Affixed form	English
atíp	apt-án	roof; roofed
taním	tamn-án	to plant; plant on
baníg	baŋg-án	mat; lay a mat
ganáp	gamp-án	fulfill, do duty
liníb	limb-án	to close
panaginíp	panagimp-án	to dream
datíŋ	datn-án	to arrive

Apart from a few unexplained exceptions such as *halík : hagk-án* 'sniff, kiss', no change occurs for similar clusters that are not coronal-noncoronal (*higít : higt-án* 'haul, pull', *kápit : kapt-án* 'grasp, embrace', *táŋan : taŋn-án* 'grasp'), or for coronal-noncoronal clusters that do not consist of two stops, two nasals, or a stop and a nasal in either order, *gísiŋ : gisŋ-án* 'awake', *hasík : hask-án* 'sow'). A similar avoidance of coronal-noncoronal clusters through multiple repair strategies is seen in other languages of the central Philippines, such as Cebuano (*atúp* 'roof' : *atp-án ~ apt-án* 'put a roof on', *tanúm* 'to plant' : *t<al>amn-án-an* 'area prepared for planting'). In Chamorro avoidance of coronal-noncoronal clusters apparently applies only to segments with the same value for nasality, thus allowing the irregular phonological relationship seen in *tanom : tatm-e* 'to plant seeds or seedling', without adopting the more drastic measure of metathesis seen in *atof* 'roof' : *aft-e* 'to roof, cover with a roof'. Many languages that create such clusters through syncope tolerate the disfavoured sequence, as with Bontok *atəp : atp-an* 'roof thatched with *cogon* grass', or Mansaka *atub* 'trap for wild pigs' : *kyaka-atb-an* 'was caught in a trap', but this does not detract from the significance of the cross-linguistic agreements, since marked features may or may not be unmarked in a given language, and coronal-noncoronal clusters are avoided in other languages, including English.

Other alternations triggered by affixation seem to be motivated by avoidance of certain consonant clusters, but paradoxically these clusters are found in unaffixed bases. Blake (1925:300ff) lists more than 230 Tagalog bases that syncopate under affixation, and where this results in -lC- the segments frequently metathesise, as in *bílin* : *binl-án* 'commission, charge', *habílin* : *habinl-án* 'deposit', *halíli* : *halinl-án* 'substitute', *kilála* : *kilanl-ín* 'be acquainted with', *silíd* : *sidl-án* 'put into', *súlid* : *sudl-án* 'spin', or *taláb* : *tabl-á*n 'penetrate'.[38] Certain subregularities are clear. First, metathesis never occurs in -lC- if C = ʔ or *h* (19 examples). Second, in the two known examples metathesis does not occur if C is velar: *halík* : *hagk-ín* 'sniff, kiss', *kalág* : *kalg-ín* 'untie, loosen'. Although a single example does not constitute a regularity, the cluster -*lt*- also remains unchanged in *palít* : *palt-ín* 'exchange, barter'. The metathesis of derived -*ln*- clusters in Tagalog, and the failure of -*lʔ*- or -*lh*- clusters to change is consistent with the morpheme structure of the language, since -*ln*- and preconsonantal laryngeals are disallowed in unaffixed bases (Table 4.27). The other facts noted here, however, are less easily explained. Unaffixed bases show no clusters of *l* + velar consonant, but the form *kalg-ín* tolerates such a cluster, although *hagk-ín* does not. A generalising statement therefore might be that -lC- metathesises when C is labial or alveolar. This generally works, but fails with *palt-ín*. Since the cluster -*tl*- is universally marked, metathesis is not expected in this form, yet the cluster -*tl*- oocurs in such unaffixed Tagalog bases as *bitlág* 'seat in boats made of cane or rattan', *butlíg* 'wen', or *sutláʔ* 'silk'. Most surprising of all is the metathesis of -*ld*-, and -*lb*- in three forms, since both -*lb*- and -*ld*- occur in a number of unaffixed bases: *albáy* 'support, stand', *bulbóg* 'bruised', *malbás* 'k.o. medicinal plant', *paldák* 'hardened or flattened by repeated treading or trampling', *paldás* 'faded, discoloured', *tuldók/tudlók* 'period mark'.

4.6   Accidental complementation

Theoretical phonology has made remarkable progress over the past 40 years, but this has been confined mostly to the treatment of phonological alternations, leaving the study of complementation almost untouched. Traditional phonemic theory held that phones that are phonetically similar and in complementary distribution should be united into single phonemes. In most cases this procedure makes sense both synchronically and diachronically, since complementation usually arises from the influence of neighbouring speech sounds on one another, and the process of uniting allophones into phonemes simply 'undoes' the conditioned changes that gave rise to allophony in the first place. The key to following such procedures 'correctly' has always been phonetic similarity: where phones are in complementary distribution but are phonetically dissimilar, as with English *h* and *ŋ*, they can be assigned to different phonemes, but otherwise they cannot. However, close attention to particular cases shows that this procedure can lead to error. As anyone who must use the results of a phonemic analysis for linguistic comparison soon discovers, the most important consideration in uniting complementary phones is not phonetic similarity, but whether the distribution is due to environmental conditioning or to adventitious causes.

By almost any criterion [s] and [h] must be regarded as phonetically similar, since in many languages, including New World dialects of Spanish and Portuguese, and Chamic languages such as Jarai, [h] is a word-final allophone of /s/. In Tring of northern Sarawak a

---

[38]   Both -*h* and -*n* appear in Tagalog as thematic consonants after some stems that end in a vowel when unsuffixed. The former never appears as a surface final consonant, but the latter does.

similar situation holds: [h] occurs in final position and [s] elsewhere: *siaʔ* 'red', *sikuh* 'elbow', but *laʔih* 'male', *matah* 'eye', or *təluh* 'three'. Tring *s*, however, reflects *s (occasionally *t) before *i, and *-h* was added after word-final vowels. Unlike most languages which exhibit such complementation, then, the phones [s] and [h] in Tring have no historical connection, and if they were united in accordance with the same procedures that require us to unite [s] and [h] in languages such as Spanish or Jarai, we would have to posit underlying forms such as *sikus* 'elbow', *laʔis* 'male', *matas* 'eye', or *təlus* 'three' for words that originally ended in a vowel (*siku, *laki, *mata, *telu), with disastrous consequences for statements about sound change. Cases such as this show that complementation is not always harmlessly accidental as with English *h* and *ŋ*, but under certain historical circumstances it can mimic phonological conditioning in ways that are only fully comprehensible through a comparative analysis.

4.7 Double complementation

Although accidental complementation is rare for consonants, accidental complementation among vowels is found in a number of AN languages. Proto Austronesian had a four-vowel system, the vowel 'triangle' plus schwa. Of these vowels the schwa is the default in many historically independent cases of epenthesis, and of prepenultimate neutralisation. The schwa is also noteworthy for its distributional limitations: it could not occur before a glide, another vowel or word-finally. Many languages in island Southeast Asia retain these constraints, and some of these have developed new mid vowels *e* and *o* through the monophthongisation of *-ay and *-aw. In such languages, of which Kayan of central Borneo can serve as an example, the schwa is in complementary distribution with both *e* and *o*. Since all three are mid vowels, and since the historical change of schwa to *e* or *o* is repeatedly attested in AN languages, the schwa must be regarded as phonetically similar to *both e* and *o*. As with [s] and [h] in Tring, the complementation and phonetic similarity of schwa, *e* and *o* is accidental. However, in Kayan the relationship of schwa to *e* and *o* creates a quandary even without considering diachronic consequences: if *e* and *o* have equal claims to being the same phoneme as schwa, on what basis can the schwa be combined with either of them? Note that this situation differs from that of a phoneme with three allophones, since *e* and *o* clearly contrast with one another (*ate* 'liver' : *lavo* 'rat').

4.8 Free variation

The term 'free variation' was coined in American Structuralism to signal an unconstrained interchangeability of two or more phones that are normally, but not necessarily, subphonemic. As noted by Labov (1972:xiv), a tacit assumption of post-Bloomfieldian linguistics in America was that 'free variation could not in principle be constrained.' Much of Labov's early work showed that phonological variation, which arises through non-social causes, is often recruited for purposes of social identity, and so acquires an emblematic value that drives the implementation of sound change. To call such variation free, then, ignores the important ways in which it may be constrained by socially expressed meaning. However, even within quantitative sociolinguistics it is commonly assumed that interchangeable phones have an equal likelihood to occur in any given morpheme, since if variants were favored by particular morphemes it would become difficult to distinguish variation from contrast. The data from some AN languages

nonetheless suggests that 'free' variation can involve differential probabilities of occurrence in a given environment or a given lexical item without any obvious relationship to social determinants. Schütz (1994:119ff) has shown, for example, how the pronunciation of Hawaiian /w/ has caused problems for native speakers of English since the early decades of the nineteenth century. Opinions have differed as to whether /w/ varied between [w] and [v], or whether what Westerners were mishearing was phonetically intermediate between the two (but not [β]). In modern Hawaiian /w/ has been resolved into English-like allophones [w] and [v] which are in 'free' variation, but apparently have different probabilities of appearing in a given position in the word (thus Waikīkī [waiki:ki:], but Hawai'i [haváiʔi]). Variation of this kind does not appear to be related to social dynamics, but purely to phonological environment.

5   *The lexicon*

5.0   Introduction

Scholars with divergent research interests have viewed the vocabulary of a language in very different ways. In early Generative Grammar lexical information was conceived as a depository of exceptions: any feature of language that could not be produced by rule and so assigned to the grammar was condemned to the lexicon. An unfortunate consequence of this view was that vocabulary appeared to be little more than a mass of particularities devoid of theoretical interest. Historical linguists and anthropological linguistics have long maintained a more sanguine view of the value of lexical data. Not only are there important features of structure in the lexicon of any language, but social, cultural or technological history is often 'captured' in lexical data. This can be seen both in cognate sets that have a non-universal semantic content, as reflexes of *pajay 'riceplant, rice in the field' in many AN languages, and through the etymology of words in individual languages, as with English 'pen' (< Latin *penna* 'feather', reflecting the former use of feather quills as writing implements). In addition, quantitative measures of cognation have been used, albeit problematically, in subgrouping by lexicostatistics, and widespread skepticism among linguists about this use of lexical data has recently been challenged by biological taxonomists who believe that methods which make use of Bayesian inference as developed in biological phylogenetics can overcome the problems encountered with traditional approaches to cognate counting (Greenhill and Gray 2009).

This chapter examines both synchronic and diachronic aspects of lexical data in AN languages, and relates these to wider issues in linguistics where this appears appropriate. The lexicon touches all features of linguistic structure, and for practical purposes it will be necessary to confine the discussion to a small number of topics. Those chosen are: 1) numerals and numeration, 2) numeral classifiers, 3) colour terms, 4) demonstratives, locatives, and directionals, 5) pronouns, 6) metaphor, 7) language names and greetings, 8) semantic change, 9) lexical change, and 10) linguistic paleontology. Because of its multiple connections with other aspects of linguistic structure, lexical data cannot always be clearly separated from concerns which belong more properly to morphology or syntax. For example, an inclusive/exclusive distinction in pronouns has no known syntactic consequences, and so is best treated as part of the lexicon. The description of different pronoun sets, on the other hand, is inseparable from the syntax, and is best treated in the discussion of that part of a language. As a result, the discussion of pronouns must be divided between different chapters. The core of what is covered in this chapter can thus be characterised roughly as 'lexical semantics', although some topics fall outside the range of this label. In the interest of coherence both synchronic and diachronic material is handled together. The general organisation thus appeals to lexical domains that show a high degree of semantic coherence, but additional sections are included to allow some latitude in addressing other aspects of lexical/semantic structure or history.

5.1   Numerals and numeration

Numerals form a well-defined subsystem within the lexicon of any language. While all languages have numerals, systems of counting vary considerably, as do the methods of deriving non-primary numerals from the primary set of cardinal forms. In addition, some languages rely heavily on numeral classifiers, while others do quite well without them. Austronesian languages vary considerably in the complexity of their numeral systems, although a basic decimal system used in serial counting and for many other functions is found in hundreds of the modern languages.

5.1.1 Structurally intact decimal systems

PAN had a decimal system of counting that has been retained in most of its descendants. Table 5.1 gives the PAN forms for 1-10 with supporting evidence from five widely separated daughter languages (forms within parentheses are lexical innovations):

Table 5.1 The numerals 1-10 in Proto Austronesian and five descendants

PAN	Paiwan	Cebuano	Malagasy	Tetun	Hawaiian
*esa/isa	ita	usá	ísa	ida	(ʔe-kahi)
*duSa	ḍusa	duhá	róa	rua	ʔe-lua
*telu	tjəlu	tulú	télo	tolu	ʔe-kolu
*Sepat	səpatj	upát	éfatra	hat	ʔe-hā
*lima	lima	limá	dímy	lima	ʔe-lima
*enem	ənəm/unəm	unúm	énina	nen	ʔe-ono
*pitu	pitju	pitú	fíto	hitu	ʔe-hiku
*walu	valu	walú	válo	walu	ʔe-walu
*Siwa	siva	(siyám)	sívy	sia	ʔe-iwa
*sa-puluq	ta-puluq	púluʔ	fólo	sa-n-ulu	(ʔumi)

The languages in Table 5.1 represent almost the entire geographical range of the AN language family. The numeral systems of these languages show few innovations, and those that occur affect neither the decimal basis of counting, nor the monomorphemic, non-derived character of the numerals themselves. These languages can therefore be characterised as having structurally intact decimal systems. The great majority of AN languages have systems of this general type.

Before illustrating deviations from the norm a few comments on Table 5.1 will be useful. First, the variation in PAN *esa/isa is an instance of doubleting, a phenomenon that will be treated in a later section (the variation in Paiwan *ənəm/unəm*, on the other hand is dialectal). Second, as seen in Chapter 4, stress is phonemic in most languages of the northern and central Philippines. In the numerals, however, it is fairly common for stress to fall on the ultima in 1-9, but on the penult in 10. In addition to Cebuano this pattern is found in Ilokano, Isneg, Bontok, Pangasinan and in a modified form in other languages such as Bikol (where 'one' and 'ten' carry penultimate stress, but all other numerals are oxytone). Considerations of rhythm and phonological linking have long been known to be important in serial counting, and a rhythmic shift at the conclusion of a natural unit signals consummation. In decimal systems this shift of rhythm would be expected on the numeral '10'. In languages such as English which lack lexical stress (as opposed to morphological stress), and where most numerals are monosyllabic, the consummation of a decimal unit is

expressed by intensity or loudness. In languages that have contrastive lexical stress the same effect can be achieved by a rhythmic shift, and there can be little doubt that such considerations have played a part in skewing the prosody of this lexical domain in Philippine languages. Not all Philippine languages adhere to this pattern. In Tagalog, for example, the numerals '4' and '6' are stressed on the penult, while *puóʔ* 'group of ten (obsolete)', *sa-m-púʔ* '10' are stressed on the final. However, Old Tagalog had *sa-m-púoʔ* (Jason Lobel, p.c.) and its trisyllabic form may have rendered the use of rhythmic alternation redundant. The Malagasy system, by contrast, descends from one in which stress was predictably penultimate, but the addition of a supporting vowel *-a* after word-final consonants produced a superficial contrast of initial stress in some forms vs. penultimate stress in most others (like English 'ten', the form *fólo* presumably is pronounced, with greater intensity than the numerals 1-9).

Third, although it is barely visible in the data given here, Tetun *sa-n-ulu* reflects a numeral ligature *ŋa that is widespread in AN languages outside Taiwan. Although the PAN word for 'ten' is reconstructed as *sa-puluq, then, the PMP word is *sa-ŋa-puluq. In both cases *sa- is a clitic form of *esa 'one', and the meaning of the morphologically complex form can best be translated as 'one group of ten'. Finally, the Hawaiian system uses what Elbert and Pukui (1979:158) call a "general classifier *ʻe-* (or rarely *ʻa-*)." Elsewhere (Pukui and Elbert 1971:1) they describe *ʻa-* (/ʔa/-) as a "refix to numbers from one through nine, especially for counting in series." A number of other Oceanic languages use an apparently related numeral marker, and something very similar also appears in a few non-Oceanic languages, as in Palauan, where the most basic set of numerals can be represented as 1) *taŋ*, 2) *e-ruŋ*, 3) *e-dey*, 4) *e-waŋ*, 5) *e-yim*, 6) *e-loləm*, 7) *e-wid*, 8) *e-ay*, 9) *e-tiw*, 10) *təruyəʔ*.[39] What is noteworthy about such systems is that the numeral 'ten' is treated differently from other basic numerals (in many languages 'one' does not seem to matter). In effect, then, the use or withholding of a preposed numeral marker in Palauan and many Oceanic languages serves the same function as rhythmic alternation in Philippine languages: it marks the consummation of a decimal round. Many other languages, such as Paiwan or Tetun, lack either a preposed numeral marker or contrastive lexical stress, but in many of these the word for 'ten' is morphologically complex, and is therefore either trisyllabic or quadrisyllabic, thus marking the consummation of a decimal round through its deviant canonical shape, as with Mussau 1) *sesa*, 2) *lua*, 3) *tolu*, 4) *ata*, 5) *lima*, 6) *nomo*, 7) *itu*, 8) *ualu*, 9) *sio*, 10) *sa-ŋaulu*.

5.1.2 Structurally modified decimal systems

In numeral systems of this type (sometimes called 'imperfect decimal' systems) the basic decimal structure is retained, but some numerals are innovative products of addition, multiplication, or subtraction (never of division). Several Formosan languages, for example, have decimal systems in which '6' = 2x3 and/or '8' = 2x4, either alone or in combination with additive numerals:

---

[39]   For essentially the same reasons given by Dyen (1971c) my transcriptions of Palauan deviate from the standard orthography as represented by, e.g. Josephs (1975).

Table 5.2 Structurally modified decimal systems in Formosan languages

Seediq	Thao	Saisiyat (Taai)	English
kiŋal	tata	æhæ	one
daha	tusha	roʃa	two
təru	turu	toLo	three
spat	pat	ʃəpat	four
lima	rima	Lasəb	five
matəru	makalhturuturu	ʃayboʃil	six
mpitu	pitu	ʃayboʃilo æhæ	seven
maspat	makalhshpashpat	kaʃpat	eight
məŋari	tanacu	Læʔhæ	nine
mahal	maqcin	laŋpəz	ten

It is striking that Seediq and Thao, which are neither closely related nor in contact, show innovations whereby 6 = 2x3 and 8 = 2x4, but no other numeral is derivative. In Saisiyat, which is not closely related to either of these nor contiguous to them, a similar innovation appears in the word for '8', but 6 is unanalyzable, 7 = 6+1, and 9 = 10-1. Several other Formosan languages, including the extinct Favorlang, Taokas and probably Siraya, also show an innovation whereby 8 = 2x4, although the quality of the available materials makes definite conclusions difficult. These historically secondary numerals in various Formosan languages thus appear to be an areal feature which may have begun with 8 = 2x4. The present distribution of the languages that show this feature is not compatible with an areal hypothesis, but this discrepancy may provide a clue to past population movement within the island. However, innovations of this type may arise independently, since Motu of southeast New Guinea has a structurally similar (but historically unrelated) set of innovations: 1, 2, 3, 4, 5, 2x3, 7, 2x4, 2x4+1, 10 (*ta, rua, toi, hani, ima, tauratoi, hitu, taurahani, taurahani-ta, g^w auta*).

A second type of structurally modified decimal system that is found in widely separated languages is one that makes use of subtractive numerals. This can be illustrated by Malay, Yapese, and many of the languages of the Admiralty Islands, of which Levei and Penchal may be taken as representative:

Table 5.3 Austronesian decimal systems with subtractive numerals

Malay	Yapese	Levei	Penchal
satu	reeb	eri	sɨw
dua	l'agruw	lueh	lʊp
tiga	dalip	toloh	tʊlʊp
əmpat	qaniŋeeg	hahup	talɨt
lima	laal	limeh	rurɨn
ənam	neel'	cohahup	ʊnʊp
tujuh	meedlip	cotoloh	karutʊlʊp
dəlapan	meeruuk	colueh	karulʊp
səmbilan	meereeb	coeri	karusɨw
sə-puluh	ragaag	ronoh	saŋahul

Yapese, Levei and Penchal share a synchronically transparent use of subtraction to form numerals between '6' and '10'. In Yapese and Penchal only 7-9 are subtractive, while '6'

is an unanalyzable base. In Levei all numerals from 6-9 are subtractives. Since each of these languages has an independent word for '10', their counting systems must be classified as decimal. Malay differs from them in that its use of subtractive numerals is only apparent from comparative data. In Malay the numerals 7-9 are innovative, but have been formed on two different principles. The word for '7' derives from *tuzuq 'to point', evidently from the position of the index in finger counting. The words for '8' and '9' are believed to derive from *dua alap-an, and *sa-ambil-an, where *alap is a widely reflected verb meaning 'to fetch', but is no longer found in Malay, and *ambil* is the contemporary Malay equivalent. The earlier meanings of Malay *dəlapan* and *səmbilan* were thus 'two taken away' and 'one taken away', although this analysis is synchronically opaque. Similar innovations are shared with Sundanese of west Java, and the Chamic languages of mainland Southeast Asia, where the morphology in '8' and '9' is curiously reversed, as in Jarai *səpan* (< *sa-alap-an) '8', and *dua rəpan* (< *dua alap-an) '9'. A few languages of eastern Indonesia have counting systems of the form 1-8, 10-1, 10, as with Buru (Grimes 1991:293), and Soboyo (Fortgens 1921).

A number of languages in eastern Indonesia have modified decimal systems that use addition together with some other arithmetical operation. Adequate data on counting systems is available for few languages of this area, but fragmentary information suggests that in several of the languages of Flores, including Keo, Ngadha, Lio and Ende 6 = 5+1, and 7 = 5+2, and in several of the languages of western Sumba, including Kodi and Lamboya 8 = 2x4, while '10' is unanalyzable. Systems on which the fullest information is available include Lio of central Flores, with 1, 2, 3, 4, 5, 5+1, 5+2, 2x4, 10-1 (where '10' is not expressed), and 10: *əsa, rua, təlu, sutu, lima, lima əsa, lima rua, rua-m-butu, təra əsa, sa-m-bulu*, and Kédang of the Solor Archipelago, with 1, 2, 3, 4, 5, 6, 7, 8, 5+4, 10: *udeʔ, sue, talu, apaʔ, ləme, əʔnəŋ, pitu, buturai, ləme apaʔ, puluh*, (where *puluh* is a Malay loan). The Kédang system is particularly unusual in making use of addition only for the number '9'. Languages that have a structurally modified decimal system based purely on addition include Pazeh of north-central Taiwan: 1) *ida*, 2) *dusa*, 3) *turu*, 4) *supat*, 5) *xasəp*, 6) *xasəp uza*, 7) *xasəp i dusa*, 8) *xasəp i turu*, 9) *xasəp i supat*, 10) *isit*, Manam of the north coast of New Guinea (Lichtenberk 1983:337): 1) *teʔe*, 2) *rua*, 3) *toli*, 4) *wati*, 5) *lima*, 6) *lima teʔe*, 7) *lima rua*, 8) *lima toli*, 9) *lima wati*, 10) *ʔulem^wa*, Kilivila of the Louisiade Archipelago southeast of New Guinea (Senft 1986:77ff): 1) *tala*, 2) *yu*, 3) *tolu*, 4) *vasi*, 5) *lima*, 6) *lima tala*, 7) *lima yu*, 8) *lima tolu*, 9) *lima vasi*, 10) *luwatala* (= 'one group of ten'), Tigak of New Ireland (Beaumont 1979:105): 1) *sakai*, 2) *pauak*, 3) *potul*, 4) *poiat*, 5) *palmit*, 6) *palmit sakai*, 7) *palmit pauak*, 8) *palmit potul*, 9) *palmit poiat*, 10) *saŋauluŋ*, and Pije of New Caledonia (Haudricourt and Rivierre 1982:261): 1) *heec*, 2) *haluk*, 3) *hien*, 4) *hovac*, 5) *nim*, 6) *ni-b^weec/hamen heec*, 7) *ni-b^waluk/hamen haluk*, 8) *ni-b^wien/hamen hien*, 9) *ni-bovac/hamen hovac*, 10) *paidu*. A similar system is also found in Ilongot of the northern Philippines, although it is unique in this area. According to the data in Reid (1971) Ilongot has innovated additive numerals 5+1, 5+2, 5+3, and 5+4: 1) *sit*, 2) *duwa*, 3) *təgu*, 4) *qəpat*, 5) *tambiaŋ*, 6) *tambiaŋ nu sit*, 7) *tambiaŋ nu duwa*, 8) *tambiaŋ nu təgu*, 9) *tambiaŋ nu qəpat*, 10) *tampo*.[40]

Finally, a few lexical innovations in numerals that have no structural consequences are interesting for their semantic/cultural content. The number 'five' is homophonous with 'hand' in many AN languages, but this was already true in PAN, and should occasion no surprise. In Sundanese *gənəp* 'six' reflects *genep 'complete, sufficient'. Initially, the

---

[40] The Ilongot word for 'ten' does not appear in Reid (1971). This form was supplied courtesy of Steve Quakenbush, Summer Institute of Linguistics, Philippine Branch.

source of this semantic change is puzzling, but as Gonda (1975:444) notes, reflexes of *genep in languages such as Malay, Toba Batak, and Makasarese include the meaning 'even, of numbers'. Barnes (1974:76, 1980) has observed that in many of the languages of Indonesia odd numbers are regarded as 'incomplete' (and propitious), and that even numbers are regarded as 'complete'. In some languages, such as Kédang of the southwest Moluccas, this idea has wider symbolic connections with material culture (especially house construction) and with matters of life, death, and the transmission of souls. While this establishes a connection between *genep and the numeral system, it is not clear why 'six' should be seen as the prototypical even number. Gonda (1975:444) speculated that this choice was made in Sundanese because, as Dahl (1981b:50) put it "The hand was full at 5, therefore this word is used to designate the next numeral." But if all that were involved symbolically were completion in finger counting the number 'five' would certainly be a better candidate than 'six' for this distinction. Moreover, (Donohue 1999:107) reports *gana* 'four' in northern dialects of Tukang Besi of southeast Sulawesi. Although the expected Tukang Besi reflex of *genep is **gono, and *gana* may be a borrowing of Malay *gǝnap* 'completing; rounding off; even (of numbers)', the evidence from this language makes it clear that the number 'six' has no special status vis-à-vis the idea of completion. Rather than finger counting, what appears more relevant to understanding these semantic changes is the importance of the distinction between even and odd numbers, and the way this distinction is associated with symbolic aspects of social organisation or material culture in a number of AN-speaking societies. By contrast, finger counting does appear to lie behind reflexes of *tuzuq 'seven' which are found over a large swath of western Indonesia, including nearly all languages of Borneo, Malay and its closest relatives (Iban, Minangkabau, etc.), the Chamic languages, Sundanese, and in a few other languages that have borrowed heavily from Malay, such as Rejang of southern Sumatra. As noted by Wilkinson (1959:1242) this numeral is etymologically associated with *tuzuq 'index finger; to point'), and the change may have resulted from decimal counting on one hand with the index being the seventh finger.

5.1.3 Non-decimal counting systems

As seen in the previous examples, some features of quinary counting systems appear here and there in the AN language family. However, fully developed quinary systems are rare, and most of these are found in Melanesia. A number of languages spoken on Efate, in the Shepherd Islands, and on Epi and Paama in central and south-central Vanuatu have quinary systems with the structure 1, 2, 3, 4, 5, 6, 5+2, 5+3, 5+4, 2x5, as in Bonkovia (Epi Island) 1) *ta*, 2) *cua*, 3) *tolu*, 4) *veri,* 5) *cima*, 6) *wora*, 7) *oko-lua*, 8) *oko-rolu*, 9) *oko-veri*, 10. *lua-lima*. In addition, Anejom and most languages of Tanna in southern Vanuatu have quinary systems with the structure 1, 2, 3, 4, 5, 5+1, 5+2, 5+3, 5+4, 5+5, as in Lenakel 1) *karena*, 2) *kiu*, 3) *kǝsil*, 4) *kuvǝr*, 5) *katilum*, 6) *katilum karena*, 7) *katilum kiu*, 8) *katilum kǝsil* , 9) *katilum kuvǝr*, 10) *katilum katilum* (Lynch 2001:139ff). Seimat, alone among the languages of the Admiralty Islands, also has a true quinary system, with the structure 1, 2, 3, 4, 5, 5+1, 5+2, 5+3, 5+4, 2x5. Although this may seem like a drastic departure from the decimal system that these languages inherited from a remote common ancestor, even more drastic innovations in numeral systems are found in some AN languages of New Guinea, where they clearly reflect Papuan contact influence. A striking example of such an innovative system is seen in Gapapaiwa of Milne Bay, New Guinea, which has 1) *sago*, 2) *rua*, 3) *aroba*, 4) *rua ma rua*, 5) *miikovi* (*ima ikovi* = 'hand

finished'), 6) *miikovi ma sago*, 7) *miikovi ma rua*, 8) *miikovi ma aroba*, 9) *miikovi ma rua ma rua*, 10) *imarua*, hence 1, 2, 3, 2+2, 5, 5+1, 5+2, 5+3, 5+2+2, 2x5 (McGuckin 2002). The apparent instability of such systems is seen in the variant system that the present writer recorded for the same language: 1) *sago*, 2) *rua*, 3) *rua ma sago*, 4) *rua ma rua*, 5) *rua ma rua ma sago*, 6) *aroba ma aroba*, 7) *aroba ma aroba ma sago*, 8) *aroba ma aroba ma rua*, 9) *aroba ma aroba ma aroba*, 10) *aroba ma aroba ma aroba ma sago*, hence 1, 2, 2+1, 2+2, 2+2+1, 3+3 (where '3' is not the basic numeral used in this sense), 3+3+1, 3+3+2, 3+3+3, 3+3+3+1.

It is difficult to characterise the Gapapaiwa system in positive terms, since it has features of both ternary and quinary methods of counting. It is also noteworthy that multiplication has no role in this system, a factor that must place a severe practical limitation on the use of numerals much beyond '10'. Similar counting systems are common among the Papuan languages of New Guinea, as noted by Laycock (1975:222), who speaks of "the widespread binary (or 'Australian') system, where only the first two numerals are expressed by separate roots." Table 5.4 provides a bird's-eye view of types of numeral systems found in AN languages. Much of the data for Melanesia is drawn from Lynch, Ross and Crowley (2002), but the table as a whole has been compiled from multiple sources. This typology is based only on the manner of representation of the first ten numerals ('basic numerals'). A consideration of higher numerals would introduce additional types for some languages:[41]

Table 5.4 Types of basic numeral systems in Austronesian languages

1.	1-10	:	most AN languages
2.	1-5, 5+1, 5+2, 5+3, 5+4, 10	:	Pazeh, Ilongot, Sobei, Kairiru, Manam, Arop-Lokep, Kilivila, Tigak, Bali-Vitu, Sakao, Neve'ei, Port Sandwich, Pije, Cèmuhî, Xârâcùù
3.	1-5, , 5+1, 5+2, 5+3, 5+4, 2x5	:	Dehu, Nengone
4.	1-5, 5+1, 5+2, 2x4, 10-1, 10	:	Lio
5.	1-5, 2x3, 7, 2x4, 9-10	:	Seediq, Thao
6.	1-5, 2x3, 7, 2x4, 2x4+1, 10	:	Motu
7.	1-5, 2x3, 2x3+1, 2x4, 2x4+1, 10	:	'Ala'ala
8.	1-5, 10-4, 10-3, 10-2, 10-1, 10	:	Levei
9.	1-6, 5+2, 5+3, 5+4, 10	:	Mwotlap, Sye
10.	1-6, 6+1, 2x4, 10-1, 10	:	Saisiyat
11.	1-6, 10-3, 10-2, 10-1, 10	:	Yapese, Penchal
12.	1-8, 5+4, 10	:	Kédang
13.	1-8, 10-1, 10	:	Buru, Soboyo
14.	1-2, 2+1/3, 2+2, 5, 5+1, 5+2, 5+3, 5+2+2, 2x5	:	Gapapaiwa
15.	1-5, 5+1, 5+2, 5+3, 5+4, 5+5	:	Anejom, Lenakel
16.	1-5, 5+1, 5+2, 5+3, 5+4, 2x5	:	Takia, Yabem, Kaulong, SE Ambrym, Iaai
17.	1-6, 5+2, 5+3, 5+4, 2x5	:	Lamenu, Bonkovia
18.	Unclassifiable	:	Tobati, Banoni

---

[41] Lynch (2009b:394) provides data for some Vanuatu systems that may be Type 2, but are difficult to analyse. In Merei *ese* '1', *ruwa* '2', *tolu* '3', *vat* '4', *lima* '5', *maravo* '6', *ravorua* '7', *raptol* '8', *raitat* '9', *saŋavul* '10', for example, the numerals 7-9 can be seen as having the structure 5+1, 5+2, 5+3 and 5+4, except that the value of *ravo* is unclear.

As seen even in this sketchy account, the most common type of innovative numeral system in AN languages is Type 2, an imperfect decimal system that makes use only of addition in forming the numerals 6-9. The rarest types of innovative systems are those that 1) use subtraction, either alone or in conjunction with addition or multiplication, 2) use addition arbitrarily for certain numerals from '6' to '9' but not others, as with Kédang, and 3) those with a base lower than 'five', as Gapapaiwa.

The Tobati system is 1) *tei*, 2) *ros*, 3) *tor*, 4) *aw*, 5) *mnyiam*, 6) *rwador*, 7) *mandosim*, 8) *rughondu*, 9) *rwador mani* or *tei am,* 10) *jer roj foghonjam*, and the Banoni system used for counting anything other than round objects is said to be 1) *kadaken*, 2) *toom*, 3) *dapisa*, 4) *tovatsi*, 5) *ghinima*, 6) *bena*, 7) *bena tom*, 8) *bena kapisa*, 9) *visa*, 10) *manogha*. The first of these is difficult to classify because *rwador* '6' reappears together with an unknown morpheme as one of two forms representing '9'. Given the difficulty of relating six and nine through any kind of transparent arithmetical operation there is no obvious way to interpret this recurrence of *rwador* in the numeral system. The second option for '9' appears to be 10-1, but the word for '10' evidently is a descriptive phrase for which no translation is given. The Banoni numeral system is equally puzzling and difficult to classify. The first six numerals are represented by independent morphemes, but '7' appears to be 6+2 and '8' appears to be 6+3.

It is clear from these observations that many AN languages have transformed an inherited decimal system to one that is computationally more cumbersome. Since the computational efficiency of numeral systems is sometimes correlated with level of cultural advancement (Greenberg 1978b:290ff), such trajectories of change must either provide fodder for theories of cultural devolution, or cast doubt on the proposed correlations. It is true that the most extreme deviations from the original decimal system, as that of Gapapaiwa, are clear products of culture contact in which small AN-speaking populations have become culturally and linguistically assimilated to their more numerous Papuan-speaking neighbours. However, the distribution of many other AN languages with structurally modified numeral systems is more difficult to explain by contact, as these are found in Taiwan, the northern Philippines, western Indonesia and mainland Southeast Asia, the Lesser Sundas and southern Moluccas, and various parts of Melanesia where they are not contiguous with Papuan languages.

5.1.4 Onset 'runs'

Matisoff (1995) has shown that many Tibeto-Burman counting systems have 'runs' of up to 7 or 8 consecutive numerals with identical onsets. This phenomenon does not appear to be so highly developed in AN languages, but turns up occasionally. In Thao, for example, regular sound change would have produced 1) *ta*, 2) *shusha*, 3) *turú*, but this rhythmically awkward beginning was altered to 1) *tata*, 2) *tusha*, 3) *turu*, with three consecutive forms that have a *t*-onset, and repeated penultimate stress. The most extreme form of numeral onset run for an AN language is found in Buma (or Teanu), spoken on Vanikoro Island in the southern Santa Cruz Archipelago, where all numerals from 2-9 begin with *t*: 1) *iune*, 2) *tilu*, 3) *tete*, 4) *teva*, 5) *tili*, 6) *tuo*, 7) *tibi*, 8) *tua*, 9) *tudi*, 10) *saŋaulu/uluko*. Attention to the second syllable suggests that this pattern arose through prefixation with *tV (perhaps a borrowing of the Polynesian article *te*) to the first syllable of an inherited Oceanic lexical base: POC *rua 'two', *tolu 'three', *pati 'four', *lima 'five', *onom 'six', *pitu 'seven', *walu 'eight', *siwa 'nine'. Mwotlap and Neve'ei

(Vinmavis) of Vanuatu have double onset runs that are individually shorter, but collectively as long as the eight-term run of Buma: Mwotlap 1) *vitwaɣ,* 2) *voyo,* 3) *vetel,* 4) *vevɛt,* 5) *tevelem,* 6) *lɛvɛtɛ,* 7) *liviyo,* 8) *lɛvɛ,* 9) *lɛvɛvɛt,* 10) *sɔŋwul,* Neve'ei 1) *sefax,* 2) *iru,* 3) *itl,* 4) *ifah,* 5) *ilim,* 6) *nsouh,* 7) *nsuru,* 8) *nsutl,* 9) *nsafah,* 10) *naŋafil.* The *v*-initial run in Mwotlap apparently results from prefixation with an unknown morpheme, and the *l*-initial run from compounding on a quinary base of variable vocalism (5+1, 5+2, 5+3, 5+4). In Neve'ei the *i*-initial run probably results from prefixation with a numeral marker or predication marker of the type seen in Hawaiian *Ɂe-kahi, Ɂe-lua, Ɂe-kolu, Ɂe-hā, Ɂe-lima,* or Palauan *taŋ, e-ruŋ, e-dey, e-waŋ, e-yim*; the *n*-initial run apparently results from compounding on a quinary base (5+1, 5+2, 5+3, 5+4).

5.1.5 Higher numerals

Table 5.1 gives PAN 1-10. While these numerals are straightforward, the reconstruction of a base for '100' rests on slender foundations, although such a form must have existed. Most Malayo-Polynesian languages in island Southeast Asia and a minority in the Pacific reflect *Ratus '100'. In sharp contrast, Formosan languages have many unrelated forms for this meaning. One of these, *matala gasut* '100' (*matala* = 'one'), recorded by the Japanese scholar Naoyoshi Ogawa in 1909 from a Hoanya speaker born in 1834, appears to be cognate with forms outside Taiwan, and points to PAN *RaCus '100' (Tsuchida 1982:40). However, this is the only known Formosan word that might be related to PMP *Ratus, and it is upon this single lexical item from a language that is now extinct that a PAN word for '100' must be based. Words for '1000' are found in a number of languages, but these often appear to be borrowed. In southern Taiwan Tanan Rukai *koḍólo,* Paiwan *kuzuɭʸ,* and Puyuma *kuḍul* appear to reflect PAN *kuduN '1000'. However, since these languages are contiguous, and share many known loanwords this comparison must be treated with caution. Similarly, a number of languages in insular Southeast Asia share forms similar to Malay *ribu* '1,000', but these often appear to be Malay loans, as with Ilokano *sa-ŋa-ríbo,* Tagalog *sa-n-líbo,* Muna *riwu,* or Tetun *rihun* '1,000'. Nonetheless, a few languages have a related word that appears to be native, as with Maranao (southern Philippines) *sa-ŋ-gibo,* Kadazan (Sabah) *iɓu,* or Long Terawan Berawan (northern Sarawak) *gikəw* '1,000'. There is thus some evidence for a protoform *Ribu '1,000', although it cannot be assigned to PAN, or even with any certainty to PMP.

In insular Southeast Asia words for '10,000' and higher multiples of 10 are often borrowed from non-AN sources, as with Thao, Saisiyat *ban,* Atayal *maŋ* '10,000' (from Taiwanese), Malay *laksa* '10,000', or *juta* 1,000,000 (from Sanskrit). Higher numerals borrowed by Malay from a non-AN source were then sometimes transmitted to other languages, as in Sasak *laksa,* Makasarese *lassa,* Muna *lasa,* Maranao *laksaɁ,* Cebuano *láksaɁ,* Tagalog *laksáɁ,* Ilokano *sa-ŋa-laksá* '10,000', or Makasarese *juta,* Muna *dhuta* '1,000,000'. Where terms for 10,000 or higher are native they appear to be innovations in individual languages, as with Tagalog *sa-ŋ-áŋaw* 'one million', *sa-n-libo-ŋ-áŋaw* 'one billion', *sa-ŋ-áŋaw na áŋaw* 'one trillion', or Ilokano *riwríw* 'one million,' *sa-ŋa-púlo a riwríw* 'one billion', or *sa-ŋa- riwríw a riwríw* 'one trillion'. These words generally have no other meaning, and the age of the innovation is indeterminate. Since no numeral base above 1,000 can be reconstructed even with a shallow time depth, it seems safe to infer that numeral bases above 1,000 are late innovations. In Southeast Asia numeral bases above 1,000 tend to be confined to languages that have formed part of extensive trade networks prior to Western contact.

In most AN languages the numerals 11-19 are formed by addition (10+1, etc.), and those from 20-90 by multiplication (2x10, etc). Multiples of ten sometimes include a ligature reflecting PMP *a or *ŋa, as in the following data from Ilokano (northern Philippines), Kelabit (northern Sarawak), and Tondano (northern Sulawesi):

Table 5.5 Formation of higher numerals in Ilokano, Kelabit, and Tondano

	Ilokano	**Kelabit**	**Tondano**
1	maysá	ədʰəh	əsa
2	duá	duəh	rua
3	talló	təluh	təlu
4	uppát	əpat	əpat
5	limá	liməh	lima
6	inném	ənəm	ənəm
7	pitó	tuduʔ	pitu
8	waló	(w)aluh	walu/ualu
9	siám	iwaʔ	siow
10	sa-ŋa-púlo	puluʔ	ma-puluʔ
11	sa-ŋa-púlo ket maysá	puluʔ ədʰəh	ma-puluʔ wo-osa
12	sa-ŋa-púlo ket duá	puluʔ duəh	ma-puluʔ wo-rua
20	duapúlo	duəh ŋəh puluʔ	rua-ŋa-puluʔ
30	tallopúlo	təluh ŋəh puluʔ	təlu-ŋa-puluʔ
40	uppát a púlo	əpat ŋəh puluʔ	əpat-ŋa-puluʔ
50	limapúlo	liməh ŋəh puluʔ	lima-ŋa-puluʔ
60	inném a púlo	ənəm ŋəh puluʔ	ənəm-ŋa-puluʔ
70	pitopúlo	tuduʔ ŋəh puluʔ	pitu-ŋa-puluʔ
80	walopúlo	(w)aluh ŋəh puluʔ	walu-ŋa-puluʔ
90	siam a púlo	iwaʔ ŋəh puluʔ	siow-ŋa-puluʔ
100	sa-ŋa-gasút	ədhəh ŋəh ratu	ma-atus
200	dua gasút	duəh ŋəh ratu	rua-ŋa-atus
1,000	sa-ŋa-ríbo	ədhəh ŋəh ribuh	ma-riwu

Structurally similar systems for forming higher numerals are found in many eastern Indonesian languages, but with certain syntactic differences. In the systems seen above the numerals from 11-19 differ from 20-90 in two respects. First, the smaller set is based on addition, and the larger set on multiplication. Second, the additive set places '10' first (10+1, etc.), whereas the multiplicative set places '10' second (2x10, etc.). From roughly central Flores eastward through the Lesser Sunda chain, and the southwest Moluccas as far as western Melanesia both of these numeral sets place '10' first, and in some languages the ligature and base have fused into a single morpheme, as with Lio *sa-mbulu* '10', *mbulu rua* '20', *mbulu təlu* '30', etc., Kédang *puru-n sue* '20', *puru-n təlu* '30', etc., or Manam *ʔulemʷa* '10', *ʔulemʷa rua* '20', *ʔulemʷa toli* '30', etc. Since morpheme order is one way that the words for '12' and '20', or '13' and '30' are distinguished in many AN languages, some additional features of morphology are needed to prevent homophony in languages that use the same order for additive and multiplicative values of '10'. In Lio this is achieved by using *sa-mbulu* 'one group of ten' for 11-19, but *mbulu* 'group of ten' for 20-90: *sa-mbulu əsa rua* '12' : *mbulu rua* '20', *sa-mbulu əsa təlu* '13' : *mbulu təlu* '30', etc. In Kédang the same effect is achieved by using a base *pulaʔ* that is otherwise unattested for additive values of '10', but *puru-* for multiplicative values: *pulaʔ sue* '12' : *puru-n sue*

'20', *pulaʔ təlu* '13' : *puru-n təlu* '30', etc. In Manam it is accomplished by use of the conjunction *be* 'and' with additive values of '10': *ʔulemʷa be rua* '12' : *ʔulemʷa rua* '20', *ʔulemʷa be toli* '13' : *ʔulemʷa toli* '30', etc.

The data in Table 5.5 is representative of many languages in western Indonesia and the Philippines, and preserves certain features of numerals that must be attributed to Proto Malayo Polynesian. The PMP numerals '10', '20' and '30' can be reconstructed as *esa ŋa puluq, *duha ŋa puluq and *telu ŋa puluq. However, it is not clear that '40' was *epat ŋa puluq. Ilokano uses the ligature *ŋa* in '10', '100', and '1,000', but in multiples of 2-9 there are two patterns: *a* after bases that end with a consonant, and zero after bases that end with a vowel. This partly corresponds to a pattern found in the Kaili-Pamona and South Sulawesi languages, Wolio, Old and modern Javanese, Balinese and Sasak, where a reflex of *pat-aŋ is found in '40', '400', and '4,000'. These differences in the form of the ligature correlate with the canonical shape of numeral bases: bases that end with a vowel take the ligature *ŋa* in multiples of '10', while bases that end with a consonant take something else. One possibility is that a strategy has been employed to avoid a cumbersome *-Cŋ-* cluster; in Ilokano this has been accomplished by reduction of *ŋa* to the vocalic nucleus, while in languages with *-aŋ* it has been accomplished by metathesis. Some observations suggest another explanation. In languages that have *pat-aŋ* as a combining form for 'four' in multiples of ten the ligature is a simple velar nasal after vowel-final bases, as in Javanese *təlu-ŋ puluh*, or Tae' *tallu-ŋ pulo* '30'. This suggests that in Javanese *papat* '4' : *pat-a-ŋ puluh* '40', or Tae' *aʔpaʔ* '4' : *pat-a-ŋ pulo* '40' the vowel preceding the velar nasal is not part of the ligature, but an epenthetic segment motivated by rhythmic requirements in consecutive serial counting of multiples of '10'. Given the full form of the ligature *ŋa* after consonant-final bases in languages such as Kelabit or Tondano the history of the form *pat-a-ŋ becomes problematic: was epenthesis present in the PMP word for '40', or is it a product of multiple convergent innovations?

The Javanese numeral system (Robson 2002) contains idiosyncracies in connection with multiples of '10' that deserve special mention. We need only consider the Ngoko register here, since the Krama register is largely derived from it. Although the numerals 1-10 form a transparent decimal system [(1) *siji*, 2) *loro*, 3) *təlu*, 4) *papat*, 5) *lima*, 6) *nəm*, 7) *pitu*, 8) *wɔlu*, 9) *saŋa*, 10) *sa-puluh*], some features of numeration suggest a vigesimal system, but one that is overlaid with other peculiarities. The numerals 20-90 in Javanese are: *ro-ŋ puluh* '20', *təlu-ŋ puluh* '30', *pata-ŋ puluh* '40', *sɛkət* '50', *sawidak* '60', *pitu-ŋ puluh* '70', *wɔlu-ŋ puluh* '80', *saŋa-ŋ puluh* '90'. Nothing in these numerals suggests a system based on twenty; rather, what stands out is the forms for '50' and '60', which are not the expected multiples of '10', but unanalyzable morphemes that appear nowhere else in the numeral systems. However, a separate morpheme for 'twenty' is seen in the words for 21-29: *sa-likur* '21', *ro-likur* '22', *təlu-likur* '23', *pat-likur* '24', *salawe* '25', *nəm-likur* '26', *pitu-likur* '27', *wɔlu-likur* '28', *saŋa-likur* '29'. Javanese thus contains unanalyzable morphemes meaning '20' '25', '50' and '60'. Of these idiosyncratic forms the base for '20' stands out for two reasons: 1) it does not occur as the numeral '20' itself (which is 2x10), and 2) additive values for 21-29 (except 25) are formed by *preposing* the primary digit (2+20 = '22'), while for all other decades additive values are formed by *postposing* the primary digit (30+2 = '32', 40+3 = '43', 50+4 = '54', etc.).

The Tondano numeral prefix *ma-* raises another point. Sneddon (1975:108) states that *ma-* replaces the sequence *əsa ŋa* before *-puluʔ* 'ten', *-atus* 'hundred', and *riwu* 'thousand'. Similar forms of '10' and '100' are found in other languages reaching from the northern Philippines to southern Borneo, as in Atta *ma-pulu* '10', *ma-gatuʔ* '100', Botolan Sambal

ma-poʔ '10', *ma-gato* '100', Koronadal Bilaan *m-latuh* '100', Timugon Murut *m-atus* '100', Lun Dayeh *mə-ratu* '100', Narum *mə-rataw* '100', and in Bornean place names such as the Sarawak Kenyah Lio Matoh ('Hundred channels'), and Pegunungan Meratus ('Hundred mountains') in southeast Kalimantan. Reflexes of *ma-Ratus appear to be much more common than reflexes of *ma-puluq, and the latter may be secondary formations based on the word for '100'. Since this numeral prefix is homophonous with PAN *ma-'stative', *ma-Ratus may have been used attributively in the sense of 'myriad'.

Until recently evidence for the PAN numerals 20-90 was overlooked, but Zeitoun, Teng, and Ferrell (2010) have convincingly shown the need to posit *ma-puSa-N '20', *ma-telu-N '30', *ma-Sepat-eN '40', *ma-lima-N '50', *ma-enem-eN '60', *ma-pitu-N '70', *ma-walu-N '80', and *ma-Siwa-N '90'. Although the first of these was recognized in Blust (2009a), it was misanalyzed as *ma-puSaN, which led to the erroneous conclusion that there was a base *puSaN that "plays no part in the construction of higher numerals" (Blust 2009a:278). The basis for the irregular allomorphy in *duSa 'two', but *ma-puSa-N '20', which partially contributed to this misanalysis, remains unexplained.

In addition, Mead (2001) has shown that Balinese, Sasak, and a number of languages in Sulawesi share a series of temporal adverbs that reflect *i-puan 'two days ago/two days hence', *i-telu-n 'three days ago/three days hence', and *i-pat-en 'four days ago/four days hence'. It is possible that these derive from PAN *puSa-N, *telu-N and *Sepat-eN, as the first of these affixed bases occurs not only in *ma-puSa-N '20', but also in some attested languages in the meaning 'two', as in Thao *lhim-pushaz-in* 'be sifted a second time (as rice in a winnowing tray)', *makim-pushaz* 'second from the bottom, as stair steps, storeys of a building, etc.; have two levels), *mu-pushaz* 'twice', or *pushaz-an* 'two, of paces'.

Finally, Harrison and Jackson (1984:61) have shown that in some Micronesian languages "there exist monomorphemic numbers for multiples of ten" which reach 10^9 in the Ponapeic languages (Pohnpeian, Mokilese, Pingilapese, Ngatikese), and 10^{10} in the Polynesian Outlier of Nukuoro. They counter the suggestion of some researchers that higher numerals in Oceanic languages are used only "poetically as nouns indicative of great numbers" with tangible evidence of linguistic forms, and claim that speakers of Micronesian languages will recite them in order much as English speakers will recite the numbers 1-10 "making it clear through the counting procedure itself that each member of the series is agreed upon as a ten-multiple of the immediately lower member." Table 5.6 presents a selection of the evidence they give for this claim:

Table 5.6 Numerals 10^1 to 10^9 in three Nuclear Micronesian languages

Gilbertese	Pohnpeian	Woleaian	
te-bʷiina	eisek	seig	10^1
te-bubua	epʷiki	sebiugiuw	10^2
te-ŋaa	kid	saŋeras	10^3
te-rebu	nen	sen	10^4
te-kuri	lopʷ	selob	10^5
te-ea	rar	sepiy	10^6
te-tano	dep	seŋit	10^7
te-toki	sapʷ	saŋerai	10^8
	lik		10^9

Harrison and Jackson suggest that ten-power bases in excess of 1,000, or possibly even 100, were innovated in the separate histories of these languages. This is supported in some

cases by evidence that the morphemes used in larger numerals are drawn from nouns meaning 'sand', 'end/completion', 'outside' and the like. Because the languages inherited a system of numeration which included at least 10^1 and 10^2, the potential for expansion to higher powers of ten was present from an early time. Since the same precondition existed in all AN languages, however, it is unclear why such an elaboration of a ten-power system would take place in one small set of languages, but not in others. Harrison and Jackson favor the hypothesis that enumeration by countable bases (numeral classifiers), a feature that is prominent in attested Micronesian languages for counting objects in the real world, provided the springboard for innovating higher multiples of ten.

In addition, a number of Polynesian languages have monomorphemic ten-power bases for higher numerals. In Hawaiian these are based on 4×10^n, where n must be at least 2: *lau* '400', *mano* '4,000', *kini* '40,000', *lehu* '400,000'. However, cognates in other languages have different numerical values, are verbs of surpassing, or are nouns that suggest vast quantities, as Maori *rau* '100', *mano* '1,000', *tini* 'host, myriad', *rehu* 'haze, mist, spray, fine dust', Rarotongan *rau* '200', *mano* '1,000', *tini* 'host, myriad', *reu* 'dust, powder, ashes', Samoan *se-lau* '100', *mano* '100,000 (or any very large number)', *tini* 'pass the finishing post, reach the goal', *lefulefu* 'ash', Rennellese *gau* '100'; *haka-gau* 'to count, as mats, tapa, etc.', *mano* '100 (as of piles consisting each of four banana bunches)', *tini* '10 (piles of panna, bags)', *gehu* 'dust', Tongan *te-au* '100', *mano* '10,000; myriad'.

5.1.6 Derivative numerals

The term 'derivative numeral' will be used to cover all non-cardinal numerals formed by affixation, including reduplication. The most important of these are: 1) numerals for counting humans, 2) ordinal numerals, 3) distributive numerals, and 4) frequentative or multiplicative numerals.

5.1.6.1 *Numerals with human referents*

Apart from the cardinal numerals (Set A), some languages have a second set (Set B) that is used in counting [+human], or less commonly [+living] referents. Set B numerals are formed by Ca- reduplication, a process of word formation that copies the base-initial consonant followed by the fixed vowel *a*. Although both sets can be reconstructed for PAN/PMP, very few languages make active use of this distinction, and most of these are in Taiwan. Table 5.7 illustrates the contrast of Set A and B numerals in two Formosan languages, in the Iraralay dialect of Yami, a Philippine language spoken on Botel Tobago Island off the southeast coast of Taiwan, and in Chamorro of western Micronesia. In addition to Sets A and B, some languages also distinguish the interrogative of quantity in the same way, as with Thao *piza* : *pa-piza*, 'how much/how many?', where the longer term is [+human], or Chamorro *fiʔa* : *fa-fiʔa* 'how much/how many', where the longer term is [+living].

Table 5.7 Contrast of Set A and Set B numerals in four Austronesian languages

	Thao		Puyuma	
	A	B	A	B
1.	tata	(same)	sa	sa-sa
2.	tusha	ta-tusha	Tuwa	za-zuwa
3.	turu	ta-turu	Təri	ta-təru
4.	pat	shpa-shpat	pat	a-apat
5.	rima	ra-rima	rima	la-luwaT
6.	makalhturuturu	(same)	ʔənəm	a-ʔnəm
7.	pitu	pa-pitu	pitu	pa-pitu
8.	makalhshpashpat	(same)	waru	wa-waru
9.	tanacu	(same)	iwa	a-iwa
10.	maqcin	(same)	puruH	pa-puruH

	Yami (Iraralay)		Chamorro	
1.	asa	(same)	haca	maysa
2.	dowa	ra-rowa	hug^{w}a	(same)
3.	atlo	ta-tlo	tulu	ta-to
4.	apat	pa-pat	fadfad	(same)
5.	lima	la-lima	lima	la-lima
6.	anəm	na-nəm	gunum	g^{w}a-gunum
7.	pito	pa-pito	fitu	fa-fitu
8.	wawo	wa-wawo	g^{w}alu	g^{w}a-g^{w}alu
9.	siyam	sa-siyam	sig^{w}a	sa-sig^{w}a
10.	poo/aŋaŋalenan	(same)	maʔnud	maʔonod

Even in languages that still make active use of them, the Set B numerals seem to be on the decline. Contrasts such as *tusha wa qali* (two **lig** day) 'two days' and *ta-tusha wa azazak* (**red**-two **lig** child) 'two children' were elicited from elderly speakers of Thao, but these speakers were hesitant in using Ca- forms, and out of context even denied that they exist. As seen in Table 5.7, Set B numerals often are defective. Although Puyuma has Set B forms for 1-10, they do not always match the shapes of their Set A counterparts. In all other languages for which Set B forms are known the set is incomplete. Puyuma is the only language known to use Ca- reduplication with the number '10', and it is likely that this is innovative. In the first three languages Set B numerals refer to human referents. In Chamorro the situation is somewhat more complex. Costenoble (1940:259ff), whose knowledge of the language was acquired between 1905 and 1913, reported that in daily use the native Chamorro numerals had already been replaced decades earlier by their Spanish equivalents. In 1913 they were remembered passively by only a few elders on Guam and Saipan, but were still used actively by the oldest generation of speakers on the island of Rota. The material for Chamorro was thus acquired at a time when this feature was moribund. Costenoble distinguishes 1) the basic numeral set (*Grundzahlen*) from several other sets, including 2) numerals for living things (Set B), 3) numerals for lifeless things, and 4) numerals for elongated objects. Chamorro Set B numerals thus refer to human referents, but unlike the Set B numerals in the other languages, they include other living things within their scope of reference.

In a number of languages, particularly in the Philippines, Sets A and B have been integrated into a single historically syncretic system of general numerical reference. That these types of systems must be the product of multiple independent historical changes is shown by differences in the individual numerals that are selected from each set, as seen with the following data from Tagalog, Ata, and Tigwa Manobo:

Table 5.8 The historically composite numerals of Tagalog, Ata, and Tigwa Manobo

	Tagalog	**Set**	**Ata**	**Set**	**Tigwa Manobo**	**Set**
1	isá	(<A)	isa	(<A)	sabəka	(<A)
2	dalawá	(<B)	dadawa	(<B)	dadua	(<B)
3	tatló	(<B)	tatolu	(<B)	tatəlu	(<B)
4	ápat	(<B)	hopʔat	(<B)	həpʔat	(<B)
5	limá	(<A)	lalima	(<B)	lalima	(<B)
6	ánim	(<B)	honʔom	(<B)	hənʔəm	(<B)
7	pitó	(<A)	papitu	(<B)	pitu	(<A)
8	waló	(<A)	wawalu	(<B)	walu	(<A)
9	siyám	(<A)	sasiam	(<B)	siam	(<A)
10	sa-m-púʔ	(<A)	sa-m-puluʔ	(<A)	sa-puluʔ	(<A)

Such systems show elements of Ca- reduplication that are more visible in some languages (Ata, Tigwa Manobo) than in others (Tagalog), but in all cases they suggest a derivational process that is fossilised. In other languages a distinction between numeral sets that refer to humans and those that do not is maintained, but this is expressed through innovative morphology, as in Paiwan of southeast Taiwan, which requires a numeral set marked by *ma-* or *manə-* in counting people (Tang 2004).

5.1.6.2 Other derivative numerals

Some AN languages have extremely elaborate numeral systems, with many types of derived numerals. In an exceptionally thorough study Andersen (1999) devotes 72 pages to describing the numeral system of a single language (Moronene) in southeast Sulawesi. We will have to content ourselves with a few passing remarks.

In PAN ordinal numerals were derived by prefixation with *Sika-. This derivational process is preserved in many daughter languages throughout the family, as in Paiwan *tjəlu* 'three' : *sika-tjəlu* 'third', Tagalog *ápat* 'four' : *ika-ápat* 'fourth', Malay *lima* 'five' : *kə-lima* 'fifth', Bolaang Mongondow *opat* 'four' : *ko-opat* 'fourth', Gilbertese *tenua* 'three' : *te ka-tenua* 'the third', Pohnpeian *riau* 'two' : *ka-riau* 'second', or Fijian *vā* 'four' : *ka-vā* 'fourth'. In languages generally the ordinal form for 'one' often deviates morphologically from other ordinal numerals, and this is also true of AN languages, as with Ilokano *maysá* 'one' : *um-uná* 'first', next to *duá* 'two' : *maika-duá* 'second',[42] Pangasinan *isá* 'one' : *uná/primera/primero* 'first', next to *duá* 'two' : *koma-duá/segundo* 'second', Bolaang Mongondow *intaʔ* 'one' : *intaʔ duŋkul* 'first', next to *doyowa* 'two' : *ko-doyowa* 'second', or Malay *satu* 'one' : *pərtama* 'first' (from Sanskrit), next to *dua* 'two' : *kə-dua* 'second'. There are, however, some exceptions, as with Maranao *isa* 'one' : *ika-isa* 'first', Fijian *dua* 'one' : *i ka-dua* 'first', or Pohnpeian *e:u* 'one' : *ke-ieu* 'first'. As in other language

---

[42] Rubino (2000:lviii) supposes that Ilokano *uná* is borrowed from Spanish, but the form regularly reflects PAN *SuNa 'first, before, anterior in time'.

families, non-numerical words based on 'one' may refer both to unity and to isolation. The word for 'alone' in a number of AN languages is a reflex of *ma-isa, composed of *isa 'one' plus what appears to be the prefix of stative verbs.

Distributive numerals typically are formed by full reduplication, as in Paiwan *ma-ita-ita* 'one by one, one after the other', Tagalog *ápat-ápat* 'four at a time', Javanese *pat-pat* 'by fours', Bolaang Mongondow *opat-opat* 'four-by-four, four at a time', Makasarese *appaʔ-appaʔ* 'four at a time', Rotinese *esa-esa* 'one by one, each in its turn', Yamdena *fate-fate* 'four by four, four at a time', or Manam *wati-wati* 'four each, four at a time'. Forms such as Kambera *pa-patu* 'by fours, in groups of four' presumably are reductions of earlier full reduplications. Agreements between distantly related languages, as with Bolaang Mongondow *opat-opat*, and Fijian *vā-vā* 'all four' suggest that full reduplication may have been used to form both distributive and collective numerals. In addition, a fully reduplicated form of 'two' in many languages means 'uncertain, of two minds'.

Frequentative or multiplicative numerals are usually formed with a reflex of the causative prefix *paka-: Chamorro *faha-unum* 'six times', Arosi *haʔa-hai* 'four times', Fijian *vaka-ono* 'six times', Rennellese *haka-ono* 'do six times'. As will be seen, PAN marked non-stative causatives with *pa- and stative causatives with *pa-ka- (> *paka- in most daughter languages). It is noteworthy that where frequentative numerals take a reflex of the causative prefix they are invariably marked like stative verbs. Reflexes of *pa- + numeral are most commonly found with 'two', and mean 'divide into two'.

Finally, in western Indonesia fractions are often expressed with a reflex of *paR-+ numeral, as with Malay *sə-pər-əmpat*, Toba Batak *sa-par-opat*, Makasarese *parapaʔ* 'one fourth', or Wolio *parapa* 'fourth part, quarter'. Given the long history of borrowing from Malay over this area this distribution may well be a product of diffusion, but reflexes of *maR-numeral are more widely distributed in the meaning 'become four, divide into four', and since *maR- probably derives from *p<um>aR- both forms appear to have a considerable history in AN.

5.2   Numeral classifiers

Numeral classifiers (which go by an exasperating variety of names in the literature) are found in many of the languages of East and Southeast Asia, and AN languages are no exception. Languages like English use classifiers under certain conditions, in particular for mass nouns ('a handful of sand', 'two cups of rice', 'ten gallons of gasoline'), for counting some discrete count nouns ('fifty head of cattle'), and for collective count nouns, usually if these form an uncountable mass, in which case they normally occur only as an indefinite singular ('a school of fish', 'a herd of buffalo' or 'a flock of birds'). Yet English is not commonly regarded as a language that has a full-blown system of numeral classifiers. In reality, as noted by Tang (2004), there seems to be no sharp line between 'classifier' and non-classifier languages: some languages have no trace of numeral classifiers, others have elaborate systems of auxiliaries used in conjunction with numerals, and many others lie between these extremes. AN languages are distributed all along this continuum, but many tend to have fairly elaborate systems of classifiers used in counting. Classifiers used with count nouns are marginal in the AN languages of Taiwan and the Philippines.[43]   They are

---

[43]   Among Formosan languages Thao offers one known example: *shakish tata wa bagkir* (camphor laurel one **lig** trunk) 'one camphor laurel' (Blust 2003a:206). Tang (2004) calls Paiwan a classifier language, and suggests that some other Formosan languages may be as well. However, all her examples of numeral classifiers (as opposed to [+human] numerals), involve measure words or mass nouns, and this

poorly reported in the languages of Sabah, and only begin to be well-documented further south in Borneo, reaching a high degree of elaboration in traditional Malay and in some of the languages of Sumatra and Java. Further east they are reported sporadically in eastern Indonesia, and then become highly elaborated in parts of the Pacific, where they often take on a different formal aspect and somewhat different semantics from the systems of western Indonesia.

Numeral classifiers in island Southeast Asia tend to be common nouns that serve secondarily in counting. In fieldnotes for Bintulu of coastal Sarawak, collected by the writer in 1971, eight classifiers were recorded (class of referents appears in parentheses):

1.   *apəh* (animals, including at least fish, birds, pigs and cats)

2.   *lambar* (sheets of paper)

3.   *əmbaŋ* (flattish sheet-like objects, including leaves and sheets of paper)'

4.   *puʔun* (trees)

5.   *təŋən* (trees)

6.   *tukuŋ* (chunk-like objects, including at least stones and pieces of meat)

7.   *uŋ* (fruits)

8.   *usa* (people)

Examples of usage are: *lima apəh bakas* (five **clas** pig) 'five pigs', *ləw lambar/əmbaŋ kərtəs* (three **clas** paper) 'three sheets of paper', *nəm əmbaŋ raʔun* (six **clas** leaf) 'six leaves', *ba əmbaŋ raʔun sigup* (two **clas** leaf tobacco) 'two nipa palm leaves (for rolling cigarettes)', *ba puʔun/təŋən kazəw* (two **clas** tree) 'two trees', *pat tukuŋ batəw* (four **clas** stone) 'four stones', *ba tukuŋ dagiŋ* (two **clas** meat) 'two pieces of meat', *jiʔəŋ uŋ pakən* (one **clas** white durian) 'one white durian', *pat usa anak* (four **clas** child) 'four children', *lew usa redu* (three **clas** woman) 'three women'. Of these numeral classifiers at least five occur as common nouns in Bintulu: *puʔun* 'base of a tree', *təŋən* 'tree, trunk of a tree', *tukuŋ* 'clod, lump of earth', *uŋ* 'fruit', and *usa* 'body'. Examples such as *tujuʔ uŋ* 'seven fruits' reflect a property exhibited by many systems of numeral classifiers in western Indonesia: a classifer cannot be used with the noun from which it is derived. One other noun, *madiʔ bələd* (eight seed) 'eight seeds' was recorded with zero classifier. In this case it is possible that the noun is a classifier for small seed-like objects which happened to be recorded only with its source noun. Finally, as seen above, some nouns allow more than one classifier, as with *lambar/əmbaŋ* or *puʔun/təŋən*. In the first set of duplicate terms *lambar* is a borrowing of Malay *ləmbar* 'classifier for sheetlike objects', and perhaps for this reason was recorded with the Malay loanword *kərtəs* 'paper', but not with the native term *raʔun* 'leaf'. Both members of the second set of duplicate terms appear to be native, and although a semantic distinction between the independent nouns is clear no difference is apparent in their use as numeral classifiers. In Bintulu only one order of numerative expressions with classifiers was recorded, namely **num-clas noun**, but in Mukah Melanau, spoken further south on the coast of Sarawak, such expressions may occur either in this order, or as **noun num-clas**: *lima usah kayəw* (five **clas** tree) 'five trees', but *kayəw dua awaʔ* (stick two **clas**) 'two sticks'.

___

is not true of numeral classifier systems in Indonesia or the Pacific. Lopez (1967) claimed that numeral classifiers are present in Philippine languages, but Gonzalez (1973a) argued that these are 'quantifiers', not numeral classifiers in the sense this term is normally understood.

Although Standard Malay and Standard Indonesian have greatly reduced the use of numeral classifiers in the spoken language since Malay/Indonesian became the language of politically independent states, older spoken and written forms of Malay are very rich in these elements. Maxwell (1907:70ff) lists the following classifiers that were common in colloquial peninsular Malay during the second half of the nineteenth century. The literal meaning of the source noun is in single quotes and the class of referents in parentheses; the original orthography has been retained, apart from the substitution of *ŋ* for the velar nasal, *ə* for the schwa, and *ʔ* for the glottal stop: 1) *oraŋ* 'person' (mankind), 2) *ekor* 'tail' (animals), 3) *buah* 'fruit' (fruit, houses, ships, places, etc.), 4) *biji* 'seed' (small objects, mostly round)', 5) *halei/lei* 'leaf, blade' (tenuous objects such as hair, feathers, leaves, wearing apparel, etc.), 6) *bataŋ* 'stem' (long objects), 7) *puchuk* 'young shoot' (letters, muskets, cannon, elephant tusks, etc.), 8) *kəpiŋ* 'piece, slice' (pieces of wood, metal, etc.)', 9) *taŋga* 'ladder' (houses), 10) *pintu* 'door (houses), 11) *lapis* 'fold' (clothing), 12) *rawan* (nets and cordage), 13) *bilah* 'thin strip or lath'(cutting weapons), 14) *buntoh* (rings, fish-hooks, etc.), 15) *bidaŋ* 'spacious' (things spread out), 16) *butir* 'grain, particle' (fruit, seeds, and other small round objects). Maxwell (1907:70) calls these "*numeral affixes*, some one or other of which is always used as a co-efficient to the numeral," and he illustrates them with examples such as *China tiga oraŋ, Malayu saʔoraŋ* (Chinese three **clas**, Malay one-**clas**) 'three Chinese and a Malay', *kuda bəlaŋ dua ekor* (horse piebald two **clas**) 'two piebald horses', *rumah dua taŋga* (house two **clas**) 'two houses', and *rumah batu ənam pintu* 'six brick houses'. The examples that he gives show only the order **noun num-clas**, and the apparent vacillation between *taŋga* and *pintu* as classifiers for houses almost certainly correlated with different types of dwellings, the first with traditional country houses raised on piles (hence entered by ladders), and the second with urban dwellings built directly upon the ground.

Winstedt (1927:129ff), writing some thirty years later, but still during the British colonial period, gives a longer, alphabetically ordered list of 'numeral coefficients' that were used in colloquial peninsular Malay early in the twentieth century: 1) *bataŋ* (trees, poles, spears, teeth), 2) *bəntok* (rings), 3) *bidaŋ* (widths of cloth, matting, sails, ricefields), 4) *biji* (eyes, eggs, small stones, coconuts, caskets, chairs, fruits, fingers, bullets, tombstones), 5) *bilah* (daggers, knives, needles), 6) *buah* (fruits, countries, islands, lakes, ships, houses), 7) *butir* (coconuts, grain, jewels, cannons), 8) *charek* (scraps of paper and linen), 9) *ekor* (animals, birds, insects, and contemptuously of men), 10) *həlai/ʔlai* (hair, leaves, cloth, paper), 11) *kaki* (insects, umbrellas, long-stemmed flowers), 12) *kampoh* (pieces of fish, roe), 13) *kayu* (cloth), 14) *kəpiŋ* (blocks of timber, metal, and bunches of bread, meat, cake), 15) *kuntum* (flowers), 16) *laboh* (hanging objects: curtains, necklace, etc.), 17) *oraŋ* (persons), 18) *patah* (words), 19) *pəraŋgu* (sets of betel-boxes, buttons), 20) *pintu* (houses), 21) *taŋga* (houses), 22) *potoŋ* (slices of meat and bread), 23) *puchok* (guns, letters, needles), 24) *rawan* (nets), 25) *utas* (nets), 26) *taŋkai* (flowers), 27) *urat* (thread). Winstedt makes it clear that his list gives only 'the commoner' members of this class of words, and indeed one of the classifiers cited by Maxwell (*buntoh*) is omitted. Furthermore, he notes (1927:130) that "The numeral always stands immediately before its coefficient," and that "before a coefficient *sa-* is used instead of *suatu*" (the independent form of 'one'). Unlike Maxwell, whose examples reflect only the order **noun num-clas**, Winstedt gives a more intricate set of word-order rules for the use of Malay numeral classifiers: 1) *sa-***clas** precedes a noun, 2) other numerals + classifier follow the noun, 3) if emphasis falls on the numeral the order just stated is reversed. From this we find divergent orders in e.g. *sa-biji piŋgan* 'a plate', but *piŋgan dua biji* 'two plates', or *sa-oraŋ anak-ku*

'one (or a) son of mine' vs. *anak-ku tiga oraŋ* 'my three sons, or my sons, they are three' vs. *anak-ku sa-oraŋ* 'my one and only son, or I have only one son'. Although he does not list *piŋgan* 'plate' as a numeral classifer, Winstedt contrasts *sa-biji piŋgan* 'a plate' with *sa-piŋgan* 'a plateful', implying that the domain of this word class in languages such as Malay is essentially open-ended. Perhaps for this reason the exact number of numeral classifiers in Malay has never been stated explicitly. Adam and Butler (1948:38), who call them 'numeratives' list 38 numeral classifiers for Malay. On the other hand, Macdonald and Darjowidjojo (1967:132ff), describing Standard Indonesian (Bahasa Indonesia) as used in the 1960s list twenty classifiers (called 'counter nouns'), and state that only *oraŋ*, *ekor* and *buah* are commonly used, but that "Most speakers avoid them on ordinary occasions, except after the number one, which frequently occurs in its proclitic alternant *se-*. Thus, 'one student' will probably occur as *seoraŋ mahasiswa*, but 'two students' is much more likely to occur as *dua mahasiswa* rather than *dua oraŋ mahasiswa*."

The picture that emerges from these descriptions is that numeral classifiers were richly developed in traditional Malay rural society, but that urbanisation and the creation of national languages has had a simplifying effect on their use. Since most other regional languages have been less dramatically subjected to modernising influences, and have not become the vehicle of national life, they presumably would be more conservative in retaining what can be characterised as non-essential, elaborative features of the grammar. Surprisingly, however, few other languages of western Indonesia appear to have a system of numeral classifiers that rivals that of Malay.

No reliable mapping of the distribution of AN numeral classifiers is available. Coolsma (1985:167ff) lists 37 forms for Sundanese. Some of these have Malay cognates but most do not, as *papan sə-bebek* (board one-**clas**) 'a board', *wəlit sa-jalon* (roofing grass one-**clas**) 'a roof shingle of *Imperata cylindrica*', or *bənaŋ sa-kukular* (thread one-**clas**) 'a length of thread'. Woollams (1996:131) lists 21 'measure nouns' for Karo Batak ('a pinch of something', 'a handful', 'a mouthful', etc.), and distinguishes these from 'noun classifiers' which he compares with the numeral classifiers of Malay. However, the numeral classifier systems of these three languages may be unusually elaborate. Comparison of the eight known classifiers of Bintulu with the nearly forty for Malay and Sundanese is problematic, since the first language is not well-described, but even where full grammars are available these often indicate a less richly developed system. In a brief account of Toba Batak numeral classifiers (called 'auxiliary numerals'), van der Tuuk (1971:214) gives twelve forms, some of which occur with the prefix *ka-*. As in other languages, many of these transparently derive from independent nouns, as with *napuran tolu-k ka-baba* (betel three-**clas**) 'three betel quids' (*baba* = 'mouth'). Language-specific semantic categories represented by numeral classifiers such as Toba Batak *ka-baba* 'mouthful' suggest that numeral classification is a fluid part of the grammar, which may be sensitive to individual improvisation. As already noted, numeral classifiers thus appear to form a more-or-less open-ended word-class, but the size of this class varies widely across languages. Donohue (1999:109ff) gives just twelve numeral classifiers for Tukang Besi of southeast Sulawesi, and notes that one of these (*ʔasa* 'general counter') is taking over for all referents among younger speakers of the language. Similarly, for Timugon Murut of Sabah, Prentice (1971:175ff), gives only eight 'metrical nouns': 1) *ŋa-ulun* 'persons' (human beings and spirits), 2) *ŋa-inan* 'bodies' (living things and large fruits), 3) *ŋa-n-taun* 'sticks' (long, cylindrical things such as blowpipes, houseposts, snakes, eels), 4) *ŋa-unor* 'kernels' (small—especially round—things, such as fish, insects, small fruits, eggs, nails, grains of rice), 5) *ŋa-m-pilaq* 'breadths' (thin flat things, as doors, letters, cloths, mats, winnowing

trays, shallow gongs, locally made hats), 6) *ŋa-uat* 'sinews' (long, thin, thread-like things, as hairs, blades of grass, thin worms), 7) *ŋa-n-dapak* 'pieces?' (deep, open vessels such as cooking pots, bowls, deep gongs), 8) *ŋa-m-puun* 'trunks' (trees and earthenware jars). As can be seen from the recurrent partial in these forms, the ligature *ŋa* is analyzed as part of the classifier. Hardeland (1858:92), who limits his discussion to a short paragraph, gives only three numeral classifiers in Ngaju Dayak, *biti* (lit. 'body'), used for humans), *koŋan*, used for animals, and *kabawak*, used for roundish objects, as in *aso lima koŋan* (dog five-**clas**) 'five dogs', or *eñoh telo kabawak* (coconut three-**clas**) 'three coconuts'. Walker (1976:16), who also limits his discussion of this topic to a few lines, states that numeral classifiers ('counter-nouns') in Lampung of southern Sumatra are 'used much less than in Indonesian and in other closely related languages.' He gives just three classifiers, two derived from independent nouns and one from an independent verb: *biji* 'seed', *cintil* 'bunch' and *ikoʔ* 'tie up': *təlu ŋam-biji manuʔ* (three **lig-clas** chicken) 'three chickens', *xua ŋa-ikoʔ maŋga* (two **lig-clas** mango) 'two mangos', *paʔ ŋa-cintil xambutan* (four **lig-clas** rambutan) 'four bunches of rambutans'. Similarly, for Manggarai of west Flores, Burger (1946:97) identifies only three words of this type: 1) *moŋko* (inanimate things, and some animals), 2) *tau* (humans), 3) *ŋata* (humans at the beginning of a story). All three are said to be optional. Finally, Robson (2002:64) gives a handful of Javanese classifiers that correspond functionally to the 'measure nouns' of Karo Batak, but only a single 'general classifier', *iji* 'piece', as in *jəruk lima-ŋ iji* (orange five-**lig** piece).

Numeral classifier systems in AN languages of eastern Indonesia and the Pacific differ from those typical of western Indonesia in at least four ways: 1) in some languages numerals cannot occur without a classifier, which may become fossilised on the base, 2) classifiers are sometimes limited to certain semantic domains, whereas in western Indonesia classifiers are almost always general purpose category labels, 3) there may be partial fusion of the numeral and its associated classifier, or paradigmatic irregularity, 4) in Oceanic languages numeral classifers are sometimes based on multiples of ten, while the use of classifiers to mark multiplicative values is virtually unknown further west.

The first of these distinguishing characteristics is seen in South Halmahera-West New Guinea languages. Some languages in this group have a 'prefix' that is obligatory on the numerals 1-9. Table 5.9 illustrates the occurrence of this element in three languages of southern Halmahera:

Table 5.9 Numerals with obligatory 'prefix' in three South Halmahera languages

Weda	**Gane**	**Taba**	
p-uso	p-so	p-so	'one'
pe-lú	p-lu	p-lu	'two'
pe-tél	p-tol	p-tol	'three'
pe-fót	p-hot	p-hot	'four'
pe-lím	p-lim	p-lim	'five'
pe-wonem	p-wonam	p-wonam	'six'
pe-fĭt	p-fit	p-hit	'seven'
pe-wál	p-wal	p-wal	'eight'
pe-pupet	p-siw	p-sio	'nine'
yofesó	yagimsó	yo-ha-so	'ten'

Maan (1951) showed that this element is separable in Buli, but more information on the South Halmahera languages became available in Bowden's (2001) grammar of Taba, or

Makian Dalam, where it is noted (2001:244) that "Taba numerals must always co-occur with … the default classifier *p-* (probably derived ultimately from PAN *buaq 'fruit')." This suggestion is supported by data in Anceaux (1961) on West New Guinea languages, where the corresponding element is *bo-* (cf. Ambai *boŋ*, Kurudu *bo-* 'fruit'):

Table 5.10 Numerals with obligatory 'prefix' in three West New Guinea languages[44]

Ambai	Munggui	Kurudu	
bo-iri	bo-hiri	bo-sandi	'one'
bo-ru	bo-ru	bo-ru	'two'
bo-toru	bo-toru	bo-toru	'three'
bo-a	bo-ati	bo-at	'four'
ri/riŋ	bo-rim	bo-ve-rim	'five'
wona	bo-wonam	5+1	'six'
itu	bo-itu	5+2	'seven'
5+3	bo-waru	5+3	'eight'
5+4	bo-hiun	5+4	'nine'
sura	saura	sur	'ten'

Surprisingly, although the numeral base is clear from contrast, it never occurs alone. This would be equivalent to a Malay speaker using *sə-buah*, *dua buah*, *tiga buah*, etc. in serial counting—something totally alien to the numeral classifier systems of western Indonesia. The languages of the Admiralty Islands may show a similar incorporation of a reflex of *buaq 'fruit'. In these languages classifiers follow the numeral base, and the numerals 1-9 (except '4'!) have an unexplained 'suffix' that reflects *-puV:

Table 5.11 Numerals with obligatory 'suffix' in four languages of the Admiralty Islands

Penchal	Ere	Sori	Seimat	
sɨw	sih	sip	te-hu	'one'
lʊp	ruoh	huop	hũõ-hu	'two'
tulʊp	tulah	tarop	tolu-hu	'three'
talɨt	hahuw	papuw	hinalo	'four'
rurɨn	limoh	limep	te-panim	'five'
ʊnʊp	onah	gonop	tepanim tehu	'six'
karutulʊp	drotulah	ehetarop	tepanim hũõ-hu	'seven'
karulʊp	droruoh	anuhuop	tepanim toluhu	'eight'
karusɨw	droasih	anusip	tepanim hinalo	'nine'
saŋahul	saŋul	saŋop	hũõ-panim	'ten'

Semantic specialisation is seen in the numeral classifiers of Samoan. Mosel and Hovdhaugen (1992) give fifteen 'food classifiers', all of which apply to raw foodstuffs or prepared foods. Of these *fua*, used for breadfruit, coconuts, fowls and some shellfish is familiar from the classifier systems of insular Southeast Asia (Malay *buah*): *tolu ŋa fua niu* (three **lig clas** coconut) 'three coconuts'. Others correspond semantically, but are not cognate, as with Samoan *tau*, used with bunches or clusters of coconuts or other fruits: *tau*

---

[44]   Anceaux (1961:74-75) does not give the Ambai forms for '8' and '9', or the Kurudu forms for 6-9, but states in each case that they "use compounds" based on five plus a smaller numeral.

fia niu (**clas** how many coconut) 'how many coconuts?', *tau lua popo* (**clas** two ripe coconut) 'a pair of coconuts'. Still others are associated with ideas not usually expressed by numeral classifiers further west, as with Samoan *aea* 'a score (of coconuts)', or *afi* 'packages of small fish wrapped in leaves'. Mosel and Hovdhaugen give just one non-food classifier, *toʔa* (persons). Unlike the food classifiers this word does not require an associated noun: *toʔafia* (X) 'how many (people)?', *toʔalima* (X) 'five (people)'.

   Fusion of numeral and classifier can be illustrated from a number of languages in eastern Indonesia and the Pacific. Klamer (1998:136ff) lists five classifiers for Kambera of eastern Sumba: 1) *wua/mbua* (spherical objects), 2) *puŋu/mbuŋu* (oblong objects), 3) *wàla/mbàla* (flat, thin objects), 4) *iu/ŋiu* (animals), 5) *tau* (people). She also notes that in some forms 'the number/prefix and the classifier are merged.' This is illustrated with *hau kajawa* (one-**clas** papaya) 'one papaya' : *dàmbu kajawa* 'two papayas' : *tailu mbua kajawa* 'three papayas', as against *ha-puŋu pena* (one-**clas** pen) 'a pen' : *dua mbuŋu pena* 'two pens' : *tailu mbuŋu pena* 'three pens', where the forms *hau* and *dàmbu* evidently originate from earlier *ha + wua* (one + **clas**), and *dua + mbua* (two + **clas**) respectively. In addition, a number of languages in the Admiralty Islands have rich systems of numeral classification that illustrate partial fusion of numeral and classifier, as well as categories based on multiples of ten. For Loniu of eastern Manus, Hamel (1994:54ff) lists some 30 numeral classifiers. In effect, the fusion of numeral and classifier has given rise to 30 distinct sets of numerals, although these show a number of recurrent partials. Table 5.12 gives the full sets for three of these to illustrate the workings of the larger system:

Table 5.12 Three sets of numeral classifiers with morphological fusion in Loniu

	can/calan	**kew/kɛwan**	**kɔʔɔt**
1	hacan	hekew	hɔkɔʔɔt
2	maʔucɛn	maʔakew	maʔakɔʔɔt
3	maculucan	maculukew	maculukɔʔɔt
4	mahacan	mahakew	mahakɔʔɔt
5	malimɛcan	malimɛkew	malimɛkɔʔɔt
6	mawɔnɔcan	mawɔnɔkew	mawɔnɔkɔʔɔt
7	maʔaruculucan	maʔaruculukew	maʔaruculukɔʔɔt
8	maʔaruʔucɛn	maʔaruʔukew	maʔarukɔʔɔt
9	maʔarusacan	maʔarusekew	maʔarusɔkɔʔɔt
10	macalansɔʔɔn	makɛwansɔŋɔn	masɔŋɔn

   The independent noun *can* (unpossessed/alienably possessed) or *calan* (inalienably possessed) heads the first column, and this numeral set is used in counting roads, paths, and boundaries (as section markers in gardens). The second numeral set, headed by the independent noun *kew/kɛwan*, of unknown meaning, is used to count strings of valued objects such as beads, dogs' teeth, tambu shells, or fish: *lɛhɛ mʷi masaŋat tɔ hekew* (tooth dog 100 **stat** one-string) 'there are 100 dogs' teeth on one string'. The third set, headed by the independent noun *kɔʔɔtan* 'bundle', is used to count bundles of long thin items such as spears, sugarcane, bamboo, firewood, or palm thatch. In this set the number 'ten' is the bare numeral without a classifier. This set raises another interesting point. Since some classifiers refer to individual entities and others to groups of ten, an expression such as 'ten spears', which is culturally ambiguous to a speaker of Loniu can be disambiguated as *ñah makɔʔɔsɔŋɔn* 'ten (individual) spears' vs. *ñah hɔkɔʔɔt* 'ten spears (in a bundle)'.

Similar complexities are found in the counting systems of Nuclear Micronesian languages. Rehg (1981:125), for example, states that there are "at least thirty ways" to count in Pohnpeian. By this he means that there are at least thirty distinct numeral sets formed by the fusion or partial fusion of a numeral-classifier sequence. Three of these numeral sets are illustrated in Table 5.13 (*-u* < *puaq, *-men* < *manuk, *-um*^{w} < *qumun):

Table 5.13 Three sets of numeral classifiers with morphological fusion in Pohnpeian

	I	II	III
1	e:u	emen	oum^w
2	riau	riemen	rioum^w
3	silu:	silimen	silu:m^w
4	pa:ieu	pa:men	pa:um^w
5	limau	limmen	limoum^w
6	weneu	wenemen	wenoum^w
7	isu:	isimen	isu:m^w
8	walu:	welimen	welu:m^w
9	duwau	duwemen	duwoum^w
10	eisek	e:k	ŋoul

Set I is said (1981:125) to be the "general counting system, often used to count things which have the property of being round." Set II is used with animate referents, and Set III for certain foods that are baked. Historically, these sets consist of numeral + POC *puaq 'fruit' (Set I), *manuk 'non-marine animal' (Set II), and *qumun 'earth oven' (Set III).

Finally, some Oceanic languages use classifiers for collective nouns, rather like English 'school', 'herd', or 'flock'. Davis (2003:71) has shown that Hoava of the western Solomons uses collective nouns without numerals to designate "large quantities of animals and fish": *sa rovana boko* (**art clas** pig) 'the large number of pigs', *sa rovana lipa* 'the school of *lipa* fish', *sa puku nikana* (**art clas** man) 'the group of men', *sa puku igana* (**art clas** fish) 'the group of fish', *sa topatopa boko* (**art clas** pig) 'the large number of wild pigs'. As seen in the last example, the distinction between domesticated and wild pigs, which many languages of insular Southeast Asia encode in the independent noun (Puyuma *verek* 'domesticated pig' : *vavuy* 'wild pig', Kelabit *bərək* 'domesticated pig' : *baka* 'wild pig') is encoded in Hoava by the collective noun marker.

Almost no comparative work has been done on numeral classifiers in AN languages. However, the first thing that emerges clearly even from the limited data considered here, is the centrality of *buaq as a classifier. This is the only form that can be attributed to PMP with possible use as a numeral classifier. In addition, although they are commonly associated with roundish objects, reflexes of *buaq are given as the 'default' classifier in a number of widely separated languages (Malay, South Halmahera-West New Guinea languages as a whole, Pohnpeian, Woleaian). Both of these observations suggest that the history of numeral classifiers in AN languages is critically dependent on this single form. In other words, PMP may have had no numeral classifiers other than *buaq, and from this limited beginning many of the attested systems were elaborated.

To see how this might have happened it should be noted that in many AN languages the names of specific fruits must be accompanied by the general marker that reflects *buaq: Kelabit *buaʔ buyo* 'citrus fruit', *buaʔ datuʔ* 'durian', *buah laam* 'mango', Malay *buah ñiur* 'coconut', *buah maŋga* 'mango', *buah pisaŋ* 'banana', Tetun *hudi fua-n* 'banana', Soboyo *nuo mfuai-n* 'coconut'. These collocations produce an ambiguous constituent structure. In

Malay/Indonesian, expressions such as *dua buah ñiur* (two-fruit coconut) 'two coconuts' and *dua buah rumah* (two-**clas** house) 'two houses' appear to be structurally parallel, but since names of fruits are normally preceded by *buah* there is a bracketing ambiguity in quantitative expressions for fruit names that is absent in similar expressions for other nouns, whether they take *buah* or another numeral classifier:

---

buah 'fruit; classifier for large spherical objects'

rumah 'house' (**buah rumah)
[dua buah] [rumah] 'two houses'

buah ñiur 'coconut' (**ñiur)
[dua buah] [ñiur]/[dua] [buah ñiur] 'two coconuts'

Figure 5.1 Bracketing ambiguity in Malay quantitative expressions for names of fruits

Expressions such as *dua buah rumah*, in which the independent noun 'fruit' appears as a classifier, almost certainly derive from constructions such as *dua buah ñiur*, where the constituent structure was ambiguous. But if numeral classifiers did not already exist, it is difficult to see what could motivate a misanalysis of constituent structure so as to reinterpret PMP *buaq or its reflexes as a classifier. As seen in the Timugon Murut, Lampung and Samoan examples given above, a reflex of *ŋa (shortened to ŋ in some languages) is required between a numeral and a classifier, just as in higher numerals based on multiplicative values of '10' or '100'. A structural parallel must therefore have existed in PMP numerative expressions of the type *telu ŋa puluq (ma) esa (three **lig** group-of-ten (and) one) '31', and *telu ŋa buaq niuR (three **lig** fruit coconut) 'three coconuts'. Since the constituent structure of multiplicative numerals with a numeral adjunct clearly was {telu ŋa puluq} {ma esa} a model existed for {telu ŋa} {buaq niuR} to be reanalyzed as {telu ŋa buaq} {niuR}, freeing the independent noun to occur without a class name, as it does in many attested languages, and forcing a reinterpretation of *buaq as an incipient numeral classifier. From this point the system presumably would have grown by accretion, an inference that is consistent with the wide variety of forms and functions exhibited by attested systems of numeral classifiers in AN languages.[45]

One last point is perhaps worth making. There is some marginal evidence for *tau 'person' as a PMP numeral classifier, but the need to reconstruct Set B numerals for PAN and PMP makes it almost certain that PMP *tau was not a numeral classifier for [+human] referents, since this distinction was already marked in the form of the numeral.

5.3   Colour terms

In their often questioned, but foundation-laying study of cross-linguistic regularities in colour nomenclature, Berlin and Kay (1969) made use of eleven basic colour categories.

---

[45]   PMP *buaq evidently was also found in some figurative expressions such as *buaq na bities 'calf of the leg' or *buaq na lima 'finger', and the preexistence of such non-literal uses of this morpheme may have faciltitated the transition from *buaq 'fruit' to *buaq 'fruit; numeral classifier.'

These categories appear in Table 5.14 in the order of their implicational relationships, together with the forms that can be reconstructed for them in Proto Austronesian, Proto Malayo-Polynesian, and Proto Oceanic:

Table 5.14 Reconstructed colour terminology for PAN, PMP and POC

	PAN	PMP	POC
White	ma-puNi	ma-putiq	ma-puteq
Black	ma-CeŋeN	ma-qitem	ma-qetom
Red	ma-taNah	ma-iRaq	meRaq
Green			karakarawa
Yellow			aŋo
Blue			karakarawa

As seen in Table 5.14, only the three most basic colour terms can be reconstructed for PAN (c. 5,500 BP), and PMP (c. 4,500 BP), and only the five most basic colour terms (conflating green-blue as 'grue') can be reconstructed for POC (c. 3,500 BP). No colour term that is widely distributed outside Taiwan appears in any Formosan language. It is difficult to know how to interpret the first of these observations. Given the evolutionary schema of Berlin and Kay it could easily be concluded that no more than three colour terms were used by AN speakers as recently as 4,500 years ago, but that two additional terms were innovated by speakers of Proto Oceanic as they moved outward into the Pacific. The problem with this line of reasoning is that it links terminological elaboration to time without reference to any other factor. The second observation suggests that the colour terminology of PAN was completely relexified by speakers of PMP, who retained the stative prefix *ma-, but replaced the lexical bases meaning 'white', 'black' and 'red'.

Many colour terms in attested languages have been recruited from common nouns, and some of these reconstructed forms show similar points of contact with nominal bases. The last syllable of PMP *ma-iRaq is identical to that of PAN/PMP *daRaq 'blood', and suggests a historical derivation, although there is no transparent morphological basis for such a hypothesis. POC *aŋo meant both 'yellow', and 'turmeric', and it is certain that the colour term derives from the botanical one.

The colour nomenclature of most attested AN languages is considerably richer than the bare minimal that can be reconstructed for earlier stages, and the source of additional colour terms is a matter of some interest. Table 5.15 gives the full colour terminology of four languages, Ilokano of the northern Philippines, Malay of western Indonesia, Chuukese of central Micronesia, and Hawaiian:

Table 5.15 Colour terminology of Ilokano, Malay, Chuukese and Hawaiian

	Ilokano	**Malay**
White	na-púdaw	putih
Black	na-ŋísit	hitam
Red	na-labága/ na-labásit	merah
Green	na-laŋtó	hijaw
Yellow	duyáw	kuniŋ/jiŋga
Blue	balbág	biru
Brown	madkét	coklat
Pink	paŋ-in-dará-en	merah muda
Purple	morádo/púrpura	uŋu/jiŋga[46]
Orange	kiáw	jiŋga tua
Gray	kolordapó	abu-abu/kəlabu

	Chuukese	**Hawaiian**
White	pʷech	kea/keʔo(keʔo)
Black	chón	ʔeleʔele
Red	par	ʔula/mea
Green	énúyénún fetin/araw	ʔōmaʔo, uli
Yellow	reŋ/ón/ram	melemele/ʔolenalena
Blue	araw	uli
Brown	kuŋ/ffach/-móów	kamaʔehu
Pink	rowarow	ʔakala/ʔōhelo
Purple	énúwén foorket	poni
Orange	ram	melemele ʔili ʔalani
Gray	topʷ	hina

Based on implicational relationships and historical change the Berlin-Kay basic colour terms can be divided into three tiers: 1) white, black and red, 2) green-blue (sometimes called 'grue', since many languages do not distinguish the two), and yellow; and 3) the rest. Since PAN, PMP, and POC etyma exist for the first tier, we would expect attested terms to reflect the reconstructed forms, or to be semantically opaque lexical replacements. In other words, it would be unexpected for terms like 'black' or 'red' to be replaced by terms which earlier meant 'charcoal', or 'blood', since the concept of an abstract colour category already existed, and there would be no need to re-create it by derivation from a concrete noun. However, this expectation is sometimes violated. Ilokano *na-púdaw* 'white' reflects an innovation common to the Northern and Central Cordilleran languages of Luzon, where it usually means 'white' (a reflex of PMP *putiq is retained in Southern Cordilleran languages such as Ibaloy and Pangasinan). However, Newell (1993:462) gives Batad Ifugao *pudaw* 'light-skinned, as … an albino or Caucasian', and Yamada (2002:213) gives Itbayaten *poraw* 'gray, grayish'. In Proto Philippines *ma-putiq continued to mean 'white', and *ma-pudaw had some other, related sense, perhaps 'light-skinned/albino'. The

---

[46]   Wilkinson (1959:472) glosses Malay *jiŋga* as 'dark yellow; yellow mixed with red or purple; light purple'. Although such non-canonical colours which cover a wide range of the spectrum may at first appear baffling, it is possible that they derive from the colouration of natural vegetative objects which changes through maturation, and so passes through mixed colour phases, as with certain varieties of bananas.

basic colour term *ma-putiq was then replaced in the ancestor of many of the languages of northern Luzon by this non-colour term. Goodenough and Sugita (1980) give Chuukese *pᵂech* 'white', *pᵂeech* 'powdered lime made by burning coral limestone', and it is clear that the colour term derives from the concrete noun, as the derivational relationship between concrete noun and colour adjective is transparently preserved in other Nuclear Micronesian languages: Kosraean *fasr* 'coral lime, limestone' : *fasrfasr* 'white', Pohnpeian *pᵂe:t* 'lime made from coral' : *pᵂetepᵂet* 'white; grey hair'. Similarly, Hawaiian *kea* 'white' reflects Proto Central Pacific *tea 'pale, albino'. The PMP monomorphemic colour base *putiq was thus replaced by terms meaning 'pale-skinned, albino' both in the northern Philippines and in Polynesia, and by terms meaning 'coral lime' in Micronesia. Many other languages have innovative terms for 'white', 'black', or 'red', but the source is generally obscure.

When we move to the second tier of colour terms the derivation of colour words from concrete nouns is even more apparent. A number of Philippine languages reflect PPH *dulaw 'turmeric' in the meaning 'yellow', as with Kalamian Tagbanwa *ma-dulaw* or Mamanwa *ma-dolaw* (Ilokano *duyáw* may also reflect this term, but if so it is irregular). A very similar situation is found with Malay *kuñit* 'turmeric' and *kuniŋ* 'yellow', both of which apparently reflect *kunij 'turmeric', the former directly and the latter through an early borrowing of Karo Batak *kuniŋ* 'turmeric'. The same set of semantic relationships recurs in Oceanic languages with a different base. Hawaiian *ʔolenalena* 'yellow; dye made from the *ʔolena* plant' is derived from *ʔolena* 'the turmeric: *Curcuma domestica*', and Goodenough and Sugita (1980:309) gloss Chuukese *reŋ* as '(be) yellow, yellow-green, saffron coloured (associated with turmeric)', a form that is cognate with e.g. Fijian *re-reŋa* 'turmeric prepared from the root of the *caŋo* and used for besmearing the body of a new-born infant'. The term *caŋo* in turn is glossed 'the turmeric plant, *Curcuma longa*', and is cognate with the word for 'yellow' in many other Oceanic languages, as Lou *aŋo-an*, Talise, Mafea *aŋo*, Nggela *aŋoaŋo*, Merlav *aŋaŋ* or Anejom *in-yaŋ* 'yellow'. In short, words for 'yellow' have been derived repeatedly throughout the AN language family from the name of the turmeric, a plant traditionally valued for its yellow dye. The distinction between 'green' and 'blue' in Malay is relatively recent, as *hijaw* earlier covered the entire blue-green range of the spectrum, and probably referred to a colouration pattern in fighting cocks, as seen in Tagalog *híraw* 'cock with metal-green feathers', Iban *ijaw* 'dark colours on yellow or white, as a colour of fighting cocks'. In Chuukese *araw* covers blue-green, and the use of *énúyénún fetin* (lit. 'colour of grass') to distinguish green from blue almost certainly postdates western contact.

Tier 3 terms are created 1) by affixation or compounding of a concrete noun, 2) by borrowing from a European language, 3) by nuancing of a basic colour term, or 4) by a descriptive term 'colour of X'. The first type of formation is seen in Ilokano *paŋ-in-dará-en* 'light red, pink', which Rubino (2000:154) derives from *dára* 'blood', and in Malay *abu-abu* and *kəlabu* 'gray', both derived from *abu* 'ash'. The second is seen in Ilokano *morádo* and *púrpura*, from Spanish *morado* 'purple, mulberry-coloured', and *púrpura* 'purple, purple cloth', and in Malay/Indonesian *coklat*, borrowed from English or Dutch. The third is seen in Malay *merah muda*, and illustrates the use of the terms *muda* 'young' and *tua* 'old' to nuance colour terms (comparable to 'light' and 'dark' in English). The fourth method used to create colour terms is seen in Ilokano *kolordapó* 'gray' (lit. 'colour of ashes'), in Chuukese *énúwén foorket* 'purple' (lit. 'colour of Forget-me-nots'), and in Hawaiian *melemele ʔili ʔalani* 'orange' (lit. 'orange-peel yellow'). The references to introduced plants in the last two terms is a clear indication of their relative newness.

Terms for 'gray' in languages of insular Southeast Asia often derive from a word for 'ashes', but in the Pacific they appear to be more commonly created from words meaning 'faded, discoloured, blurred; absence of colour', or less often, from the word for 'gray hair' (a semantic category that is lexically distinguished from the colour terms in many or most AN languages). Words for 'brown' rarely refer to the prototypical brown of Indo-European languages, but often indicate some admixture of qualities, as in Hawaiian *kamaʔehu* 'brownish, reddish-brown', or the unusual occurrence in Chuukese of three terms for 'brown', *kuŋ* '(be) brown', *ffach* '(be) light brown', and *-moow* 'reddish-brown' (in compounds only). Likewise, a number of Oceanic languages have a word that appears to translate better as 'reddish-brown' or 'reddish-yellow' than as any of these single terms, and this is sometimes associated with the colour of fish, or with reddish-brown earth: Lau *mela* 'light brown, reddish brown; sp. of fish', Gilbertese *mea* 'reddish-yellow colour; rust; grey', Tongan *mea* 'light red or light brown, reddish, brownish: especially in names of plants, fish, etc.', Hawaiian *mea* 'reddish-brown, as water with red earth in it; yellowish-white, of feathers'. The source of the monomorphemic terms for 'purple' in Malay and Hawaiian, and of 'orange' in Ilokano or Chuukese, is unknown.

A special feature of colour terms in Oceanic languages is the widespread use of fossilised reduplication (Blust 2001c). To illustrate, Pukui and Elbert (1971:37) give Hawaiian *ʔele* 'black', but state that this is less common than *ʔeleʔele*, which has the same meaning and syntactic distribution. A similar nonfunctional use of reduplication in colour terms is found in many other Oceanic languages where a reduplicated term either has no corresponding simplex base, as with Manam *botiboti* 'blue' (no **boti), or is derived by reduplication from a simplex concrete noun (Kairiru *kietkiet* 'black', from *kiet* 'black paint'). This pattern may have arisen from the use of reduplication to mark colour terms as inexact equivalents of their canonical meanings: whitish, blackish, reddish, etc. Since real world referents generally depart to some extent from canonical types the reduplicated variants of colour terms presumably would have been more frequent in daily use than their simplex counterparts. In time, the overuse of such attenuative or approximative forms led them to become semantically unmarked.

Most grammars contain little information on such features as nuanced colour terms ('light red', 'dark green'). As noted earlier, in Malay/Indonesian these are marked by the terms for 'young' and 'old', but it is unknown how widespread this feature is. Similarly, little information is available on conjunctive colour terms (red and white, black and blue). In Malay/Indonesian these tend to follow the same order as English, thus *merah-putih* 'red (and) white', in which the darker or more intense of two colours tends to come first.[47]

In a number of the societies of insular Southeast Asia where cockfighting is an important social activity a special set of words is used to describe distinctive patterns of colouration in fighting cocks. Richards (1981:207) lists no fewer than 30 named colour

---

[47]   A somewhat similar, but not identical problem, is discussed by Shen and Gil (2007), who address the issue of word order in synaesthetic metaphors, with special reference to Indonesian. Building on earlier work, they appeal to a directionality principle which determines the order of lexical elements in a synaesthetic metaphor in relation to a universal hierarchy of sensory modalities (sight > sound > smell > taste > touch) such that "With greater than chance frequency, synaesthetic metaphors involve mappings upwards on the Hierarchy of Sensory Modalities (with the exception of sight and sound, which behave in a similar fashion)" (Shen and Gil 2007:3). It is unclear how this proposal would deal with the widespread use of reflexes of *deŋeR 'to hear' in AN languages for e.g. 'smell', as in Nume (Banks Islands) *rongo-mbun* 'to smell', or its equivalent *harem* in Bislama for e.g. 'taste', 'smell', 'feel', etc., as in *harem kava* 'feel the effects of kava', *harem smell* 'smell an odor', *harem traot* 'nauseous, feel sick' (Crowley 2003).

varieties for this lexical domain in Iban of southwest Borneo. The main colour groups are *adoŋ* 'white with red wings', *banda* 'yellowish red', *biriŋ* 'deep red-brown', *burik* 'speckled, streaky (black and white)', *ijau* 'dark colours on yellow or white', and *kəlabuʔ* 'light brown,' but each of these may be further subdivided, as with *adoŋ burik* 'white with light brown patches', *banda pipit* 'red with black feather tips', *biriŋ cərəkup* 'dark red and green-black', *burik kəsaʔ* 'red-speckled', *burik kəsulai* 'speckled with light brown', *burik mənaul* 'large speckles', *burik paŋgaŋ* 'lightly speckled', etc.

One last feature of colour terms that might be mentioned is the symbolic value of colours in describing mood, temperament or character, as with English blue = 'unhappy', yellow = 'cowardly', green = 'inexperienced', and the black/white contrast generally signaling negative vs. positive valuation. In AN languages colour terms usually mark human character traits only when combined with some other morpheme, normally the word for 'liver' (the seat of the emotions), as in Tausug *atay-itum* ('black liver') 'treachery', and in the following expressions, all of which literally mean 'white liver': Tausug *atay putiʔ* 'sincerity, blamelessness', Mapun *poteʔ atay* 'pure heart/blameless', Mansaka *ma-potiʔ na atay* 'kind heart, tender heart', Malay *puteh hati* 'sincerity', Madurese *pote ate* 'upright, honest', or Motu *ase kuro tau-na* (lit. 'man with a white liver') 'a brave man, one who is not afraid'.

5.4   Demonstratives, locatives, and directionals

The way that a language encodes directional information, or uses directional terms to encode other types of information can be subtle and far-reaching. In recent years this aspect of the lexicon in AN languages has attracted increasing attention, as seen in such general collections as Senft (1997) and Bennardo (2002), in the penetrating work of François (2003a, 2004), and in the valuable overview of the linguistic expression of location and direction in Oceanic languages by Ross (2003). In order to facilitate the discussion it will be useful to divide the material into systems of micro-orientation, and systems of macro-orientation. Under the former I include two subcategories: 1) the spatial and temporal location of referents in relation to the speaker, and 2) the location of referents in relation to their surroundings (above, below, inside, outside, etc.). Under macro-orientation I include directional systems used to orient oneself to the wider physical environment. Systems of micro-orientation thus contain information that is commonly included under the rubric 'deixis', although it is both wider than this category in some respects and narrower in others. Although systems of micro-orientation and macro-orientation are complementary in most languages, in some AN languages they reportedly interpenetrate in unexpected ways.

To begin with demonstrative systems, unlike modern English, which has a two-way distinction in the determiners *this : that* or the adverbs *here : there*, most AN languages divide this semantic space into a proximal deictic and two distal deictics. While the semantic specification of the proximal deictic is relatively fixed across languages, the distal deictics are distinguished in different ways. Reid (1971) provides comparable data on the demonstrative pronouns of 41 Philippine languages, most of which share three general features. First, of the 41 languages in his sample (43 minus two for which glosses were not available), 25 have a three-way deictic distinction that closely matches the person-marking distinction in personal pronouns: 1) this, 2) that (by you), 3) that (by third person). Examples include Guinaang Kalinga 1) *siyaná*, 2) *siyanát*, 3) *siyadí* in northern

Luzon, and Siocon Subanen 1) *koni*, 2) *koyon*, 3) *kituʔ* in western Mindanao.[48]   Eleven other languages make a four-way distinction: 1) this, 2) that (by you), 3) that (by third person), 4) far distant/out of sight. Examples include Agta 1) *yən*, 2) *ənə*, 3) *yewən*, 4) *yen* in northern Luzon, and Koronadal Bilaan 1) *ani*, 2) *aye*, 3 *atuʔ*, 4) *ayə*, in southern Mindanao. Five languages have a two-way distinction, either in the entire demonstrative system, or in one part of it. An example of a symmetrical two-term demonstrative system is Itbayaten, with 1) *niaʔ* 'this', 2) *nawiʔ* 'that', and 1) *diiʔ* 'here', 2) *dawiʔ* 'there'; an example of an asymmetrical two-term demonstrative system is Cotabato Manobo, with 1) *ini* 'this', 2) *ia* 'that', but 1) *dahini* 'here', 2) *dahia* 'there (by you)', 3) *kaʔədoʔ* 'there (by third person)', 4) *dahədoʔ* 'there (distant or out of sight)'. No information is given about the semantic specification of category 4) in any given system, but the meanings 'far distant' and 'out of sight' never occur in the same language. Second, in many languages there is a phonemic similarity or overlap in the shapes of words that translate as English 'this' vs. 'here' or 'that' vs. 'there'. Examples include Pamplona Atta (Northern Cagayan Negrito) 1) *yawe* 'this', 2) *yine* 'that (by you)', 3) *yuke* 'that (by third person)', 4) *yu:rin* 'that (distant or out of sight)' next to 1) *sawe* 'here', 2) *tane* 'there (by you)', 3) *tuke* 'there (by third person)', 4) *tu:rin* 'there (distant or out of sight)', or Guinaang Kalinga 1) *siyaná* 'this', 2) *siyanát* 'that (by you), 3) *siyadí* 'that (by third person) next to 1) *siná* 'here', 2) *sinát* 'there (by you), 3) *sidí* 'there (by third person). Third, as these examples illustrate, many of these forms begin with the same phoneme or phoneme sequence, which sometimes can be identified as a fossilised generic marker of location (at-here, at-there).

Many of the languages of Borneo have similar systems of demonstrative reference, but with differences of detail. The Uma Juman dialect of Kayan in central Borneo, for example, has a three-way deictic system that evidently correlates with person reference: 1) *anih* 'this', 2) *anan* 'that (by you)', 3) *atih* 'that (by third person)', and 1) *hinih* 'here', 2) *tinan* 'there (by you)', 3) *hitih* 'there (by third person)'. Long Lamai Penan also has a three-way deictic system, but one that maps the third and fourth terms of some Philippine systems onto the second and third terms: 1) *itəuʔ* 'this', 2) *inəh* 'that (visible)', 3) *itay* 'that (out of view)', 1) *sitəuʔ* 'here', 2) *sinəh* 'there (visible)', 3) *sitay* 'there (out of view)'. Still another variant occurs in Mukah Melanau, where a three-way contrast is based partly on proximity to speaker and partly on the speaker assuming the hearer's previous knowledge of the location mentioned: 1) *itəw* 'this', 2) *iən* 'that' (addressee already knows the place mentioned, whether near or far), 3) *inan* 'that' (indefinite location), 1) *gaʔ gitəw* 'here', 2) *gaʔ giən* 'there'(addressee already knows the place mentioned, whether near or far), 3) *gaʔ ginan* 'there (indefinite location)'. As in the Philippines, some two-term deictic systems in western Indonesia are asymmetrical. A well-known example is Malay, which has 1) *di-sini* 'here', 2) *di-situ* 'there (near you)', 3) *di-sana* 'there (distant, whether in view or not)', but 1) *ini* 'this', 2) *itu* 'that' (no **ana). A number of languages in western Indonesia and the Philippines have what appears to be a historically double layer of affixation marking a generic locative in terms that translate as 'here' or 'there'. In some cases the generic locative marker is the same morpheme, with one layer fossilised and the other active, as with Mukah Melanau *itəw* : *gaʔ gitəw*, where *g-itəw* presumably reflects *gaʔ *itəw*. In other cases, the fossilised generic marker of location differs from the marker that is now active, as seen in comparing the following double-prefixed Malay forms with the single-prefixed forms of other languages:

Table 5.16 Historically double prefixation in the Malay demonstrative adverbs[49]

	Malay	**Ifugaw**	**Sangil**	**Aborlan Tagbanwa**
This	ini	tue	íni	ini
That (2nd person)	itu	naye	éne	itu
That (3rd person)	itu	die	éne	ian
That (distant)				iti
Here	di-s-ini	hitú	s-íni	s-ini
There (2nd person)	di-s-itu	hiná	s-éne	s-itu
There (3rd person)	di-s-ana	hidí	s-éne	asan
There (distant)				duʔun

In Malay the bound forms *s-ini* and *s-itu* contain a fossilised element that appears to be cognate with the initial phoneme of such Philippine forms as Amganad Ifugaw *hitú* 'here', *hiná* 'there (by you)', *hidí* 'there (by third person)', Sangil *síni* 'here', *séne* 'there (by you)', or most strikingly with Aborlan Tagbanwa *s-ini* 'here', *s-itu* 'there (by you)', where the entire form is cognate and *s-* marks generic location in Aborlan Tagbanwa just as *di-* now does in Malay. It thus appears that *s-* once had a similar function in the Malay bound demonstratives *-sini* 'here', *-situ* 'there', and *-sana* 'yonder'. Structurally similar systems of deictic reference are widespread in the AN language family, being found in Loniu of the eastern Admiralty Islands (Hamel 1994:99), in Mekeo and other closely related languages of southeast New Guinea (Jones 1998:157), in Mokilese of central Micronesia (Harrison 1976:77ff), and in Fijian of the central Pacific (Schütz 1985:378ff).

Somewhat more elaborate systems of demonstrative reference are found in some parts of Indonesia. For Kambera of eastern Sumba, Klamer (1998:55-56) lists four deictics that differ from the preceding systems in having a two-way distinction for the proximal member: 1) *ni* 'this', 2) *nai* 'that (near speaker, but further away than *ni*)', 3) *na* 'that (near addressee)', 4) *nu* 'that (far from speaker and addressee)'. Woollams (1996:122ff), reports five demonstrative pronouns in Karo Batak of northern Sumatra, again with what appears to be a two-way proximal distinction (1 and 5): 1) *enda* 'this (relatively close to speaker)', 2) *ena* that (relatively close to addressee)', 3) *ah*, *adah* 'that (over there, outside the immediate proximity of both speakers and addressees)', 4) *oh* 'that (in the far distance, possibly out of sight)', 5) *e* 'that (within view of speaker and addressee, or something just referred to)'. Each of these deictics corresponds to what Woollams called a 'locative pronoun', some of which are preceded by j- (probably < *di-): 1) *enda* 'here', 2) *ena* 'there (by addressee), 3) *jah*, *jadah* 'there (away from both speakers and addressees)', 4) *joh* 'there (some distant place)', 5) *e* 'there (the place just referred to)'.

Some Sulawesian languages have a deictic system that is Philippine-like or Borneo-like. Himmelmann (2001:98ff), for example, notes that most Tomini-Tolitoli languages have a three-way demonstrative distinction which he describes as 1) proximal, 2) medial, and 3) distal. He observes that while closeness to hearer appears to play a role in determining the use of medial forms, the distal form can also be used under similar circumstances. However, in other languages of Sulawesi the semantic structure of the demonstrative

---

[49]   Jason Lobel (p.c., August 3, 2007) has pointed out that Aborlan Tagbanwa *ini* and *sini* both mean 'here (by speaker, not by listener)', *itu* and *situ* mean 'here (by both speaker and listener)', *ian* and *asan* mean 'there (by listener, not by speaker)', and *itu/duʔun* both mean 'there (far from both speaker and listener)'. The evidence for historically double affixation in Malay, however, is unaffected by this observation.

system diverges widely from the types seen so far. All AN systems of demonstrative reference considered previously can be expressed in terms of 1) degrees of distance in relation to speaker or hearer, or 2) visibility. The system of spatial deixis in Muna of southeast Sulawesi appeals to these semantic parameters, but in addition it makes crucial reference to relative *height* and *audibility* of referent. According to van den Berg (1989:85) Muna has six basic demonstrative pronouns that occur in two sets:

Table 5.17 The demonstrative pronouns of Muna, southeast Sulawesi

	Set 1	**Set 2**
1	ini	aini
2	itu	aitu
3	maitu	amaitu
3	watu	awatu
3	tatu	atatu
3	nagha	anagha

The first pronoun is used for whatever is within reach of the speaker. The second is used for what is nearer to the hearer than to the speaker, although it need not be within reach. It may also be used for something near the speaker under certain circumstances, as in distinguishing two objects that are at the same distance ('not this one, but that one'). The third pronoun typically refers to an object that is not far away. Van den Berg (1989:86) considers the fourth pronoun 'the most neutral form in the third person series.' It contrasts with the fifth pronoun in that the latter has an additional component of relative height. The last demonstrative is reserved for referents that cannot be seen by speaker or hearer, but are audible (as a crying child or barking dog). Unlike the other systems seen earlier, where only one set of forms is usually referred to as 'demonstrative pronouns', and the second set carries some other grammatical label, van den Berg calls all of the above forms 'demonstrative pronouns', and shows that the longer forms have a contrastive function: a) *bhai-ku aini* (friend-my this) 'this friend of mine (but not the others here/there)', b) *bhai-ku ini* (friend-my this) 'my friend here/this friend of mine (already mentioned)', a) *ne Raha aini* (loc Raha this) 'in this Raha (capital of Muna—implies that there are more places called 'Raha')', b) *ne Raha ini* (loc Raha this) 'here in Raha (uttered by someone who is in Raha)'.

Other structurally atypical deictic systems are found in the Pacific. Josephs (1975:465ff) describes Palauan demonstratives in terms of three core parameters: 1) whether the referent is a person, animal, or thing, 2) whether the referent is singular or plural, and 3) the relative distance of the referent from speaker and hearer. Only the last parameter, which distinguishes a) near speaker and hearer, b) near hearer but far from speaker, and c) far from speaker and hearer, corresponds closely to those used to structure the deictic systems of insular Southeast Asia, animacy and grammatical number being irrelevant in all known systems of the Philippines and western Indonesia. The deictic systems of some Nuclear Micronesian languages also refer to grammatical number, as in Ulithian where the relevant parameters are 1) the relative distance of the referent from speaker and hearer, 2) the visibility of the referent, and 3) whether the referent is singular or plural (Sohn and Bender 1973:217). The most complex known demonstrative systems in AN languages are found in Vanuatu. François (2002:69ff) reports no fewer than fourteen deictic markers (28 if combined with the 'assertive suffix' -ni ~ -n) in Araki, spoken on a small island off the coast of Espiritu Santo. Relevant semantic parameters include 1) absolute distance and

spatial direction from the starting point/from the speaker, 2) spatial/abstract relationship with a particular person in the situation, 3) the syntactic function of the deictic, and 4) the pragmatic value of the phrase.

Superficially, locative expressions such as 'above', 'below', 'in front', or 'behind' appear to share many of the same semantic properties as their translation equivalents in English. However, patterns of semantic change suggest that this surface similarity in semantic composition may be misleading as a result of differences in the most common contexts of usage. Table 5.18 illustrates locative expressions from four languages:[50]

Table 5.18 Locative expressions in four Austronesian languages

	Paiwan	**Malay**	**Chamorro**	**Samoan**
Above/On top	i-vavaw	di-atas	gi hiloʔ	i luŋa
Below	i-təku	di-bawah	gi papaʔ	i lalo
In front	i-qayaw			
	i-saŋas	di-muka	gi meʔnan	i luma
Behind	i-likuz			
	tja-i-viliľ	di-bəlakaŋ	gi tatten	i tua
Inside	i-taladj	di-dalam	gi halom	i totonu
Outside	i-tsasaw	di-luar	gi hiyoŋ	i fafo
Beside	i-gidigidi	di-sampiŋ	gi fiʔon	i tafatafa

In most languages the term that is glossed 'above' appears to be ambiguous for whether the referent is in contact with a surface or not: Malay *buruŋ itu di-atas rumah* (bird that above house) could mean either 'the bird is on top of the house', or 'the bird is flying/hovering over the house' and can only be disambiguated by including verbs of perching, flying, etc. in the sentence. In general, lexical bases meaning 'above/below' are also used for 'up/down'. Like some other languages, Paiwan lexically distinguishes 'in front' and 'behind' with reference to motion: *i-qayaw* and *i-likuz* refer to stationary location, while *i-saŋas* and *ča-i-viliľ* refer to bodies in motion, as people walking single-file. Finally, in all of these languages locative categories have the structure generic locative + locative noun, where *i* and *di* carry the general glosses 'in, at, on'.

Certain locative expressions have undergone semantic changes that are at first surprising Blust (1997a). In particular, PMP *dalem 'inside; deep' and *babaw 'above, on top' became Proto Polynesian *ralo 'below' and *fafo 'outside'. These changes make little sense in relation to the field of reference that guides the use of such terms in Indo-European languages, namely a three-dimensional bounded space such as a house or box. However, they do make sense with reference to a two-dimensional planar surface such as the surface of the sea or ground, where 'inside' = 'under', and 'above' = 'outside'. Observations such as this suggest that although terms like *dalem or *babaw may be glossed with fairly exact English equivalents, the contexts of use determined a different set of semantic associations (hence semantic changes) in AN languages. Ultimately these differences appear to correlate with differences in the physical environment: for people who spend much of their time on or near the sea and consequently use many locative

---

[50] The neutral term 'locative expression' is used here, since words such as Malay *di-atas* contain a preposition plus a noun, yet the entire expression is preposed to another noun, and used like a morphologically complex preposition.

expressions that relate to it, 'inside' and 'under' or 'above' and 'outside' are easily interchangeable notions, whereas for land-oriented populations this is not the case.

A second feature of AN locative expressions that at first appears similar to the translation equivalents in English but which closer attention shows to be fundamentally different, is seen in words composed of generic locative + common noun. While the locative expressions just discussed contain a noun which specifies relative location (above, below, etc.), prepositional phrases that contain non-locative nouns are often fossilised as the simple noun. This appears to be most common in the words for 'forest' (PAN *Salas, PMP *halas) and 'sea' (PAN *tenem, PMP *tasik). Blust (1989a) called cases such as these instances of an 'adhesive locative':

Table 5.19 The 'adhesive locative' in words for 'forest' and 'sea'

	Forest	**Sea**
Pazeh	rizik binayu (**loc** forest/mountain)	awas
Cebuano	lasáŋ, ihálas (**loc** forest)	dágat
Maranao	daləm a kayo (**loc lig** tree)	ragat
Bisaya	manalam talon (**loc** forest)	tasik
Kelabit	ləm kura (**loc** primary forest)	baŋət
Dali'	alam asau (**loc** tree)	laud
Murik	taləm uru? (**loc** grass/primary forest)	baŋət
Maloh Kalis	ləm tuan (**loc** primary forest)	roŋ jawa
Masiwang	ai lalan (tree **loc**)	tasi
Soboyo	kayu liañ (tree **loc**)	gehe
Buli	ai lolo (tree **loc**)	olat
Chamorro	halom tano? (**loc** soil, ground)	tasi
Lindrou	lo-key (**loc** tree)	dras
Puluwat	leewal (**loc** forest)	lehet (**loc** sea)
Pohnpeian	nanwel (**loc** forest)	nansed (**loc** sea)
Marshallese	buḷōn mar (**loc** bush)	lǫjet (**loc** sea)

A few languages in Melanesia show equivalent fossilisations of locative markers in the words for 'men's house', and two others have an expression for 'sky, heaven' which yields a similar analysis. In most languages the nouns meaning 'forest' and 'sea' may occur without a locative marker, but the preposition and noun are so closely associated that the equivalent of a prepositional phrase is given as the common expression for the independent noun. In other cases, as with Cebuano *ihálas* or Marshallese *lǫjet*, the form is historically bimorphemic but synchronically unanalyzable. This phenomenon suggests a different conceptualisation of locative relationships than is typical of Indo-European languages such as English. One lead to explaining this behavior is the use of generic locative marker + specifier noun in expressions such as 'above', 'below', and the like. Given the pervasiveness of this type of construction it is likely that expressions like *i babaw (lit. at + upper surface) 'above' served as a model for *i Salas (lit. at + forest) = 'forest', even though the resulting words belong to different grammatical categories. However, why 'forest' should be preferred for this type of reanalysis rather than e.g. 'house' or 'cultivated field' is not clear.

Finally, demonstrative pronouns in many AN languages have both spatial and temporal reference. In general this is not true of locative expressions, but there is one notable exception. In a number of AN languages the periphrastic expression of future time (i.e. the

marking of future by use of independent words rather than verbal affixation) is expressed by a word that literally means 'back, behind'. This is seen, for example, in the Malay expression *di-kəmudi-an hari* 'at a future day/time', where *kəmudi* means 'rudder of a boat', and *kəmudi-an* means 'position behind or after; subsequently; astern' (Wilkinson 1959:553). Similar usages appear in other languages, as in Kwaio (southeast Solomons) *buli* 'after, behind' : *buli ʔai* 'at a later time' (cf. *foolu* 'young, new' : *foolu ʔai* 'at the same time, right away'). From an Indo-European perspective, where the future clearly lies ahead, such constructions pose a 'back to the future' paradox: why should future time be indicated by a word that means 'behind'? As with many 'paradoxes', this one is generated by implicit assumptions of the observer. To an English speaker the future is 'ahead' because the observer is moving into it: we approach the future, the future does not approach us. However, from an AN perspective, where verbs often focus on the undergoer rather than the agent, it is the future that approaches the observer. This is seen in expressions like Cebuano *ulahi* 'last in a group to do something; late; later, not in the early days of one's life', where events near the end of one's life are lexically encoded like the last member of a troupe or file. Speakers of a language like English find this conceptual structure easier to process if the troupe or file is passing laterally across the observer's field of vision. When it is coming directly toward the observer, however, the concept may appear contradictory, since everything that lies ahead of the observer is in the 'future', while from an AN perspective what is immediately ahead of the observer is present and what lies behind it (and not behind the observer) is in the future.

The system of macro-orientation in AN languages is fundamentally different from that in languages such as English. Although a number of AN languages today make use of the cardinal directions with a four-point or even an eight-point system (Adelaar 1997b), this is not the inherited system, nor the one which is most widely represented in attested languages. Without a doubt the most general principle of macro-orientation in AN languages is the land-sea axis, associated with the PAN terms *daya 'toward the interior' and *lahud 'toward the sea'. Either with reflexes of these terms or with lexical innovations that represent the same semantic distinction, the land-sea axis is part of the directional terminology of languages from Taiwan to Polynesia:

Table 5.20 Lexical expressions of the land-sea axis in selected Austronesian languages

	Inland	Seaward
PAN	*daya	*lahud
Pazeh	daya 'upstream; east'	rahut 'downstream; west'
Thao	saya 'uphill, upstream'	raus 'downhill, downstream'
Paiwan	zaya 'upland, upriver'	lauz 'seaward, downriver'
Ilokano	dáya 'east'	láud 'west'
Cebuano	iláya 'away from coast/town'	iláwud 'near coast/town'
Mansaka	saka 'ascend; go upstream'	lawud 'downstream, seaward'
Kelabit	dayəh 'upstream'	laʔud 'downstream'
Malay	hulu 'upstream'	hilir 'downstream'
Kambera	dia 'upstream'	lauru 'downstream'
Paulohi	lia 'landward'	lau 'seaward'
Manam	auta 'landward'	ilau 'seaward'
Lakalai	-ilo 'landward'	-lau 'seaward'
Pohnpeian	peiloŋ- 'landward'	peiei- 'seaward'
Hawaiian	mauka 'landward'	makai 'seaward'

On larger landmasses where the sea is not in view, or even part of common experience, the land/sea axis is expressible as an upriver/downriver or uphill/downhill distinction. In extreme cases of isolation from the sea or habitation at high elevations the inherited forms that expressed this distinction may undergo unusual semantic changes. In a some of the Central Cordilleran languages of northern Luzon, for example, including at least Bontok, Kankanaey and Ifugaw, the reflex of *daya means 'sky; heaven'; having attained the highest elevations around the headwaters of the river systems on which they lived there was no further earthly 'upstream' for these groups, and the meaning of *daya was thus displaced to the heavens. On smaller landmasses a single distinction may exist in the minds of speakers, but be split into multiple translation equivalents by Western observers. On the relatively small island of Bali, for example, *kaya* (< *ka-daya is sometimes glossed 'south (in north Bali)', and 'north (in south Bali)', while *kalod* (< *ka-lahud) is glossed 'north (in south Bali)', and 'south (in north Bali)'. The actual meaning of these terms is 'toward the interior' and 'toward the sea', and their apparent polysemy is due to the imposition of a Western conceptual structure. Similarly, among the Pazeh, who live on the western flank of the Central Mountains in Taiwan, or the Ilokano, who live in the coastal strip between the west coast of northern Luzon and the Cordillera in the Philippines, upstream and downstream correspond to 'east' and 'west' respectively, and such cases could easily be multiplied. In some other languages, however, reflexes of *daya and *lahud appear to have evolved true cardinal direction senses, as in Chamorro, where *haya* (< *daya) means 'south' on Guam and Rota, but 'east' on Saipan, and *lagu* (< *lahud) means 'north' on Guam and Rota, but 'west' on Saipan. Unlike the situation on islands like Bali or Lombok, then, the meaning of these terms is constant on any given island, but differs between islands, implying that the current senses took shape on the north of Guam and Rota, but on the west of Saipan. Other features of Table 5.20 that merit a brief comment are: 1) in Cebuano the generic locative marker *i-* has become phonologically attached and fossilised in reflexes of both *daya and *lahud; semantically these terms have further evolved senses of 'near town' and 'away from town' that derive from the coastal settlement pattern of most Cebuano speakers, and the common lowland Filipino association of rural backwardness with the mountain populations; 2) Malay *hulu* and *hilir* are based on terms for 'head/headwaters' and 'flow' (which must naturally be downward); 3) in most Oceanic languages PMP *daya has been replaced by a reflex of *qutan 'scrubland, bush'.

The second major principle of macro-orientation in AN languages is confined to the monsoon regime of insular Southeast Asia and western Melanesia. In this area the prevailing winds that blow with the seasonal rains are critical to voyaging. As a result of this relationship between weather and voyaging, terms for the west and east monsoons form the second axis of macro-orientation (Ross 1995b). Moreover, because of their inherent association with approximations of the cardinal directions, these terms have more commonly taken on cardinal senses than have reflexes of *daya and *lahud:

East of New Guinea the monsoon regime fades out, but reflexes of PAN *SabaRat and *timuR persist with changes of meaning, as in Fijian *cavā*, Tongan, Samoan *afā*, Anutan *apaa*, Rennellese *ahaa* ''hurricane, gale, storm', and Samoan *timu* 'be rainy', Futunan *timu* 'blast of strong wind', Rennellese *timu* 'to rile, devastate, as by wind and storm', Anuta *timu* 'rain lightly, drizzle'. In some languages an eight-point compass system is built up in part by combining terms from the land-sea and monsoon axes, as with Malay *utara* 'N', *timur-laut* 'NE', *timur* 'E', *təŋgara* 'SE', *səlatan* 'S', *barat daya* 'SW', *barat* 'W' : *barat-laut* 'NW'. It seems clear that the land-sea axis was traditionally more important on land, and the monsoon axis at sea, although the latter was also relevant to the agricultural cycle.

In addition to these widely shared coordinates in the directional systems of many AN languages there are particular sets of coordinates or combinations of coordinates that appear to characterise a single language or small set of languages. In Manam, spoken on the small islands of Manam and Boesa off the north coast of New Guinea, the land-sea axis (manifested by *auta* 'inland', *ilau* 'seaward') is central, but a secondary axis based on it has also developed, as seen in the terms *ata* 'to one's right when facing the sea; to one's left when facing inland', and its complement *awa* 'to one's left when facing the sea; to one's right when facing inland' (Lichtenberk 1983:572ff). Barnes (1974:87ff) has argued that on the island of Kédang in eastern Indonesia the notions of up and down, right and left, and north and south are inextricably bound to the same lexical expressions, and as a result terms that may refer to the cardinal directions are commonly used to refer to objects in the immediate physical environment. In such cases the systems of micro-orientation and macro-orientation appear to interpenetrate in surprising ways, although it is unlikely that speakers of Kédang think of the orientation of objects a few feet from them in terms of the cardinal directions. To choose just one of many more possible examples of directional systems in AN languages that differ in fundamental respects from those of English or most other European languages, François (2003a) has described a complex system of spatial directions used on the island of Mwotlap in the Banks Islands of northern Vanuatu which appeals to three sets of coordinates: 1) personal coordinates (toward/away from speaker), 2) local coordinates (up/down, in/out), and 3) what he calls 'geocentric' coordinates (landward/seaward, and east/west); the east/west coordinates happen to correspond to a direction parallel to the shore, but are true cardinal points.

Table 5.21 Lexical expressions of the monsoon axis in Austronesian languages

	West monsoon	**East monsoon**
PAN	*SabaRat	*timuR
Kavalan	balat 'east wind'	timuR 'south wind'
Amis	safalat 'south wind'	timol 'south'
Tagalog	habágat 'W or SW wind'	tímog 'south; south wind'
Bikol	habágat 'south wind'	tímog 'E or SE wind'
Cebuano	habágat 'SW monsoon'	tímug 'east monsoon'
Maranao	abagat 'SW monsoon'	otaraʔ 'NE monsoon'[51]
Malay	barat 'west; west wind'	timur 'east; east wind'
Sasak	barat 'storm'	timuʔ 'east'
Kambera	waratu 'west, west wind'	timiru 'east, east wind'
Lamaholot	aŋĩ warat 'west monsoon'	aŋĩ timũ 'east monsoon'
Wetan	wartayani 'west monsoon'	tipriyani 'east monsoon'
Buli	pāt 'west monsoon'	morla 'east monsoon'
Numfor	wām-barek 'west monsoon'	yampasi 'east monsoon'
Motu	lahara 'NW wind and season'	mairiveina 'east wind'

As noted already, the comparative evidence suggests that true cardinal direction terms in AN languages invariably have evolved from other senses. Some of the sources of these terms have been shown in tables 5.20 and 5.21. In addition, PAN *qamiS 'north wind' has given rise to terms for 'north' in some languages of the Philippines and northern Indonesia. Other sources of cardinal direction terms are words that mean 'place of sunrise' and 'place

---

[51] From Malay '(ultimately Sanskrit) *utara* 'north'.

of sunset' for the east/west axis, or occasionally words that are equivalent to 'upriver/downriver' in terms of local orientation. Many Philippine examples of the first type can be found in Reid (1971) under the annotations for 'east' and 'west', and similar sources lie behind POC *sake 'east' (< ascend, rise), and *sipo 'west' (< descend) as noted by Ross 1995b). For the second type, Pangasinan, spoken on the plains west of the mountains of north-central Luzon has *bokíg* 'east; eastern section of town or province' < *bukij 'mountains, forested inland mountain areas'. Finally, Brown (1983) has shown how cardinal direction terms have evolved from the meanings 'up' and 'down' in many languages globally. Among AN languages this type of development can be seen in Chamorro *papaʔ* 'down, below, bottom, southward', Komodo *wawa* 'under, below; west', Kambera *wawa* 'beneath; west', Fordat *vava* 'beneath; south', Kei *wāv* 'beneath; north', Lonwolwol *fa* 'under, below; deep; far out at sea; north' (< *babaq 'below'), and Niue *lalo* 'below, under; bottom; west', Rarotongan *raro* 'beneath, under; leeward; west', Maori *raro* 'bottom, underside; down, under; underworld; north, north wind', Hawaiian *lalo* 'down; low; under; leeward; southern' (< *dalem 'inside'). The latter term has a rather complex semantic history, as it began with the meaning 'inside; deep', evolved the sense 'beneath, below', and then developed associations with various cardinal directions.

In addition to the foregoing uses of directional terms a number of Oceanic languages use directional auxiliaries with a main verb to indicate movement toward or away from the speaker. Because these elements have arisen from full verbs and so exhibit a history reminiscent of that of serial verbs they will be treated in Chapter 7.

5.5   Pronouns

Some aspects of the social parameters that determine pronoun usage were discussed in Chapter 3, and the morphosyntactic properties of AN pronouns will be treated in Chapter 7. Our major concern here is with the semantic features that define pronoun sets in typical AN languages rather than with their syntactic behavior or their value as markers of self in relation to others. The principal parameters that require discussion are inclusion/exclusion, number, and gender.

Table 5.22 Proto Austronesian and Proto Malayo-Polynesian personal pronouns

	PAN	PMP
Sg.		
1	i-aku	i-aku
2	i-Su, i-kaSu	i-kahu
3	si-ia	si-ia
Pl.		
1in	i-(k)ita	i-(k)ita
ex	i-(k)ami	i-(k)ami
2	i-kamu	i-kamu, ihu
3	si-ida	si-ida

Two sets of pronouns that can be called the 'long set' and the 'short set' were reconstructed for PAN and PMP in Blust (1977a). This reconstruction was revised and considerably expanded by Ross (2006), but for expository convenience it is easiest to begin

a discussion of the history of pronouns in AN languages with the earlier forms. These are shown in Table 5.22.

The first thing to note about these forms is that they are bimorphemic: each word consists of a pronominal base plus a nominative case marker (*si in the third person, *i elsewhere) that was syntactically distinct from the pronoun, but phonologically attached to it, and hence often fossilized on the pronoun in historical change. Reflexes of both nominative case markers, often called 'personal articles', also occur with personal nouns or with words that function as personal nouns in languages reaching from Taiwan to the Pacific. In many daughter languages these case markers have been lost, either in general, or before pronouns, leaving the bare pronominal stem. In others they continue to function actively, and in still others they have become fossilised randomly, affecting one or another pronoun in one language, but different pronouns in other languages. For this reason Dempwolff (1934-1938) reconstructed *ia '3sg', and *sida '3p', although many languages reflect a form with *s- in the singular (Atayal *hia*, Yami *sia*, Tagalog *siyá*, Pamona *sia*, Bimanese *sia*) and others reflect a form without *s- in the plural (Pangasinan *irá*, Subanen *ila-n*, Limbang Bisaya, Dohoi *iro*, Kenyah *ida*, Long Labid Penan *irəh*).

The second thing to note about these forms is that they show an inclusive/exclusive distinction in the first person plural, which is nearly universal in AN languages. Table 5.23 illustrates this distinction in the 'long form' pronouns for a representative set of languages reaching from Taiwan to Polynesia:

Table 5.23 The first person inclusive/exclusive distinction in non-singular pronouns

	Inclusive	**Exclusive**
Pazeh	ita	yami
Thao	itan	yamin
Amis	kita	kami
Ilokano	(da)tayó	(da)kamí
Bikol	kitá	kamí
Tiruray	(be)tom	(be)gey
Kelabit	tauh	kamih
Malagasy	isika	izahay
Jarai	ɓiŋ ta	ɓiŋ gəməy
Malay	kita	kami
Toba Batak	híta	hámi
Muna	intaidi-imu	insaidi
Makasarese	(i)katte	(i)kambe
Tetun	ita	ami
Taba	tit	am
Palauan	kid	kəmam
Chamorro	hita	hami
Manam	ʔita	ʔeʔa
Anejom	akaja	ajama
Woleaian	giish	gaamam
Tongan	kitautolu	kimautolu
Hawaiian	kākou	mākou

Although very few AN languages lack this contrast, most that do have collapsed the pronominal number distinction. Javanese, for example, has pronouns for first, second, and

third person which differ in speech register, but not in number, as in the Ngoko (low register) forms *aku* 'I/we', *kowe* 'you/ye', and *dhɛwɛke* 'he, she, it/they' (Robson 2002). Uhlenbeck (1960) cautions that in actual usage *aku* designates only the speaker unless a plural meaning is unambiguous from the context; otherwise the pronominal stem must be disambiguated by adding *kabɛh* 'all; every one', *padha* 'alike, the same', *kətəlu* (< Ngoko *təlu* 'three'), or the like. Robson (2002:25) notes that *awake dhewe* (Ngoko) and *kita* (Ngoko/Krama) are also used in the sense of 'we', although the first of these is a circumlocution, and the second is a borrowing from Indonesian.[52]   In either case no consistent use of an inclusive/exclusive distinction appears to be present in the Javanese pronominal system. Stevens (1968:207) observes a similarly anomalous state of affairs in Madurese, where the first person singular can be expressed as *sinkuʔ*, *bula*, *kaúla*, or *bhadhan kaúla* depending upon speech level (coarse, ordinary, refined, etc.). As though this use of personal deixis were not sufficiently qualified already, first person pronouns may also be replaced by proper names or titles, or the non-personal deictic marker 'this', and the literal translation of *bhadhan kaúla* is a circumlocution meaning 'my body'. Similarly, one of the second person forms apparently derives from a verb 'to stand', and names, titles, or titles plus names may also be used for the second person singular. Madurese has no third person pronoun, but uses nominal substitutes such as the name or title, title plus name of the referent, or the expression 'his body'. There reportedly are no native plural pronouns. The function of plural pronouns can be expressed in Madurese in either of two ways: 1) by combining native singular pronouns (*sinkuʔ ban baʔna kabbhi* 'I and you all' = 1p in, *sinkuʔ kabbhi* 'I all' = 1p ex, *abaʔ-na kabbhi* 'he all' = 'they, them', or 2) through the use of loanwords from Indonesian (*kita*, *kami*, *məreka*). These pronoun systems show extreme distortion of the AN norm. It can hardly be accidental that they have arisen in societies that pay minute attention to rank and social etiquette, and it seems safe to conclude that they have been conditioned by a social context in which hyper-sensitivity to social differences has made the inherited AN pronouns unworkable. The result is quite remarkable: although it can be argued that Javanese is used in the most complex of all AN-speaking societies, it has denotationally one of the most impoverished pronoun systems (the Madurese system can be seen as a reflection of intensive Javanese influence over many centuries). The elimination of a singular/plural distinction, the use of non-personal deictics as personal pronouns, and the use of circumlocutions such as 'your body' or 'his body' for pronominal reference all share a common function, namely to de-individuate personal deixis and so create a system of what might be called 'insinuative reference' rather than one of determinative reference. The major mechanism in achieving this result probably was the collapse of the number distinction, and once this happened an inclusive/exclusive distinction in the first person pronouns was no longer possible.

In Tukang Besi of southeast Sulawesi (Donohue 1999:114) the 7-way contrast in the PAN pronouns has been preserved, but the inclusive/exclusive distinction has been transformed into a distinction of number: *ikami* 'we (paucal)' vs. *ikita* 'we (plural)'. The logic of this innovation is not entirely clear, but since inclusive pronouns must include at least one human referent more than the exclusive equivalent, this implied difference of quantity may have been abstracted out and used to replace the original distinction.

In addition to a singular/plural distinction, the two Polynesian examples in Table 5.23 contain elements that are identical to or very similar to the number 'three' (Tongan *tolu*, Hawaiian *kolu*). Many Oceanic languages have not only singular and plural numbers, but

---

[52]   The first of these pronouns is said to mean literally 'ourselves, oneself', but despite the problem of person agreement, *awak-é* seems clearly to be *awak* 'body' + *-é* '3sg possessor'.

also a dual number in the pronouns, and sometimes a paucal ('3-10') form derived from a historical trial. Table 5.24 shows the subject pronouns of two Oceanic languages with a three-number pronominal system and two with a four-number system:

Table 5.24 Three-number and four-number pronominal systems in Oceanic languages

	Kilivila	**Hawaiian**	**Manam**	**Fijian**
Sg				
1	yegu	(w)au/aʔu	ŋa(u)	au
2	yokʷa/yoku	ʔoe	ʔai(ʔo)	o
3	mtona/minana	ia	ŋai	e
Dl				
1 in	yakida	kā-ua	ʔita-ru	(e)daru
ex	yakama	mā-ua	ʔe-ru	keirau
2		ʔo-lua	ʔan-ru	(o)drau
3		lā-ua	di-a-ru	(e)rau
Pc				
1 in			ʔita-to	((e)da)tou
ex			ʔe-to	keitou
2			ʔan-to	(o)dou
3			di-a-to	(e)ratou
Pl				
1 in	yakidasi	kā-kou	ʔita	(e)da
ex	yakamesi	mā-kou	ʔeʔa	keimami
2	yokʷami	ʔou-kou	ʔaŋ/ʔaʔamiŋ	(o)nī
3	mtosina/minasina	lā-kou	di	(e)ra

Structurally similar systems are widespread in Oceanic languages, and it can safely be inferred that Proto Oceanic distinguished at least singular, dual and plural pronominal numbers. In most languages the dual is derived by adding the numeral 'two', and the trial/paucal by adding the numeral 'three' to the pronominal stem (Hawaiian *lua*, Manam, Fijian *rua* 'two', Hawaiian *kolu* (> *kou*), Manam *toli*, Fijian *tolu* 'three'). In many languages this numeral is irregularly shortened, either by loss of a phoneme (Hawaiian and Fijian), or by loss of a syllable (Manam). Occasionally, idiosyncratic variations appear in pronominal number markers, as with Hawaiian *ʔo-lua* (not **ʔo-ua), or Fijian *-ru ~ -rau* in the dual, and *-tou ~ -dou* in the trial forms of the pronouns. Three other features observable in Table 5.24 merit brief notice. First, Senft (1986:46ff) describes Kilivila as having a dual number only in the first person, a typological pattern that is attested in many Philippine languages (Liao 2008, Reid 2009), and in other language families (Ingram 1978:243), but not in other Oceanic languages.[53]   Second, the plural pronouns of Kilivila are unusual in that the first person forms are derived by suffixing *-si* to the corresponding duals (in most languages duals are formed by suffixing plurals), while the second and third person forms are derived by either suffixing or infixing the same element to the corresponding singulars. Third, like other Polynesian languages, the Hawaiian plural set was historically a trial/paucal. Discrepancies between morphology and function such as this suggest that the

---

[53]   Dual pronouns in Philippine languages occur only for the 1st person inclusive, the inclusive/exclusive distinction being confined to the plurals. Kilivila differs from this in marking the inclusive/exclusive distinction for both duals and plurals.

Polynesian languages once had a four-number pronoun system, and that the trial/paucal set came to be used so much more frequently than the plural set that it eventually replaced the unmarked plural pronouns.

Although South Halmahera languages apparently have only a singular/plural distinction (Maan 1951, Bowden 2001), several languages in the West New Guinea branch of South Halmahera-West New Guinea have either a dual (Numfor-Biak, Kurudu, Windesi), or both a dual and trial number (Waropen, Ambai) in the personal pronouns (Anceaux 1961:150ff). With a minor exception to be noted below, the only other AN languages known to recognise more than a singular/plural pronominal number distinction are found in central and western Borneo. Almost all languages in this area have a dual, many have a trial, and some Kenyah languages go to the rare extreme of distinguishing singular, dual, trial, quadral and plural numbers morphologically in addition to having an inclusive/exclusive distinction. Table 5.25 shows the long form pronouns of four languages of central and western Borneo:

Table 5.25 Number marking in the long form pronouns of four Bornean languages

	Melanau (Mukah)	Kayan (Uma Juman)	Kelabit (Bario)	Kenyah (Long Anap)
Sg				
1	akəw	akuy	uih	akeʔ
2	kaʔaw	ikaʔ	iko	Ikoʔ
3	siən	hiaʔ	iəh	Ia
Dl				
1 in	tua	ituʔ	kitəh	tua
ex	mua	kawaʔ	kədiwəh	ameʔ dua
2	kədua	kuaʔ	məduəh	Ikəm dua
3	dua iən	dahuʔ	diwəh	ida dua
Tr				
1 in		təluʔ	təluh	təlu
ex		kaluʔ	kətəluh	ameʔ təlu
2		kəluʔ	mətəluh	ikəm təlu
3		dəhaluʔ	dətəluh	ida təlu
Qd				
1 in				təpat
ex				ameʔ pat
2				ikəm pat
3				ida pat
Pl				
1 in	tələw	itam	tauh	ilu
ex	mələw	kamiʔ	kamih	ameʔ dini
2	kələw	ikam	muyuh	ikəm muŋ
3	(də)ləw iən	dahaʔ	idəh	ida muŋ

The relevant numerals in these languages are MM *dua*, UJK *duaʔ*, BK *duəh*, LAK *dua* 'two', MM *tələw*, UJK *təluʔ*, BK *təluh*, LAK *təlu* 'three', *pat* 'four'. A comparison of pronominal number marking in Bornean and Oceanic languages shows both similarities and differences. First, in examples such as Mukah Melanau, Long Anap Kenyah *tua* (< *kita dua) '1dl in', Mukah Melanau *mua* (< *kami dua) '1dl ex', or Uma Juman Kayan

kuaʔ (< *kamu duaʔ) '2dl', the pronominal stem and numeral show greater phonological fusion than is typical of Oceanic languages, where bound numerals often irregularly lose consonants or even syllables, but do not fuse phonologically with the pronoun. Second, the plural pronouns of Mukah Melanau clearly align with the trial set in the other three languages. As in the Polynesian languages, then, there is morphological evidence that Mukah once had a four-number system that was reduced to a three-number system, presumably because the trial/paucal forms became the pragmatically unmarked plural set. Third, the first person plural inclusive in Mukah and the first person trial inclusive in the other languages is simply the number 'three' with no pronominal stem, a pattern that is all but unknown in Oceanic languages. Fourth, where a pronominal stem and numeral have fused, it is sometimes the first syllable of the stem (MM *kədua* < *kamu dua, BK *kətəluh* < *kami təlu), and sometimes the last syllable (MM *tua* < *kita dua, *mua* < *kami dua, BK *dətəluh* < *ida təlu) which forms the complex pronoun. Fifth, it is clear that the quadral pronouns of Kenyah are not nonce-forms, but are an integral part of the pronoun system. As such they represent a type of system that was unattested in the sample of 71 languages used by Ingram (1978), where the maximum number of distinctions, found in two languages, was fifteen. A close examination of details shows clearly that while the dual and trial/paucal pronominal numbers of West New Guinea and Oceanic languages may have a common historical origin, the similarity between these and the comparable phenomena in Bornean languages is a product of convergence.

One last complication must be mentioned in connection with pronominal number. Many languages in the Philippines have a dual pronoun that distinguishes 'I and thou' from 'I and you'. Of the 41 languages in Reid (1971) 30 exhibit this distinction in the long form pronouns, and 29 in the short form pronouns. In most cases the first person dual inclusive reflects PAN *kita, and its plural counterpart is formed by suffixing a second pronoun to this form, as shown in Table 5.26:

Table 5.26 Dual and plural inclusive pronouns in languages of the Philippines

	1st dual incl.	1st plural incl.
Central Cagayan Agta	ikitə	ikitam
Balangaw	díta	díta:w
Casiguran Dumagat	sikita	sikitam
Amganad Ifugaw	díta	dita??ú
Isneg	daʔta	daʔtada
Binongan Itneg	ditá	ditayó
Kalagan	kita	kitadun
Guinaang Kalinga	ditá	ditaʔó
Samal	kita	kitam
Botolan Sambal	hitá	hitámo
Palawan Batak	kita	kitami
Aborlan Tagbanwa	kita/ta	tami
Tausug	kitah	kitaniuh
Sangir	i kadua ini	i kiteʔ

A priori there seems to be solid evidence for a dual/plural distinction in the Proto Philippines first person inclusive pronoun, but certain observations weigh against this interpretation. First, the dual form in most cases reflects PAN *kita '1pl in'. If this is taken as evidence for modifying PAN (or PMP) *kita to a first dual inclusive there will be no

etymon for first plural inclusive, leading to a system that is both typologically implausible and unattested among extant languages. Second, the first person plural inclusive forms in Philippine languages have been formed in many different ways by a common process of suffixing some other (generally short form) pronoun to the reflex of *kita (Liao 2008). This suggests drift rather than common innovation, an inference that is further supported by the occurrence of *kitǝh* '1st dual inclusive' in Bario Kelabit, where it is embedded in a very different type of pronominal system. Third, no language is reported as having a first person dual *exclusive* pronoun. The most likely explanation of these special dual forms arises from the pragmatics of the speech act: most conversations take place between a speaker and a single hearer. As a result, the use of an inclusive pronoun would normally involve only the conversational dyad of speaker and hearer, whereas this would not necessarily hold for the exclusive form, since speakers commonly refer to themselves and *others* rather than a single *other*. Frequency of usage alone would lead reflexes of *kita to become de facto duals, creating a need for new plural inclusive forms, which were then cobbled together from the existing reflex of *kita plus parts of other pronouns (-*ihu*, *n-ihu*, -*m(u)* '2sg.', -*da* '3pl', etc.). In Sangir, a Philippine language spoken in northern Sulawesi, the reflex of *kita is 1st plural inclusive, and the corresponding dual pronoun has been formed by a transparent use of the numeral 'two' (*dua*). In Sangil, a dialect of the same language spoken in southern Mindanao the dual number has been extended throughout the paradigm, creating an eleven-term system, with singular, dual and plural forms of the first, second and third person, plus an inclusive/exclusive distinction for the dual as well as the plural. It has thus regularised an asymmetric system of number marking by extending the dual number to cover the same range of distinctions marked in the plural.

Finally, AN languages rarely mark pronominal gender. Where distinctions are made two possibilities appear to be realised: 1) animate/inanimate, and 2) masculine/feminine. Anceaux (1961:155) has drawn attention to an animate/inanimate distinction in the Windesi dialect of Wandamen, in western New Guinea. Since the AN languages of this area are poorly described, it is possible that this feature is much more widespread than this single example indicates. A masculine/feminine distinction in the third person singular full form pronoun is reported for Kilivila (Senft 1986:47). However, it does not exist in the possessive or emphatic pronouns. Both of these traits are found in many Papuan languages, which have pronominal gender, and make 'different distinctions in pronominal systems in different parts of the grammar' (Foley 1986:67). While Kilivila is not currently in contact with any Papuan language, the typological divergence of such traits from the AN norm and their similarity to those of many Papuan languages suggests sustained contact between pre-Kilivila speakers and speakers of some Papuan language or languages. At least two other typologically unusual pronominal systems in AN languages distinguish masculine and feminine gender. Austin (2000a:8) notes that in the Mataram-Selong dialect of Sasak, spoken on the island of Lombok just east of Bali, the low register second person distinguishes male and female addressee as *ante* (male), and *kamu* (female). The first of these is a lexical innovation, the second a reflex of PAN *kamu '2s'. In addition, Sellato (1981) has reported a 'three gender' system of personal pronouns used by several groups of nomadic Punan who inhabit the Muller-Schwaner Mountains of southeast Borneo. This is reproduced in Table 5.27 where (M) = male speaker, and (F) = female speaker.

Table 5.27 Three-gender pronoun system of the Muller-Schwaner Punan

		Seputan	Kereho	Nanga Ira'	Aoheng
(M)	he	ana	ana	ana	ana
	she	isɔ	sɔ	sɔʔ	hɔ
	it	hɔ	hɔ	hɔ	hɔ
(F)	he	isɔ	sɔ	sɔʔ	hɔ
	she	isɔ	sɔ	sɔʔ	hɔ
	it	hɔ	hɔ	hɔ	hɔ

These pronominal systems exploit animate/inanimate and masculine/feminine parameters, but in a very different way from AN languages in the New Guinea area. Although both male and female speakers of all languages except Aoheng consistently distinguish animate from inanimate, only male speakers distinguish male from female pronominal referents. These systems are thus sensitive to sex of speaker, or to a parameter of relative sex. In this respect they are similar to the sibling terminologies of many AN languages, although a parameter of relative sex in sibling terminology is almost always symmetrical (i.e. works for both male and female speakers).

5.6   Metaphor

As Lakoff and Johnson (1980) showed in a book that laid the foundations for cognitive linguistics, metaphor is not merely a literary device, but plays a central role in the ordinary use of language. Although most discussions of metaphor confine their topic to transparent uses of concrete imagery (argument is war, time is money, communication is sending, etc.), many words that have no imagistic basis in contemporary languages are etymologically metaphorical, as with English 'alleviate' (lit. 'lighten a load'), or 'subject' (lit. 'what is thrown down'). While these examples are relatively easy to categorise (to most English speakers words like 'alleviate' or 'subject' are imagistically opaque), not all languages present such a clear divide between historical and synchronic imagery, and this raises a fundamental question, namely where to draw the line between language that is genuinely metaphorical, and language that is not. In the following discussion 'metaphor' refers to any use of concrete imagery to convey abstract information. One of the principle theses advanced by Lakoff and Johnson (1980:7) is that "Because the metaphorical concept is systematic, the language we use to talk about that aspect of the concept is systematic." In effect, their claim is that the physical properties of a concrete model are exploited systematically to order more abstract sets of relationships, and that these relationships should therefore be consistent with one another.

Many AN languages are rich in metaphorical or quasi-metaphorical expressions of various types. At least in general form, most of these are shared with languages in other families, but some are unique or at least unusual. The following discussion will be confined to four specific sources of metaphorical expressions: 1) body parts, 2) kin terms, 3) plants, and 4) animals.

5.6.1 Body part terms and their extensions

The human body is a model for much that is conceptualised in the natural world. Some aspects of this model, as 'foot of the bed', 'headland', or 'give someone a hand' (either in

applause or assistance) involve transparent uses of metaphor. Other extended uses of body part terms draw on cognitive processes that are much less obvious (e.g. 'to foot the bill'). Although the relationship does not appear to have been explicity discussed in such earlier studies of body-part terminology such as Blacking (1977), Andersen (1978), or Matisoff (1978), metaphors based on external body parts such as the head, eyes or ears usually refer to the world of physical sense impressions, while those based on internal organs refer to qualities of temperament or character. Table 5.28 summarises metaphorical extensions of body-part terminology that are common in AN languages. Note that English body-part metaphors such as 'guts' or 'balls' (raw courage), 'heart' (courage in adversity; sympathy), or 'nerve' (impudence), are not known to occur in any AN language:

Table 5.28 Types of metaphorical extensions of body-part terms in AN languages

Body part	Metaphor
A.	**External body parts**
head	top/summit; chief/leader; handle (of knife, etc.); upper course of river; prow of a boat; first-born child; earlier, before.
ear	lug handle (of jar, etc.); bracket fungi
eye	sun; budding part of plant; spring of water; mesh of net; thread hole of needle; center of storm; etc.
nose	cape of land; prow of boat
navel	center
afterbirth	younger sibling
B.	**Internal body parts**
gall	common sense/judgement; courage
liver	sensibility/emotions

In many AN languages the morpheme meaning 'head' (often a reflex of PAN *qulu) is extended to other meanings, of which the most common are 1) top or summit, 2) chief, leader, 3) hilt or handle of a knife, axe, or paddle, 4) upper part (headwaters) of a river, 5) prow of a boat, 6) first-born child, and 7) earlier, before. Most of these are in need of no special comment, but the last two merit a brief discussion.

In several widely separated languages the first-born child is called by an expression that means 'head child', or that contains a reflex of *qulu 'head' which has undergone semantic change, as in Acehnese *aniʔ uleë bara* (child + head + ?), Bare'e *ana uyu-e* (child + first, but etymologically < *qulu 'head'), Rembong *anak ulu* (child + head), and Roti *ana ulu-k* (child + before/eldest < *qulu 'head'). The corresponding term for 'last-born child', however, is not 'tail child' (with an animal body serving as the metaphorical model) or 'foot child' (with a human body serving as the metaphorical model). Rather, it is formed from the word for 'child' plus a base that means 'lastborn', as in Acehnese *aniʔ buŋsu* (child + lastborn), or Rembong *anak sopo* (child + lastborn), or is formed from the word for 'child' plus a word having the more general meaning 'last, end; young' or the like, as with Bare'e *ana ka-supu-a* 'lastborn child' (< *ka-supu* 'end, completion'), or Roti *ana muli-k* (child + young). The Rotinese expression raises a comparative issue, since *muli-k* reflects PMP *(ma)-udehi 'last; come after or behind; late, later; future; stern of a boat' and 'youngest child'. The basic sense of this term in many languages is 'bring up the rear, be last in a procession'. Since it also refers to the stern of a boat, and the word for 'head' often

refers to the prow of a boat, this suggests that historically the metaphor for first- and lastborn children may have been that of the prow and stern of a boat.

Reflexes of *qulu 'head' usually refer to what has gone before only when they are preceded by a reflex of the locative marker *di: Iban *d-uluʔ* 'before, first; formerly, of old', Malay *da-hulu/dulu* 'before, in advance of, ahead', Sasak *j-uluʔ* 'passed by; do something before another', Uma *ri-ʔulu* 'the first'; *me-ri-ʔulu* 'precede', Makasarese *ri-olo* 'in front, go before; earlier', Roti *d-ulu* 'east'. Again, the metaphorical opposition is not with another body-part term, but generally with a reflex of *(ma)-udehi: Iban *udi* 'after, later', Malay *kə-mudi-an* 'later, after', Sasak *mudi/muri* 'after, later', Uma *muli* 'what follows or comes after', Roti *muli* 'west'. Although the reconstructed expression *di qulu uses the image of the (presumably human) head to indicate temporal precedence (it also means 'upstream' in many languages), the exploitation of a metaphorical model in this case appears unsystematic, since the opposed term for future time is not a body part, but rather a verbal expression meaning 'to bring up the rear, be last in a procession'. The metaphorical basis of the latter expression was explained in §5.4: the association of 'back, behind' with future time is not through an allusion to the human body, but rather through reference to the oncoming sequence of impinging events. The further extension of reflexes of PMP *qulu and *(ma)-udehi to cardinal direction terms, on the other hand, appeals to a metaphorical model of directional orientation in which the observer is facing east (the direction of beginnings and life).

Other metaphors that employ the word for 'head' include the unsurprising expression 'stone head' meaning 'stubborn', as in Malay *kəpala batu* (head + stone), or Rembong *ulu watu* (head + stone), and the more puzzling expression 'head of the knee'. In some languages the morpheme boundary in this expression has been lost, as with Malay *lutut* (< PMP *qulu tuhud) in comparison with e.g. Kavalan *tusuz*, Casiguran Dumagat *tod*, Tagalog *túhod*, Tiruray *ətur*, or Yakan *tuʔut* 'knee'. In others, however, it is clear, as with Sulu Samal *tuʔut* 'knee' : *kook tuʔut* (head + knee) 'front of the knee', or Long Terawan Berawan *ulo ləm* (head + knee) 'knee'. PMP *qulu tuhud apparently meant 'kneecap', although it has this meaning in too few languages, and is too poorly described in most dictionaries to warrant certainty.

In several widely separated AN languages words for 'ear' (generally < PAN *Caliŋa) also refer to projections such as lug handles on larger pots or similar wooden structures, as with Kanakanabu *ʔiŋa* 'ear; handle, lug', Ilokano *taliŋa* 'ear; handle', Iban *kəliŋa* 'ear (poetic); ear of a jar', Muna *poŋke* 'ear (on the head, also on pots, baskets, etc.)', or Hawaiian *pepeiao* 'ear; lugs or blocks inside a canoe hull to which the *ʔiako* (outrigger) booms are fastened'. The most striking extensions of the word for 'ear', however, are seen in the names for various fungi, under the literal names 'rat ears', 'tree ears', 'ghost ears' and 'thunder ears' (Blust 2000b). Whether these are properly metaphorical is open to question, since the shape of a bracket fungus is transparently ear-like, and is reflected in the names for similar fungi in English and other languages.

The largest and most intriguing set of external body part metaphors is seen with the word for 'eye', generally reflecting PAN *maCa. In many AN languages the word for 'eye' is used in the sense of 'center', 'nucleus', or 'most important part', much as it is in English expressions such as 'eye of a potato', or 'eye of a storm'. There have been two surveys of these extended uses of the morpheme meaning 'eye' in AN languages (Barnes 1977, Chowning 1996). Neither of these studies compares the AN data with data in other language families to show that many extended uses of the morpheme 'eye' are language universals, and so must be motivated by perceptual paths that are common to all humans

(Blust n.d. (e)). The use of 'eye' metaphors in AN languages can be illustrated with data from Malay, together with a few comments on other languages. Due to the number of such expressions the following list is selective:

Table 5.29 Extended uses of *mata* 'eye' in Malay metaphors

mata air ('eye of water')	spring of water
mata alamat ('eye of sign, portent')	bull's eye of target
mata aŋin ('eye of wind')	point of the compass, direction
mata bajak ('eye of plough')	plough-share
mata bantal ('eye of pillow')	pillow-end of stiff embroidery, etc.
mata bədil ('eye of firearm')	muzzle of firearm
mata bəlanak ('eye of mullet')	center of hairwhorl
mata bəlioŋ ('eye of adze/hatchet')	adze or hatchet blade
mata bənda ('eye of article')	valuables
mata bisul ('eye of boil')	head of a boil
mata buku ('eye of knot')	knot in wood; center of knot
mata daciŋ ('eye of scale')	marks on balance scale
mata hari ('eye of day')	sun
mata ikan ('eye of fish')	corn on foot; wart
mata jalan ('eye of path')	scout
mata jarum ('eye of needle')	eye of a needle
mata kaki ('eye of foot')	ankle
mata kəris ('eye of kris')	blade of a kris
mata kuliah ('eye of higher study')	subject of study
mata luka ('eye of wound')	orifice of a wound
mata panah ('eye of arrow')	point of an arrow
mata pədoman ('eye of compass')	compass needle
mata piano ('eye of a piano')	piano keys
mata pukat ('eye of net')	mesh of a net
mata susu ('eye of breast')	nipple of the breast
mata taŋga ('eye of ladder')	rung of ladder
mata uaŋ ('eye of money')	unit of monetary value; currency

Many of these uses are found in other AN languages, as with Wayan (Western Fijian) *mata ni caŋi* ('eye of wind') 'direction of the wind', *mata ni sā* ('eye of spear') 'point of a spear', *mata ni siŋa* ('eye of day/daylight') 'sun', or *mata ni wai* ('eye of water') 'spring, source of a river; rain clouds'. This body part differs from all others in representing a more abstract notion than that conveyed by the physical properties of the organ itself. One can argue that the sun, the nipple of the breast, or even a corn on the foot are 'eye-like', but this clearly is impossible with points of the compass, blades of cutting implements, or piano keys. A number of these expressions are fully cognate and permit PMP reconstructions, although the operation of universal cognitive processes in the formation of such expressions makes it difficult to completely rule out convergence.

Little needs to be said about the metaphorical extension of other external body parts. The morpheme for 'nose' is extended to capes of land and canoe prows in various parts of the Pacific, as in Nggela *ihu-* 'nose; beak; cape of land', Lau *isu-* 'prow and stern erections of a canoe', Rotuman *isu-* 'nose; projection, cape of land', or Maori *ihu-* 'nose; bow of a canoe'. The word for 'navel' is a metaphor for 'center' in widely separated languages, as in

Malay *pusat* 'navel; center', or in the famous Rapanui self-description as *te pito ʔo te henua* (**art** navel **gen art** land) 'the center of the world'. The last 'external' body part listed in Part A of Table 5.28 is exceptional in that it begins as an internal body part and becomes external with birth. In a number of widely separated AN-speaking societies the placenta is regarded symbolically as a newborn's younger sibling.

Unlike metaphors based on external body parts, metaphors based on internal body parts often refer to qualities of temperament or character. Two of these are of particular significance. The first is the word for 'gall' or 'bile', generally reflecting PAN *qapeju. In English 'gall' is a metaphor for arrogant boldness, as in 'He had the gall to call me a liar' (he being less than truthful himself). Here it signals a negatively evaluated personality trait, but in a number of AN languages gall is a metaphor for good sense or sound judgement, and is used in personal criticism only when negated, as with Bario Kelabit *naʔəm pədʰuh* (**neg** + gall) 'idiotic, inconsiderate, flouting custom or common sense', Karo Batak *la ər-pəgu* (**neg** have-gall) 'said of people: without substance, senseless, referring to speech and actions', Tontemboan *raʔica ni-aparu-an* (**neg** have-gall) 'silly, simple-minded; one who cannot grasp anything', Tae' *taeʔ paʔdu-nna* (**neg** gall-**3sg)** 'he has no understanding, he has poor judgement', Manggarai *atat toe maŋa pəsu-n* (person-who **neg** have gall-**3sg)** 'unthinking, oblivious to custom', or Kambera *tau nda niŋu kapídu-ŋu* (person **neg** have gall) 'a person who is thoughtless or inattentive'. Positive uses of the same metaphor are seen in Tiruray *fədəw* 'bile; feelings of an intellectual sort' (as opposed to visceral feelings); *fədəw-an* 'wise and sensible, doing what one feels is best without being directed', and Tontemboan *ni-aparu-an* 'having good understanding and self-control.' Since gall is seen as a repository of common sense and good judgement in many AN languages (possibly because of the widespread traditional practice of augury from examination of animal gall bladders and the consequent deliberation by experts in divination) other metaphors based on gall should be consistent with this model, and this generally appears to be the case, as with Tae' *bosi paʔdu-nna* (rotten gall-**3sg)** 'he is not shrewd, he doesn't know how to discuss or deliberate', or Yakan *pessaʔ peddu-nu* (broken gall-bladder-**2sg)**   'you forget things (said in scolding a person)'. Apart from this use a few AN languages also associate gall with courage, as seen in Thao *uka sa qazpu* (**neg** *sa* gall) 'cowardly', Cebuano *ispísu ug apdú* (thick *ug* gall) 'brave', or Marshallese *at* 'gall bladder; seat of brave emotions; seat of ambition'.

A second internal organ that is extraordinarily rich in metaphorical expressions is the liver. Although the heart is associated with the mind or character in some Formosan languages, as in Amis *falocoʔ* 'heart (physical); attitudes, personality, soul, character', in most AN languages, as in Southeast Asian languages generally, it is the liver rather than the heart which is the seat of the emotions. PMP had distinct terms for 'heart' (*pusuq), and 'liver' (*qatay). However, the word for heart was not an independent body-part term. Rather, it referred both to the heart as a body part and to the 'heart of banana', the pendent purplish bud at the end of the fruiting stalk which is widely used as a cooked vegetable (Merrill 1954:150). Attempts to elicit a word for 'heart' in fieldwork situations often produce vague results, and confusion with the word for 'liver', although elicitation of a term for 'heart of banana' is not problematic. What all of this points to is the cultural centrality of the liver among the internal organs of the body, a centrality reflected in many linguistic expressions, of which the following are merely representative:

1. **big liver** = bold: Iban *ati bəsai* (liver + big) 'boastful, talking big', Malay *bəsar hati* (big + liver) 'presumptuous', Sundanese *gədeʔ hate* (big + liver) 'brave; dare to do something', Tetun *ema ate-n boot* (person + liver-**3sg** + big) ' brave person', Rarotongan *ate nui* (liver + big) 'brave, stout-hearted; arrogant, cheeky, impudent', BUT Fijian *yate levu* (liver + big) 'coward'

2. **small liver** = afraid; resentful: Malay *kəcil hati* (small + liver) 'bear a grudge; feel afraid', Sundanese *lətik hate* (small + liver) 'afraid, alarmed', Ngadha *ate kədhi* (liver + small) 'careworn, worried; faint-hearted; envious, jealous', Kambera *màrahu eti* (small + liver) 'afraid, sad, sorrowful'

3. **burning liver** = angry: Malay *bakar hati* (burn + liver) 'angry emotion', Toba Batak *m-ohop ateate* (burn + liver) 'become angry', Lakalai *la-hate-la mamasi* (**art**-liver-**3sg** + burning) 'he is angry'

4. **rotten liver** = ill-will, full of malice: Malay *busok hati* (rotten + liver) 'ill-nature, malice; dirty feelings', Ngadha *ate zeʔe* (liver + rotten) 'take everything the wrong way, be quickly offended'

5. **sick/hurt liver** = hurt feelings, offended; angry: Malay *sakit hati* (hurt + liver) 'resentment, annoyance; anger, ill-will', Sundanese *ñəri hate* (hurt + liver) 'grief, sorrow, heartache', Komodo *bəti ate* (hurt + liver) 'angry', Kambera *hídu eti* (hurt + liver) 'heartsick, offended, hurt; disgruntled', Gitua *ate yayap* (liver hurt) 'angry, frustrated'

6. **white liver** = pure hearted: Mansaka *ma-potiʔ na atay* (white + **lig** + liver) 'kind heart, tender heart', Malay *puteh hati* (white + liver) 'sincere', Madurese *pote ate* (white + liver) 'upright, honest', Motu *ase kuro tau-na* (liver + white + man) 'a brave man, one who is not afraid'

7. **fetch the liver** = win someone's affection: Malay *ambil hati* 'loveable', Bahasa Indonesia *ambil hati* 'please someone so as to win their affection', Sundanese *ŋ-ala hate* 'win someone's affection', Madurese *ŋ-alaʔ ate* 'please someone', Ngadha *ala go ate* 'win someone's affection' (all lit. 'fetch the liver').

More puzzling are uses of 'liver' for other body-part terms, as with 8) <u>liver liver</u> = calf of the leg: Marshallese *aj* 'liver' : *ajaj* 'calf of the leg', Tongan *ʔate* 'liver' : *ʔate ʔi vaʔe* (liver of leg) 'calf of the leg', Maori *ate* 'liver' : *ateate* 'calf of the leg', or 9) <u>liver of hand/foot</u> = palm/sole: Tagalog *atáy* 'liver; arch of the sole of the foot', Cebuano *atay-átay* 'the hollow or fleshy part of the palm and its analogue in the foot', Malay *hati taŋan* (liver + hand) 'hollow of the palm of the hand', Javanese *ati* 'liver; fleshy part of the hand/foot at the base of the fingers/toes', Bimanese *ade edi* (liver + foot/leg) 'sole of the foot', *ade rima* (liver + hand) 'palm of the hand', Gedaged *nie-n ate-n* (foot-**3sg** + liver-**3sg**) 'the sole of his foot', *nima-n ate-n* (hand-**3sg** + liver-**3sg**) 'the palm of his hand'. It is possible that the calf of the leg is seen as resembling the liver in shape (for the same reason that words for 'lung' reflect PAN *qaCay* 'liver', with or without a modifier. The connection of the liver with the palm of the hand or sole of the foot is more elusive, but apparently has to do with the notion of an 'inner' part, and for much the same reason the word for 'liver' in some languages also refers to the pith of bamboo or other plants. All of these usages are found in languages that are spoken thousands of miles apart. Other metaphorical uses of the word for 'liver' have, so far as is presently known, a more restricted distribution in AN

languages, as with Lakalai *la hate-la raga* (**art** liver-**3sg** leap) 'he is startled', or Gitua *ate-ŋgu mutu* (liver-**1sg** broken) 'I am surprised'.

5.6.2 Kin terms and their extensions

Another terminological set used to form metaphors in many AN languages is the vocabulary of kinship. Extensions of two kinship terms will be considered here: 1) child, and 2) mother. In addition, extended uses of the more general terms 'male' and 'female' will be discussed briefly.

Words for 'child' or their morphological derivatives (generally reflecting PAN *aNak) are found in a number of purely kinship-related meanings. In addition to these meanings, however, the same word in many languages carries the sense of 'smaller part of a larger whole' (in relation to inanimate referents), or 'member of a group' (in relation to humans). In the first of these senses it is used in various expressions, including:

1. **child of fire** = sparks: Kayan *anak apuy* (child fire), Simalur *anaʔ axoe* (child fire) 'sparks of a fire'

2. **child of ladder** = rung or step of a ladder: Bintulu *anak kəjan* (child ladder), Malay *anak taŋga* (child ladder), Simalur *anaʔ aeran* (child ladder) 'rung of a runged ladder', Wolio *ana-na oda* (child-**3sg** ladder) 'step of a notched log ladder'

3. **child of mortar** = pestle: Sasak *anak lisuŋ* (child mortar), Wolio *ana na nosu* (child of mortar), Kei *luhun yana-n* (mortar child-**3sg**) 'pestle'

4. **child of eye** = pupil of the eye: Malagasy *anaka ndri maso* (child of eye), Malay *anak mata* (child eye), Toba Batak *anak ni mata* (child of eye), Bare'e *ana mata* (child eye) 'pupil of the eye'

5. **child of bow** = arrow: Malay *anak panah* (child bow), Nias *ono fana* (child bow), Wolio *ana na fana* (child of bow), Kambera *na ana-na pana* (**art** child-**3sg** bow) 'arrow'.

In the second sense it appears in:

6. **child of land/village** = 'fellow villager, fellow community member': Nias *ono mbanua* (child + village) 'common man, villager', Old Javanese *anak wanwa* (child + inhabited area) 'person belonging to the *wanwa*-community', or Ngadha *ana nua* (child + inhabited area/village) 'villager', and in expressions such as child of the male group = 'brother of a woman' and child of the female group = 'sister of a man' which require a more detailed discussion of cultural presuppositions in relation to hierarchically-ordered social groups that function in the regulation of marriage (Blust 1993d, and end of this chapter).

In most AN languages the word meaning 'mother' does not have as many types of extended meanings as 'child'. However, as in many other language families, this morpheme often represents the largest member of a set, prototypically the thumb or big toe (Brown and Witkowski 1981:601ff): Hanunóo *ʔinaʔ* 'mother' : *ʔinaʔiná* 'thumb or great toe', Malay *ibu jari* (mother finger), Lamaholot *lima inã* (mother of fingers) 'thumb', Soboyo *ina* 'mother; most prominent or largest member of a set', *kiki* 'finger' : *kini-n ina* ('mother of fingers') 'thumb', or Nggela *tina* 'mother; large, of its kind'. Less commonly the morpheme for 'father' or 'parent' may serve this metaphorical purpose, as with Tona Rukai *t-ama* 'father' : *tama-tamanə* 'thumb', Batad Ifugaw *ama* 'father, uncle' :

am-ʔama-ʔʔa 'thumb or big toe of a person or monkey', or Samoan *lima-matua* (parent of the fingers) 'thumb'. In a number of languages the word for 'mother' (of humans) is also similar to the word for 'female' (of animals), presumably because the distinction between 'female' and 'mother' is not as clearly drawn for animals as it is for humans.

The words for 'male' and 'female' involve a number of interesting complications in AN languages. Many languages distinguish these terms for humans and animals, as with Western Bukidnon Manobo *məʔama/bahi* 'male/female (humans)' : *lumansad/upa* 'male/female (fowls)', or Malay *(lə)laki/wanita, pərəmpuan* 'male/female (humans)' : *jantan/bətina* 'male/female (animals). Some languages make further distinctions among categories of animals, as Paiwan *uqaɬʸay/va-vai-an* 'male/female (humans)', but *vaɬʸas* 'male (animals)', *rukut* 'sow, female pig', *djiɬʸaq* 'doe, female deer', *djumu* 'female muntjac (barking) deer', *parimukaw* 'oldest female monkey in pack'. Although it is debatable whether this type of extension is metaphorical, the terms 'male' and 'female' are associated with the right and left hands in some AN languages, following a pattern found in many traditional societies that have systems of dual cosmological classification (Needham 1973). Given the cultural importance of these associations the words for 'male' (POC *maRuqane) and 'female' (POC *papine) have become polysemous for sex and handedness in some languages, as in Chuukese *mʷáán* 'male; right hand or side', *feefin* 'female; left hand or side', or Carolinian *mʷáál* 'man, male' : *peighi-mʷáál* 'right side', *schóóbʷut* 'female, woman' : *peighi-schóóbʷut* 'left side'. What is surprising is that the right : left opposition, which is generally regarded as the pivot for dual symbolic classifications, apparently is subordinate to the male : female opposition, since these forms show a development from male : female to right : left rather than the reverse.

5.6.3 Plants and people

Plants serve as general metaphors of genealogical relationship in English expressions such as 'family tree'. In AN languages they often carry more specific types of genealogical information. Two plant-human associations are particularly important. The first of these associates a young child or descendant with seedlings or plant shoots (or sometimes the reverse), as in Thao *qati* 'bamboo shoot; grandchild', Ilokano *sagibsíb* 'young shoot of a plant (taro, banana, etc.); illegitimate children', *túbo* 'shoot, sprout, bud; k<in>aag-tu-túbo 'youth, adolescence', Tagalog *usbóŋ* 'sprout, bud, growth; offspring', Malay/Indonesian *bibit* 'rice seedling; beginning student', Tae' *taruk* 'shoot of trees and plants; descendant' Hawaiian *keiki* 'child, offspring; shoot, as of taro'. Since almost all languages have a separate term for grandchild which does not mean 'plant shoot' it is clear that these usages are figurative. Because they are figurative they generally go unrecorded even in detailed dictionaries, and in some cases become apparent only through the work of cultural anthropologists or other scholars interested in oral literature (e.g. Fox 1971).

The second metaphor which links plants and people in many AN languages is PMP *puqun, a term that lacks a monomorphemic translation equivalent in English. The basic sense of this word is 'base of a tree', the part of a tree that emerges from the ground. In addition to this physical sense it has the extended meanings 'origin', 'beginning', 'cause', 'foundation' and 'reason' in languages over a wide geographical area. In some of these languages reflexes of *puqun also refer to people who have a special social prominence based either on kinship or on political authority. Examples of such extended meanings with reference to people are seen in Ilokano *puón* 'beginning; origin, source; base; root; trunk; lower part; parentage, ancestry; unit for counting trees; cause, reason', Tagalog *púnoʔ*

(with consonant metathesis) 'trunk of a tree; beginning; chief', Hanunóo *púʔun* 'base of plant stem, base, trunk; an officiating elder at a *panagdáhan* ritual (feast and ritual of propitiation to certain unseen spirits)', Masbatenyo *púnoʔ* 'tree trunk; head, chief, leader', Cebuano *púnuʔ* 'trunk of a tree; tree as a unit for numeration; base, lowest part of something; point of attachment of a part of the body; officials in charge of an office', *pan-púnuʔ* 'president, governor', Maranao *ponoʔ-an* 'tree trunk; mainline (ancestry)', Tiruray *fuʔun* 'beginning; root, base; ancestor', Palauan *uʔúl* 'base of tree; reason; cause; basis', *uʔəl-él* 'beginning, start, origin; ancestors', Ngadha *puu* 'base of a tree; beginning; foundation; origin; basis; true, genuine', *sao puu* (house origin) 'clan origin house', *puu taŋi* (origin ladder) 'clan ancestor', Tetun *hun* 'base, foot, bottom, the lower part of flank; source; trunk of any tree'. Although he does not list *hun* by itself with a human sense, Morris (1984:89) also gives *lalutuk ain hun* (pigsty tree-**3sg** base) 'people who live next to royalty and render service to them', which provides indirect evidence that *hun* refers to the upper classes. In some languages only the extended personal sense of this term has survived, cutting it off from the botanical metaphor in which it originated, as with Karo Batak *bəru puhun-na* (woman base-**3sg**) 'a man's 'true' wife: the mother's brother's daughter' (in a system of preferential matrilateral cross-cousin marriage)', or Bolaang Mongondow *punuʔ* 'title of rank; lord, prince'. In some Oceanic languages roughly the same set of semantic relationships is paired with an etymologically distinct base, as with Puluwat *pʷopʷulepán* 'base, starting point, base of a tree': *pʷopʷulepán aynaŋ hamʷol* 'original and leading chiefly clan', or Rarotongan *tumu* 'foundation, root, cause, origin, source; reason; trunk or main part of anything' : *tumu enua* (base land) 'the original land, the head or chief of the land, the leaders or people of note of the land, the nobility or aristocracy'. In short, the idea of emergence or origin that is associated with the base of a tree in many AN languages provides a model for notions not only of precedence in time or causal relationship, but also for the founders of genealogical units and the social rank accorded to them (Fox 1995, Blust and Trussel ongoing).

5.6.4 Animals and people

Where plant metaphors refer to human beings in AN languages they generally model genealogical or hierarchical relationships. By contrast, human-animal comparisons appear to model behavior, and metaphor, which appeals to an implicit model, often trails off into simile, which appeals to a more explicit comparison. One recurrent example of an animal image that is associated with a medical condition is seen in reflexes of PAN *babuy* 'pig' meaning 'epilepsy': Maranao *bəboy* 'pig' : *babo-baboy* 'epilepsy', Iban *babi* 'pig' : *gila babi* (lunacy pig), Malay *babi* 'pig' : *gila babi* (lunacy pig), *pitam babi* (vertigo pig), *sawan babi* (convulsion pig) 'epilepsy', Sangir *bawi* 'pig' : *sakiʔ u wawi* (sickness **gen** pig) 'epilepsy'. This association, which is also found in Sino-Tibetan languages, appears to derive from the victim's rolling on the ground during a convulsion, a behavior that evidently is compared with the wallowing of pigs. The Maranao forms, with their divergent vocalism, suggest that the historical association of the words for 'pig' and 'epilepsy' has been broken in this language.

As in many other cultures, man's closest animal companion is exploited for purposes of derogatory reference to humans. Wilkinson (1959:36-37) notes that in Malay 'proverbs about dogs are usually uncomplimentary,' but all of the examples he cites are in fact about humans, as with *anjiŋ dəŋan kuciŋ* (dog with cat) 'cat and dog life', or *baŋsa anjiŋ* (race dog) 'man of the dog-class, smelling of filth even when not eating it'. Similar examples are

found in many other languages, as with Old Javanese *asu* 'dog' : *aŋ-asu* 'low, vile' ('like a dog') , Cebuano *íruʔ* 'dog' : *íruʔ ŋa daug* (dog **lig** yield) 'become gluttonous; become a slave to an overpowering emotion (become oblivious to shame, like a dog)', or Ilokano *áso* 'dog' : *áso-áso* 'henchman of a politician'.

5.7   Language names and greetings

Two features of the lexicon that rarely receive attention will be addressed briefly here. The first of these is the origin of language names, a topic selected for its inherent interest. The second is the shape of formulaic greetings, a topic which is of basic practical importance.

5.7.1 Language names

The source of many AN language names is unknown, but those that are polysemous generally fall into a small set of categories which includes 1) language name = 'person, human being', 2) language name = proper name or descriptive term for location, 3) language name = negative identification, or just the negative marker. Cross-cutting these three categories is the autonym/exonym distinction, where autonyms are self-designations and exonyms are designations imposed from without.

In the first category, a language name means 'people, human beings', presumably a consequence of the manner in which speakers identified themselves to outsiders when asked who they were. Examples include Thao, Bunun, and Tsou in Taiwan, all of which derive from the common noun in these languages meaning 'person, human being', Iban, which derives from the common noun meaning 'person, layman (not shaman or master of myth/poet)', Nias (self-designation Niha) 'person; someone; pagan (non-Muslim)', and 'Āre'āre (< *ʔāre* 'thing, person). While each of these language names probably is an autonym, many of the languages spoken by Negrito populations in the Philippines are called by a reflex of PMP *qaRta 'outsiders, alien people. This category includes Agta, Alta, Arta, and Atta of northern Luzon, Ayta (sometimes written Aeta) of central Luzon; Inati of Panay Island and Inata of Negros Island in the central Philippines; and possibly Ata Manobo of Mindanao. To non-Negrito Filipinos the terms Agta, Ayta and the like mean 'Negrito'. Cognate terms in languages outside the Philippines range over the meanings 'outsider, alien person', 'slave', and 'person, human being' (Blust 1972). Somewhat surprisingly, this term has been expropriated as a self-designation by a number of these groups within historical times. Headland and Headland (1974:3), for example, note that the Casiguran Dumagat (or Casiguran Agta) in northeast Luzon use the word *ágta* to refer to a Negrito person (self-reference), and as a verb meaning 'to speak in the Dumagat language'.

In the second category a language is called by a place name, which may be either a proper noun, or a descriptive term designating relative location. Examples include Itbayaten, the name of an island north of Luzon in the Philippines, and the language spoken there; Ifugaw, probably from *i-* 'inhabitants of' + *pugáw* 'cosmic earth', hence 'people of the earth'; Kapampangan, from *ka- -an* + *paŋpaŋ* 'river bank', hence 'people of the riverbank'; Tagalog, from *taga-* 'hailing from' + *ílog* 'river', hence 'people of the river'; Mansaka, from *man-* + *saka* 'upstream', hence 'upstream people'; Mandaya, from *man-* + *daya* 'interior', hence 'people of the interior'; Maranao, from *ma-* 'stative' + *ranaw* 'lake', and Tondano *to-* 'people' + *dano* 'lake', hence in both cases 'people of the lake';

Tausug, from *tau* 'person' + sūg 'current', hence 'people of the current'; Lun Dayeh, from *ulun* 'people' + *dayəh* 'interior', and Toraja from *to-* 'people' + *raja* 'interior', hence 'people of the interior'; Toba Batak, from *toba* = lake margin (the Toba Batak living around Lake Toba), and Tetun, from *tetun* 'coastal plain', hence 'people of the coastal plain'. Some of these terms, especially those for interior languages, probably are exonyms or reactions to the views of lowlanders, since Lun Dayeh speakers in Sarawak often call themselves Lun Bawang 'people of the country', and Toraja is a generic term for highland people in Sulawesi, not a self-designation. In much the same way the name 'Tonga' (*toŋa* 'south') must have been given to the Tongans in prehistoric times by their neighbours to the north either in Fiji or Samoa (cf. Samoan *toŋa* 'south', a term not found in Fijian, although it may have been in the past).

In the third category a language is identified negatively or by the general negative marker, as with Dusun Deyah (Dusun **neg**) 'not Dusun' (presumably to distinguish themselves from the neighbouring Dusun Malang, Dusun Witu or some similar group). This type of language name is particularly common in Sulawesi, being found in e.g. Lauje, Uma, Tae' (also known as Southern Toraja), and Bare'e, all of which take their names from the negative marker 'no, not'.

Other ways of forming language names are also used in more restricted areas. Many Philippine languages, for example, use the infix *-in-* (prefixed to vowel-initial bases) to form language names from ethnic names, as with *ibaloy* 'Ibaloy person' : *in-ibaloy* 'the Ibaloy language', or *bisáya?* 'Bisayan person' : *b<in>isáya?* 'the Bisayan language'. In the Bismarck Archipelago language names are sometimes formed from the word for 'younger parallel sibling' or specifically from the **1sg** possessed form, as with Tigak of New Ireland *tiga-k*, or Nali of eastern Manus *nali*, both meaning 'my younger sibling'.

5.7.2 Greetings

Greetings constitute perhaps the most universal type of formulaic speech, and as such can be regarded as part of the lexicon of a language. The common greeting of encounters over much of the AN world is literally 'Where are you going?' The response may be 'Just strolling' or 'To X (place)'. A common variant is 'Where are you coming from?', in which case 'From X (place)' is the only appropriate answer. There are, however, many other variations, some of which seem almost certainly to be calqued on European or other models. A small sampling should suffice to illustrate:[54]

Thao: 1) *a mu-ntua ihu* (**fut** go-where **2sg)** 'where are you going?', 2) *k<m>in-an iza ihu* (eat-**av-perf** already **2sg)** 'Have you eaten yet?'. Only fifteen active Thao speakers are known, all but one born before 1937, and the daily language they use to interact with the larger community around them is Taiwanese (Minnan Chinese). As a result the greeting that is first offered during elicitation, and the one that is preferred today is *k<m>in-an iza ihu* 'Have you eaten yet?', a clear calque from Taiwanese.

Paiwan (southeast Taiwan): *pa-djavay* 'to do much work on land; to work much land', *djava-djavay* 'common greeting (implies one has a lot of bother)'.

Malagasy: *manao ahoana ianao* (do what you) 'What are you doing?'

Malay/Indonesian: 1) *(mau) kəmana* (want to-where) 'Where are you going?'; *jalan-jalan saja* (walk-walk just) 'Just strolling'; *ke-X* (to-X) 'I'm going to X', 2) *dari mana* (from where) 'Where are you coming from?'; *dari* X (from X) 'I'm coming from X',

---

[54]   It is worth noting that the common form of AN greetings has also been calqued in Tok Pisin as 'yu go we' ('Where are you going?') or 'yu stap we na yu kam?' ('Where are you coming from?').

3) *sudah mandi*? (already bathe) 'Have you bathed yet?', all equivalent to 'How are you?' and responses to this question.

Nguna: 1) *malip̃ogi wia* (morning good) 'Good morning!' (and similar greetings for other parts of the day). Possibly a product of Western contact, but this remains unclear.

Fijian: 1) *o(nī) lai vei* (you leave where) 'Where are you going?' (*o* [**sg**] is casual or familiar; *onī* is formal), 2) *bula* (live/life) 'Greetings!', 3) *bula vinaka* (live/life good), 4) (*sā*) *yadra* (already awaken) 'Good morning'.

Tongan: *ʔalu ki fee* (go to where) 'Where are you going?'; *ʔeva pee* (stroll just) 'Just strolling'.

Pohnpeian: 1) *kasele:lie/kaselel* (lit. 'most fine/perfect/precious') 'Hello!', 2) *ke pa:n ko:la ia* (you **irr** go-there where) 'Where are you going?', 3) *ke ko:saŋ ia* (you come-from where) 'Where did you come from?'.

Marshallese: 1) *lǫkʷe* 'love (equivalent to Hawaiian 'aloha'), 2) *lǫkʷe ia ṇe* (love where that-**2sg)** 'Aloha, where are you going?', 3) *kʷōj etal ṇan ia* (**2sg.prog** go to where) 'Where are you going?'. The first of these is said to be the most common greeting, and is often prefaced to the others. *lǫkʷe ia ṇe* would be used by those who have been on the same island all along, and is addressed to a single individual. *kʷōj etal ṇan ia*, which is less common, can be inflected for dual, trial or quadral number, but would normally be pluralised for numbers greater than two. For someone who has just come to your island the normal greeting would be 'When did you (sg) get here?' (Byron W. Bender, p.c.).

Even where greetings differ in form, the general rule is that they do not inquire about one's personal state, but rather about one's relationship to events preceding or to come. A number of Philippine languages have borrowed the Spanish 'Como esta' as as single morpheme *kumusta*, and in these the general rule appears to be broken, as with Ilokano *kumustá* 'interrogative used in asking condition of a person', *kumustaen* 'to greet; hail; inquire about someone's well-being', Tagalog *kumustá* 'How are you' : *kumustahín* 'to inquire about the condition of something or the health of someone', or Cebuano *kumustá* 'how is (was, are, etc.)', *paŋumustá* 'ask how someone is, send regards'.

5.8   Semantic change

Bloomfield (1933:426ff) summarised the work of earlier generations of researchers on semantic change by presenting a typology with nine major divisions: 1) *narrowing of meaning*: Old English *mete* 'food' > meat 'edible flesh', 2) *widening of meaning*: Middle English *bridde* 'nestling' > bird, 3) *metaphor*: Proto Germanic *bítraz 'biting' > bitter 'harsh of taste', 4) *metonymy* (meanings near to each other in space or time): Old English *cēace* 'jaw' > cheek, 5) *synecdoche* (meanings related as part to whole): Proto Germanic *tú:naz 'fence' > town, 6) *hyperbole* (from stronger to weaker meaning): pre-French *ex-tonāre 'strike with thunder' > French *étonner* 'astonish', 7) *litotes* (from weaker to stronger meaning): pre-English *kwálljan 'torment' > Old English *cwellan* 'kill', 8) *degeneration*: Old English *cnafa* 'boy, servant' > knave), 9) *elevation*: Old English *cniht* 'boy, servant' > knight). Much additional work has subsequently been done, especially within the framework of cognitive linguistics, but this still remains a useful starting point for discussions of semantic change.

5.8.1 Prototype/category interchange

Like most subsequent ones, Bloomfield's typology is Eurocentric, and although it has some carry-over value, it appears to offer both too much and too little to accommodate the data from AN languages. Moreover, some semantic innovations in AN languages suggest that distinctions such as 'widening' and 'narrowing' may be instances of the same type of change. To illustrate, PMP had *hulaR 'snake (generic)', and *sawa 'python', but reflexes of *sawa sometimes represent the broader category, as in Yakan *sawe*, Dampelas *saa*, Balaesang *ule saa*, Ratahan, Boano, Simalur, Bimanese *sawa* 'snake'. It would be easy to force this change into Bloomfield's typology as an example of widening of meaning, but to do so without further discussion would obscure the fact that these examples represent at least five historically independent semantic innovations (one in Yakan, another in Ratahan, a third Dampelas, Balaesang and possibly Boano, a fourth in Simalur, and a fifth in Bimanese). Recurrent semantic changes suggest that some aspect of perception or psychology has been active for generations, favoring one type of change as opposed to others that might have occurred but didn't. The reticulated python is the largest and psychologically most impressive snake in insular Southeast Asia, and the equation python = snake suggests that it was conceived as the prototype of the snake category, the snake par-excellence. Although concealed by a complex historical phonology, Palauan *ŋúis* 'green tree viper, the Palau tree snake: *Dendrelaphis lineolatus*' reflects *hulaR. Five snakes are found in Palau. Two of these are venomous but non-aggressive, with no history of human snakebite casualities from either. A third is well-camouflaged and rarely seen, and the fourth is the Brahminy blind snake, a tiny species that spends much of its time underground. Since the python is absent, it is a reasonable inference that the venomous green tree viper, which is described as 'a very fast, nervous snake' common in small trees and shrubs has taken its place as the most dangerous and hence psychologically most salient snake in Palau. In Bloomfield's terms this semantic change (which is unique) could be classified as an example of narrowing of meaning, but given the larger context both this and the preceding changes can be seen as sharing a common basis. Whether a change is from prototype to category (as with *sawa), or from category to prototype (as with *hulaR) the category/prototype boundary becomes blurred and eventually lost. In one case this appears to be an instance of semantic widening, and in the other an instance of narrowing, but in both cases a semantic change arguably occurs because a prototype comes to be identified with the entire category of which it is a part.

5.8.2 Change of physical environment

Another type of semantic change in AN languages that does not fit neatly into the categories that appear in most discussions is change of meaning triggered by change of physical environment. Speakers of Proto Malayo-Polynesian, probably located somewhere in the northern Philippines, had the following general category terms for animals (Blust 2002b):

Table 5.30 General category terms for animals in Proto Malayo-Polynesian

*qayam	domesticated animal
*manuk	domestic fowl, chicken
*manu-manuk	bird
*hikan	fish
*hulaR	snake

Surprisingly, there is no reconstructed term for 'animal', the closest equivalent being *qayam 'domesticated animal', a word that retained this meaning or the meaning 'pet' in several widely separated languages, but came to represent specific domesticated animals in others, as in Tagalog *áyam* 'dog', Murik (northern Sarawak) *ayam* 'domesticated pig' (cf. *mabi* 'wild pig'), or Malay *(h)ayam*, Sundanese *hayam* 'fowl, cock, hen'. It is also noteworthy that *manuk meant 'chicken', and the general term for 'bird' was derived from it by reduplication (*manu-manuk or *manuk-manuk). In addition, there were terms for dog and pig, and many terms for various types of non-domesticated placental mammals, including monkey, deer, pangolin, squirrel, rat, bat, flying fox, bear, clouded leopard, civet cat, and dugong. As AN speakers moved ever eastward into the Pacific they encountered progressively fewer mammals, with the result that a tripartite typology into 'creatures of the sea', 'creatures of the air' and 'creatures of the land' tended to develop from a much more complex inventory of faunal terms found in PAN and PMP. In many languages of the central and eastern Pacific PMP *manuk 'chicken' eventually came to represent virtually all creatures that fly, as in Arosi (Solomons) *manu* 'a creature that flies, insect, bird, angel, etc.', Fijian *manumanu* 'bird; sometimes also animal or insect, but these are so few that the actual name will generally be used', Hawaiian *manu* 'bird; any winged creature'. Similarly, PAN *Sikan, PMP *hikan 'fish' came to represent virtually all creatures that swim, such as Samoan *iʔa* 'fish, turtle, whale' and Hawaiian *iʔa* 'fish or any marine animal, such as eel, oyster, crab, whale'. Land mammals or reptiles other than the domesticated dog or pig were rare in precontact times. Since there was no inherited PMP term for 'animal' it was necessary to coin a variety of terms for land animals, as with Samoan *mea* 'thing; animal', or Hawaiian *holoholona* 'animal, beast, insect' (also 'a trip, excursion, ride' evidently < *holo* 'to run, sail, ride, go; to flow, of water'). In still other languages reflexes of PMP *manuk and *hikan came to represent (with certain exceptions) a two-way contrast into animals of the land and air vs. animals of the sea or just animals that swim, as in Chuukese *maan* 'living creature of land or air (other than human)', *iik* 'fish', Rennellese *manu* 'fauna except human beings and fish, including birds and flying insects, hairy animals, reptiles except turtles, germs and other creeping insects, sea cucumbers, slugs', *ika* 'fish, turtle', or Tongan *manu* 'animal: especially bird ('flying animal'), but applied also to quadrupeds ('four-footed animals'), reptiles, insects, etc., but not to fish', *ika* 'fish, turtle, whale, but not eels, cuttlefish or jellyfish'.

As noted by Clark (1982a), and Biggs (1994), other types of semantic change triggered by movement into a new environment are seen in Maori, which had to adapt its lexicon to a radically altered inventory of plants and animals in moving from the tropical Pacific to a group of islands lying between roughly 34 degrees and 47 degrees south latitude. In many cases older terms were retained, but applied to new referents. This is especially evident in bird names, as PPN *kea 'hawksbill turtle' > Maori *kea* 'large carnivorous dull green parrot endemic to New Zealand', PPN *kiwi 'shore bird species' (possibly the curlew or sandpiper) > Maori *kiwi* 'flightless bird of the genus *Apteryx*', PPN *kaalewa(lewa) 'long-tailed cuckoo' > Maori *kaarewarewa* 'the New Zealand falcon: *Falco novaeseelandiae*', and most strikingly with PPN *moa 'domesticated fowl' > Maori *moa*, designating eleven species of the genus *Dinornis*, a group of endemic flightless birds that varied greatly in size, reaching their maximum in *Dinornis maximus*, at about ten feet in height and 500 pounds. Most of these semantic changes apparently took place because the original referent was not found in New Zealand, or a new referent was encountered for which a name was needed. More difficult is the task of determining why a retained term was applied to the

particular new referent it came to designate. In the case of *kea* there was a physical similarity in the beak of a hawksbill turtle and the endemic parrot. In the case of *kiwi* Clark (1982a:130) speculates that the change may have been motived by a similarity in the cry of shoreline waders and the indigenous flightless bird.

A particularly striking example of semantic change motivated by change of physical environment is seen with reflexes of PMP *taRutuŋ 'porcupine fish'. Where a language community has remained in contact with the sea reflexes of this word generally have not changed meaning. Languages meeting this description reach from the central Philippines through eastern Indonesia and Palau in western Micronesia, to Polynesia. In inland areas where the porcupine fish is not found, the word is sometimes retained but applied to land animals. This is true of several languages in Borneo, as Lun Dayeh *tərutuŋ*, Katingan *tahatuŋ*, and Ma'anyan *tetuŋ* 'porcupine', as well as some languages of the Lesser Sundas (e.g. Manggarai *rutuŋ* 'porcupine'). Somewhat more surprisingly, in the Batak languages of interior northern Sumatra a likely reflex of this term has come to refer to fruits that resemble the rounded, prickly body of an inflated porcupine fish, as with Toba Batak *tarutuŋ* 'durian; soursop', Karo Batak *tarutuŋ* 'kind of durian'.

5.8.3 Reduced importance of the referent

A third type of semantic change in AN languages that departs from established typologies is change motivated by reduced importance of the referent. To chose one striking example, PAN distinguished three terms for rice: *pajay 'riceplant, rice in the field', *beRas 'husked rice/rice in storage', and *Semay 'cooked rice'. Many languages in Taiwan, the Philippines and western Indonesia retain these distinctions but in eastern Indonesia, where the importance of rice diminishes in moving from the Lesser Sundas to the Moluccas and New Guinea, only a single term is found. Words for 'rice' in West New Guinea languages such as Biak, Dusner *fas*, or Serui-Laut *fa* thus have essentially the same meaning as the corresponding word in English. As with 'python' > 'snake', one could speak here of 'widening of meaning' (from 'rice in the field/riceplant' to 'rice'), but the mechanism is completely different, since in the first case it results from the identification of a prototype with its inclusive category, while in the second it results from the reduced cultural importance of the referent, and hence with the reduced need for a highly ramified terminology.

5.8.4 Semantic fragmentation

A fourth type of semantic change that is not well served by Bloomfield's typology or subsequent proposals is 'semantic fragmentation'. In this type of change a meaning composed of clearly distinguishable subparts separates into different components in different daughter languages. Bloomfield described the relationship of German *Zaun* 'fence' to English town as one mediated by synecdoche, where a part is taken to represent the whole. But, given Dutch *tuin* 'garden' and Old Irish *dun* 'fortified place' it appears far more likely that the original meaning of this term was 'settlement'. Virtually all early settlements in Europe, including medieval towns, were fortified by an encircling enclosure, and the notion of 'town' thus implicitly included a town wall. From the closely associated complex of ideas in this concept English preserved the idea of a collection of buildings and their inhabitants, German the idea of the encircling enclosure, and Dutch the idea of an area of cultivation surrounded by an encircling enclosure. Much the same type of

fragmentation appears in English *thatch* : Dutch *dak* 'roof', German *Dach* 'roof'; since the only roofs known to speakers of Proto West Germanic were thatched, the idea of the material was retained in English, while that of the structure was preserved in Dutch and German (which now refer to the material by further specification as *dakriet* 'roof reeds' or *Dachstroh* 'roof straw').

Several striking examples of semantic fragmentation are found in AN languages. Two of these are terms that originally referred to the seasonal monsoons, the rain-bearing winds that blow predominantly from the west or east at different times of the year in insular Southeast Asia and the western Pacific. PMP *habaRat designated the west monsoon, and *timuR the east monsoon. In many languages these meanings are retained with little change, the only difference being nuances of directionality between the major points of the compass, where a reflex may be glossed as northwest, southwest, southeast, or northeast monsoon, varying with the latitude of the community of speakers. However, in other languages reflexes of these terms preserve only elements of the original complex meaning associating wind, rain, and directionality in a single unified concept. Examples include Hiligaynon *bagát-nan* 'south', Tiruray *barat* 'rainy season', Malagasy *avaratra* 'north', Malay *barat* 'west', Sasak *barat* 'storm, to storm', Tae' *baraʔ* 'big, terrific, violent, of rain and wind', Ngadha *vara* 'wind, storm, stormy', Kamarian *halat*, Numfor *barek* 'west', Fijian *cava* 'hurricane, windstorm', Tongan *afā* 'gale or very severe storm', and Maori *awhaa* 'gale, storm; rain' from *habaRat, and Itbayaten *timuy* 'rain', Tagalog *tímog* 'south', Timugon Murut *timug* 'water', Malay *timur* 'east', Rennellese *timu* 'rile, devastate, as by wind and storm', and Samoan *timu* 'rain' as reflexes of *timuR.

5.8.5 Semantic chaining

Other kinds of semantic change in AN languages fit more neatly within Bloomfield's typology. PAN *Rumaq 'house' > Proto Philippines *Rumaq 'sheath (of knife, machete)' can be seen as metaphorical, and changes such as PMP *laŋit 'sky' > Mono-Alu *laiti*, Talise *laŋi*, Arosi *raŋi* 'rain', Roria *laŋet* 'sky; cloud', Titan *laŋ* 'light' (*kole-laŋ* 'place of light' = 'sky') as instances of metonymy.

Semantic changes that result from cultural innovation are well-known. In all such cases a sign-referent relationship has been altered, sometimes quite drastically, as a result of a cultural practice in which earlier and later referents were closely associated. Changes of this kind can be described as semantic chaining, since the earlier and later referents are linked by a period in which they are culturally juxtaposed. Examples of this type of change in English and other Indo-European languages include *pen* 'writing implement' (< Latin *penna* 'feather', from the earlier use of feather quills as writing implements), and *clock* (< Anglo Saxon *clugge* 'bell', from the medieval practice of marking public time by the ringing of church bells). Bloomfield (1933) characterised the form of such changes in a way that can be schematised as in Figure 5.2:

Stage	1	2	3
	A	A+B	B

Figure 5.2 Schematic representation of semantic change motivated by cultural change

In stage 1 A represents the original meaning ('feather; bell'); in stage 2 this and a second meaning are closely associated through a cultural practice ('feather = writing implement; bell = public time-keeping device'), and in stage 3 the derived meaning is separated from its source ('writing implement; time-keeping device'). Similar types of semantic changes are found in AN languages, and can be taken as evidence for changes in various cultural practices over time. To illustrate, PMP *liaŋ meant 'cave', but in a number of the societies of northern Sarawak reflexes of this word refer to burial places, as in Long Anap Kenyah *liaŋ* 'grave', Long Wat Kenyah *liaŋ* 'cemetery, burial place', Baram Kayan *liaŋ* 'burial post or grave; cemetery (modern)', Kelabit *liaŋ tanəm* 'grave', Batu Belah Berawan *ləjaŋ* 'wooden house-shaped coffin raised on pillars', Long Teru Berawan *lijeŋ* 'single-use post or pillar tomb'. In the most extreme cases the physical resemblance between a cave and the burial structure named by a reflex of *liaŋ is nil. Among the Berawan of northern Sarawak, for example this word refers to an elaborate vertical mortuary post into which a coffin is inserted for the deposition of the bones of the (upper class) deceased. These words could be treated as unrelated, but various clues point to them having a common origin (Blust 1986/87). Since the mortuary posts of the Berawan bear no similarity to caves it can be inferred that these peoples, who have not practiced cave burial within the ethnographic present once did, hence in stage 1 *liaŋ = 'cave', in stage 2 *liaŋ = 'cave/burial place', and in stage 3 *liaŋ = 'burial place'.

5.8.6 Avoidance

It is well known that many of the languages of Europe have developed avoidance terms for certain dangerous animals, most notably the 'bear' ('honey-eater', etc.). A similar psychology of avoidance based on fear that names attract their referents is found in AN languages. This is seen in replacements for PAN *daRaq 'blood'. The most common lexical innovations meaning 'blood' derive from words for 'sap' or 'juice', a change that has happened repeatedly with different bases. Examples include PMP *zuRuq 'sap, juice, gravy' > Tagalog *dugóʔ*, Maranao *rogoʔ*, Bolaang Mongondow *duguʔ* 'blood', PMP *liteq 'sap of tree or plant' > Bilaan *litəʔ*, Tboli *litoʔ* 'blood', Proto Central Philippines *tagek 'sap of plants, trees, or fruits' > Palawan Batak *tagək* 'blood', PMP *pulut 'breadfruit sap' > Sebop *pulut* 'blood', and Javanese, Balinese *gətih* 'blood', presumably from a doublet of *getaq 'tree sap'. A second source of innovative 'blood' words in Philippine languages is PPH *laŋesa 'having a fishy smell or the taste of blood' > Binukid *laŋesa*, Ata *laŋosa* 'blood', and a third is PMP *baseq 'wet' > Northern Kankanaey *basa* 'blood'. In each of these cases it is reasonable to suppose that the innovation is an avoidance term: blood is universally a sign of danger and a cause for alarm, and the colour of blood carries similar symbolic overtones. In Borneo, when traveling in the jungle terms for 'vine' and the like serve as substitutes for 'snake' in some languages. Although this is not known to have resulted in permanent semantic change in any language, the psychological mechanism is essentially the same: words that are associated with dangerous referents are avoided so as not to attract the danger they connote. In some other cases a dangerous referent is shown respect rather than avoidance, as where the crocodile is addressed as 'grandfather' in some of the languages of Indonesia, a practice in which fear and respect are inseparably mixed.

Probably the most extreme, or at least the most systematic form of lexical avoidance behavior is seen in word taboo, a phenomenon that straddles the boundary between semantic change and lexical change. There are various forms that word tabooing may take,

all of which appear to be motivated by fear of natural or supernatural retribution. Simons (1982) provides a useful overview of this phenomenon in AN languages.

A common type of word taboo forbids the use of words resembling the name of a person of high rank or superior genalogical status. In either case violation of the taboo evidently is seen as an affront to hierarchy. Examples mentioned by Simons include Malagasy, where the use of a word that resembles the name of a living chief is forbidden, and this prohibition continues even after the chief's death, Labuk Kadazan, where use of the name of a parent-in-law will be automatically punished by supernatural swelling of the abdomen (the 'busung taboo' in Blust 1981c), and Tahitian, where cognate densities with other AN languages have decreased more rapidly than normal because the use of any word that resembled the name of a high chief was forbidden, often causing the word to disappear from the language.

5.9   Doubleting

Lexical doublets are not unknown in languages such as English. However, virtually all discussions of this phenomenon attribute it to borrowing between related languages. Pairs such as *shirt* and *skirt*, for example, are explained as divergent developments from the same protoform, the first variant part of the native vocabulary, and the second a loanword acquired during the Scandinavian occupation of northeast England after the sound change *sk- > sh-* had taken place in English. Occasionally, doublets may also arise by borrowing from the same language at different historical periods, as with English *wine* and *vine*, both from Latin *vīnum*, the first word borrowed earlier and the second later. In this type of case doublets may arise even if the borrowing language and the source language are unrelated.

Doubleting in AN languages is extremely common (Blust 2011a). Some doublets are of the familiar type just described, as in Tiruray of the southern Philippines, where PMP *Ratas > ratah* 'human breast milk', next to *gatas* 'store-bought milk', the latter a loanword from a neighbouring and socially dominant Danao language. Doublets that result from borrowing between related languages are probably as common in AN languages as they are in English and other well-studied languages of Europe. Far more common, however, are phonemically and semantically similar words in the vocabulary of a single language that cannot plausibly be attributed to a native/nonnative distinction. To cite one example, Wilkinson (1959:142) gives Malay *biŋah* and *biŋar* 'a shell: *Voluta diadema*'. There is no evidence that one of these words is a dialect form borrowed into standard Malay, and to dismiss this possibility completely, Cebuano Bisayan in the central Philippines also shows doublets for this form (Wolff 1972:138): *biŋá, biŋáʔ, biŋág* 'kind of baler (volute) shell, used to crush cocoa seeds'. Although the first of these Cebuano forms does not correspond to either Malay word, the last two correspond regularly with Malay *biŋah* and *biŋar*, and so indicate proto doublets *biŋaq, *biŋaR. This is one of the rare cases in which both proto doublets are reflected in two or more languages. Far more common is a pattern in which widely separated languages A, B and C reflect variant (1), while widely separated languages D, E and F reflect variant (2), thus requiring the reconstruction of doublets for a given morpheme, even though no attested language is reported as having more than one variant. Two well-known examples of this pattern are PMP *ijuŋ/ujuŋ 'nose', and *ma-tiduR/ma-tuduR 'sleep'. Reflexes of the first pair are seen in e.g. Agta *iguŋ*, Malay *hiduŋ*, Samoan *isu*, as against Kalamian Tagbanwa, Kayan *uruŋ*, Fijian *ucu* 'nose'; reflexes of the second pair are seen in e.g. Casiguran Dumagat *tidug*, Malay *tidur*, Makura *matir*, as against Isneg *ma-túdug*, Javanese *turu*, Nguna *maturu*

'sleep'. It is clear that in each case both forms have an equally wide distribution, and that there is no areal bias: reflexes of both *ijuŋ and *ujuŋ 'nose' are found in the Philippines, western Indonesia and the central Pacific, and reflexes of both *tiduR and *tuduR are found in the Philippines, western Indonesia and central Vanuatu.

An 'obvious', but clearly incorrect solution to this problem would be to argue that only *ijuŋ and *tiduR are valid, and that the forms with penultimate *u* are products of sporadic vowel assimilation. However, there are many other reconstructed disyllables of the form *CiCuC which are richly attested, but show no evidence for a *CuCuC variant, as with PMP *hiRup 'sip', PMP *ikuR 'tail', PAN *iluR 'river channel', PMP *qi(m)pun 'collect, gather', *likud 'back' or *pitu 'seven'. If apparent reflexes of *ujuŋ 'nose', and *ma-tuduR 'sleep' actually reflect *ijuŋ and *ma-tiduR with sporadic vowel assimilation one must ask why such irregular changes have targeted these morphemes repeatedly in different languages, but left many others untouched. This is not to deny that sporadic assimilation sometimes occurs, but one would naturally expect it to be confined to a single language or set of languages descended from an immediate common ancestor, while doublets for 'nose' and 'sleep' are widely distributed throughout the AN language family.

Even if we could explain apparent reflexes of *ujuŋ, and *ma-tuduR as products of parallel vocalic assimilations from PMP *ijuŋ and *ma-tiduR this would not provide a general explanation for lexical doubleting in AN languages, since many other patterns of variation exist, most of which cannot be explained as due to sporadic sound change or borrowing. A few of these are listed in Table 5.31:

Table 5.31 Sample patterns of lexical doubleting in AN languages

1. d/n : *adaduq (north Philippines, central Borneo) : *anaduq (north Philippines, north Borneo, east Indonesia) 'long, of objects'
2. -aw/u: *qali-maŋaw (Sumatra, Sulawesi, west Micronesia, Solomons, west Polynesia) : *qali-maŋu (central Philippines, west Melanesia) 'mangrove crab'
3. a/e : *añam (south Philippines, Borneo, Sulawesi, east Indonesia, west New Guinea, Micronesia, Fiji : *añem (south Philippines, Sulawesi, east Indonesia) 'to plait, as mats'
4. R/w : *banaR (central and south Taiwan, north and central Philippines, Malay Peninsula) : *banaw (south Taiwan, Malay Peninsula) 'a useful plant: *Smilax* species'
5. b/l : *baŋaw (east Taiwan, central Philippines) : *laŋaw (Taiwan, Philippines, Borneo, Malay Peninsula, Sulawesi, Melanesia, Polynesia) 'blowfly, housefly'
6. q/t : *beriq (south Taiwan, south Philippines, central Borneo, central Sulawesi, east Indonesia) : *berit (south Philippines, central Borneo) 'burst, tear open'
7. i/u, a/e : *bileR (central and south Philippines, Malay Peninsula, Java, central Sulawesi) : *bulaR (e. Taiwan, central and south Philippines, Malay Peninsula) 'ocular cataract'
8. Ø/ŋ : *bakuku (central Philippines, Malay Peninsula) : *bakukuŋ (Malay Peninsula, south Sulawesi) 'a fish: sea bream'
9. Ø/l : *esuŋ (central Philippines, south Philippines, north Borneo, central Borneo) : *lesuŋ (north Taiwan, south Taiwan, north Philippines, Malay Peninsula, west Indonesia, west Micronesia)
10. b/p, e/a : *betuŋ (south Taiwan, central Philippines, south Philippines, Borneo, Malay Peninsula, west and east Indonesia, west Melanesia) : *patuŋ '(central Philippines, Malay Peninsula, Sulawesi) 'large bamboo: *Dendrocalamus* species'

Even confining ourselves to this table, which represents only a small fraction of the known range of phonemic variation in reconstructed doublets, it is clear that this phenomenon is different in kind from doubleting of the type seen in English *shirt*/*skirt* or *wine*/*vine*. In most cases there is no basis for assuming that doublets have arisen from early dialect borrowing, since the phonemic differences between doublets are often not of a type likely to be due to sound correspondences between dialects or even languages, and there are far too many patterns of variation for a borrowing explanation to work. Why such exuberant variability should have existed at earlier times in the history of the AN languages, or why it exists in many of the modern languages is very difficult to say. It should be noted that true lexical doubleting is distinct from families of words that share a common monosyllabic root (cf. 6.2).

5.10 Lexical change

Word taboo, noted in §5.8, raises the issue of factors that might govern the rate of lexical change. Quantitative aspects of lexical reconstruction will be treated in Chapter 8, 'Reconstruction', and this section will therefore be devoted entirely to lexical change. Two different measures of the rate of lexical change are of potential interest: 1) the rate of change of the entire basic vocabulary in a given language, as in lexicostatistics, and 2) the rate of change of a given lexical item across languages, or lexical stability indices.

5.10.1 Lexicostatistics

Based on thirteen languages for which written records were available spanning at least a millennium, Lees (1953) concluded that the retention rate of basic vocabulary is governed by a universal constant such that 90% of all languages will have rates expressible as 80.5% +/- 1.8%/mill. In other words, the retention rate of basic vocabulary for all but 10% of the world's languages will cluster within the very narrow range of 78.7% to 82.3% per millennium. Although neither Lees nor Swadesh used the term, this can be called the 'universal constant hypothesis' or UCH (Blust 2000a).

In view of its inevitable selection bias, the UCH was a very strong claim. Eleven of the thirteen languages in the Lees database are Indo-European, and six of these involve changes from Plautine Latin to various of the modern Romance languages. As noted by Guy (1983), to a statistician the sample bias in this database was roughly equivalent to measuring the heights of "13 people, 11 related, 4 cousins, 2 brothers, most of them male adults of the same age group," and after various statistical manipulations concluding "that the heights recorded are representative of the individual heights of all human beings, irrespective of age, sex, or race." Guy's criticism is clearly valid, and the boldness with which Lees advanced his claim must have stemmed in part from his awareness of the difficulty of falsifying it. Bergsland and Vogt (1962) believed they had falsified the UCH, since Icelandic, Georgian and Armenian show retention rates that are well above 82.3%. However, Hymes (1962) dismissed these examples as selective, since 10% of all languages are expected to have retention rates that fall outside the predicted range.

Granted the reality of sample bias in the database used by Lees, one might ask what choice did he have? Hymes (1960:3) distinguished two types of cases in measuring rates of change in basic vocabulary: "If the languages are different stages in a single line of development, it is a *control case*. If the languages are the outcomes of different lines of

development from a single ancestor, it is a *case of application*." Control cases require languages with at least a millennium of documentation, and so are severely limited. It would seem, then, that the theory is at the mercy of an unavoidably inadequate database.

Blust (2000a) showed that the calculation of lexical retention rates need not depend entirely upon languages with at least a millennium of documentation. If the comparative method is trustworthy, it should be possible to reconstruct the most basic part of the lexicon of a language -- that is, its basic vocabulary. When this is done for a given proto language retention percentages can be calculated by comparing reconstructed forms with their reflexes in the modern languages. By assuming a range of separation times these percentages can be converted to rates, and by listing a sufficiently large number of rates it can be determined whether these rates cluster as tightly around a mean value as the UCH predicts. This was done for more than 230 languages. As a result of constructing this database many new 'control cases' in the sense used by Hymes were created, opening the possibility for doing 'vertical' lexicostatistics on a much larger scale than was previously imagined possible. As a result of following this procedure it was discovered that there is a very wide range of retention percentages in AN languages, from 58% (standard Malay) to barely 5% (Kaulong of New Britain). Regardless of how these percentages are converted into rates, no more than 26.5% of all languages in the sample can be fitted into the range of variation predicted by the UCH. Moreover, there is marked variation in mean retention rates across major Malayo-Polynesian subgroups, as shown in Figure 5.3:

| WMP: 40.5 | CMP: 38.9 | SHWNG 25.6 | OC: 23.6 |

Figure 5.3 Mean retention percentages for major Malayo-Polynesian subgroups

Impressionistically it had long been noted that many Oceanic languages, especially in Melanesia, are lexically 'aberrant', meaning that only a small part of the vocabulary shows evidence of cognation with other languages. This led Dyen (1965a) to infer a Melanesian homeland for the AN family, a conclusion that is radically at odds with the qualitative evidence from historical linguistics, and the archaeological record that has accumulated over the past 40 years. As first noted by Grace (1966), the lexical diversity of Oceanic languages thus appears to be a product of accelerated lexical replacement rates rather than a function of greater separation time. A likely factor in this process is contact with Papuan languages, as this feature is also shared by South Halmahera-West New Guinea languages in the northern Moluccas. However, as noted by several writers (Pawley 2006, Lynch 2009b) some of the most aberrant AN languages (those with very low cognate densities and a high level of irregularity in sound correspondences) are found in parts of Melanesia that are far from the nearest Papuan language.

5.10.2 Lexical stability indices

One of the first discoveries made in experiments with lexicostatistics was that individual test-list items have different probabilities of replacement. This is why Swadesh developed a variant 100-word list which selected those items from the 200-word list that appeared least likely to be borrowed or replaced. In the hope of improving lexicostatistics as a tool of subgrouping Dyen, James and Cole (1967) calculated differences in word retention rates for a number of AN languages, using maximum likelihood estimates. The

results showed that the ten most stable form-meaning associations on the Swadesh 200-word list (reduced to 196 words) were 1) five, 2) two, 3) eye, 4) we, 5) louse, 6) father, 7) to die, 8) to eat, 9) mother, and 10) four, and that the ten least stable form-meaning associations were 187. to think, 188. some, 189. there 190. to squeeze, 191. smooth, 192. to fall, 193. to hold, 194. to say, 195. how and 196. to play. The form that this study took was largely determined by its intended use as a tool of lexicostatistics. It is possible, however, to calculate stability indices for lexical data in other ways that are potentially of more general interest. In particular, it might prove valuable to determine differences of stability over time for terms that belong to a common semantic field, and to further compare these differences across geographical regions. Rather than measuring the stability of lexical forms, then, the goal is to measure the stability of lexical representation for given categories of meaning.

An index of stability for the lexical representation of meaning can be calculated in the following way: 1) let Y be the number of languages in the sample, 2) let X be the number of etymologically distinct sets (including isolates), 3) then X-1 equals the number of changes over time, called 'C', 4) hence (Y-C)/Y = the stability index, with lower values representing more labile forms and higher values more stable forms. For 20 languages in which all terms are cognate we have 20-(1-1)/20 = 1.00, or maximum stability; for 20 languages in which a given meaning is represented by 16 cognate sets we have 20-(16-1)/20 = .25, or a 25% stability index, etc. It is assumed that one term in the list continues the protoform, whether or not this can be demonstrated on the basis of the comparative method.

As an example, this procedure can be applied to colour terminology, using the five basic terms black, white, red, green and yellow. Ferrell (1969) provides data for 17 Formosan languages plus Yami, a Philippine language spoken within the political region of Taiwan. Two of these are dialects of Atayal and a third is the closely related language Seediq. Other than these the remaining languages are all quite distinct from one another. Reid (1971) contains comparative wordlists for 43 minor languages of the Philippines, representing all parts of the archipelago. Tryon and Hackman (1983) give a comparative vocabulary for 63 languages of the Solomon Islands, but seven of these are non-AN, and consequently will not be considered here. Finally, Tryon (1976) provides a comparative vocabulary for 105 languages of Vanuatu. Table 5.32 summarises the diversity of terms in these four geographical regions:

Table 5.32 Stability indices for colour terms in four major geographical regions

	Taiwan	Philippines	Solomons	Vanuatu
Black	15/18 = .22	12/43 = .74	46/103 = .56	61/178 = .66
White	15/18 = . 22	17/43 = .63	39/103 = .63	77/179 = .58
Red	17/18 = .11	28/43 = .37	41/100 = .61	27/178 = .85
Green	18/18 = .06	21/40 = .50	30/95 = .69	59/138 = .58
Yellow	18/18 = .06	20/43 = .56	45/103 = .57	60/175 = .66

The calculation of lexical stability indices in this way is novel and as yet untested, but certain patterns stand out with reasonable clarity. First, Formosan languages show no widely distributed cognate sets for colour terms: most languages have a term that has no demonstrable relationship to any other (the 18 terms recorded for each meaning fall into a minimum of 15, and a maximum of 18 etymologically distinct sets, most of which are isolates). Second, in all areas outside Taiwan language gaps occur for 'green', either

because a term cannot be elicited, or because it is borrowed from a non-AN source, as in Atta *be:rdi*, Northern Kankanaey *bildi*, or Mamanwa *grin*. Finally, stability values for colour terms in Oceanic languages show little variation, while in the Philippines terms for 'black' are exceptionally stable and terms for 'red' exceptionally labile. No theory to account for these differences is currently available, but the observations themselves may stimulate thinking toward that end. Before that can be done, however, similar measures of lexical stability should be taken for other semantic fields, such as body part terms, numerals, pronouns, or flora and fauna.

Pawley (2011) has recently done this for Oceanic fish names, and reached the following tentative conclusions. First, binomials tend to be highly unstable, possibly because competing modifiers crowd one another out, leaving little that is fully comparable over long intervals of time. Second, other factors that seem to play a part in enhancing stability are economic importance, and danger. The common element in these factors appears to be psychological salience, which in turn probably is related to frequency of use, since both fish that are important as food sources, and those that must be avoided because of the danger they present are more likely to receive frequent mention than fish that are not useful sources of protein or for which warnings are unnecessary.

5.11 Linguistic paleontology

The term 'linguistic paleontology' was first employed Saussure (1959:224) to describe the use of comparative linguistic data to draw inferences about homelands and the content of prehistoric cultures. Such inferences are ultimately dependent upon subgrouping, a topic that will be treated in Chapter 10. The focus here is the use of reconstructed lexical data for culture-historical inference. Before proceeding it is necessary to consider the problem of semantic reconstruction, and thus the issue of categorial non-correspondence.

5.11.1 Categorial non-correspondence

Semantic categories may not correspond between languages, and for this reason it is impossible to predict the meaning of a term that may be the translation equivalent of an English word. The lack of correspondence to Indo-European semantic categories was the basis of one of the primary objections raised in early critiques of lexicostatistics, as by Hoijer (1956), who showed that many items of basic vocabulary in Navaho are difficult to match in any exact way with a list of terms derived from European languages. AN languages are no exception, as many items of basic vocabulary fail to correspond to the boundaries of English semantic categories. A good place to begin is the word for 'hair'. Malay is typical of many AN languages in distinguishing head hair (*rambut*) from body hair (*bulu*). However, *bulu* isn't just body hair, since it also includes feathers exclusive of tail-feathers (*lawi*), as well as fur and the fine floss on plant stems:

Table 5.33 Category boundaries for English *hair*, *feathers*, *floss*, *fur*
and Malay *rambut*, *bulu*, *lawi*

English	Malay
hair	rambut, bulu
feathers	bulu, lawi
floss	Bulu
fur	Bulu

Malay	English
rambut	Hair
bulu	hair, downy feathers, floss, fur
lawi	tail feathers

Although some English terms relating to hair have a single translation equivalent in Malay (*floss = bulu*, *fur = bulu*), there is no unique correspondence between any English term with a Malay term, or any Malay term with an English term. Malay *rambut* is also applied to hair-like appendages on natural objects, as in the name of the *rambutan*, a hairy-skinned grape-like fruit (called *buluan* in other languages, such as Sundanese and Tae'). In addition, most AN languages encode the meanings 'facial hair' ('beard/moustache'), and 'gray hair' with separate morphemes, and several of the Formosan languages have a distinct term for 'pubic hair' (Ferrell 1969:216). A second item of basic vocabulary that may seem straightforward when presented on a lexicostatistical test-list is 'water', but reflexes of PMP *danum and *wahiR both mean 'fresh water' (or less commonly 'river'), and reflexes of *tasik mean both 'salt water' and 'sea' in hundreds of daughter languages.

In some AN languages, especially in island Southeast Asia, verbs of cutting and carrying are highly differentiated. While English may come close to matching the diversity of the former category with distinctions such as butcher, cut, chop, hack, sever, slash, slice, snip, split and the like, it has nothing to match the patient-based distinctions seen in e.g. Long Anap Kenyah *l<əm>ase* 'split bamboo' vs. *m-əpeʔ* 'split firewood', Bario Kelabit *ŋ-upa* 'split bamboo' vs. *nəpak* 'split wood; kill an animal with a machete', Limbang Bisaya *ŋ-utob* 'cut string' vs. *mutul* 'cut wood; break', Murik *ŋələŋ* 'cut string or wood' vs. *nətək* 'cut meat' or Squliq Atayal *h<m>obiŋ* 'cut meat (as in butchering an animal)' vs. *k<m>ut* 'cut flesh (of a living person)'. Verbs of carrying are notoriously differentiated in the languages of both mainland and insular Southeast Asia. An example from one language should be sufficient to make the point. For Yakan, a Sama-Bajaw language of the southern Philippines, Behrens (2002) lists nineteen verbs meaning 'to carry' (in some cases glosses have been simplified): 1) *abet* 'carry or hold something in a sarong', 2) *abit* 'carry or hold something fixed to one's side (by a string or hook)', 3) *aŋkut* 'carry in several trips', 4) *baluŋ* 'carry something or someone on the upper back', 5) *bimbit* 'carry something suspended in the hand', 6) *boʔo* 'bring, carry (from one place to another), take something along', 7) *duwaʔ* 'carry a load, of vehicles or animals (the load is often suspended on either side)', 8) *komoŋ* 'carry something in the claws, as an eagle does', 9) *limbit* 'carry a load', 10) *lutu* 'carry something on the head', 11) *panaŋkit* 'carry on the shoulder (especially something long)', 12) *pippi* 'hold or carry something in one's arms, especially a child; to hug (something that is lifted off the ground)', 13) *sabley* 'wear or carry something over the shoulder (either straight or diagonally over the chest)', 14) *sugelley* 'carry something on one's shoulder, or hanging over the shoulder', 15) *taŋaʔ* 'hold, carry, or take in the mouth

(like cats and dogs with their young)', 16) *taŋguŋ* 'carry something on the shoulder suspended from a pole, either balanced on both sides, or between two persons', 17) *teppik* 'carry something under the arm; carry a child in front and to one side on the arm or hip', 18) *tumpey* 'carry something on the back', 19) *usuŋ* 'use or be carried in a sedan chair'. Yakan probably is not exceptional in the number of terms that translate English 'carry'. Ferrell (1982), for example, lists eighteen terms for carrying in Paiwan, and Topping, Ogo and Dungca (1975) list twenty three for Chamorro. Since there is only partial overlap in the semantic categories lexicalised by these languages the potential exists for some language to have an even larger number of 'carry' verbs than is reported here. Given these observations it seems likely that if dramatically fewer terms are reported in a language spoken in insular Southeast Asia, this probably is a shortcoming of the lexicographer rather than an indication of relative poverty in this area of the lexicon.

Many other examples of categorial non-correspondence between English and AN languages could be cited, but these would not serve to further advance the point already made. Some of these will appear as a matter of course in discussing the history of kinship terminology and some other topics. One last example will be given to drive home the sometimes overlooked point that, unlike typological traits, which are to some extent constrained by limited possibilities, semantic categories in the lexicon seem almost endlessly creative. While technical discussions in English distinguish various types of lightning based on their form (sheet lightning, lightning bolt, etc.), some languages in the Philippines distinguish types of lightning on an entirely different basis, as with Koronadal Bilaan *kilət* 'lightning during the day' vs. *siloʔ* 'lightning at night', Botolan Sambal *kímat* 'heat lightning' vs. *kílat* 'lightning with rain and thunder', or Siocon Subanun (Reid 1971) *kilat* 'lightning' vs. *glotiʔ* 'close thunder and lightning'

The discussion of categorial non-correspondence leads naturally to the topic of semantic reconstruction. While lexical and phonological reconstruction will be treated together in Chapter 8, semantic reconstruction and change appear to fit better into the discussion of the lexicon.

5.11.2 Semantic reconstruction

Unlike phonological reconstruction, for which a well-developed method has existed since at least the middle of the nineteenth century, no generally accepted method is available to reconstruct the meanings of linguistic forms. In many cases the need for reconstruction is superfluous, as in reflexes of PAN *maCa 'eye', *lima 'five', or *asu 'dog', which have identical meanings in most daughter languages. In other cases, however, the meaning of a protoform may be less obvious.

Pawley (1985) and Blust (1987b) address the issue of how to reconstruct the meanings of forms whose reflexes overlap semantically. To illustrate from the second of these publications, reflexes of five PMP etyma (*lepaw, *kamaliR, *balay, *Rumaq, *banua) have the meaning 'house' in two or more primary branches of MP. Since no language reflects more than one of these terms with the same sense it seems clear that most of them must have had some other meaning.

Three principles are critical in inferring proto meanings from semantically diverse reflexes. The first is that semantic categories are not given a priori. The second is that meanings are assigned competitively. The third is that synonymy should be accepted only as a last resort. On this basis *Rumaq is glossed 'house, family dwelling', since this is the only meaning that its reflexes have, apart from transparent metaphorical extensions to

'nest; web of a spider', or in some languages 'sheath of a knife'. In addition, there are indications that it referred to a low-level segmentary descent group, a type of application to social structure that is also known in other parts of the world. Since *Rumaq covered this semantic range it follows that the other terms probably did not. While reflexes of *balay mean 'house' in languages reaching from the Philippines to Polynesia, other reflexes in a diverse collection of languages refer to public or communal buildings (village meeting house, storage shed for yams, house for women during menstruation and after childbirth, men's clubhouse, boathouse). In many cases these are open-sided structures, but this does not always appear to be the case. PMP *balay, then, apparently designated a building reserved for public use, as in community meetings. Reflexes of *banua are even more diverse, including 'land,' 'village,' 'house', 'country', and even 'sky, heaven'. It is clear that PMP *banua designated neither a private dwelling nor a communal meeting house. Part of its sense included the notion of land, and part that of human habitation. The glosses in two languages turn out to be especially illuminating: Iban (Sarawak) *mənoa/mənua* 'area of land held and used by a distinct community, espcially longhouse (*rumah*), including house, farms, gardens, fruit groves, cemetery, water and all forest within half a day's journey. Use of the *menoa* is only gained and maintained by much effort and danger, and by proper rites to secure and preserve a ritual harmony of all within it and the unseen forces involved', *mənoa laŋit* 'the heavens, abode of *Petara* and other deities'; 'Āre'āre (southeast Solomons) 'land, as opposed to *asi* sea; district, place, country, island; the territory, area where a person lives, where his possessions are, such as food, bamboo, trees, pigs, water and graves is called his *hanua*'. Iban and 'Āre'āre are only distantly related, and are spoken thousands of miles apart. It is unlikely that such detailed agreement in the glosses given for these reflexes of PMP *banua is convergent. Rather, it suggests that the meaning of *banua did not correspond to any one semantic category of English. Words like 'home', 'abode', 'place', 'district', 'country' and 'region', which co-occur with the Iban gloss, and 'land', 'district', 'place', 'country', 'island' in the 'Āre'āre gloss show a kind of shotgun approach that tries to capture the sense of the term in English semantic categories. The real meaning comes out in the more discursive glosses which mention land, habitation, food, water, and graves. PMP *banua, then, presumably designated the life-support system of a human community, including the land on which dwellings are erected and food grown, water for drinking and bathing, and the graves of the ancestors. Its extension to 'sky, heaven' in widely separated languages further indicates that in the mythology of PMP speakers there probably was a celestial *banua for the deities that matched that of humans on earth. The remaining two terms are of less interest; one (*lepaw) almost certainly meant 'granary', while the other (*kamaliR) may have meant 'bachelors' clubhouse'.[55]

---

[55]   This interpretation is rejected by Green and Pawley (1999:34) on the grounds that "A building specifically for men alone to use as a meeting and sleeping place and for ritual activities is common among Austronesian societies only in Melanesia." However, special structures commonly called 'men's houses' that serve as dormitories and as meeting houses for exclusively male activities are actually reported for many Austronesian-speaking societies outside Melanesia, including at least the Saaroa, Amis, Tsou, and Puyuma of Taiwan (Ferrell 1969:36, 38, Chen 1988:271-278), the Bontok and the Kankanaey of Lepanto in northern Luzon (Keesing 1962:95), the Modang of East Kalimantan (Guerreiro 1998), various Land Dayak groups, where they are generally called 'head houses' (Geddes 1961, Lebar 1972:195), the Toba Batak of northern Sumatra (Lebar 1972:20), the Mentawei of the Barrier Islands west of Sumatra (Lebar 1972:42), the Ende of eastern Flores (Lebar 1972:87), the Lamaholot of Adonara (Lebar 1972:92), the Yamdena of the Tanimbar archipelago (Lebar 1972: 112),

5.11.3 Linguistic approaches to Austronesian culture history

The preceding remarks show that diverse semantic reflexes need not present an intractable problem to reconstruction, particularly where cognate sets can be treated within a larger semantic field. This concern is irrelevant to many inferences about AN culture history, but in some significant cases it is of central importance. The potential contribution of lexical reconstruction to AN culture history is enormous, as demonstrated by Ross, Pawley and Osmond (1998, 2003, 2008, 2011, 2013). In these first five volumes of a planned seven volume work on Proto Oceanic society these authors have already devoted over 2,200 pages to exploring the material culture and physical environment of Proto Oceanic speakers, as inferred from comparative linguistic data. Subsequent volumes will treat basic human conditions and activities, social organisation, and grammatical categories. When this work is completed it will certainly rank among the most sophisticated and detailed studies of linguistically-based culture history ever undertaken. The remainder of this section briefly illustrates some ways in which linguistic approaches to culture history are relevant to 1) the concerns of Southeast Asian and Pacific archaeologists, and 2) social anthropologists concerned with diachronic aspects of social organisation.

5.11.3.1 *Historical linguistics and archaeology*

Linguistics and archaeology had little contact in the first half of the twentieth century, but since the 1970s many scholars in both disciplines have come to see the benefits of interdisciplinary cooperation. This is reflected in such collaborative volumes as the June, 1976 number of *World Archaeology*, devoted to the topic 'archaeology and linguistics', the influential and provocative book *Archaeology and language: the puzzle of Indo-European origins* (Renfrew 1987), the multi-volume proceedings of a 1994 conference on *Archaeology and Language* (e.g. Blench and Spriggs 1997), the two-volume *Time depth in historical linguistics* hosted by the McDonald Institute for Archaeological Research at the University of Cambridge (Renfrew, McMahon and Trask 2000), and in *Hawaiki, Ancestral Polynesia* (Kirch and Green 2001), a book written by two of the most prominent Pacific archaeologists, but guided in its approach by the cultural riches of POLLEX, the Proto Polynesian lexicon initiated by Walsh and Biggs (1966) and expanded and refined continuously over the past 40 years. Not all archaeologists have shared this vision, and some have reacted almost with territorial defensiveness. Nonetheless it seems fair to say that a working relationship has arisen between historical linguists and archaeologists that is likely to continue into the future to the benefit of both disciplines.

Given the material with which archaeologists normally work, it is natural that the culture historical contributions of linguistics that would be of greatest interest to scholars in this field are those relating to material culture. The following are a few examples that serve to illustrate how linguistic evidence may corroborate, complement or contradict the archaeological record.

Pottery

Pottery is one of the mainstays of Neolithic archaeology for obvious reasons: it is often abundant, and is durable under highly variable soil conditions. PAN *kuden 'clay cooking pot' can be reconstructed on the basis of cognates distributed from eastern Taiwan to Fiji.

the Chuukese of the eastern Caroline islands (Fischer 1970:101-102), and the prehispanic Chamorro (Thompson 1945:37, Alkire 1977:21).

A sample cognate distribution includes Vakon Amis *kuren*, Hanunóo *kurún*, Maranao *kodən*, Kelabit *kudən*, Toba Batak *hudon*, Kuruti *kur*, Likum *kuh*, Lindrou *kun*, Nauna *kul*, Motu *uro*, Fijian *kuro*, and Tongan *kulo* (lost in Tonga and then reacquired from Fiji), all of which refer to clay cooking pots in general. The archaeological record shows pottery of a distinctive decorative style and set of cultural associations turning up suddenly in the western Pacific around 3400-3500 BP (Kirch 2000:91). This 'Lapita ware' almost certainly was a continuation of the tradition of pottery-making that can be traced back to PAN via the cognate distribution given here, although the linguistic evidence permits no more detail than the reconstructed meaning 'clay cooking pot' in association with the form *kuden. However, a variety of other terms relating to pottery manufacture and use have been reconstructed for POC, providing a much fuller picture of ceramic culture than we yet have for earlier stages of Austronesian (Ross 1996d).

Pile dwellings

With reference to PAN society Blust (1976b:36) suggested that "Dwelling units were evidently raised on posts." The basis for this inference was a cognate set that reflects PAN *SadiRi 'housepost', seen in e.g. Nataoran Amis *salili* 'housepost', Ilokano *adígi*, Tagalog *halígi* 'post, pillar', Tiruray *liləy* 'the main posts in house construction', Uma Juman Kayan *jihiʔ*, Mukah Melanau *dii*, Ma'anyan *ari*, Malagasy *andry*, Mentawai *arigi*, Manggarai *siri*, Atoni *ni*, Leti *riri*, Yamdena *diri*, Asilulu *lili*, Numfor *rir*, Manam *ariri*, Numbami *alili*, Lau *lili* 'house post'. Wherever sufficient material is available regarding house types reflexes of *SadiRi refer to the main house posts, which normally support the entire structure two meters or more above the level of the ground. In the better-known parts of the Pacific, such as Polynesia, houses are built directly on the ground, and for Pacific archaeologists with a Polynesian focus there thus was little expectation of finding evidence that early Lapita-associated dwellings were raised on piles. Nonetheless, Kirch (1997:171ff) was able to provide archaeological confirmation of this linguistic inference during a 1985 excavation of waterlogged house sites at Talepakemalai in the Mussau Islands north of New Ireland, dated to about 3,500 BP, and hence near the beginning of the Lapita colonisation of the western Pacific. The inference that PAN speakers used pile dwellings, and that this house type continued in use as early AN speakers expanded out into the Pacific implies that ladders were used to enter the dwellings. Reflexes of PMP *haRezan 'notched log ladder' are distributed from northern Luzon to south Halmahera, where they are associated with pile dwellings that are entered vertically from the ground. Reflexes of this term are unknown in Formosan and Oceanic languages. In Taiwan this is because pile dwellings are uncommon. In the Pacific, on the other hand, it may be because pile dwellings are normally built in shallow lagoon waters, and entered by a gradually sloping plank walkway that connects them to the beach terrace (cf. Kirch 1997:174 for a Mussau example). This is consistent with POC *tete 'ladder, bridge', a term that can be reconstructed for PMP (*taytay) only in the meaning 'bamboo suspension bridge', suggesting that in Proto Oceanic the earlier meaning was extended to rickety plank walkways for entering pile dwellings because of the similarity in general structure between these and traditional suspension bridges.

The bow

Linguistic evidence allows us to attribute both pottery and pile dwellings to speakers of PAN, and to infer that these cultural attributes were carried into the Pacific as part of the

AN expansion out of insular Southeast Asia. In both cases archaeological corroboration is available (for decades with regard to pottery, much more recently with regard to pile dwellings). For some other features of material culture the linguistic evidence continues to stand alone, and the chances of archaeological corroboration are slim. It is cases such as these that show with greatest clarity the need to combine linguistic evidence with archaeological evidence in order to 'round out' the picture of the life of prehistoric communities, since perishable referents are rarely preserved in archaeological sites, and when they are preserved they tend to be of certain types (large or stationary objects such as beached canoes, or the posts of pile dwellings).

In addition to practicing agriculture and collecting wild plants PAN speakers hunted game, and perhaps fought with the bow and arrow. Two terms are particularly important in this complex: *busuR 'bow' and *panaq 'flight of an arrow'. The first of these is reflected from northern Taiwan to Vanuatu, as seen in Mayrinax Atayal *buh<in>ug*, Bunun *busul*, Thao *futulh*, Casiguran Dumagat, Tagalog *busóg*, Malay *busor*, Yamdena *busir*, Asilulu *ha-husul*, Paulohi *husule*, Buli *pusi*, Mota *us*, Baetora *vusu*, Wailengi *fuhu*, Tasmate *wusu*, Neve'ei *ne-vis*, Makura *na-vih* 'bow and arrow'. In insular Southeast Asia reflexes of this form can usually be glossed 'hunting bow', but in the Pacific there was little to hunt, and the bow either became an instrument of warfare (as in Vanuatu), or atrophied to a toy (as in Polynesia). What is important in this comparison is the clear evidence it provides that the 'Lapita people', as they are known to Pacific archaeologists, were armed with bow and arrow as they sailed into the western Pacific in their single-outrigger canoes to settle new lands on the margins of territory that had been discovered at least 40,000 years earlier by ancestral Papuans.

The reconstruction of PAN *busuR and the study of its reflexes permits another useful inference. Hunting with bow and arrow is not practical in dense jungle, where the profusion of vegetation increases the risk of deflecting an arrow in its flight. For this reason the bow disappeared over much of Borneo, even though the hunting of wild pigs and sambhur deer was an important part of traditional Dayak cultures. Here, in parts of Sulawesi, and in a few other areas, the bow was replaced by the blowpipe, since the smaller dart can more easily be directed at a target in thick foliage without high risk of deflection. Reflexes of *sumpit 'blowpipe' are unknown in Taiwan, but are distributed from the central Philippines to western Indonesia and parts of the Lesser Sundas. In general this correlates with an increasingly tropical environment, and suggests that the blowpipe may have been invented as a supplement to the bow as AN speakers moved into areas with denser jungle cover.

The *bubu fish trap

Traditional AN-speaking societies hunted and fished using a wide variety of methods. In addition to the bow and arrow, the blowpipe, and fishing by line and hook, a number of different types of traps were used in hunting for protein foods. One of the most widespread and common of these is a conical wickerwork cage trap about one meter in length with a mouth of converging bamboo splints that can be pressed open by an entering fish seeking the bait inside, but cannot be pressed open from within. Very similar forms of this trap are found over much of the AN-speaking world. Since closely similar traps are also found in some other parts of the world a consideration of the distributional evidence alone does not rule out the possibility of independent invention. Here linguistic evidence plays a crucial role, as almost all AN-speaking groups that use this type of trap call it by a reflex of PAN *bubu: Kavalan *bubu*, Amis *fofo*, Tagalog *búbo*, Kelabit *bubuh*, Bintulu *buvəw*, Malagasy

vovo, Malay *bubu*, Javanese *wuwu*, Makasarese *buwu*, Palauan *bub*, Hawu *wuwu*, Roti *bufu*, Yamdena *bubu*, Asilulu *huhu*, Lou *pup*, Raluana *vup*, Manam *u*, Arosi *huhu*, Pohnpeian *uu*, Chuukese *wuu*, Fijian *vuvu*.

Given this cognate distribution there is no question that PAN speakers made and used this type of trap, and that some of their descendants took it with them into the Pacific as far as Polynesia (although surprisingly, given its stability in other parts of the AN world, it appears to be called by a number of lexical innovations in Polynesia). Like the bow, it is unlikely that direct archaeological evidence of the *bubu fish trap will ever be found in early settlement contexts, but in view of the linguistic evidence such documentation, while welcome, would clearly be redundant.

The bamboo nose flute

In some cases the linguistic evidence for an item of prehistoric material culture is tenuous, but undeniable. Words for the traditional bamboo nose flute are somewhat variable in shape, but clearly point to the use of this musical instrument by speakers of PAN, and to its transport into the Pacific at least as far as Fiji. This is seen in Kavalan *tulani*, Ilokano *tuláli*, Toba Batak *tulila*, Bare'e *tuyali*, Tae' *tulali*, Fijian *dulali*. The Kavalan word points to PAN *tulani or *tulaNi, and all other forms to PMP *tulali (with irregular *n/N > *l), if vocalic metathesis is assumed in Toba Batak *tulila*. Some dictionaries gloss this only as 'flute'. Where more information is available it appears as 'nose flute', and where maximal information is available it becomes clear that it is a bamboo nose flute. Fijian *dulali* is the only known reflex in any Oceanic language, but it provides unmistakeable evidence that the 'Lapita peoples' who left a highly visible trail of pot sherds in their colonisation of the Pacific also played the bamboo nose flute.

5.11.3.2 Historical linguistics and social anthropology

While the relationship between historical linguistics and archaeology has blossomed in recent decades (at least in some parts of the archaeological community), the relationship between historical linguistics and social anthropology has been more distant. Nineteenth century anthropology was characterised by three salient traits: 1) it was comparative, 2) it often incorporated implicit assumptions about unilineal evolution, and 3) it relied on second-hand data collected by non-professional observers. Beginning with the work of Malinowski in the second decade of the twentieth century, fieldwork by trained professionals came to replace second-hand data. At about the same time the Eurocentric nature of thinking regarding social evolution in most nineteenth-century work became a source of intellectual embarrassment, leading to a general rejection of evolutionary notions in social thought until the issue was revisited with more sophisticated tools several decades later. This still left open the possibility for meaningful comparative work. However, various excesses in attempts to reconstruct social and cultural history on the basis of ethnographic data without a generally accepted method (e.g. Rivers 1914) led to an abandonment of 'conjectural history' by A.R. Radcliffe-Brown and subsequent social anthropologists. Although some social and cultural anthropologists maintained an interest in history, the general view was that reliable results are possible only were documentary evidence is available. In the absence of a generally accepted comparative method in anthropology inferential history thus became 'conjectural history' or 'pseudo history'. As a result of this deeply-ingrained attitude, few social and cultural anthropologists are trained in historical linguistics, or prepared to understand how historical inference works with

linguistic data. Moreover, when they try to use historical inference in dealing with comparative data, there is a pervasive confusion of typological and genetic comparison, with resulting errors in the inferences reached. This problem was recognised in Dyen and Aberle (1974), where an attempt was made to marry the methods of historical linguistics with those of social anthropology in order to avoid substituting typological comparison for genetic comparison. Unfortunately, however, this book in turn was plagued by a failure to distinguish lexical reconstruction from semantic reconstruction (Blust 1987b).

As Pawley (1982:39) has noted, "Typological resemblances across culture are open to at least three different historical explanations: parallel evolution (convergent development), borrowing, and common origin… Structural resemblances between social institutions, unless underpinned by cognate terminologies, cannot be proved to be of common origin." This methodological point is fundamental, but to date few social or cultural anthropologists have shown an understanding of the issues involved in drawing valid historical inferences regarding the social organisation of prehistoric societies (see e.g. Marshall 1984, and the accompanying discussion). A notable exception was the social anthropologist Per Hage, who combined kinship theory and historical linguistics in a number of penetrating studies of early Austronesian social organisation, including Hage and Harary (1996), Hage (1998), Hage (1999) and Hage and Marck (2003). A few sample issues that have been the subject of controversy in the literature are given here.

Hereditary rank

From a purely distributional standpoint a strong case can be made for hereditary rank in AN-speaking societies. Hereditary chieftainship, usually based on genealogical seniority, is reported for the Paiwan and Rukai in Taiwan, for various parts of Indonesia, for most Micronesian and Polynesian societies, and for scattered regions in Melanesia reaching from the Mekeo of southeast New Guinea to New Caledonia, the Loyalty Islands and Fiji. Moreover, a system of three social ranks (nobles, commoners, slaves) was common to many Bornean societies, to Nias, to a number of societies in Timor and other parts of eastern Indonesia, and to the traditional Chamorro. However, without cognate terms, where the arbitrary association of sound and meaning almost always rules out convergence as a likely explanation of similarity, such resemblances could be ascribed to parallel developments from an ancestral society that lacked hereditary status distinctions. To date little linguistic material has been found that could help to resolve this issue on the PAN or PMP levels. However, Pawley (1982) presented evidence that Proto Polynesian *qariki 'hereditary noble' continues a distinction between POC *qa-lapa 'chief' ('the big one') vs. *qa-riki 'eldest son of a chief' ('the little one'). Some aspects of this argument were questioned by Lichtenberk (1986), but together with other reconstructions such as *mʷala 'commoner' ('rubbish man' in Pacific pidgins), it is difficult to avoid the conclusion that several terms representing early stages of Oceanic linguistic history refer to distinctions of hereditary rank. Until relevant linguistic comparisons are discovered, however, the reconstruction of hereditary status distinctions in AN-speaking societies cannot be pressed further back in time.

Descent and marriage

Synchronically, the analysis of kinship is the province of the ethnographer or ethnologist rather than the linguist. In fact, the structural analysis of kinship systems is indifferent to the linguistic content of the terms, which could be represented by purely

abstract markers, such as letters of the alphabet (and in pedagogical texts often are designated by kin category abbreviations such as M = mother, MB = mother's brother, F = father, FZ = father's sister, or MBD = mother's brother's daughter). All that truly matters in synchronic analyses is how distinct genealogical relationships are conflated under common linguistic labels. The phonemic composition of the labels themselves plays no role at all in the outcome. Diachronically, however, the matter is totally different. Whereas the synchronic analysis of kinship terminologies is concerned exclusively with structure, the diachronic analysis is concerned primarily with content, since it is the reconstructed terminology which forms the sole basis for valid inferences about change. As noted already, this critical distinction has caused no end of confusion among cultural anthropologists, who often believe that it is possible to base historical inferences about change in kinship systems on a direct comparison of the structures without reference to the linguistic labels that mark them.[56] Synchronically, then, linguists usually have little to say: the analysis of attested terminological systems is the domain of the ethnographer. Diachronically, however, the situation is reversed, as the comparison of kinship terminologies among linguistically related peoples, and their use to reach historical inferences depend critically on use of the comparative method of linguistics. One particular subset of kinship terminology turns out to have far-reaching implications for the history of social organisation in AN-speaking societies.

   Proto Malayo Polynesian had four sibling terms: 1) *kaka 'elder same sex sibling', 2) *huaji 'younger same sex sibling', 3) *ñaRa 'brother of a woman', 4) *betaw 'sister of a man' (Blust 1993d). The parallel/cross distinction in these terms is important, as it is associated in large data samples with ancestor-based descent groups, and especially with matrilineal descent (Murdock 1968). To the extent that these associations have been established with adequate statistical controls, then, the reconstruction of a relative sex parameter for PMP sibling terms makes it very likely that PMP society was organised around ancestor-based descent groups rather than bilateral kindreds or other possible types of descent principles. This is a striking example of how the comparative method of linguistics may interact with generalisations in social anthropology to yield historical inferences that do not follow directly from the linguistic data.

   Although these four sibling terms must be reconstructed for PMP, they have been replaced in widely separated languages by terms that translate literally as 'male/female' or 'male/female child'. Subgrouping considerations and details of form make it clear that these secondary terms are the result of a number of historically independent innovations—in effect, a drift. From a linguistic standpoint the cross-sibling substitution drifts, as they may be called, are mysterious, as they suggest a priori that there were alternative ways to designate cross-siblings, but not parallel siblings. Moreover, these secondary terms are descriptively opaque, as seen in Table 5.32; Z (m.s.) = sister, man speaking, B (w.s.) = brother, woman speaking, F = female, M = male, C = child, x = cross-sibling:

---

56   Classic examples of this kind of confusion more than three decades apart are seen in Murdock (1949:323ff, and Marshall 1984), where historical inferences about change in kinship systems are based on a 'least moves' approach relating structural types, without any reference to cognate relationships among the terms themselves.

Table 5.34 The cross-sibling substitution drifts in Austronesian languages

	B (w.s.)	Z (m.s.)	Literal meaning
PMP	*ñaRa	*betaw	none
(A)			
Bontok	ka-lalaki-an	ka-babai-an	M/F
Maranao	laki	bəbay	M/F
Tiruray	lagəy	libun	M/F
Middle Malay	moanay	kəlaway	M/F
Sangir	mahuane	bawine	M/F
Mongondow	lolaki	bobai	M/F
Sika	nara	wine	none/F
Solorese	naa	bine	none/F
Bileki	hata male	hata vile	xM/xF
Erromangan	man	veven	M/F
Chuukese	mʷääni	feefiney	M/F
(B)			
Malagasy	ana dahy	ana bavy	CM/CF
Tae'	anak muane	anak dara	CM/CF[57]
Kambera	ana mini	ana wini	CM/CF

It is not obvious why *ñaRa and *betaw would be replaced repeatedly by terms that translate literally as 'male/female' or 'male/female child'. First, nothing similar is known for the parallel sibling terms. Second, although brothers are male and sisters female, this appears irrelevant to the changes, since the terms themselves do not refer to absolute sex, but rather to relative sex. Likewise, in cross-cultural perspective it is highly unusual for brothers and sisters to call each other by terms with the literal meanings 'male/female' or 'man/woman', which are far more commonly applied to spouses than to cross-siblings. Third, a literal translation of the set B terms as 'male child', or 'female child' is obscure. Linguistic analysis is insufficient to make sense of this pervasive set of innovations, although one point should be noted. Reflexes of PAN *aNak often mean not only 'child', but also 'smaller part of a larger whole', or 'member of a group', as in Malay *anak buah* 'followers; tribesmen; clansmen', Toba Batak *anak bua* 'subjects, serfs', Simalur *anaʔ banwo* 'subject (of state), serf; native of a place', Old Javanese *anak wanwa* 'person belonging to the *wanwa* community', Makasarese *anaʔ bua-na paʔrasaŋaŋa* 'inhabitant of an area', Ngadha *ana nua* 'villager', or Erai *ana hira* 'the children, often used in the meaning of: villagers, village population'.

There is no need to enter into extensive details here, but as noted in Blust (1993d), the key to understanding the cross sibling substitution drifts derives from the form of marriage in a number of culturally conservative societies in northern Sumatra and eastern Indonesia, where systems of political alliance are based on asymmetric exchange, such that descent group A provides wives to B, B provides wives to C, and C provides wives to A. In such systems wife-givers are symbolically 'male' (superior), and wife-takers 'female' (inferior), a distinction reflected in the terminology for wife-giving and wife-taking groups (literally 'male/female' or 'male/female child'. Given this terminology the historically secondary cross-sibling terms become intelligible as replacements of the cross-sibling terms by terms for 'wife-giver', and 'wife-taker', where the brother-sister relationship was viewed from

---

[57]   Tae' *anak dara* is literally 'child + maiden/virgin'.

the standpoint of their children's marital expectations (the sister's son being expected to marry the brother's daughter). The cross-sibling substitution drifts, then, have a double theoretical interest. On the one hand they provide linguistic evidence that PMP society practiced asymmetric exchange, and that this system persisted in many areas outside northern Sumatra and eastern Indoneisa until after the historically secondary cross-sibling terms had been innovated. On the other hand, they provide the first well-documented case of a linguistic drift that is powered not by features of language structure, but rather by features of social organisation.

6   *Morphology*

6.0   Introduction

In many AN languages it is difficult to separate morphology and syntax, since key information regarding participant roles, tense-aspect and the like is marked by verbal affixation. This is especially true of 'Philippine-type' languages, which have elaborate systems of verbal and nominal morphology. However, even in these languages many affixation processes, (including reduplication) function mainly in word-formation, and so can be treated without reference to the larger syntactic contexts in which they occur. In addition, many AN languages in insular Southeast Asia show abundant evidence of submorphemic sound-meaning correlations. These provide an important addition to the general literature on phonesthemes, and probably are best discussed in the context of how morphemes are identified and classified. Finally, many AN languages contain fossilised or semi-fossilised affixes which fall outside the range of phenomena usually addressed by general linguistic theory, and these raise important questions about the interrelationship of language and cultural presupposition. A separate treatment of morphology in AN languages thus appears to be justified. This chapter has the following general structure: 1) morphological typology, 2) submorphemes, 3) affixes important for word formation, 4) circumfixes, 5) ablaut, 6) suprasegmental morphology, 7) zero morphology, 8) subtractive morphology, 9) reduplication, 10) triplication, 11) compounding, and 12) morphological change.

6.1   Morphological typology

Before considering specific affixes and their interrelations it will be useful to characterise the morphological traits of AN languages as a whole. Because this discussion touches on issues in general linguistics that are familiar to scholars outside the AN field, they will require only a brief discussion here. In terms of the classic typology of agglutinative vs. fusional, and isolating vs. synthetic vs. polysynthetic systems of morphology, most AN languages can be characterised as agglutinative-synthetic. In other words, there is a relative abundance of affixes (especially in Philippine-type languages, and in some non-Philippine-type languages of Borneo and Sulawesi), and at the same time the morpheme boundaries are usually clear. This is illustrated in Table 6.1 for Thao of central Taiwan, Ilokano of the northern Philippines, Kadazan of Sabah, and Makasarese of southwest Sulawesi:

Table 6.1 Agglutinative-synthetic morphology in four sample languages

Base	Affixed forms
Thao (Blust 2003a)	
danshir 'scarecrow'	pia-danshir-i 'put up a scarecrow (imper.)'
	p\<in>u-danshir-an 'was protected by a scarecrow'
kan 'eat'	kilh-a-kan-in 'search for food'
	p\<in>a-ka-kan-ak 'I used (it) to feed (something)'
lhufu 'embrace'	m\<in>apa-lhufu 'embraced one another'
	pash-lhufu-an 'place where one broods or sits on eggs'
parbu 'bake, roast'	p\<in>arbu-rbu-an 'place where something was baked or roasted'
	parbu-n 'be baked by s.o.'
qtu 'by surprise'	k\<in>ilh-qtu-ak 'I happened upon it (after searching)'
	q\<m>\<iN>-qtu 'found something by accident (stumbling upon)'

Note: pia- = causative of stative verbs; -i = imperative; pu-X -an = circumfix usually meaning 'wear X'; -in-/-iN- = perfective; kilh- = 'search for, seek'; a- = ?; -in/-n = patient voice; pa- 'causative of dynamic verbs'; full reduplication = repetitive; -ak = 1sg; mapa- = reciprocal; (pash-) X -an = 'place where X is performed; suffixal reduplication = repetitive; -m- = actor voice

Base	Affixed forms
Ilokano (Rubino 2000)	
báyad 'pay'	maka-bayád-en 'able to be paid'
	ba-bayád-en 'remainder of a debt'
bayág 'late rice'	i-bay-bayág 'procrastinate'
	maka-pag-bayág 'able to stay long'
liwá 'amusement'	liw-liwa-én 'to comfort, console, cheer up'
	pag-pal-pa-liwa-án 'comfort, solace; amusement'
saludsód 'question'	m\<ann>aki-saludsód-da 'they are always asking questions'
	pag-saludsúd-an 'ask s.o.'
túrog 'sleep'	ma-turóg-an 'fall asleep while on watch'
	m\<in>aka-pa-túrog 'something that caused s.o. to sleep'

Note: maka- = intransitive potentive; -en = patient voice; CV reduplication = ?; i- = theme; CVC- reduplication = durative; pag- = instrumental; -an = nominaliser; maki- = social prefix; -ann- = gerund (manner); -da = 3pl; -an = directional; ma- + -an = involuntary action; -in- = perfective; pa- = causative

Base	Affixed forms
Kadazan (Kadazan Dusun Cultural Association 1995)	
giot 'tighten'	k<in>o-giat-an 'act of having tied tightly'
	p<in>oki-giat-an 'was asked to tighten something'
otuŋ 'fall on s.t.'	ko-po-otuŋ-an 'act of making something fall on s.t.'
	mod-tig-ko-otuŋ 'almost fell on s.t.'
patay 'finish'	ka-pa-mataz-an 'act of killing'
	mod-tig-ka-patay 'at the point of death'
tukod 'support'	k<in>o-tuu-tukod 'reason for having supported something'
	p<in>o-po-tukod 'caused or allowed to support'
vanit 'poison'	ko-vonit-an 'act of poisoning'
	m<in>a-manit 'poisoned'

Note: ko/ka- = nominaliser; -in- = perfective; -an = referent voice; locative; poki- = petitive; po/pa- = causative; mod- = affected by; tig- = immediate; ko- = uncertainty; ka/ko-X-an 'act of Xing'; CVV reduplication = iterative

Base	Affixed forms
Makasarese (Cense 1979)	
kanre 'eat; food'	aŋ-ŋanre 'to eat'
	paŋ-ŋanre-aŋ 'place where one eats'
mate 'die, dead'	an-tu-mate-aŋ-i 'take responsibility for someone's burial'
	na-ka-mate-aŋ-i 'will die because of something'
moroŋ 'sit'	na-passi-moroŋ-aŋ 'will sit together (ceremonially, of bride and groom'
	pa-si-moroŋ 'let bride and groom sit down together (invite them to sit)'
pinawaŋ 'follow'	na-na-pa-pinawan-toŋ 'he took (something) along'
	pam-minawaŋŋ-aŋ-ku 'follow because of me'
tawa 'share'	at-tawa-əŋ-i 'a half share of this, and a half share of that'
	pat-tawa-tawa-əŋ 'share or portion of s.t.'

Note: aŋ/an- = active verb; (paŋ-)X-aŋ = place where one X; tu- = someone who; -i = 3p; locative; na- = future; passi- = reciprocal?; pa- = causative; si- = reciprocal; na- = 3p; na- = ?; -toŋ = also, in addition; paŋ-X-aŋ = ?; -ku 1sg; at- = ?; -əŋ = passive nominalisation; -i = intensifier?; pat-X-əŋ = nominaliser; full reduplication = distributive

These examples give some idea of the complexity of affixation processes in Malagasy, Palauan, Chamorro, and many of the languages of Taiwan, the Philippines, and western Indonesia. At the same time they raise a number of questions. Despite the heavy use of affixation morpheme boundaries generally are clear, and are marked by a hyphen. Angled

brackets are used to set off infixes, since otherwise the discontinuous parts of bases might appear to belong to different morphemes, but this is a matter of visual representation rather than a cognitive or analytic obscurity. As in any typological category, however, 'pure' types are difficult to find, and there are some conditions under which morpheme boundaries are difficult to determine. Kadazan Dusun *ka-pa-mataz-an* 'act of killing' (base: *patay*) shows a regular phonological alternation between final *-y* and intervocalic *-z-*, but this does not affect recognition of the boundary. However, where phonological alternations do not involve a one-to-one phoneme correspondence boundaries may be blurred. This is often seen in connection with the Malagasy passive suffix *-ina*, as with *lefitra* 'endurance, bearing with' : *man-defitra* 'to endure', but *lefer-ina* 'be endured, be borne', *zaitra* 'needlework, stitching' : *man-jaitra* 'to sew, to stitch', but *zair-ina* 'sewn, stitched', *iraka* 'messenger' : *man-iraka* 'send a messenger', but *irah-ina* 'be sent as a messenger', or *tafika* 'invasion, plundering' : *ma-nafika* 'to invade', but *tafih-ina* 'be invaded, be attacked'. Here, as in Dusun, there are consonant alternations (*tr* is a single phoneme), but in addition non-suffixed bases that end with *-a* often drop the final vowel before *-ina*. Historically this alternation reflects the addition of *-a* after pre-Malagasy final consonants, and where *-a* was part of the base, no alternation occurs, as in *buka > voha* 'open' : *ma-moha* 'to open' : *voha-ina* 'be opened', or *panas > fana* 'heat' : *ma-mana* 'to heat, warm up' : *a-fana-ina* 'be warmed up'. Synchronically, these alternations present problems typical of fusional languages (e.g. Romance): where a morpheme boundary is imposed to isolate a base or affix of invariant shape it leaves a morphological residue that does not occur independently in the language. The principal difference is that in fusional languages of the Romance type single phonemes may be portmanteau morphemes, while in Malagasy the stem that co-occurs with *-ina* does not match the stem found under prefixation. If the historical supporting vowel *-a* is treated as part of the synchronic phonology this problem disappears, but given a sufficiently abstract level of phonological representation this solution probably could be used to eliminate fusional morphology in most languages.

Malagasy morphology has more fusional properties than most AN languages, but is hardly unique. Many languages in the Admiralty Islands show fusional characteristics in possessive noun morphology, or substantial phonetic variation in the form of suffixed nouns, as with Nali *moro* 'my eye' : *mara-m* 'your eye' : *mara-n* 'his/her eye', or Levei *moto-k* 'my eye' : *muto-ŋ* 'your eye' : *mʷato-ŋ* 'his/her eye'. Other languages present global difficulties in determining morpheme boundaries, as with Palauan, where the problem of trying to reconcile verb forms such as *ʔəsimáll* 'is to be turned, wound, screwed' and *məŋəsóim* 'to turn, wind, screw', or *ʔiltutíi* 'has put on headwear' and *məŋətíut* 'to put on headwear' is commonplace. Pre-Palauan evidently had a very rich morphology, and this language underwent a large number of consonant shifts and stress-sensitive vowel reductions or deletions. Consequently, although several theoretically sophisticated studies of Palauan syntax exist, there still is no adequate description of the morphology of the language, and the fairly substantial dictionary of McManus and Josephs (1977) lists hundreds of affixed forms as dictionary entries, apparently because the authors were unsure of the shape of the base. In terms of what Whaley (1997:133) calls the 'index of fusion', then, AN languages tend to lie toward the agglutinative end of the continuum, although individual languages that have undergone exceptionally complex sound changes may show more fusional traits.

The most pervasive morpheme boundary problem in AN languages is found with certain nasal-final prefixes. As seen in Chapter 4, most languages of the Philippines and

western Indonesia, as well as Malagasy, Palauan and Chamorro, show homorganic nasal substitution under prefixation with reflexes of *maŋ- and *paŋ-: Malay *pukul* : *mə-mukul* 'to hit', *tanam : mə-nanam* 'to plant', *satu* 'one' : *mə-ñatu-kan* 'to unite', *kupas* 'shelling' : *mə-ŋupas* 'to shell, to husk'. Since this prefix invariably surfaces as *məŋ-*before vowel-initial bases (*ambil* 'taking over; receiving' : *məŋ-ambil* 'to fetch', *idam* 'morbid craving' : *məŋ-idam* 'to crave s.t.'), it is clear that words such as *mə-mukul* are underlyingly *məŋ-pukul*, the surface form resulting from place assimilation of the nasal and deletion or replacement of a voiceless obstruent (in Malay, but not all AN languages, bases with initial voiced obstruent show only the first of these processes, as with *bantu* 'help' : *məm-bantu* 'to help'). Since place assimilation is conditioned by an obstruent that deletes, the morpheme boundary is stranded between prefix and base: *məm-ukul* preserves the nasal in the prefix where it originates, but leaves only a partial base, while *mə-mukul* preserves the canonical form of the base and the place features of the initial consonant, but leaves a prefix without the nasal that triggered substitution. This type of boundary problem differs from classic discussions of fusional languages, where the imposition of a morpheme boundary sometimes leaves non-occurring residues, but rarely if ever requires the analyst to posit a morpheme boundary within a phoneme.

No adequate statistical data is available on morpheme:word ratios of the sort carried out by Greenberg (1954), but it is clear from Table 6.1 that many AN languages allow two, three or even more affixes to co-occur in the word-formation process. Where this process combines infixation with prefixation, suffixation, and perhaps reduplication the results can be quite complex. Some of the examples in Table 6.1 show elements of polysynthesis—that is, evidence that an entire sentence can be expressed as a single word, as in Ilokano *m<ann>aki-saludsód-da* (petitive/social prefix *maki-* + frequentative infix *<ann>* + noun 'question' + **3p**) 'they are always asking questions', or Makasarese *na-na-pa-pinawan-toŋ* (3rd person subject **na-** + linking element **na-** + causative **pa-** + verb base 'follow' + modifier roughly = 'also') 'he took (something) along'. Such examples are natural results of elaborate systems of affixation, but it should be kept in mind that the great majority of sentences in all of these languages must contain more than one word. The difference between the use of polysynthesis in true polysynthetic languages and in languages such as these is ultimately a matter of frequency: in languages such as Yupik or Oneida sentence-level words are the norm, while in AN languages—even those with the most elaborate systems of affixation—sentence-level words occur only occasionally.

Not all AN languages have such rich systems of affixation. Most languages of eastern Indonesia and the Pacific have much less elaborate systems of affixation than is typical of Taiwan, the Philippines or western Indonesia. Elbert and Pukui (1979:64ff), for example, give fewer than 25 affixes in Hawaiian (they list more, but many of these are clearly variants of the same morpheme), whereas Thao of central Taiwan has about 200 (Blust 2003a:91ff), and Rubino (2000:xviii) lists over 400 for Ilokano. Typical Oceanic languages differ from those of Taiwan and the Philippines not only in size of affix inventory, but also in the index of synthesis, that is the extent to which a language 'piles up' affixes on a single base. In Hawaiian, for example, morphologically complex words typically contain a single affix, as with *ʔai* 'eat' : *aka-ʔai* 'eat slowly', *ola* 'life' : *pā-ola* 'quick recovery', or *ikaika* 'strong, powerful' : *hoʔo-ikaika* 'strengthen; make a great effort'. Bases with two affixes are possible but not common, and bases with three or more affixes such as those seen in Table 6.1 are virtually unknown. A similar situation holds for Polynesian languages as a whole, and for most other Oceanic languages.

Although Central-Eastern Malayo-Polynesian languages appear to have the lowest index of synthesis, some of the most extreme cases of morphological impoverishment are found further west. The Chamic languages of mainland Southeast Asia, for example, have severely reduced affix inventories as a result of extensive typological adaptation to their Mon-Khmer neighbours, most of which make limited use of morphology. Thurgood (1999:238ff) presents evidence for just five affixes in Proto Chamic: 1) *pə- 'causative', 2) *tə- 'inadvertent action', 3) *mə- 'verb prefix', 4) *-ən- 'instrumental', and 5) *-əm- 'nominaliser'. Not all of these are found in every daughter language, and even where they are preserved their use may be restricted (*pə- 'causative' and the two infixes appear to be by far the most productive). A similar level of morphological impoverishment is found in many languages of eastern Indonesia. Verheijen (1977) states that Kambera (eastern Sumba) has just four prefixes, *a-*, *ka-*, *ma-* and *pa-*, but Klamer (1998:26, 266ff) lists *ha-*, *ka-*, *la-*, *ma-*, *pa-*, *ta-* and prenasalisation (*pata* 'break X' : *mbata* 'be broken, as a chair', *tutu* 'stay close to X' : *ndutu* 'follow X'). Infixes and suffixes evidently are absent. Verheijen notes that Bimanese has just two active prefixes, *ka-* 'causative' and *ma-* 'active participle', although some others can be identified by internal reconstruction. Walker (1982) describes Hawu (Sawu) morphology as minimal: reduplication and *he-* 'one' with nouns, causative *pe-*, reciprocal *pe-* and reduplication with verbs. The most extreme morphological impoverishment is found in some of the languages of western and central Flores. Verheijen (1977) claims that Manggarai has no affixes at all, and suggests that the same probably is true of Ngadha and Lio. Baird (2002) confirms this for Kéo of east-central Flores. Unlike the situation in Chamic, the isolating morphological typology in this area cannot easily be explained by contact. As Verheijen notes, there are remnants of earlier morphological processes in these languages, suggesting that the use of bound morphemes to signal grammatical information declined gradually throughout the region, reaching its nadir in the languages of central and western Flores. By contrast, Manggarai and other languages of western Flores are lexically very conservative (Blust 2000a:329). In terms of what general typologists call the 'index of synthesis', then, AN languages tend to be synthetic, although elements of polysynthesis appear in heavily affixing languages, and some languages approach the isolating end of the typological continuum.

6.1.1 Types of morphemes

The discussion so far has been concerned primarily with morphological processes, but there is also a need to address the classification of morphemes into types. In discussing this and other related material I will use the terms 'base' or 'stem' to refer to independent morphemes that are capable of being affixed, and will reserve the term 'root' for a smaller, submorphemic unit that is defined by recurrent association but not by contrast. In quoting material from other writers the term 'root' may appear as equivalent to my use of 'base' or 'stem'.

For reasons of space no attempt can be made here to survey the entire range of proposals about types of morphemes in AN languages, and the following discussion is clearly incomplete. However, in addition to bases a number of writers recognise three other categories of morphemes, namely clitics, particles and affixes. Not all morphemes in all languages fall clearly into one or another of these categories, and, as shown below, the categories themselves are sometimes problematic. Nonetheless, these are among the most frequently mentioned types of morphemes, and they provide what is probably the best basis for a general discussion of morphological typology in AN languages.

Bowden (2001:85) recognises four categories of morphemes in Taba of south Halmahera: Words (bases), clitics, particles and affixes. He defines these categories in terms of a matrix that makes reference to morphosyntactic independence (MSI) on the one hand, and phonological independence (PI) on the other. Figure 6.1 converts this matrix to a tabular format using binary oppositions:

	MSI	PI
Words	+	+
Particles	-	+
Clitics	+	-
Affixes	-	-

Figure 6.1: Matrix defining morpheme types in Taba (after Bowden 2001)

Bowden proposes the following 'defining characteristics' for these morpheme types: 1) a word attracts primary stress and can occur as a free form, 2) a particle attracts primary stress but cannot occur as a free form, 3) a clitic never attracts stress and attaches itself to phrases, and 4) an affix never attracts stress and attaches itself to words or roots. Although this schema is proposed only for Taba, it provides a useful springboard for discussing types of morphemes in AN languages more generally.

The distinction between word and affix is not problematic, but disagreements arise in distinguishing clitic from affix and particle. Some scholars do not recognise particles, which they see as irrelevant to determining word classes. Bowden's usage avoids this problem, as it treats particles as elements in a family of morpheme types without reference to word classes.

In general, atonic monosyllabic pronouns that attach to an independent word are treated as clitics rather than affixes, although the basis for this decision is rarely explicit. Klamer (1998:27) states that Kambera "has enclitics marking aspect and mood, proclitics marking coordination and subordination, and pronominal pro- and enclitics." To this she adds the statement "I consider the Kambera conjunctions, prepositions and articles, and the negation *nda*, to be (phonological) clitics as well, because these items do not conform to the minimal word requirement … and only occur with a phonological/syntactic host." However, this definition does not serve to distinguish clitics from affixes. To distinguish clitics from affixes she appeals to distributional criteria. Both pronominal clitics and modal and aspectual clitics, for example, are said to attach to a phrase rather than to a specific morpheme as affixes do, and some specific clitics are further distinguished from affixes in that "their prosodic host differs from their syntactic host." This analysis agrees closely with Bowden's distinction, and undoubtedly applies to many AN languages. It is, however, not problem-free as a set of guidelines for all languages, and vacillation over whether to classify postposed possessive pronouns as clitics or suffixes is not uncommon in the literature (e.g. Himmelmann 2001:92, fn. 35 where pronominal clitics and suffixes have the same shape, but the former may be separated from the stem by an epenthetic *o*).

Bowden distinguishes clitics from affixes in that clitics are morphosyntactically independent, while affixes are not. Needless to say, morphosyntactic independence is a matter of degree. Because most affixes probably begin as independent words that become phonologically compressed and dependent on adjacent morphemes, there are stages in the historical development of any language in which the distinction between clitic and affix is

likely to be unclear for particular morphemes, though not necessarily for all. In Thao of central Taiwan, for example, the base *kan* 'step, tread; pace, go, come' is clearly an independent word in expressions such as 1) or 2):

1.   *kan ta-tusha yamin*
     step **hum**-two we.**excl**   'The two of us are going'

2.   *cicu kan (ma)-mamuri*
     he step alone   'He is coming by himself'

Yet in expressions such as 3) - 5) *kan* is closely bound to the morphemes *qca* and *tup*, which never occur alone:

3.   *m-ihu a kuskus kan qca-k*
     your **lig** foot step.on-I   'I stepped on your foot'

4.   *kan tup-ik cicu*
     follow-I him   'I'm following him'

5.   *pasay-in cicu qnuan pa-kan qca diplhaq*
     use-**pv** him carabao **caus**-step.on mud   'He used buffaloes to trample the mud'

The morpheme sequences *kan qca* and *kan tup*, then, are essentially words meaning 'step on', and 'follow' respectively, and in these words *kan* is no less bound than *qca* or *tup*. Although frequency data is not available, it seems impressionistically that the free morpheme *kan* is most common in constructions such as *kan qca* and *kan tup*, in which it is bound. In effect, the independent verb base *kan* appears to be losing its independence in specific environments, making it an independent word in some expressions, but a bound element in others. While 'cliticisation' normally is applied to functors that become phonologically attached to free content morphemes, the free morphemes /kan/ and /tup/ apparently are becoming bound to one another in a fixed expression, and hence are in a generalised sense 'cliticised' to one another.

With regard to the clitic/affix distinction, the claim that clitics attach to phrases and affixes to words may work with many languages, but is difficult to reconcile with Thao examples such as 6), where /min/- attaches to the following morpheme, next to 7), where /min/- attaches to the possessive phrase *nak a hulus*:

6.   *mim-binanauʔaz iza cicu*
     become-woman already she   'She has already grown into a woman'

7.   *takcat-i i-zay a maqusum a maqa a min-nak a hulus*
     cut-**imp** this **lig** cloth **a** in.order.to **a** become-my **lig** clothes   'Cut this cloth to make clothes for me' (lit. 'Cut this cloth to become my clothes')

Thao *min-* 'inchoative' can thus be called an affix when it attaches to a following morpheme, but a clitic when it attaches to a following phrase. Similarly, *sa* functions like a particle when it is preposed to a noun or pronoun in sentence-initial position, as in 8), but

like a postclitic in some environments when it is destressed, loses the vowel, and become phonologically attached to the preceding morpheme, as in 9):

8.   *sa azazak a m-ihu ma-cuaw ma-ania*
      sa child **lig** your very intelligent 'As for that child of yours s/he is very intelligent'

9.   *yaku sa p<in>acay sa i-zaháy a shput* [jakus pinaθajs iðaja ʃpu:t]
      I **sa** was-beaten **sa** that **lig** person 'I'm the one who was beaten by that person'.

Examples of completed change in the categorisation of the same morpheme are also visible in some historical data. In Malay prepenultimate vowels within a word have normally merged as schwa (written *e* in the standard orthography). This merger has affected the vowels of prefixes (*maŋ- > *məŋ- 'active transitive verb', *maR- > *bər- 'active intransitive verb', *taR- > *tər- 'unintended action', *ka- > *kə- 'ordinal numeral', etc.). In Malay the morpheme *si* marking personal names or words used as personal names (nicknames, etc.) is best described as a particle, since in collocations such as *si Ahmad* 'Ahmad', or *si Gəmuk* 'Chubby' the vowel is unchanged, and the same is true where the morpheme boundary is less apparent, as with *siapa* 'who?' (cf. *apa* 'what?'). In Mukah Melanau of coastal Sarawak, which has also merged prepenultimate vowels as schwa, the cognate marker of personal names, which appears to function essentially the same as in Malay, is *sə-*, as in *sə Tugau* 'Tugau' (a Melanau culture hero). Both languages have merged prepenultimate vowels as schwa, and so have a synchronic constraint against vowels other than schwa in prepenultimate syllables, yet they treat the reflex of *si differently: Malay *si* is a particle, while Mukah *sə-* arguably is an affix.

Another example of completed change in morpheme category is seen in *ña '3sg possessor', from PMP *ni 'genitive of singular personal nouns' + *-a '3sg', with contraction (Blust 1977a). For contraction to occur the pronoun must have carried stress, as it still does in forms like Tagalog *niyá* '3sg. genitive'. Given Bowden's criteria for distinguishing particles from clitics a possessive phrase such as PMP *mata ni-á 'his/her eye' would contain a free morpheme *mata 'eye', but *ni 'genitive', which is normally considered a particle, would be classified somewhat awkwardly as either a clitic (although it does not attach to a phrase), or an affix (although it does not attach to a free morpheme), and the possessive pronoun *-a would have to be classified unconvincingly as a particle. After contraction to *-ña the new possessive pronoun was unstressed (*matá-ña), and so became an affix. Although the schema in Figure 6.1 probably carries some general validity for many languages, then, it does not appear unusual for category boundaries to be blurred or ambiguous. Since particles are rarely confused with affixes, the most problematic term in this set of four is 'clitic', a category that is arguably intermediate between particles and affixes, and that shares some properties of both.

The second problem found in many descriptions of AN languages is how to distinguish derivation from inflection. Klamer (1998:58) notes that "the intuitive understanding of this distinction has not (or not yet) resulted in objective criteria for a discrete division between derivational and inflectional morphemes in all languages." Nonetheless, she suggests that in general 'inflectional morphology is a way to express relations between syntactic constituents.' Under this interpretation the voice-marking affixes of Philippine languages would be inflectional since they clearly express relations between syntactic constituents. Other writers, however, have pointed out that the distribution of these affixes is far less predictable than e.g. the plural suffix *-s* on English nouns, and that for this and other

reasons they should be regarded as derivational (Starosta 2002). This issue can only be resolved by reaching agreement on how to weight the criteria that have been proposed in the general linguistics literature for distinguishing inflection from derivation: is e.g. productivity more important than change of word class?

The ability of an affix to change the word class of a base is often considered the most unambiguous characteristic of derivational morphemes. In languages like English this trait often co-occurs with low productivity, but in AN languages the productivity of an affix and its ability to change the word class of a base do not appear to be strongly associated. To illustrate, *-um-* 'actor voice' and *-in-* 'perfective' are highly productive in Philippine-type languages, and for this reason they might be considered inflectional. In many languages of the Philippines, however, *-um-* competes with *mag-*. For Tagalog Ramos (1971:57-58) describes the difference between these affixes as follows: 'The *mag-* affix indicates deliberation and comprehensiveness of the action. It usually has an added feature of transitiveness. Most of the sentences where *mag-* is used or where it contrasts with *-um-* have an obligatory object … The *-um-* affix, in comparison to the *mag-* affix, is more casual, involuntary, and suggests internal action. It is mostly intransitive because an object is not necessary to complete the sentence … *Mang-* has a special use indicating plurality or distributiveness of action or habitual, repeated action.'

In illustration she gives *mag-tayóʔ* 'to erect, build' vs. *t<um>ayóʔ* 'to stand'. This characterisation works for the examples given, but it is impressionistic, and does not agree well with examples such as *b<um>ilí* 'to buy', which hardly seems casual, involuntary or based on internal action. Some Tagalog verbs take both *mag-* and *<um>*, while others take only one or the other of these affixes (McFarland 1976). Since they are in competition for the same bases these two actor voice affixes reduce one another's productivity, making them appear more derivational than would otherwise be the case. This tendency is even more apparent in cross-linguistic perspective, since *-um-* has been all but completely replaced by a reflex of *maR- or *maŋ- in some languages. Malagasy, for example, usually marks the actor voice with *maŋ-* or *mi-* (*faka* 'a root' : *ma-maka* 'send forth roots', *leha* 'path, way' : *man-deha* 'go', *safotra* 'flooded' : *ma-nafotra* 'overflow, flood', *toto* 'act of pounding' : *mi-toto* 'to pound'). However, *-om-* still appears in a few verbs. In *tany* 'crying, lamentation' : *t<om>any*/*mi-t<om>any* 'weep, lament, complain' the infix is productive, but the redundant affixation of *mi-* suggests that *-om-* has little psychological force as an affix, and that the form *t<om>any* is therefore treated as a base. Even in Bikol, a close relative of Tagalog, the functions once expressed by *-um-* have been almost completely usurped by *mag-* (Lobel 2004). Neither *mag-* nor *-um-* typically triggers a change of word-class, and for this reason both affixes might be considered inflectional, yet the diminishing uses of *-um-* in some Philippine-type languages make this affix appear more nearly derivational.

The infix *-in-* presents a different set of problems for the inflection/derivation distinction. In most Philippine-type languages *-in-* is a highly productive marker of perfective aspect, as in Thao *k<m>an* 'eat' (actor voice) : *k<m><in>an* 'ate' (actor voice, perfective), or *mu-saran* 'go out onto the road' : *m<in>u-saran* 'went out onto the road'. In these contexts it clearly has the hallmarks of a classic inflectional affix. However, the same infix derives deverbal nouns, as in 10)

10.   *m-ihu a s<in>aran-an yanan sapaz*
      your **lig** walk-**perf-loc** have footprint   'The place where you walked has footprints'

Since only nouns may follow a possessive pronoun + ligature, such affixed forms cannot be verbs, and although *saran* 'path, road' is itself a noun, affixed forms like *s<in>aran-ak* are clearly verbal in constructions such as 11.

11. *s<in>aran-ak iza sa i-zahay saran*
    walk-**perf-1sg** already **sa** that path 'I went that way/on that path'

In conclusion, *-in-* acts sometimes like an inflectional affix and sometimes like a derivational affix. Data of this kind may lie behind Klamer's statement that the inflection/derivation distinction can be useful in languages like English, but in many AN languages trying to operate in these terms presents more of an encumbrance than a facilitation to analysis.

6.2. Submorphemes

The preceding section addresses general questions such as the morphological type of AN languages, including the indices of synthesis and fusion. The following section surveys specific particles, clitics and affixes that can be posited for early stages of AN, and their evolution in the modern languages. Before continuing, however, it will be worthwhile to consider a phenomenon that has long been recognised in the AN languages of insular Southeast Asia. The Swiss linguist Renward Brandstetter (1916) used the term 'Wurzeln' (roots) to describe submorphemic sound-meaning correlations that are common in many languages of Indonesia and the Philippines, and in at least some of the Formosan languages. These elements usually take the form -CVC, and can be illustrated by Malay disyllables that end in *-pit* (from Blust 1988a):

Table 6.2 Malay disyllables that end in *-pit*

1.	anak ampit	fighting fish: *Betta* spp.
2.	(h)apit	>pressure between two disconnected surfaces
3.	apit-apit	a wasp, species unidentified
	buroŋ apit-apit	hornbill: *Eurylaemus ochromelas*
4.	capit	>pincers
5.	mən-cepit	>to nip
	pən-cepit	>pincers
6.	dampit	deaf to warnings, obstinate
7.	dəmpit	>pressed together, in contact
8.	(h)əmpit	>pressure between two unconnected surfaces
9.	gapit	>nipper, clamp
10.	həmpit	shy, timid
11.	(h)impit	>squeezing pressure
12.	jəpit	>to nip, catch between pincers
13.	kayu kampit	name given to the reputed seal of Alexander the Great, or to the wood of which it was made
14.	kapit	>support on each side. Of a bridegroom's 'supporters' (*peŋapit*) at a wedding; fasten on with slats, as woven grass matting to a frame
15.	kapit	name of the sixth chicky suit (in cards)
16.	kə(m)pit	earthenware water jar
17.	kəmpit	>carry under the arm
18.	kəpit	>pressure between two connected surfaces
19.	lampit	sleeping mat
20.	lapit	>lining, thin partition
21.	limpit	>in layers
22.	lipit	>a fold or twist (of thread, cotton, etc.)
	kala lipit	the common house scorpion
23.	pipit	finch, sparrow
24.	ləsoŋ pipit	small dimple in cheek
25.	pipit	>mouthpiece of a whistle
26.	pipit	penis of a very young child
27.	rəmpit	strike with a whip or cane
28.	ripit	a sweetmeat
29.	məmpəlas ropit	a plant: *Tetracera* sp.
30.	səmpit	confined (of space); cramped, shut in
31.	səpit	>nipping, to nip
32.	səpit	a creeping herb: *Sesuvium portulacastrum*
33.	simpit	>narrow, confined
34.	sipit	>half-closed (of the eye)
35.	sumpit	shooting with a blowpipe
36.	sumpit	sack of matwork for holding rice
37.	su(m)pit	>chopsticks
38.	sumpit	>narrow, confined
39.	təmpit	cheer of encouragement

Table 6.2 lists all disyllables that end with *-pit* in Wilkinson (1959), a dictionary of 1,291 double-column pages. Morphemes that clearly refer to the approximation of two surfaces have been marked with >. This applies to 21 of the 39 forms, or more than half of all words in Table 6.2. In no case is a morpheme boundary present between the initial CV(C) and the final element *-pit*. The same type of sound-meaning correlation is also found in many other languages, as in Kavalan (eastern Taiwan) 1) *ipit* 'tongs for lifting hot coals; chopsticks', 2) *kaipit* 'close together; pinched or caught between', 3) *kapit* 'sew together accidentally, as the two legs of one's pants', 4) *kəpit* 'crowd two or more things together so that there is not much room between them', 5) *k<əm>ipit* 'pinched, caught (as in a closing door)', 6) *k<əm>upit* 'stuck or glued to s.t.', 7) *pitpit* 'pick off one-by-one, as betel nuts (by pinching and twisting)', 8) *qipit* 'pin s.t., as with a clothes pin', 9) *qupit* 'stick, adhere to a surface', and 10) *sipit* 'pinch and twist'. Blust (1988a) identified 231 roots of this type that are found in at least four etymologically independent (non-cognate) morphemes. Of these, *-pit* 'press, squeeze together; narrow' is the best-supported, with 48 etymologically independent attestations, followed by *-keC* 'adhesive, sticky' with 44, *-tik* 'ticking sound' with 38, *-tuk* 'knock, pound, beat' with 36, *-kaŋ* 'spread apart, as the legs' with 34, and *-pak* 'slap, clap' with 32). Several other scholars have written on this topic since 1988, in particular Nothofer (1990) and Zorc (1990).[58]

It is noteworthy that the residue left by subtracting a root from the morpheme in which it is embedded (Malay *ha-* in *hapit*, *ca-* in *capit*, etc.) shows no clear pattern of sound-meaning association. Brandstetter (1916) called this element a 'formative', and tried to identify it with an affix. Affixes, however, are closed classes, while the residues in morphemes that contain a root form an open class. For convenience we can call this semantically empty CV- element a 'formative', with the understanding that it is not a true affix. Morphemes that contain a root thus typically consist of a formative plus a -CVC root of generalised meaning. In this respect they resemble English words such as *glare, gleam, glimmer, glint, glisten, glitter, gloss, glow*, where the residue after removing *gl-* 'light, radiance' is meaningless and non-recurrent (the formative in AN languages is occasionally recurrent simply by chance, since it is a maximally unmarked CV- syllable). For this reason AN monosyllabic roots are best regarded as phonesthemes. They differ structurally from the phonesthemes of better-known languages such as English in being whole syllables. Again, unlike the phonesthemes of Indo-European languages, if a -CVC root can be identified in a number of CVCVC bases it is likely that evidence will be found for reduplicated monosyllables such as *pitpit, *tiktik, or *tuktuk.

In addition to monosyllabic 'roots' many AN languages have more conventional types of phonesthemes, as the velar nasal that occurs initially in many non-cognate words referring to the nasal and oral region (Blust 2003d). Among the most puzzling types of submorphemic sound-meaning association is 'Gestalt symbolism' (Blust 1988a:59ff). A striking example is seen in words that mean 'wrinkled' and the like, as seen in Table 6.3 (WBM = Western Bukidnon Manobo):

---

[58]   To these we can add Potet (1995), Kempler Cohen (1999), and Wolff (1999), although these works show departures from sound method that reduce our confidence in their conclusions.

Table 6.3 Gestalt symbolism in words for 'wrinkled, creased, crumpled'

Language	Form	Gloss
Ilokano	karekkét	wrinkle
Ilokano	karenkén	crease, fold, wrinkle
Ilokano	kuretrét	wrinkle, crease, frown
Pangasinan	kumanét	crumpled
Tagalog	kulubót	wrinkled
Tagalog	kuluntóy	wrinkled
Cebuano	kulámus	squeeze or crumple
Cebuano	kulíut	grimace or distort one's face in anger or pain
Hiligaynon	kurinút	creased, wrinkled
WBM	kurərət	to wrinkle, as an aged person's skin
Kelabit	gərisət	wrinkled (skin, clothing)
Long Anap Kenyah	kərupit	wrinkle
Kayan	kəlupit	withered, shriveled; wrinkled
Kayan	kəliəŋ	wrinkled, as skin or dry leaves
Kayan	kəlubəy	half dried; shriveled, withered
Mukah Melanau	kərəñut	wrinkled
Mukah Melanau	kərəsaŋ	wrinkled
Iban	kəlapat	wrinkled, shrivelled
Malay	kələdut	much creased, crumpled, or wrinkled
Malay	kərəkut	curling or warping or shriveled up
Malay	kərəmut	puckering up the face
Malay	kəreput	crinkling or shriveling up (as the skin around and old boil)
Malay	kəresut	puckering the forehead
Malay	kərotot	deeply furrowed; shriveled up
Malay	kərutu	rough surface; corrugated, deeply lined
Toba Batak	harehut (h < *k)	creases, wrinkles in the face
Makasarese	karussu	wrinkle
Lamaholot	kəməkər	wrinkled, puckered

In each case a word meaning 'wrinkled', 'creased', 'crumpled' or the like begins with *k*-(in one case *g*-), is three syllables in length, and usually contains a liquid as the second consonant. These words are admittedly selected from a number of languages, and the sample might be regarded as biased. However, none of these words are cognate (not even Long Anap Kenyah *kərupit* and Kayan *kəlupit*), and some languages have multiple examples of the same pattern (Malay has seven). Moreover, about 90% of unaffixed word-bases in most AN languages, including most of those from which data is cited here, are disyllabic. Trisyllables are thus rare to begin with, and the occurrence of so many three syllable words in this very specific semantic domain with the combination of an initial velar stop (usually *k*-) and a liquid as the second consonant is clearly non-random. It is possible that some of these words contain a fossilised pluralising infix reflecting *<al> or *<ar>, since the notion 'wrinkle, crease' is inherently plural.

Since both roots and Gestalt symbolism are types of sound symbolism, one other aspect of this phenomenon should be noted here. Many of the 231 roots in Blust (1988a) group into 'families' that differ in such features as the voicing of the initial consonant or the quality of the vocalic nucleus. Next to *-pak 'slap, clap', for example, are the clearly

related forms *-bak 'sound of a heavy smack', and *-pik 'pat, light slap'. Sample data for each variant include: 1) *-pak: PMP *kapak 'beat the wings', PMP *tepak 'slap or beat with the hand', Bontok *dospak* 'slap someone's face with the open palm', Javanese *grapak* 'a branch snapping', Bikol *sapák* 'the sound made when animals chew', Bikol *upák* 'applaud, clap for', 2) *-bak Yamdena *ambak* 'pound into the ground; stamp with the feet', Tiruray *bakbak* 'a hammer; to hammer, pound', Kankanaey *kibbák* 'clap, clash (as parts of banana stems against each other)', Cebuano *labák* 'throw something hard on the ground', Proto South Sulawesi *tamba(k) 'hit, pound', Kambera *tumbaku* 'strike, knock against (as a buffalo butting with its horns against a fence)', 3) *-pik: PMP *lepik 'snap, break off (twigs, etc.)', Sika *kəpik* 'wing, fin', Kankanaey *kippík* 'snap, crack (as a breaking twig)', Maranao *latpik* 'crackling sound', Kankanaey *pikípik* 'drop, trickle, leak, drip, dribble', Bontok *tadpík* 'slap lightly, usually with the flat of the hand'.

The foregoing material is selective, but it exemplifies two patterns that are supported by a much richer database. The first of these is vowel variation such that: *a* = a loud, discordant or raucous sound, *ə* = a muffled or blunted sound, *i* = a high-pitched sound, and *u* = a loud or deep sound. The second is a symbolic value in which voiced consonants signal larger sound-producing objects than their voiceless equivalents.

These 'root families' are of general interest because they appear to violate a fundamental and widely accepted principle of binarity. Diffloth (1976:250) called this the property of 'lexical discreteness' (that any modification in the phonology of a root must give a different meaning or a meaningless form), and he showed that many morphemes in Semai, an Austroasiatic language of the Malay Peninsula, exhibit graduated semantics that is correlated with differences in phonological features. In AN languages the matter is somewhat different, since continuous semantic gradation is associated with submorphemic roots rather than with independent morphemes. However, it should be noted that although most monosyllabic roots in AN languages never occur as free forms, onomatopoetic roots, which constitute not quite 25% of the total, may stand alone (Karo Batak *bak* 'sound of horse's clopping hooves', Malay *tok* 'a dull knock', etc.). In this sense the association of meaning with phonological features in AN monosyllabic roots is strikingly similar to the similar association in Austroasiatic base morphemes.

One repeated misconception in relation to AN monosyllabic roots is that *every* base morpheme must contain a recurrent submorphemic partial. This is the view of Kempler Cohen (1999), and in practice Potet (1995) comes close to the same position. There is, however, little empirical support for such a view. Although the line between onomatopoetic and non-onomatopoetic roots can sometimes be difficult to draw, about 53/231, or nearly 23% of all roots identified to date are onomatopoetic. Those that are not onomatopoetic occur in diverse types of morphemes, including dynamic verbs (*-bej 'wind around repeatedly', *-buq 'fall', *-daR 'lean on', *dem 'think, ponder, brood, remember', *-kap 'feel, grope', *-kit 'join along the length'), stative verbs or adjectives (*-baw 'shallow', *-bek 'rotten, crumbling', *-kuk 'bent, crooked', *-ŋaŋ 'amazed, gaping', *-ŋel 'deaf'), and nouns (*-bir 'rim, edge', *-but 'buttocks, bottom', *-duk 'ladle, spoon', *-Neb 'door'). An examination of the 2,478 root tokens in Appendix 2 of Blust (1988a) shows almost no basic vocabulary (numerals, body part terms, terms for the natural environment, basic verbs). Despite the frequency of monosyllabic roots in morphemes representing semantic categories such as 'wind around repeatedly', 'rotten, crumbling' or 'block, stop, dam' the last syllable of forms such as PMP *zalan 'path, road', *panaw 'go, walk', *lakaw 'go, walk', *kulit 'skin', *likud 'back', *susu 'breast', *takut 'fearful, afraid', *ipen 'tooth', or *laŋit 'sky' show no evidence of recurrent sound-meaning associations.

Why some semantic categories have tended to attract submorphemes in AN languages while others have not is unknown, but the same question can be raised with regard to the phonesthemes of English, where only a few generalised meanings appear to be associated with submorphemic sound-meaning associations.

Finally, both monosyllabic roots and more abstract patterns of submorphemic sound-meaning correlation (Gestalt symbolism, etc.) raise fundamental questions about the nature of cross-generational language transmission. In learning a first language children acquire morphemes, together with patterns of word-formation, and syntax. But the type of knowledge that is transmitted with submorphemes appears to be different from any of these. If submorphemic sound-meaning correlations are distributed over a number of genetically related languages in *non-cognate* morphemes one must ask how such patterns can be transmitted independently of the forms that exemplify them. There are two logical possibilities: 1) they are transmitted in sets of morphemes which contain a recurrent submorphemic sound-meaning correlation that is then extended to neologisms, or 2) the abstract pattern itself is internalised. Much work clearly remains to be done before satisfactory answers can be brought to bear on this question.

6.3   Affixes important for word-formation

Although comparative studies of the morphology of AN languages have lagged behind phonological, lexical, and even syntactic comparison, a rich body of published material is available. Pioneering work in this area was done by the Swiss linguist Renward Brandstetter in his essays 'Common Indonesian and original Indonesian', and 'The Indonesian verb: a delineation based upon an analysis of the best texts in twenty-four languages' (reprinted in English translation in Brandstetter 1916). Blust (2003c:471-475) provides a fairly complete list of affixes and 'clitics' (most of which are better described as particles) that have been reconstructed at the Proto Austronesian, Proto Malayo-Polynesian, and Proto Oceanic levels. The results are summarised in Table 6.4. This chapter is concerned primarily with affixes that affect word-formation, although, as already noted, the line between morphology and syntax can be difficult to draw in Philippine-type languages. The discussion illustrates many of the more important affixes for early AN proto languages, but is not exhaustive:

Table 6.4 Number of reconstructed particles/clitics and affixes in PAN, PMP and POC

	PAN	PMP	POC
Particles/clitics	16	18	10
Prefixes	24	31	14
Infixes	4	4	1
Suffixes	8	13	23
Circumfixes	2	7	0
Reduplicative affixes	2?	7	4
Deletion		1	
Stress shift		1	

6.3.1 Prefixes

As seen in Table 6.4, prefixes greatly outnumbered other types of affixes in PAN and PMP. In POC there was a sharp decline in the percentage of affixes that were prefixed or infixed, and a corresponding shift of emphasis from prefixation to suffixation. It should be sufficient here to focus on a small number of the most important and widely reflected prefixes, starting with forms assignable to PAN. We will begin with some general observations on the shapes of prefixes before surveying specific forms.

Prefixes that can be reconstructed for early AN proto languages generally are monosyllabic, usually begin with a voiceless stop or a nasal, and almost always contain *a as the sole vowel, or as the first vowel in a disyllable, as shown in Table 6.5:

Table 6.5 Phonemic form of early Austronesian prefixes

Language	Prefix	Gloss
PAN	*ka-	allative; to (someone or some place)
PAN	*ka-	inchoative
PAN	*ka-	stative in negative and other irrealis constructions
PAN	*ka-	past time
PMP	*ka-	formative for abstract nouns
PMP	*ka-	manner in which an action is carried out
PMP	*ka-	past participle or achieved state
PMP	*kali-	sensitive relation to the spirit world
PAN	*ma-	stative
PMP	*maka-	consider as X (X = kin term)
PMP	*maka-	do X times (X = numeral)
PMP	*maŋ-	active verb
PAN	*maR-	relationship of parent to child, or of siblings
PMP	*maR-	intransitive verb
PAN	*pa-	causative of dynamic verbs
PMP	*pa-	divide into X (X = numeral)
PAN	*pa-ka-	causative of stative verbs
PMP	*pa-ka-	treat like X (X = kin term)
PAN	*pa-ka-	simulative, pretend to X (X = verb)
PMP	*pa-ka-	X times (X = numeral)
PWMP	*paŋ-	instrumental noun
PMP	*paR-	deverbal noun
PMP	*paR-	one Xth (X = numeral)
PMP	*paRi-	reciprocal or collective action (?)
PAN	*paSa-	verb prefix, often with verbs of direction
PAN	*pi-	causative of location
PAN	*pu-	causative of motion
PMP	*qali-	sensitive relation to the spirit world
PAN	*Sa-	deverbal instrumental noun
PAN	*ta-	verb prefix[59]
PAN	*taR-	spontaneous or accidental action

---

[59]   Rare, but cf. *likud 'back' : *ta-likud 'turn the back on someone'.

Exceptions to this pattern are PAN *mu- 'prefix of motion', PAN *Si- 'instrumental voice', and PAN *Sika- 'marker of ordinal numerals' (often reduced to *ka-*). Except for PMP *da- 'prefix on kin terms', a possible honorific marker that may be derived from the third person plural pronoun, no reconstructed prefix begins with a voiced stop, liquid or glide, or for that matter, with any nasal other than *m*. Even more strikingly, a number of prefixes form pairs that differ only in *p*- vs. *m*-, a pattern that can be called '*p/m* pairing'.

6.3.1.1   *p/m pairing*

The expression '*p/m* pairing' refers to the presence of pairs of prefixes which differ only in that one has *p*- and the other *m*-. In some cases the functions of such prefixes are transparently related, but in others they are not. Data from Thao of central Taiwan can be used to illustrate. Affixes that cannot be glossed more specifically than 'verb prefix' are left unglossed.

Table 6.6 p/m prefix pairs in Thao

p-form	m-form
pa- 'causative of dynamic verbs'	ma- 'active verb'
pak- 'intransitive verb'	mak- 'intransitive verb'
paka- 'multifunctional prefix'	maka- 'resemble X'
paka-	maka-
pakin-	makin-
paku-	maku-
pala-	mala-
pali-	mali-
palh-	malh-
palha-	malha-
palhan-	malhan-
palhi-	malhi-
palhin-	malhin-
pan-	man-
pasun-	masun-
pash-	mash-
pasha-	masha-
pashash-	mashash-
pashi-	mashi-
pashin-	mashin-
pat-	mat-
pati-	mati-
patin-	matin-
patu-	matu-
patun-	matun-
pi-	mi-
pia- 'causative of stative verbs'	mia-
pian-	mian-
pilh-	milh-
pin- 'inchoative	min-
pish-	mish-
pu- 'causative of motion'	mu- 'motion prefix'
puk-	muk-
pulha-	mulha-
pun-	mun-
put-	mut-

Examples of this pattern are often difficult to evaluate, since 1) the affix is found on only one or two forms, and 2) an adequate gloss for the affix is not available, as with *malh-* and *palh-* (each found with three bases): *muqmuq* 'chaotic' : *malh-ma-muqmuq* 'produce nonsense, as when speaking', *qarman* 'bad' : *malh-qa-qarman* 'malign, speak ill of someone', *pin-tukus-an* 'walking stick or cane': *t<m>ukus, malh-tukus* 'walk with a walking stick or cane', *palh-ma-muqmuq* 'speak nonsense, speak in a chaotic or socially unacceptable way', *palh-qa-qarman-in* 'be maligned', *palh-tuqus* 'use a walking stick or cane'. The functions of most of these affixes cannot be characterised any more precisely than with the gloss 'verb prefix'. There are, however, two patterns that emerge fairly

clearly. In one, the *p*-variant is imperative or hortative and the corresponding *m*-variant is indicative, as with 12) vs. 13):

12.   *yaku lhuan maka-rihaz mu-buhat*
       I last.night do-through work 'I worked through the whole night'

13.   *paka-rihaz ita ya ma-humhum i-nay mi-qilha*
           hort.do-through we.**incl** when night this drink 'Why don't we drink through the night?'.

    In the second, the *p*-variant is causative and the *m*-variant is not, as with *makit-na-faw* 'ascend slowly, as a mountain' : *pakit-na-faw* 'make something higher, as in stacking up chairs; go higher', or *mu-nay* 'come' : *pu-nay* 'put something here; let someone come here'. Some examples may combine imperative or hortative meanings with a causative, as in 16):

14.   *a mu-lhilhi iza yaku*
       fut mot-stand already I   'I will stand up

15.   *haya wa qrus mun-tunuq, pu-lhilhi ita*
       that **lig** post topple **caus**-stand we.**incl** 'That post has fallen; let's go put it back up!'.

    Although this pattern is particularly striking in Thao, which has at least 35 *p/m* prefix pairs out of some 201 affixes or quasi-affixes in Blust (2003a), similar data could be cited for other languages. Rubino (2000:xviiiff), for example, lists 413 simple or compound prefixes in Ilokano, and 59 of these are *p/m* pairs.

    Not all *p/m* matches are necessarily true pairs, but the sheer number of affixes that differ only in these two phonemes suggests that many *m*-initial prefixes are bimorphemic. Wolff (1973:72) suggested that nasal-initial prefixes in Philippine languages result from the infixation of forms that began with *p-, as with *paŋ- : *maŋ- (< *p<um>aŋ-), or *paR- : *maR- (< *p<um>aR-), and that the processes used to derive such nasal-initial prefixes 'go back to the protolanguage.' He does not specify which proto language, and certain qualifications appear to be in order. As Wolff suggested, and as the Thao evidence demonstrates, the derivation of *m*-initial prefixes by infixation can safely be attributed to PAN. The most likely mechanism responsible for *p/m* pairing is pseudo nasal substitution (PNS), since this disfavors prefixal sequences of the shape *pVm, but would not affect prefixes that begin with other consonants. The PNS hypothesis as the source of *p/m* pairing works well for languages like Thao, which disallows *pVm* within an affixed base, but it is problematic for Ilokano, which does not show PNS effects in *b*-initial or *p*-initial bases that are infixed with <*um*>. It is difficult to determine exactly what this means. On the one hand, it is possible that PAN had PNS only in prefixes. On the other, PNS may have been a general process in PAN that was lost as an active process in most daughter languages, but was fossilised in affixes. In addition, many Philippine languages have *p/m/n* pairing, matching *p*-initial bases with *m*-initial active indicative verbs, and their perfective *n*-initial counterparts, as with Tagalog *pag-* 'formative of verbal nouns' : *mag-* 'verb prefix indicating among other things external action: *nag-* 'past/perfective of *mag-* verbs', or *paŋ-* 'formative for instrumental nouns': *maŋ-*'active verb' : *naŋ-*'past/perfective of *maŋ-* verbs'. Traditional accounts, as that of Panganiban (1966:206) derive *p*-initial prefixes

from their *m*-initial mates (which presumably are the most frequent forms), but this reverses the direction of derivation both historically and synchronically. Although pseudo nasal substitution supplies a motivation for reduction of *p<um>aR- to *maR- and *p<um>aŋ- to *maŋ- through CV- truncation, no such mechanism is available for the reduction of *p<um><in>aR- to *naR- or *p<um><in>aŋ- to *naŋ-. While *p<um>aR- to *maR-, can be seen as a product of primary truncation driven by canonical patterns, the change *p<um><in>aR- > m<in>aR- > *naR- and the like apparently is a secondary truncation driven by a tendency to paradigmatic shape harmony, as with Tagalog *mag-walís* 'to sweep' : *nag-walís* 'swept' with matching trisyllabic shapes rather than *mag-walís* : *m<in>ag-walís* with a canonical asymmetry.

6.3.1.2 *ka-

As seen above, there were a number of affixes with this shape both in PAN and in PMP. Known reflexes of *ka- 'inchoative' are confined to Formosan languages, of which Thao provides the clearest example: *ma-bazay* 'be worn and thin, as clothing' : *ka-bazay* 'become worn and thin, as clothing'.

Reflexes of *ka- 'stative' are found in Formosan and Philippine languages, as Amis *ma-fanaʔ kako* (know I) 'I know' vs. *caay ka-fanaʔ kako* (**neg** know I) 'I don't know', or Bontok *l-om-oto* (cook-AF) 'to cook' vs. *daan ka-loto* (not.yet **stat**-cook) 'not yet cooked'.

Reflexes of *ka- 'past time' are known in only a few words, including 'when (past)' as opposed to 'when (future)', and the word for 'yesterday' in a number of AN languages. Examples from the first set include Saisiyat *inoan* 'when (future)' vs. *ka-inoan* 'when (past)', Isneg *nuŋay* 'when (future)' vs. *kan-nuŋay* 'when (past)', Waray-Waray *sanʔo* 'when (future)' vs. *ka-sanʔo* 'when (past)', and Tausug *kuʔnu* 'when (future)' vs. *kaʔnu* 'when (past)'. Contrasts in the morphology of words for 'yesterday' and 'tomorrow' also show this function, as in Paiwan *nu-tiaw* 'tomorrow' : *ka-tiaw* 'yesterday', Ivatan *ma-koyab* 'afternoon' : *ka-koyab* 'yesterday', or Minangkabau *pataŋ* 'evening' : *ka-pataŋ* 'yesterday'.

Reflexes of *ka-'corresponding one, mate; accompanied action' are widespread in Philippine languages (e.g. Ilongot *duwa* 'two' : *ka-duwa* 'companion', Tondano *wanua* 'village' : *ka-wanua* 'fellow villager, compatriot'), but are unknown elsewhere.

Reflexes of *ka- 'abstract noun formative' are found in languages reaching from Taiwan to Madagascar, as seen in Amis *tayal* 'work' : *ka-tayal* 'work as an activity not necessarily in process', Ifugaw *tagu* 'man, human being' : *ka-tagu* 'manhood', or Malagasy *tsara* 'good' : *ha-tsara* 'goodness'.

Reflexes of *ka- 'manner in which an action is carried out' are known from the southern Philippines and apparently northern Sumatra: Mapun *ka-* 'nominaliser with adjectives or stative verbs indicating the manner of doing an action', Toba Batak *ha-* 'the time at which, place whence, and even the particular way in which the content of the verb takes place'.

Finally, reflexes of *ka- 'past participle; achieved state' occur from the southern Philippines to the central Pacific: Timugon Murut *ka-* 'marker of achieved states', as in *ma-aguy* 'is tired' : *ka-aguy* 'tired', Ma'anyan *ka-* 'marker of past participle or achieved states', as in *reŋey* 'to hear' : *ka-reŋey* 'heard', or *ituŋ* 'to remember' : *ka-ituŋ* 'remembered', Nggela *ka-* 'prefix forming past participles' as in *mbihu* 'to separate, break apart' : *ka-mbihu* 'broken off', or Fijian *ka-* 'marker of achieved states', as in *basu* 'break; open a person's eyes or mouth' : *ka-basu* 'torn open'.

6.3.1.3 **ma- 'stative'*

The stative prefix **ma-* is one of the most widely attested AN affixes, although it is often fossilised. It tends to be most productive in the languages of Taiwan, the Philippines and the Philippine-type languages of western Indonesia, as in Thao *ma-bric* 'heavy' (cp. *pia-bric-ik* 'I made something heavy'), *ma-puzi* 'white' (cp. *pish-puzi* 'become white'), *ma-haha* 'furious; strong, of a current' (cp. *pia-haha* 'get angry; pretend to be angry'), Tagalog *bigát* 'weight' : *ma-bigát* 'heavy', *ínit* 'heat' : *ma-ʔínit* 'hot', *paʔít* 'bitterness' : *ma-paʔít* 'bitter', or Lun Bawang *mə-lauʔ* 'hot', *mə-tənəb* 'cold', *mə-budaʔ* 'white', *mə-siaʔ* 'red', *mə-birar* 'yellow', *mə-bəruh* 'new'. Reflexes of **ma-* can often be determined by scanning comparative vocabularies, as many words that translate as English adjectives will be marked with it, and in most cases these words are trisyllabic. However, this test may fail to distinguish reflexes of **ma-* that are fossilised from those that are still productive. Even in languages with a productive reflex of **ma-* some lexical items may show fossilisation, as with Samoan *maŋa* 'to fork, of a tree, road, etc.' (PMP **ma-saŋa* 'forked'), *maʔi* (PMP **ma-sakit*) 'sick', or *matua* (PMP **ma-tuqah*) 'mature, of fruit; adult, of people; older, elder'. In eastern Indonesia reflexes of **ma-* are often preserved only as *m-* before vowel-initial bases, or as prenasalisation or just voicing (from earlier prenasalised stops) before bases that begin with an obstruent, as with Ende *mite*, Lamaholot *mitaŋ* (PMP **ma-qitem*) 'black', Ende *mbənu*, Lamaholot *bənu* (PMP **ma-penuq*) 'full', Ende *muri*, Lamaholot *more-t* (PMP **ma-qudip*) 'living, alive', or Galoli *banas* (PMP **ma-panas*) 'warm', *buti* (PMP **ma-putiq*) 'white'.

Evans and Ross (2001) note that reflexes of **ma-* in Oceanic languages fall into four groups: 1) valency-decreasing **ma-*, 2) fossilised reflexes of **ma-* on stative verbs, 3) stative (adjectival) verbs that can be reconstructed in POC both with and without **ma-*, and 4) fossilised reflexes of **ma-* on experiential verbs. Well-known examples of **ma-* in a non-stative function include reflexes of PMP **ma-tiduR/ma-tuduR* 'to sleep', and PMP **ma-huab* (POC **mawap*) 'to yawn'. Even though reflexes of nonstative **ma-* occur in some high-frequency verbs, the vast majority of verbs with a reflex of **ma-* are stative. All of these **ma-* affixes are unusual in being among the few *m*-initial verb prefixes that do not participate in *p/m* pairing, since **pa-* can be reconstructed only as a causative prefix, and this is functionally incompatible with any of the meanings of **ma-*.

A number of languages in the Philippines and Borneo reflect **ma-* with the numerals meaning 'hundred' and 'thousand', as with Agta *mə-gətut*, Atta *ma-gatuʔ*, Botolan Sambal *ma-gato*, Koronadal Bilaan *m-latuh*, Lun Bawang *mə-ratu*, Kiput *ma-lataw*, Miri *ma-rataw*, Uma Bawang Kayan *m-atuh* 'hundred', or Agta *mə-hibu*, Atta *ma-ribu*, Koronadal Bilaan *m-libu*, Lun Bawang *mə-ribu*, Kiput *ma-libo*, Miri *ma-ribuh* 'thousand'. In the Philippines and northern Sulawesi, but apparently not in Borneo, reflexes of **ma-puluq* also occur in the meaning 'ten', as with Agta *mə-pulu*, Atta *ma-pulu*, Botolan Sambal *ma-poʔ*, Sangir *ma-pulo* 'ten'. Whether these uses of **ma-* should be regarded as stative is unclear.

Many languages allow some stative verbs to occur as a bare base. In Lun Bawang of northern Sarawak, for example, most stative verbs are given in citation form with *mə-*, and consequently stand out because they depart from the usual disyllabic canonical shape of both nouns and dynamic verbs: *məlutak* 'dirty', *məlauʔ* 'hot', *mətənəb* 'cold', *məkəriŋ* 'dry', *məbaaʔ* 'wet', *məbərat* 'heavy', *məraan* 'light in weight', *məbudaʔ* 'white', *məsiaʔ* 'red', *məbirar* 'yellow', *məbataʔ* 'green', etc. However, a few stative verbs occur as bare bases, as *suut* 'small', *rayəh* 'big', *doʔ* 'good', and *daat* 'bad'. These are reminiscent of the semantic categories of so-called 'baby adjectives' in French which form a subclass based

on distribution. Evans and Ross (2001) note similar cases in Oceanic languages, and in fact some stative verbs are zero-marked in many AN languages, although the semantic categories reserved for this treatment do not always correspond.

In some Philippine languages the perfective form of the stative prefix reflecting *m\<in\>a- is reduced to *na-, as with Bontok *na-* 'stative prefix occurring in combination with completive voice marking affixes' (Reid 1976:203). In others the abbreviated reflex of *m\<in\>a- has become the aspectually unmarked stative marker, as with Itawis *dámmat* 'weight' : *na-dámmat* 'heavy', *mit* 'sweetness' : *na-mít* 'sweet', or *dapíŋ* 'dirt' : *na-dapíŋ* 'dirty'. This development does not appear to be attested outside the Philippine group. In a few languages of western Indonesia *ma- has been truncated to *a-*, as in Old Javanese *putih* : *a-putih* 'white', *sǝlat* 'appearing in between; interruption' : *a-sǝlat* 'interspaced; interspersed with', or in Makasarese, where according to Cense (1979:429) *ma-* occurs as an 'adjectival' prefix in both older and archaising forms of the language, but shows up as *a-* in seventeenth and eighteenth century texts.

6.3.1.4 *maka- 'abilitative/aptative'

Most reflexes of *maka- are found in Philippine-type languages, including some that do not belong to the Philippine group, as with Ilokano *maka-* 'intransitive potentive prefix, corresponding to the transitive *ma-* indicating potential, abilitative, accidental, or coincidental action', Bikol *maka-* 'verbal affix, potential action series, infinitive-command form', Mapun *maka-* 'verbal prefix denoting abilitative or circumstantial action', Malagasy *maha-* 'potential prefix; it expresses the ability or power to perform any action, or what makes a thing to be what it is'. Like *ma-, *maka- does not appear to participate in *p/m* pairing, since the only *p*-initial correspondent is *pa-ka-* 'causative of stative verbs'. Prefixes that could reflect *maka- are also found in Formosan languages, as with Thao *maka-* which has various functions, none of them similar to the meaning of extra-Formosan forms, and Puyuma *maka-* 'at the side of'. These agreements are best treated as convergent, although the number of a priori convergent affixes in Formosan and Malayo-Polynesian languages raises suspicions that these were found in PAN and have diverged radically in function.

6.3.1.5 *maki/paki- 'petitive'

Clear reflexes of this affix pair are confined to the central and southern Philippines and to some Philippine-type languages of northern Borneo and Sulawesi, as with Tagalog *maki-* 'ask for, make a request for; join in company; imitate', *paki-* 'prefix forming nouns to denote favor asked or requested', or Bikol *maki-* 'fond of, in favor of', *paki-* 'verbal affix, social action series … the affix, when prefixed with *i-*, may also serve as a request without an accompanying verb base: *i-pakí mo na lang iyán diyán* 'please ask for it there' (Mintz and Britanico 1985:409), Timugon Murut *maki-* 'petitive, subject focus (= voice), future temporal', *paki-* 'petitive, atemporal', Tindal Dusun, Kadazan Dusun, Bolaang Mongondow *moki-* 'petitive prefix, actor voice', *poki-* 'petitive prefix, imperative mood', and Tondano *maki/paki-* 'petitive'. It is unclear whether Tagalog *maki-* is a single polysemous prefix or two homophonous prefixes. If the first interpretation is adopted Northern Philippine forms of corresponding shape such as Ilokano *maki-* 'participative (social) intransitive verbal prefix', *paki-* 'nominalising prefix for *maki-* verbs serving an instrumental purpose' can be included in this set. Otherwise, the evidence for *maki/paki-* is geographically more restricted.

6.3.1.6   **maŋ- 'actor voice'/*paŋ- 'instrumental noun'*

Reflexes of **maŋ-* are ubiquitous in the Philippines and western Indonesia, and are also found in Malagasy, Palauan and Chamorro. Reflexes of **paŋ-* are less widspread. Both affixes trigger nasal substitution. As noted in Blust (2004a) **maŋ-* is reflected as **ŋ-* in some languages, a truncation that evidently was motivated by disyllabic canonical pressures.

Over much of western Indonesia reflexes of **maŋ-* mark active verbs that usually take an object, but in some cases may not, as with Malay *pukul* (base) : *mə-mukul* 'to hit' : *dia məmukul saya* 'he hit me', but *tulis* (base) : *mə-nulis* 'to write' : *dia sədaŋ mənulis* 'he is busy writing'. In some languages reflexes of **maŋ-* and **-um-* mark transitive vs. intransitive forms of the same verb base, as with Kelabit *turun* (base) : *t<əm>urun* 'to descend (as a ladder)', but *nurun* 'to lower (as a ladder from a longhouse veranda to the ground)'. As noted in 6.1.1, in parts of the central Philippines reflexes of **maŋ-* are in competition with reflexes of **maR-* and **-um-* in marking actor voice, and as a result it has a narrower range of functions. This has already been illustrated for Tagalog, but a similar narrowing of functions is seen in other Philippine or Philippine-type languages, as Bikol, where "The *mang-* series is usually used intransitively to indicate an action that is somehow more encompassing than the same action would be if expressed with the affix *mag-*': *mag-bakal* 'buy' : *maŋ-bakal* 'go shopping'' (Mintz 1971:182ff). Sneddon (1975:219) notes that in Tondano of northern Sulawesi reflexes of **maŋ-* generally may substitute for reflexes of **-um-* "without discernible change of meaning." However the former affix occurs frequently "in combination with Repetitive aspect," thus suggesting that the distinction between *maŋ-*, *mag-* and *-um-* in central Philippine languages such as Tagalog preserves an older functional contrast that has been lost in many other areas. In general, reflexes of **paŋ-* create nouns, and more specifically instrumental nouns. These may be related to corresponding *maŋ-* verbs, but frequently are not. Although it might be assumed on the basis of widespread p/m pairing that **maŋ-* derives from **p<um>aŋ-*, reflexes of **maŋ-* appear to be more widely distributed than reflexes of **paŋ-*, an observation which suggests that **paŋ-* may be a later innovation.

6.3.1.7   **maR- 'actor voice'/*paR- 'instrumental noun'*

Reflexes of **maR-* and **paR-* are common in the Philippines and northern Borneo, and are found elsewhere in western Indonesia. The functions of Tagalog *mag-* have already been noted. In other languages the cognate form is said to mark one type of intransitive verb. According to Antworth (1979:15), for example, in Botolan Sambal "The affixes *-om-* and *ma-* both form intransitive change of state verbs in which the initial subject is semantically a patient … The prefix *mag-* forms intransitive verbs by verbalising nouns." As examples he gives *ganda* 'beautiful' : *g<om>anda* 'become beautiful', *bitil* 'hungry' : *b<om>itil* 'become hungry', next to *mag-baskitbol* 'play basketball', and *mag-tagalog* 'speak Tagalog'. Although *mag-* appears to be borrowed from a central Philippine source, it lacks the transitivity-marking function that Ramos (1971) reports for the related affix in Tagalog. In Tindal Dusun of Sabah *mag/mog-* almost always marks intransitive verbs, as with *mag-asu* 'hunt using dogs', *mog-gidu* 'run away from home', *mog-inum* 'to drink (liquor)', *mog-odow* 'shine, of the sun', *mog-onsok* 'cook food, boil water', or *mog-ontoluw* 'lay an egg'. Reflexes of **-um-* also mark intransitive verbs, as with *k<um>auh* 'swim', *m-odop* 'sleep', or *r<um>ikot* 'come', and the functional distinction between the two affixes is obscure. By contrast *maŋ-* sometimes requires an object as with *maŋ-anuh* 'to take or receive something', or *mo-moli* (*boli*) 'to buy', and sometimes does not allow

one, as with *moŋ-ipih* 'to dream', or *ma-manaw* (*panaw*) 'to walk'. In some cases *mog-* and *moŋ-* may each be affixed to a base, producing intransitive verbs of idiosyncratic meaning, as with *mog-inum* 'to drink (liquor)' vs. *moŋ-inum* 'to drink (water). In Malay, verbs marked with *meŋ-* generally, but not always require an object, while those marked with *bər-* (< *maR-) rarely if ever do: *bər-jalan* 'to walk', *bər-layar* 'to sail', *bər-lari* 'to run', *bər-mimpi* 'to dream', *bər-aŋkat* 'to leave, depart', and in Toba Batak of northern Sumatra *maŋ-* is said to form both transitive and intransitive verbs, while *mar-* is used to form intransitive verbs and reflexives or simple passives, to refer to a plurality of things possessing a quality, and when derived from an exclamation imitating a sudden sound, to form verbs meaning to make that sound. In general, then, it appears that *maR- marked one type of intransitive verb (the other type being marked by *-um-), while the function of *maŋ- is more difficult to characterise in terms of transitivity. Reflexes of *maR-X 'relationship of X and child' (where X = kin term) probably represent a distinct prefix.

The function of *paR- is less certain, but agreements between Philippine languages and Malay suggest that it was used to form instrumental nouns corresponding to *maR- verbs, just as *paŋ- formed instrumental nouns that correspond to *maŋ- verbs.

6.3.1.8 *mu- 'movement'

This prefix is found in Formosan languages, where it is highly productive. Examples include Thao *taipak* 'Taipei' : *mu-taipak* 'go to Taipei', *fafaw* 'top' : *mu-fafaw* 'go to the top', *qualh* 'near' : *mu-qualh* 'come near, move closer', *tantu* 'there' : *mu-tantu* 'go there', and Puyuma *darə* 'earth, ground' : *mu-darə* 'descend', *ənai* 'water' : *mu-ənai* 'enter the water', *ruma* 'house' : *mu-ruma* 'enter the house', *ləŋaw* 'sound' : *mu-ləŋaw* 'echo'. Despite some claims in the literature it is clearly distinct from *-um-, as seen in e.g. Thao *tuqris* 'snare trap' : *t<m>uqris* 'catch with a snare trap', but *mu-tuqris* 'blunder into a snare trap'. Cebuano Bisayan *mu-* 'actor voice, non-past', which can be added to nouns referring to a place, as in *grahi* 'garage' : *mu-grahi* 'go to/toward the garage', or *lawud* 'sea' : *mu-lawud* 'go to/toward the sea' may be connected, but if so it is the only reflex known outside Taiwan.

6.3.1.9 *pa/pa-ka- 'causative'

Both of these forms of the causative prefix are widely distributed in the AN family. Most languages have only one or the other, and in many cases the longer form is reflected as a single morpheme, as with Amis, Paiwan, Cebuano, Makasarese *paka-*, Kayan *pək-*, Chamorro *faha-*, Arosi *haʔa-*, Fijian *vaka-*, or Proto Polynesian *faka-. However, where a distinction is made reflexes of *pa- mark the causative of dynamic verbs, and reflexes of *paka- mark the causative of stative verbs. As noted by Zeitoun and Huang (2000), and further elaborated by Blust (2003c) the longer variant almost certainly was *pa- 'causative' + *ka- 'stative of irrealis verbs'.

As with causative affixes in many languages, reflexes of *pa/paka- mark both active intervention to insure an outcome, and desistance so as to allow a natural process to run its course, as in Kayan *su* 'far' : *pə-su* 'to separate things', *asəp* 'dirt' : *pək-asəp* 'to soil', *taŋi* 'cry' : *pə-taŋi* 'make someone cry', *urip* 'life' : *pək-urip* 'save the life of a person or animal; spare', or *həŋəm* 'cool' : *pə-həŋəm* 'let something cool, as coffee before drinking it'. What is generally regarded as the causative prefix also has a variety of extended meanings in various languages, including 'treat like X (X = kin term)', pretend to X (X = verb), and X times (X = numeral). Paiwan *pa-ka-* 'feel something to be of such-and-such a

quality' (*pa-ka-sa-ŋuaq*) 'feel something to be delicious', *pa-ka-ma-djulu* 'consider to be simple'), and Cebuano *paka-* 'consider as so-and-so, treat as so-and-so' (with the direct passive suffix *-un*) points to a function of *pa-ka- that may have been dependent upon the co-occurrence of other affixes. In other languages reflexes of *pa-ka- appear to have developed innovative functions, as with Fijian *tamata* 'human' : *vaka-tamata* 'like a human being', or Samoan *faʔa-sāmoa* 'do in the Samoan way'.

6.3.1.10 *paRi- 'reciprocal/collective action'

This affix, which is best-attested in Oceanic languages, marks collective action or reciprocity. Reflexes include Buli *fa-* 'reciprocal prefix', *fa/fai-*'prefix which indicates that the action concerns a plural subject; with nouns it also forms collectives', and related forms in many Oceanic languages, as Mussau *ai-* 'reciprocal or collective action' (Blust 1984c:167), Hoava *vari-* 'reciprocal; marker of collective action; depatientive marker' (Davis 2003:135), Roviana *vari-* 'reciprocal prefix', as in *aqa* 'to wait' : *vari-aqa-i* 'wait for one another', *manamanasa* 'to whisper' : *vari-manamanasi* 'whisper to one another', *tioko* 'to call' : *vari-tioki* 'call to one another' (Waterhouse 1949:135), Arosi *hari-* 'reciprocal prefix generally having the force of united action, rather than mutual action': *daʔa* 'laugh' : *hari-daʔa* 'all laugh together', *roho* 'to fly' : *hari-rohoroho* 'all flying together', *suru* 'lift and carry' : *hari-suru* 'to carry, of two people', *pote* 'big with child' : *hari-pote* 'two women about to give birth together', and Fijian *vei-* 'collective plural with nouns' : *vale* 'house' : *vei-vale* 'group of houses'; with names of relationships it gives a reciprocal sense: *taci* 'younger parallel sibling' : *vei-taci-ni* 'brother and sister, the *taci* relationship'.

A possible cognate of these Oceanic and South Halmahera-West New Guinea (Buli) forms is seen in Kayan *pə-* 'reciprocal', as in *lura* 'spittle, sputum' : *pə-lura* 'spit at each other', *jat* 'pull' : *pə-jat* 'pull each other', or *katəl* 'itch; scratch' : *pə-katəl* 'scratch each other'. Unlike typical Oceanic reflexes of *paRi-, Kayan *pə-* apparently never refers to collective action. In addition it may occur with intransitive verbs, as in *apir* 'either of the parts of two things joined' : *p-apir* 'stuck together, fused, as bananas grown together'. Given these differences and the brevity of the form, it is possible that the resemblance of Kayan *pə-* 'reciprocal' to the semantically similar affix in Oceanic languages is a product of chance convergence.

6.3.1.11 *qali/kali- 'sensitive connection with the spirit world'

The *qali/kali- prefixes present a set of very unusual challenges in historical morphology. First, they are almost always fossilised. Second, they show no linguistically transparent semantic or functional coherence. Third, they show great formal variability (Blust 2001d). The nearly universal fossilisation of these affixes has created word bases that are typically four or more syllables in length for certain semantic categories, which therefore deviate sharply from the predominant disyllabism of content morphemes in most AN languages. A small sample that illustrates this point includes Bunun *bulikuan*, Tamalakaw Puyuma *Halivaŋvaŋ*, Paiwan *qułⁱⁱpepe*, Ilokano *kulibaŋbáŋ*, Palawan Batak *alibaŋban*, Kayan *kələbavah*, Iban *kələbumbu*, *kələlawai*, *kələmambaŋ*, Gorontalo *alinua* 'butterfly', and Nanwang Puyuma *Haripusapus*, Paiwan *vuliłⁱawłⁱaw*, Casiguran Dumagat *alibúno*, Bontok *alipospos/dalipospos*, Bikol *aliwúswús*, Cebuano *alilúyuk*, Minangkabau *alimbubu*, Malay *kələmbubu/sələmbubu*, and Lakalai *kalivuru* 'whirlpool/whirlwind'.

This material shows a distinctive type of patterning. First, these words and non-cognate forms in many other languages that represent the semantic categories 'butterfly' and 'whirlwind/whirlpool' are much longer than the dominant disyllabic canonical shape of most base morphemes. Second, the first two syllables of many of these contain the sequence -*ali*- or its phonologically altered equivalent. Other semantic categories that show similar patterning include leech (two distinct bases), ant, bat, beetle, bumblebee, caterpillar, centipede, crab, dragonfly, earthworm, firefly, gecko, grasshopper, honeybee, millipede, aureole (lunar or solar halo), confused (of vision, sound, the mind), dizzy, dust in the air, echo, hair whorl, rainbow, restless, rustle, shadow/reflection, sparks, storm, sunshower, thick smoke or steam, turbid, clavicle, palate, pupil of the eye, scapula, and various birds, fish, and plants. Blust (2001d) argues that the *qali/kali- prefixes are most easily understood as devices used to make lexical items canonically conspicuous, a strategy that apparently was adopted for words whose referents were connected with culturally-prescribed taboos, in particular those relating to the world of spirits.

6.3.1.12 *Sa- 'instrumental voice'*

This affix has been reported from several Formosan languages (Pazeh, Rukai, Amis) and from Malagasy, where it marks the instrumental voice, much like reflexes of *Si- in other languages. In Pazeh both Ca-reduplication and *sa*- are used to form instrumental nouns, but the former process is used exclusively in this function, while bases which take *sa*- may often be used both verbally and nominally. The most likely reconstructed function for *Sa- is thus as a verbal affix marking instrumental voice, although this does not appear to differ materially from the function of *Si-. Since *Sa- is not known to mark benefactive relationships in any of the four languages in which it is attested, PAN may have distinguished *Sa- and *Si- as instrumental voice and benefactive voice respectively, with *Si- acquiring instrumental meanings secondarily.

6.3.1.13 *Si- 'instrumental voice'*

Like *Sa-, *Si- is reflected in several Formosan languages, as with Atayal *s*-, Bunun *is*-, and Paiwan *si*- 'instrumental voice'. In the Philippines it is reflected as *i*- (Itbayaten, Ilokano, Bontok, Pangasinan, Tagalog, Bikol, Cebuano *i*-), where either singly or in combination with other affixes it marks instrumental voice, benefactive voice, or sometimes other relationships. In the central Philippines the expected reflex of PAN *Si- is **hi-, but like *Sika- 'prefix of ordinal numerals', and some other high-frequency morphemes (PAN *Sepat, PMP *epat 'four') this affix shows an irregular loss of expected *h*-. Elsewhere reflexes are rare or absent. Fijian *i*-, which marks the instrument used to perform an action, as in *i-sele* 'knife' : *sele-va* 'to cut' may be cognate, but this is unclear since the role of convergence is large in creating similarity between short morphemes, and this prefix may also create nouns that mark the scene of the action, the result of an action, the method of an action, the actor, or the person or thing acted on.

6.3.1.14 *Sika- 'ordinal numeral'*

Reflexes of this affix create ordinal numerals from cardinal numeral bases in languages reaching from southern Taiwan to Kiribati. Paiwan *sika*- 'ordinal prefix', as in *ḍusa* 'two' : *sika-ḍusa* 'second', is the only known Formosan reflex, but a number of Philippine languages reflect PMP *ika- (for expected **hika). Malay *kə*- 'ordinal marker', as in *dua* 'two' : *kə-dua* 'second', could regularly reflect PMP *ika-, but Oceanic reflexes such as

Fijian, Pohnpeian, Marshallese, Gilbertese *ka-* 'ordinal of numerals' are irregularly truncated, and point to POC *ka-.

6.3.1.15 *ta/taR- 'sudden, unexpected or accidental action'*

*ta- or *taR- is reflected as an active or fossilised prefix in languages reaching from at least the central Philippines to Polynesia, generally with low prominence in the grammar. Pawley (1972:45) cites reflexes of *ta/tapa- 'stative derivative' in Polynesia, Fiji, Vanuatu and the Solomons, noting that 'With some verbs this marks a spontaneously arising condition.' Examples include Fijian *ta-* 'prefix to verbs indicating spontaneity', as *sere* 'loosen, untie' : *ta-sere* 'come loose by itself' (not very productive, and semantically rather bleached), Wayan *ta-* 'derives a stative verb, an action or process that comes about accidentally or spontaneously, without a willful agent', as *ceve* 'lifted up, overturned' : *ta-ceve* '(of skin) peel, come off' (Pawley and Sayaba 2003), and Arosi *a-* 'prefix of spontaneity' (Fox 1970). Further to the west in eastern Indonesia Klamer (1998:265-267) describes Kambera *ta-* as marking 'derived achievement verbs that are non-intentional, involuntary, accidental, sudden or unexpected,' as with *binu* 'peel something' : *ta-binu* 'be peeled', or *luŋgur* 'rub or scrape something' : *ta-luŋgur* 'be scraped/sore (skin)'. Reflexes of *ta- in western Indonesia include Iban *tə-* 'prefix denoting single complete occurrence, usually sudden', as in *tərap* 'stumble' : *tə-tərap* 'to trip', *pilok* 'crooked, twisted, lame' : *tə-pilok* 'sprain the ankle', and *gaŋgaʔ* 'peal of thunder' : *tə-gaŋgaʔ* 'to sound, of a thunderclap', and apparently the fossilised element *ta-* in Tagalog *pilók, tapilók* 'twisted, of the ankle or foot'.

A smaller set of languages appears to reflect *taR-. Macdonald and Soenjono (1967:96ff) describe six functions for *tər-* with verbs and one with adjectives in Bahasa Indonesia. The functions that appear most relevant to the present discussion are 1) *tər-* replaces *məŋ-* or *bər-* in intransitive verbs to form deverbal adjectives or verbs that often include 'a connotation of lack of control, or of being the victim of circumstances,' as with *mə-layaŋ* 'to fly, soar' : *tər- layaŋ* 'to float, wander aimlessly', or *bər-batuk* 'to cough' : *tər- batuk-batuk* 'cough repeatedly (uncontrollably)', and 2) *tər-* 'includes the implication that an action is accidental and so not performed intentionally by any agent,' as in *mə-niŋgal-kan* 'leave something behind' : *tər-tiŋgal* 'be accidentally left behind', or *məm-bawa* 'take s.t.' : *tər-bawa* 'be taken by mistake'. Although Malay *tər-* does not show regular sound correspondences with the other forms cited above (including *te-* in the closely related Iban), it seems likely that these affixes are related. This inference is supported by Makasarese *taʔ-* 'formative of transitive and intransitive verbs meaning 'brought into a certain condition', affected accidentally or suddenly' (fossilised as *tar-* in some vowel-initial bases), and the single Saaroa form *taruta* 'vomit' (Ferrell 1969:324), which presumably reflects *taR- + utaq. This affix is unusual, then, in appearing in two distinct forms that are not phonologically conditioned allomorphs.

6.3.2 Infixes

Two reconstructed infixes are especially important, and continue to be the linchpin of the verb system in many AN languages spoken today: *-um- 'Actor voice', and *-in- 'perfective aspect'. A third infix *-ar- 'plural actor' is less richly attested, although it has a wide and scattered geographical distribution.

6.3.2.1 *-um- 'actor voice'*

This is the single most important infix in languages with a Philippine-type voice or 'focus' system. In Formosan languages it does nearly all of the work of marking actor voice, while in extra-Formosan Philippine-type languages it shares this work with prefixes reflecting PMP *maŋ- and *maR-.

Transitivity

The issue of transitivity in AN languages will be treated at greater length in Chapter 7. For the present it is sufficient to work with a notional definition in which the presence or absence of a surface object can be taken as prima facie (although not necessarily conclusive) evidence that associated verbs are transitive and intransitive respectively. PAN reconstructions with *-um- are almost always intransitive, as with *Caŋis 'weeping, crying' : *C<um>aŋis 'weep, cry', *kaen 'eating' : *k<um>aen 'eat', *Naŋuy 'swimming' *N<um>aŋuy 'swim', or meteorological verbs, of which the best-attested example is *quzan 'rain' : *q<um>uzan 'to rain'. This pattern persists in some modern languages but apparently not in others. Even in the same language the functional contrast of reflexes of *maŋ- and *-um- is less than completely clear-cut. In Kelabit of northern Sarawak, for example, a reflex of *maŋ- often marks transitive verbs, while a reflex of *-um- marks verbs that are almost always intransitive, hence *kiluʔ* 'bend, curve, as in a path' : *ŋiluʔ* 'bend something, as a wire' : *k<əm>iluʔ* 'wind, meander, as a path or river', *riər* 'turn, roll' : *ŋə-riər* 'turn or roll s.t., as a log' : *r<əm>iər* 'roll without human intervention, as a log rolling down a slope', *turun* 'act or manner of descending or lowering' : *nurun* 'lower something from a height' : *t<əm>urun* 'descend, as a ladder'. However, in other verbs this distinction is less clear-cut, as with *k<um>an* 'eat', which may or may not take an object, and *araŋ* 'a dance' : *ŋ-araŋ* 'to dance', *taŋe* 'weeping' : *naŋe* 'weep, cry', *dalan* 'path' : *nalan* 'walk', or *linuh* 'thought' : *ŋə-linuh* 'think', which are affixed with /ŋ/- (< *maŋ-), but are intransitive.

Other functions

Reflexes of *-um- form inchoative verbs in widely separated WMP languages, as in Bontok *bíkas* 'energetic' : *b<um>íkas* 'he is becoming energetic', Tagalog *sakít* 'pain' : *s<um>akít* 'become painful', Tindal Dusun *gayo* 'big' : *g<um>ayo* 'become big(ger)', or Mukah *gaduŋ* 'green' : *mə-gaduŋ* 'become green; make something green'. In Chamorro the reflex of *-um- may co-occur with another affix to mark the inchoative of stative verbs, but with realignment of morpheme boundaries, as in *ma-hetok* 'hard' : *mu<ma>hetok* (< *m<um>a-hetok) 'become hard'.

Functional load

Reflexes of *-um- remain central to the verb systems of most Formosan languages and many Philippine languages, but in the latter their functional load has been reduced, since they divide the work of marking actor voice with reflexes of *maŋ- or *maR-. In other languages that preserve a Philippine-type verb system the functional load of *-um- has been reduced even further, as in Malagasy, which generally marks the actor voice with *man-*, and uses *-om-* in only a handful of verbs, as *h<om>ana* 'eat' (cp. *han-ina* 'food'), and *tany* 'crying, tears' : *t<om>any* 'weep'. A different type of reduced functional load is seen in Malay/Indonesian, which retains reflexes of *-um- only in semi-fossilised form, usually in words with strong visual or auditory symbolism: *getak-getuk* 'sound of

chattering teeth' : *gələtuk* 'shiver, as from cold' : *gəmələtuk* 'make the sound of teeth chattering because one is affected by cold', *gəməntam* 'to boom, of cannons' (< *g<em>entam?*), *gəmələtap* 'sound of feet striking the ground when someone is running' (< *g<əm><əl>ətap?*), *gilaŋ gəmilaŋ* 'shine, sparkle', *guruh* 'thunder' : *gəmuruh/guruh-gəmuruh* 'to rumble, of thunder'.

Allomorphy

Some languages have a single allomorph of *-um-* that is infixed to consonant-initial bases but prefixed to vowel-initial bases, as with Isneg *sáŋit* 'weeping' : *s<um>áŋit* 'to weep, cry', and *inúm* 'drinking' : *um-inúm* 'to drink', or Tagalog *datíŋ* 'arrival' : *d<um>atíŋ* 'to arrive', and *ulán* 'rain' : *um-ulán* 'to rain'. In many other languages *-um- is reflected as -VC- in consonant-initial bases, but as C- in bases that begin with a vowel: Ivatan, Siocon Subanun *k<um>an* 'eat', but *m-inum* 'drink', Kelabit *turun* 'manner of descending' : *t<əm>urun* 'descend, as a ladder; jump down', but *udan* 'rain' : *m-udan* 'to rain', Toba Batak *taŋis* 'weeping' : *t<um>a-taŋis* 'to weep, cry', but *inum* 'drinking' : *m-inum* 'to drink'.[60]

Although most languages have just two allomorphs of *-um-*, some have more. In Bolaang Mongondow, for example, *-um-* is realised as [m]- before vowel-initial bases (*aŋoy* 'to come' : *ma<m>aŋoy* 'be coming or going', with CV- reduplication after infixation), as pseudo nasal substitution (PNS) in labial-initial bases (*mo-patuʔ* 'warm, hot' : *u-matuʔ* < *p<um>atuʔ* 'feel warm', *bonu* 'inner part' : *u-monu* < *b<um>onu* 'tuck oneself in'), as *-im-* in bases that begin with a consonant when the first base vowel is high front (*kilat* 'lightning' : *k-im-ilat* 'to flash, of lightning', *siup* 'space under a house' : *s-im-iup* 'go under a house'), and as *-um-* elsewhere (*kuduŋ* 'bending over' : *k<um>uduŋ* 'to bend over', *tapaŋ* 'urine' : *t<um>apaŋ* 'urinate'). Muna of southeast Sulawesi shows even richer allomorphy, with at least five phonologically conditioned variants of *-um-*: 1) unchanged with most consonant-initial bases, as in *gaa* : *g<um>aa* 'to marry', 2) *m-* with vowel-initial bases, as *ala* : *m-ala* 'take', 3) PNS with bases that begin with a labial stop, as *poŋko* : *moŋko* (< *p<um>oŋko*) 'kill', 4) zero in bases that begin with *b, bh*, nasals or a prenasalised consonant, 5) *-im-* if the first base vowel is *i*, as in *limba* : *l<im>imba* 'go out', *hiri* : *h<im>iri* 'to peel', or *sikola* 'school' : *s<im>ikola* 'go to school'. Since Van den Berg (1989:29) says that this last allomorph is limited to a few villages in one district, it evidently is independent of the similar innovation in Bolaang Mongondow, spoken hundreds of miles to the north.[61]

The most exuberant allomorphy known for this infix is found in Thao of central Taiwan, where *-um-* has ten surface forms: 1) *-m-*: *canit/c<m>anit* 'to cry', 2) *m-*: *zai/m-zai* 'tell, advise', 3) zero: *fariw/fariw* 'to buy', 4) pseudo nasal substitution: *patqal/matqal* (< *p<um>atqal*) 'put a mark on s.t.', 5) *-um-*: *cpiq/c<um>piq* 'thresh grain by beating the stalks', 6) *-[om]-*: *qpit/q<um>pit* [qompɪt] 'pinch between arm and side; cut with scissors', 7) *-un-*: *ktir/k<un>tir* 'pinch with twisting motion', 8) *-[on]-*: *qtut/q<un>tut* [qontut] 'to fart', 9) *-[oɴ]-*: *tqir/t<un>qir* [toɴqer] 'protest, get angry and leave', 10) *-u-*:

---

[60]   In some analyses word bases in Philippine languages such as Isneg or Tagalog have no initial vowels, since a glottal stop automatically precedes them and remains in place with some affixation processes. However, this is not true of languages such as Paiwan, Itbayaten, or Kadazan, where the same pattern of affixation is found: Paiwan *m-alap* 'take, pick up' : *in-alap* 'one's catch in hunting', Itbayaten *ma-axap* 'be caught' : *in-axap* 'was taken', Kadazan *azak* 'funny' : *in-azak* 'hilarity'.

[61]   Jason Lobel (p.c.) reports that seventeenth century Bikol and Bisayan also had an *-im-* allomorph of *-um-* that was recorded in Spanish documents, but this has reverted to *-um-* in the modern languages.

shnara/sh<u>nara 'to burn' (Blust 1998b, 2003a). Some languages that lack pseudo nasal substitution avoid the sequences *pVm* or *bVm* in affixed words by substituting another affix for *-um-* in bases that begin with a labial stop. Tindal Dusun of Sabah, for example, uses *moŋ-*, but apparently never *-um-* in labial-initial bases. Finally, since reflexes of *-um- trigger pseudo nasal substitution in bases that begin with a labial stop, they appear in several actor voice prefixes that begin with *p- in their underlying forms, as with *maR- < *p<um>aR, and *maŋ- < *p<um>aŋ- (Wolff 1973:72ff).

6.3.2.2 *-in- 'perfective; nominaliser'

This infix differs from *-um- in at least two important respects. First, in PAN and many of its descendants *-in-> marks perfective (or in Central Philippine languages inceptive) aspect rather than one of the four voices or 'focus' potentials of the verb. Second, unless an active verb infixed with *-um- is nominalised by a case/construction marker, as in Tagalog *aŋ s<um>úlat* (**nom** write = 'the one who writes'), or combined with other morphological processes, as in Pangasinan *láko* 'merchandise' : *l<om>a-láko* 'merchant', the reflex of *-um- nearly always has exclusively verbal functions. This is not true of *-in- or other voice affixes, which are often combined alone with a base morpheme to form both verbs and deverbal nouns.

Functional load

Although the function of *-um- tends to be fairly constant across languages, as noted above its functional load varies markedly in languages that have innovated other ways to mark the actor voice. By contrast, reflexes of *-in- show more variable functions, but a more constant functional load.

Allomorphy

Reflexes of *-in- generally show less allomorphy than reflexes of *-um-. As seen above, the reflex of *-um- often appears as *m-* before vowel-initial bases. In some languages the reflex of *-in- retains its vowel in this environment even when the corresponding reflex of *-um- does not: Thao *utaq* 'vomit' : *m-utaq* 'to vomit' : *in-utaq-an* '(what) was vomited up', Paiwan *m-alap* 'take, pick up' : *in-alap* 'one's catch in hunting'. However, in other languages both *-um-* and *-in-* lose the vowel when affixed to vowel-initial disyllables, as in Kiput *m-abit* (< /um-abit/) 'hold' : *n-abit* (< /in-abit/) 'was held', or *m-itoy* (< /um-itoy/) 'push s.o. on a swing' : *n-itoy* (< /in-itoy/) 'was pushed on a swing'. Thao also allows the nasal of *-in-* to assimilate to a following stop or delete before a nasal, as in *cpiq* 'thresh grain by beating' : *c<im>piq* (< /c<in>piq/) 'was threshed', or *ta-tnun-an* 'loom' : *t<u>nan* (< /t<um>nan/) 'weave' : *t<i>-nan* (< /t<in>nan/) 'was woven by s.o.'. As with *-um-, in addition to its use as an independent affix *-in- appears as a component of actor voice affixes that have been reduced from longer underlying forms, as in Tagalog *nag-* (< p<um><in>aR-), *naŋ-* (< p<um><in>aŋ-) or *naka-* (< *m<in>aka-).

Perfective aspect

Reflexes of PAN *-in- that mark perfective aspect are common in Formosan and Philippine languages, and in some languages of western Indonesia:

Atayal (northern Taiwan). Rau (1992) glosses Atayal *-in-* 'past tense': *m-agal* 'take' : *m<in>agal* 'took', *mitaʔ* 'see' : *m<n>itaʔ* 'saw', *m-ariŋ* 'begin' : *m<in>ariŋ/n-ariŋ*

'began'. According to Rau (1992:48) 'The past verbs are used to indicate that the time of a reported event precedes the time of speaking or the time of another event.'

Ilokano (northern Philippines). Rubino (2000) calls Ilokano *-in-* a marker of 'perfective aspect: *m-apán* 'go' : *n-apán* 'went', *surát-en* 'write' : *s<in>úrat* 'wrote', *punás-an* 'wipe' : *p<in>unás-an* 'wiped'

Kelabit (northern Sarawak). Blust (1993a) shows that PAN *-in- is reflected in Kelabit as a marker of past tense or perfective aspect: *arak* 'bamboo railing' : *ŋ-arak* 'guide by the hand, as a blind person', *in-arak* 'was guided by the hand', *bulat* 'open the eyes wide' : *mulat* 'look at someone or something' : *b<in>ulat* 'was looked at', *dadaŋ* 'heat from a fire' : *nadaŋ* 'to heat by a fire' : *s<in>adaŋ* 'was heated by a fire', *pətad* 'separation' : *mətad* 'separate from something' : *pitad* 'was separated from something', *tabun* 'a heap or pile' : *nabun* 'to heap or pile up' : *s<in>abun* 'was heaped or piled up'

The 'reversed perfective' of Thao (Blust 2003a). Although reflexes of *-in- as a marker of verbal aspect generally are glossed 'perfective', subtle differences have developed in some languages. In Thao many examples appear to involve straightforward marking of perfective aspect, as in *m-apa* 'carry on the back' : *m<in>apa* 'carried on the back' : *in-apa* 'was carried on the back', *m-iup* 'blow on' : *m<in>iup* 'blew on' : *in-iup* 'was blown on', *i-tana-utu* 'over there' : *in-i-tana-utu* 'was over there', *duruk* 'stab' : *d<in>uruk* 'was stabbed', *fariw* 'buy' : *f<in>ariw* 'was bought', *kan* 'eat' : *k<m>an* 'to eat' : *k<m><in>an* 'ate' : *k<in>an* 'was eaten', *qpit* 'pinched, as between side and arm' : *q<um>pit* 'carry under the arm' : *q<m><im>pit* 'carried something under the arm' : *q-im-pit* 'was carried under the arm', *tash* 'copy, imitate' : *t<m>ash* 'to copy or imitate' : *t<m><in>ash* 'copied or imitated something' : *t<in>ash* 'was copied, was imitated'. In other examples, however, *-in-* marks a condition that resulted from a past action, but is no longer present. This can be illustrated by sentences (16)-(18):

16.   *nak a hulus shu-liqliq-in cumay*
      1sg lig clothes **shu**-tear-**pv** bear 'My clothes were torn by a bear' (and are still ragged)

17.   *nak a hulus sh<in>u-liqliq-in cumay*
      1sg lig clothes **shu-perf**-tear-**pv** bear 'My clothes were torn by a bear' (but have since been mended)

18.   *cicu p<in>an-shiz-an, ma-qitan iza, mu-qca pan-shiz-an*
      3sg   pan-perf-sick-**an**, **stat**-good already, restored sick 'She got sick, recovered, and got sick again'

In sentences (16) and (17) the focus is not on the action of the bear, but on the result to the clothes, with the result that the perfective marker in (17) acts to 'undo' the tearing. In sentence (18) this is even clearer, since the patient has fallen sick twice, but this is described with the perfective marker *-in-* only where a recovery has already taken place. Other examples of this 'reversed perfective' were recorded in Thao, but usages such as *in-iup* or *k<m><in>an* do not appear to be amenable to this interpretation.

The inceptive infix of Central Philippine languages. Although the reflex of *-in- in Tagalog is sometimes described as marking perfective aspect (Schachter and Otanes 1972:366ff), in most or all Central Philippine languages it is best glossed as an inceptive

marker, describing an action that has begun but may not yet be completed. The semantic variations seen in Thao and Central Philippine reflexes of *-in- are thus two different ways of achieving a similar effect, namely the lack of visible consequences of a past action. In Thao the trajectory is begun-completed-reversed (hence returned to begun), while in Central Philippine languages it is begun (but not necessarily completed). Similar subtleties in the way that reflexes of *-in- function probably occur in other Philippine-type languages, but are yet to be described.

Nominaliser

In addition to their use as an aspect marker, reflexes of PAN *-in- are used to derive deverbal nouns in many AN languages. In most Philippine-type languages this infix has both verbal and nominalising functions, but in some Oceanic languages it functions only as a nominaliser:

Thao: In Thao *-in-* is relatively uncommon as a nominaliser, but occurs in a few examples such as *m-acay* 'die, dead' : *m<in>acay* 'burial place, cemetery', and *saran* 'path, road' : *s<in>aran-an* 'the place where someone has walked'.

Ilokano: Rubino (2000) lists many examples of nominalising *-in-* in Ilokano. Some of these are deverbal, as with *bayu-en* 'to mill rice; crush; bruise' : *b<in>áyo* 'milled (uncooked) rice', or *mátay* 'to die' : *m<in>átay* 'corpse', but others are denominal, as in *búŋa* 'fruit' : *b<in>úŋa* 'child, offspring', *burbór* 'fur, shag' : *b<in>urbúr-an* 'kind of cotton cloth; towels', *butáy* 'coarse rice' : *b<in>utáy* 'pulverised or powdered rice', *gilíŋ-en* 'to grind' : *g<in>íliŋ* 'ground meat', or *súrat* 'letter; writing' : *s<in>úrat* 'article, essay, document'. In most cases Ilokano deverbal nouns are products that result from the action of the verb.

Tagalog: Tagalog *-in-* as a nominaliser generally forms deverbal nouns indicating a product that results from the action of the verb, as with *sáʔiŋ* 'to boil rice' : *s<in>áʔiŋ* 'boiled rice', or *tápa* 'to slice thinly, as meat' : *t<in>ápa* 'meat sliced thinly'.

In a number of the Oceanic languages of the western Solomons *-in- is reflected as a nominaliser, with no trace of verbal functions:

Roviana: (Waterhouse 1949:228ff): *ene* 'walk' : *in-ene* 'journey', *avoso* 'hear' : *in-avoso* 'news, hearing', *tavete* 'to work' : *t<in>avete* 'work', *zama* 'say' : *z<in>ama* 'saying, word', *via* 'clean' : *v<in>ia* 'purity', *salaŋa* 'heal' : *s<in>alaŋa* 'remedy, cure', *gila* 'know' : *va-gila* 'to show' : *v<in>a-gila-gila* 'a sign'.

Hoava: In Hoava, spoken on the island of New Georgia in the Solomons, Davis (2003:39) says that 'This affix is extremely productive and can be used with virtually any active or stative verb to create a noun.'   She illustrates this statement with examples such as 1) result of an action: *asa* 'grate' : *in-asa* 'pudding made from grated cassava', *guzala* 'twist bark to make string' : *g<in>uzala* 'string made from twisted bark', *bukulu* 'defecate' : *b<in>ukulu* 'feces', 2) object which undergoes an action: *babana* 'to tow' : *b<in>abana* 'towed object', *gerigeri* 'gather building materials' : *g<in>erigeri* 'logs, sticks, vines, etc. needed for building a house', *mae* 'come' : *m<in>ae* 'people who have arrived', 3) an abstract noun describing an action: *dumi* 'punch' : *d<in>umi* 'act of punching', *hade* 'wrap' : *h<in>ade* 'act of wrapping', *boru* 'to massage' : *b<in>oru* 'massaging, cure', 4) abstract nouns from experiential and stative verbs: *edo* 'happy' : *in-edo* 'happiness', *hiva* 'want' : *h<in>iva* 'wishes, desires', *to* 'alive' : *t<in>o* 'life'. In a few cases nouns are created from existing nouns, and *-in-* may be inserted within a prefix to nominalise a derived verb, as with *va-bobe* 'to fill' (with causative *va-*) : *v<in>a-bobe* 'filled object',

vari-razae 'fight each other (with reciprocal *vari-*) : *v<in>ari-razae* 'war, battle', *ta-poni* 'be given' (with passive *ta-*) : *t<in>a-poni* 'gift'.

Some non-Oceanic languages also appear to have restricted the functions of *-in-* to nominalisation. Wolff (1972:378ff), for example, lists three homophonous infixes *-in-* for Cebuano Bisayan, none of which mark perfective aspect, and two of which are exclusively nominalising, and Woollams (1996:89ff) describes Karo Batak *-in-* as an infix of low frequency that 'derives nouns from transitive verb stems, nearly all of which happen to begin with /t/.' In addition, although the use of nominalising *-in-* has not survived in the Polynesian languages, Clark (1991) has shown that it may be fossilised in PPN *faŋota 'go fishing' : *fiŋota 'shellfish' (< *f<in>aŋota 'what was obtained in fishing', with irregular loss of the syllable *-na-*).

Portmanteau voice/aspect marker

In other languages the reflex of *-in- has come to function as the marker of patient/goal 'focus' or passive voice. In most cases where adequate descriptions are available *-in-* or its phonemic equivalent in these languages is a portmanteau morpheme, simultaneously marking voice and aspect. The reason for this double function is a peculiarity in the affix potential of verbs in PAN: although *-in- co-occurred with other voice affixes, when it was attached to a base in the patient voice the suffix *-en surfaced as zero, as with PAN *k<um>ali 'dig something up (actor voice, non-perfective)' : *k<um><in>ali 'dug up (actor voice perfective)', *kali-en 'be dug up (patient voice); what is dug up' : *k<in>ali (not **k<in>ali-en) 'was dug up; what was dug up'. When *-en was realised as zero *-in- inevitably took on both aspectual and 'voice-marking' functions. This is not always explicitly recognised in published descriptions, but can sometimes be determined from the data. Topping (1973:187ff), for example, describes three homophonous infixes *-in-* in Chamorro: 1) goal focus, 2) nominalising, 3) adjectivising. He regards the goal focus function of *-in-* as primary, yet every example that he gives of a goal focus construction with *-in-* is translated as an English past. Reflexes of *-in- with portmanteau functions are common in the languages of northern Sarawak, as in Bintulu, where *-ən-* is inserted into free morphemes to mark the passive voice of many verbs, and *-in-* is inserted into *pa-* to mark the passive form of causative verbs, all of which refer to completed actions: *g<əm>azaw/mə-gazaw* 'scratch something' : *g<ən>azaw* 'was scratched by someone', *g<əm>utiŋ* 'cut something with scissors' : *g<ən>utiŋ* 'was cut with scissors', *p<in>a-təba?* 'had water poured on oneself', *p<in>a-təmbə?* 'was felled, of a tree', *p<in>a-səbut* 'was bitten'. Uniquely, Thao allows reflexes of *-in- and *-en on the same base, but this almost certainly is a historically secondary development (Blust 1998c).

Other functions

Other functions associated with reflexes of *-in- include 'intensive', as in Aklanon *káon* 'eat' : *nag-k<in>áon* 'ate and ate' (Salas Reyes, Prado and Zorc 1969:229); the formation of language names or verbs meaning 'speak language X' from ethnic names in much of the central Philippines, as Cebuano *bisáya?* 'a Bisayan person' : *b<in>isáya?* 'the Bisayan way, Bisayan language; speak Bisayan', or *tagálug* 'a Tagalog person' : *t<in>agálug* 'the Tagalog way, Tagalog language; speak Tagalog' (Wolff 1972); 'adjectivising', as with Chamorro *palao?an* 'woman' : *p<in>alao?an* 'womanish', or *a?paka?* 'white' : *in-a?paka?* 'whitish' (Topping 1973:187); 'adjective-formative affix to designate manner or state of being', as with Hiligaynon Bisayan *súgba* 'roast, broil' : *aŋ s<in>úgba-ŋ ísda?* 'the broiled

fish' (Motus 1971:147), and 'affix that derives adjectives from nouns', as in Yakan *pilak* 'peso' : *p<in>ilak* 'wealthy' (Behrens 2002:123). Given the likelihood that it is not due to convergence, the use of *-in-* to derive denominal adjectives or stative verbs in Chamorro and Yakan may reflect another secondary function of PMP *-in-. In addition to these productive uses of *-in-*, fossilised reflexes of an infix with the same shape are found in two widely distributed lexical bases, where its function is obscure: PAN *Caqi 'feces' : *C<in>aqi 'small intestine', PAN *bahi 'woman; female', but reflexes of *b<in>ahi in Malay *bini* 'wife', Numfor *bin* 'woman, female', Tongan *fine-* 'combination form in many words referring to women', and reflexes of *ba-b<in>ahi in Sangir *babine*, Palauan *babíl*, Motu *hahine*, or Hawaiian *wahine* 'woman; female' (Blust 1982a).

Finally, certain structural or distributional differences between *-um- and *-in- remain unexplained. As already noted, with vowel-initial bases the vowel of *-um- is far more likely to be lost than the vowel of *-in-. Moreover, in languages such as Kelabit, which have generally merged antepenultimate vowels as schwa, the vowel of *-in- is sometimes preserved: *zaRami > *dəramih* 'rice straw', *bituka > *bətuəh* 'stomach', *qali-matek > *ləmatək* 'jungle leech', but *layuh* 'withered' : *ŋə-layuh* 'make something wither, as by putting it near a fire' : *l<in>ayuh* 'was withered by s.o.' (cp. *l<əm>ayuh* 'to wither'), *riər* 'turn, roll' : *ŋə-riər* 'turn or roll s.t.' : *r<in>iər* 'was turned or rolled' (cp. *r<əm>iər* 'roll by itself'), *ŋə-rudap* 'put s.o. (as a child) to sleep' : *r<in>udap* 'was laid down to sleep', *tanəm* 'grave' : *nanəm* 'to bury' : *s<in>anəm* 'was buried'. Reflexes of these two infixes also contrast in their geographical distribution: both are common in Taiwan, the Philippines, most of northern Borneo, northern Sulawesi, the Batak languages of northern Sumatra, and Chamorro of western Micronesia. Reflexes of *-um-, however, are not found as an active affix in Oceanic languages or in any language of eastern Indonesia.

6.3.2.3 *-ar- 'plural'

This infix is far less visible than the preceding two. While reflexes of *-um- and *-in- are found in over 200 languages, reflexes of *-ar- appear in only a few. Recognition of this affix is also complicated by two other considerations: 1) some languages appear to reflect *-al-, or *-aR-, and it is sometimes difficult to distinguish these from *-ar-, 2) a number of languages assimilate the vowel of *-ar- to the next vowel of the base, making the location of morpheme boundaries unclear. Evidence for *-ar- is seen in the following:

Pazeh: (Ferrell 1970:78): -ar- 'instrument'. As pointed out by Li and Tsuchida (2001:18-19), this affix has been recorded only in *baranaban* 'urn', and *duŋuduŋ* : *daruŋuduŋ* 'drum'. There is no contrastive evidence for an infix in *baranaban*, and the glosses for *duŋuduŋ* and *daruŋuduŋ* are identical.

Paiwan: (Ferrell 1982). -ar- (plus Ca-reduplication) 'do on all sides, in various directions': *kim* 'search' : *k<ar>a-kim* 'search everywhere', *tjəzak* 'a drop of liquid' : *tj<ar>a-tjəzak* 'constantly dripping; dripping on all sides'.

Hanunóo: (Conklin 1953). -ar- 'plural': *ʔába* 'length, long' : *ʔar-ába* 'long (pl.)', *ʔáni* 'harvesting rice with the bare hands' : *ʔar-any-ún* 'be harvested by hand (pl.)', *daká* 'large' : *d<ar>aká* 'large (pl.)', *diʔít* 'small' : *d<ar>iʔít, d<ir>ʔít* 'small (pl.)', *táʔid* 'close together' : *t<ar>áʔid* 'close together (pl.), as of fence poles, stakes, etc.', *tarúk* 'dance, dancing' : *t<ar>arúk-an* 'dancing, i.e., much dancing', *badíl* 'gun, firearm: *b<ar>adíl-an* 'shooting of firearms'.

The Hanunóo facts are complicated in several ways. First, despite the examples given here, not all uses of *-ar-* mark plurality. In fact, the semantic contribution of *-ar-* is often obscure, as with *ʔugát* 'excreta, waste' : *ʔur-ugát-an* 'rectum, anus; privy, toilet', *gitíŋ* 'any

line modified by notches, undulations, or zigzag cuts' : *g<ar>itíŋ* 'continuous notches, as those cut on the side of a stick', *hábul* 'weaving of cloth' : *h<ar>abl-án* 'backstrap loom', *híruŋ* 'ring, circlet' : *h<ar>íruŋ* 'sacrificial circlet of beads', *sakáy* 'riding on a horse, or in a boat or land vehicle' : *s<ar>aky-án* 'vehicle', *tínduk* 'the largest local variety of plantain' : *t<ar>índuk/t<ir>índuk* 'clitoris'. Second, as seen in examples such as *kibkíb* 'coconut meat' : *k<al>ibkíb* 'that which remains adhering to the inside of a coconut shell after most of the meat has been removed by grating', *siŋát* 'position of an object in the crotch of a split stick' : *s<al>iŋát* 'sticking in the crotch of a split stick or stump', *takúp* 'cover, lid, top' : *t<al>akúp* 'door, a sliding doorway cover' or *túban* 'flesh' : *t<ag>úban* 'meat, flesh', *sumá* 'return' : *s<ag>umá* 'going back to a previous position or condition' : *s<in><ag>umá* 'retrogressed', Hanunóo has formally similar infixes *-ar-*, *-al-* and *-ag-*, and the semantic contribution of the last two is difficult to pin down. Third, as seen in examples such as *d<ir>iʔít* and *ʔur-ugát-an*, the vowel of *-ar-* sometimes assimilates to the first base vowel. This type of infix-specific assimilation, which is reminiscent of the conditions for *-im-* allomorphs of *-um-*, appears to be completely general in Central Philippine languages such as Bikol or Hiligaynon: the former has allomorphs *-ar-*, *-ir-* and *-ur-* where the first base vowel is *a*, *i*, or *u*, and the latter (which has merged *r and *l) has *-al-*, *-il-* and *-ul-*. In languages like Bikol or Hiligaynon the regular anticipatory assimilation of affixal vowels renders the boundary between stem and infix indeterminate (*-ar-* vs. *-ra-*, *-ir-* vs. *-ri-*, etc.)

Central Tagbanwa: Scebold (2005:35-36) describes an infix *-Vr-* in Central Tagbanwa of Palawan Island, Philippines, that marks 'collective aspect'. The vowel of this infix assimilates to the first stem vowel, and *r* and *l* metathesise when the infix is attached to a stem that begins with /l/: *təpad* : *t<ər>əpad* 'be next to one another', *nunut* : *n<ur>unut* 'accompany one another', *anak* : *ar-anak* 'all one's offspring together', *inəm* : *ir-inəm* 'all drink together', *laktu* : *r<al>aktu* (**l<ar>aktu) 'all run together'.

Viray (1973) argued that these infixes differ from *-um-* and *-in-* in being *-CV-* in shape. His principal evidence came from infixed reduplications such as Bikol *pintóʔ* 'door' : *piripintóʔ* 'small door', and *kandíŋ* 'goat' : *karokandíŋ* 'small goat'. The first of these is ambiguous for *p<ir>i-pintóʔ* (with an infixed reduplicant) or *pi-ri-pintóʔ* (with a sequence of prefixes that Viray wished to treat as prefix + infix), but the second appears to be analyzable only as containing a CV prefix *ro-*. However, there is reason to question this analysis. First, a number of the examples that Viray gives from Bisayan languages show a pattern of reduplication in which the vowel of the reduplicant is not a copy of the first base vowel. Regardless of where the infix boundary is placed, Cebuano *bolo-babaye* 'like a woman' (*babaye* 'woman'), and Hiligaynon *polo-panday* 'a carpenter of little experience' (*panday* 'carpenter'), for example, show a mismatch between the vowel of the copied base syllable and the vowel of the base. This allows a form such as Samar-Leyte *saro-sakay-an* 'small ship' (*sakay* 'ship') to be analyzed as *s<ar>o-sakay-an*, but not as *sa-ro-sakayan*, since there is no independent evidence for a prefix *ro-*. Second, a similar pattern of diminutive noun formation is found in Agta of northern Luzon, where the mechanism is not infixation, but rather reduplication with multiple fixed segmentism of the form *Cala-* (Healey 1960:6ff). Given the formal and functional similarity of these morphological processes, a historical connection between them thus appears likely.

Although the forms cited by Viray probably involve an unusual reduplication pattern with fixed segmentism rather than infixation, there is widespread support for infixes **-ar-*, **-al-* or **-aR-*. While Bikol and most forms of Bisayan have obligatory vowel assimilation for *-aC-* infixes, hence obscuring the morpheme boundary, this is optional in Hanunóo, as

seen in *diʔít* 'small': *d\<ar\>iʔit/d\<ir\>iʔít* 'small (pl.)', where *-ir-* must pattern like *-ar-*. In addition, reflexes of reduplicated monosyllables often appear with active or fossilized infixes *-ag-* or *-al-* in the central Philippines, as in Cebuano *b\<ag\>ukbuk* 'wood weevil' (PMP *bukbuk), *bukbuk* : *b\<al\>ukbuk* 'pound something into very fine particles or powder' (root *buk 'pound, thud'), *dukduk* 'pound something repeatedly' : *d\<ag\>ukduk* 'hammering, knocking sound' (PMP *dekdek), or *h\<al\>iphip* 'mend holes in woven material'. Similar evidence of marginal infixes in reduplicated monosyllables is found in some Formosan languages, as with Thao *mismis* 'eyelash; blink' : *mak-m\<ar\>ismis* 'keep blinking', or *pakpak* 'clap' : *p\<ar\>akpak* 'make a popping sound' (with a reflex of *-al- or *-ar-).

Malay: Malay has no active reflexes of *-al- or *-ar-, but there is considerable evidence for fossilised infixes of this shape. As in most AN languages, the great majority of Malay word bases are disyllabic. However, a large percentage of lexical entries that begin with a consonant followed by *-əl-* or *-ər-* are trisyllables. This is true both of Malay (Wilkinson 1959), and of Bahasa Indonesia (Moeliono 1989). In most cases there is no evidence that Malay trisyllables of the shape CəlVCV(C) or CərVCV(C) are bimorphemic, but their deviation from typical morpheme length suggests that they contain fossilised infixes *-al-, *-ar- or *-aR-. To adumbrate what fuller statistical information would reveal more clearly, it would not be expected that bases which begin with *gəb-*, *gəd-*, *gən-*, *gəs-*, or *gət-* would contain a fossilised infix, whereas bases that begin with *gəl-*, or *gər-* might. Figure 6.2 shows the number of bases beginning with each of these sequences that are 1) disyllabic, and 2) polysyllabic in the Malay-English dictionary of Wilkinson (1959). Known loanwords are excluded:

	Disyllables	**Polysyllables**	**Percent in Polysyllables**
gəb-	7	2	22
gəd-	17	37	68.5
gən-	44	18	29
gəs-	5	0	zero
gət-	17	1	5.5
gəl-	19	215	91.9
gər-	52	192	78.7

Figure 6.2 Canonical deviation of Malay bases with *gəl-* and *gər-*

What emerges clearly from these figures is a statistically significant association between the onsets *gəl-* and *gər-* with bases that are more than two syllables in length, suggesting in the absence of paradigmatic contrast that many of these words contain fossilised infixes *-əl-* and *-ər-*. Somewhat surprisingly, the same test provides evidence for an infix *-əd-* for which comparative evidence is unknown. Bases that begin with *gem-* have been omitted from this sample, but would show a similarly strong association with words of more than two syllables, supporting an inference that many of these contain a fossilised infix *-əm-*. Needless to say, the choice of *gəl-* rather than *bəl-*, *dəl-*, *jəl- kəl-* or other possible onset sequences is an artifact of sampling, and similar results would be expected with other *Cəl-* and *Cər-* onsets.

Toba Batak: Van der Tuuk (1971:143ff) describes infixes *-ar-* and *-al-* in Toba Batak that apparently occur only together with another affix. The affix combinations that he describes include 1) *-um-* + *-ar-* with accent on the ultima, 2) *-um-* + *-ar-* with accent on the penult, 3) *mar-* + *-ur-*, 4) *maŋ-* + -ar-, 5) *-um-* + *-al-*, 6) *mar-* + *-al-*, 7) *maŋ-* + *-al-*, and 8) *pa-* + *-al-*. Several of these are associated with plural referents, and some may co-occur with *–um-*: *siŋoh-an* 'choke while gulping' : *s<um><ar>iŋgok* 'sob loudly, of many people'. As in languages of the central Philippines, Toba Batak *-ar-* shows allomorphic vowel variation. However, the extent of variation and its conditioning are fundamentally different. Rather than three allomorphic vowel patterns (or five, as are possible in Central Tagbanwa), only *-ar-* and *-ur-* occur, the latter according to van der Tuuk (1971:145) when the next two syllables contain identical vowels, as in *ponjot* : *mar-p<ur>onjot* 'narrow, tight', or *pitik* : *mar-p<ur>itik* 'thrown away'. As in languages of the central Philippines, Toba Batak also has an infix *-al-,* and its semantic contribution to the larger word of which it is a part is equally obscure.

Sundanese: According to Robins (1959) -ar- and *-al-* may occur in Sundanese to mark 'verb plurals', that is, plural agents, patients or experiencers. The second variant 'is used with forms whose initial consonant is *l*, and with those containing a following *r*, except as initial consonant of the second syllable,' as in *hormat* 'to honor' : *h<al>ormat* 'to honor (of more than one)', *di-bawa* 'be carried' : *di-b<ar>awa* 'be carried, of more than one', *sare* 'to sleep' : *s<ar>are* 'to sleep (of more than one)'. Unlike the other cases described here, then, Sundanese *-al-* and *-ar-* appear to be conditioned allomorphs of a single infix which have diverged as a result of liquid dissimilation.

To summarise, there appears to be comparative evidence for three phonologically similar but distinct infixes *-al-, *-ar- and *-aR-. The identity and distinctness of these affixes is compromised by several considerations: 1) although all three are distinguished by languages such as Hanunóo or Bikol, *-al- and *-ar- merge in most Central Philippine languages, and *-ar- and *-aR- merge in languages such as Malay and Toba Batak, 2) in many languages these affixes are fossilised, 3) there is considerable allomorphy which includes both vocalic and consonantal alternations, sometimes (as in Sundanese) giving rise to *-al-* as an allomorph of *-ar-*, 4) even when they are productive, the semantic contribution of *-al- and *-aR- is obscure; by contrast, reflexes of *-ar- clearly mark plurality in several widely separated languages, although not all uses of this infix can be characterised in these terms. In addition, Ilokano *-an-* marks 'frequentative', and an affix of similar shape but unknown function appears to be fossilised in reflexes of PMP *kanukuh 'fingernail, claw' next to PMP *kuhkuh 'fingernail, claw'.

6.3.2.4 *Double infixation*

As noted above, the PAN perfective marker did not co-occur with *-en 'patient voice'. However, it could co-occur with other voice affixes, including *-um-, thus giving rise to double infixation in the perfective forms of actor voice verbs. This system has been preserved in most Philippine-type languages, but these languages fall into two distinct groups: 1) those that reflect the order *-umin-, and 2) those that reflect the order *-inum-. The issues associated with these different infixal orders are thoroughly discussed by Reid (1992), who documents the distribution of both types in a number of the languages of the Philippines. Table 6.7 provides an overview of the cross-linguistic distribution of these two patterns, drawing heavily on the data in Reid (1992), but adding other languages where this is possible:

Table 6.7 Distribution of languages reflecting *-umin- and *-inum- infixal orders

	*-umin-	*-inum-	*ni-C<um>
Taiwan			
Atayal	X		
Thao	X		
Saaroa			X
Kanakanabu			X
Philippines			
Itbayaten	X		
Ivatan	X		
Ibanag	X (> -imin-)		
Ilokano		X (> -imm-)	
Balangaw	X	X	
Kalinga	X	X	
Casiguran Dumagat	X	X	
Ifugaw		X	
Southern Alta	X		
Yogad		X	
Pangasinan		X	
Bolinao		X	
Kapampangan	X		
Tagalog	X		
Agutaynen	X (> -imin-)	X	
Palawano	X		
Aborlan Tagbanwa	X		
Bikol	X		
Sorsogon	X		
Waray		X	
Hiligaynon		X	
Mamanwa	X		
Mansaka		X	
Sarangani Manobo	X		
Tausug		X (> -im-)	
Borneo			
Timugon Murut	X (> -imin-)		
Kadazan	X		
Sulawesi			
Tondano	X	X (> -im-)	
	21	14	2

Which order was original? Both possibilities have been defended by different writers. Wolff (1973:73) posited PAN *-inum-. Zorc (1977:247) reconstructed *-umin- for Proto Bisayan, without further comment on the earlier history of this pattern, Reid (1992:79) proposed the same order (*-umin-) for 'Proto Extra Formosan' (= PMP), and Ross (1995a, 2002a) attributed *-umin- to PAN. The numbers in Table 6.7 generally support this

inference, in that reflexes of *-umin- occur more often than reflexes of *-inum-. However, a number of the northern Luzon patterns may be products of a single historical change.

Three features of Table 6.7 merit a brief comment. First, Saaroa and Kanakanabu reflect *-in- preceding *-um-, but the former appears as a prefix and the latter as an infix-: Saaroa *t<um>aəvə* 'to cover (AF)': *ɬi-t<um>aəvə* 'covered', Kanakanabu *k<um>aənə* 'to eat (AF)' : *ni-k<um>aənə* 'ate'. This asymmetry raises issues that will be addressed in 6.3.2.5. Second, Reid notes that *both* infix orders are found in Balangaw, Kalinga, Casiguran Dumagat and Agutaynen, but that they are complementary. A similar situation is found in Tondano of northern Sulawesi (Sneddon 1975:211). Third, Ibanag, Agutaynen and Timugon Murut show assimilation of the back vowel of -*um*- to the front vowel of a following -*in*- (hence -*imin*-). This is very similar to the assimilations in Bolaang Mongondow and Muna of Sulawesi, where -*um*- has an allomorph -*im*- if the first base vowel is *i* (note that this allomorph is distinct from e.g. Tausug -*im*- 'the begun aspect form of -*um*-', or Tondano -*im*- 'actor focus perfective in consonant-initial bases', which are portmanteau affixes formed by fusion of -*in*- + -*um*-). What is interesting about this assimilation is that it apparently is limited to the infix -*um*- and so cannot be regarded as a type of general vowel harmony.

Reid (1992) shows that -*umin*- results from infixing -*in*- before the first base vowel prior to infixing -*um*-, and that the order -*inum*- results from the opposite order of affixation. The change *-umin- > -inum- therefore presumably is morphological, as a phonological change of this type would require the unprecedented assumption of syllable metathesis. Reid's explanation for this change turns on resolving the problem of determining whether -*um*- and -*in*- are inflectional or derivational, a distinction that is sometimes very difficult to make in Philippine-type languages. He argues (1992:78) 'that as -*in*- has become more and more productive as an aspect affix on verbs, it has moved from being a derivational affix to an inflectional affix,' and hence has become more peripheral in relation to the stem. This is an interesting idea, but one that assumes, contrary to substantial distributional evidence from Formosan and Malayo-Polynesian languages, that the aspect-marking function of -*in*- is historically secondary, and that the non-derivational uses of -*in*- are more common than those of -*um*-.

6.3.2.5   *Infixes or prefixes?*

A number of examples already given to illustrate the use of -*um*- or -*in*- in word-formation processes show variants which have the shape CV- rather than -VC-. This naturally raises the question whether *-um- and *-in- or their reflexes in attested languages are better characterised as infixes or as prefixes. Table 6.8 provides an overview of infixal and prefixal reflexes of the affixes which traditionally have been reconstructed as *-um- and *-in-. Where a language has multiple allomorphs which adhere to a common pattern of either infixation or prefixation the most typical or frequent variant is taken as representative. As has been repeatedly discussed in the literature, because these affixes are inserted before the first base vowel they are infixed to consonant-initial bases but prefixed to vowel-initial bases. This distributional trait, found throughout the AN family, has no effect on the classification of *-um- and *-in- as infixes or prefixes. Rather, they are regarded as prefixes only when they precede a consonant-initial base or have the shape CV- before a vowel-initial base:

Table 6.8 Infixal and prefixal reflexes of PAN *-um- and *-in-

	*-um-	*-in-
Taiwan		
Atayal	-m-	-in-
Saisiyat	-om-	-in-
Pazeh	mu-	-in-
Thao	-m-	-in-
Kanakanabu	-um-	ni-
	mu-/__p,v,m, ŋ	-in-/__t,c,s
Saaroa	-um-	ɬi-
	u-/__p,v,m,ŋ	
Paiwan	-m-	-in-
Puyuma	-əm-	--
	mə-/__l,r	
Philippines		
Ilokano	-um-	-in-
Bontok	-om-	-in-
Pangasinan	-om-	-in-
Botolan Sambal	-om-	-in-
		ni-/__l,y
Tagalog	-um-	-in-
		ni-/__l,r (loans),w,y
Bikol	-um-	-in-
Hanunoó	-um-	-in-
Aklanon	-um-	-in-
Cebuano	mu-	-in-
Yakan	-um-	-in-
		i-/__l
Tausug	-um-	-i-
Borneo-Madagascar		
Kadazan	-um-	-in-
		ni-
Lun Dayeh	-əm-	-in-
Kiput	-əm-	-ən-
Bintulu	-əm-	-ən-/-in-
Mukah Melanau	mə-	nə-
	-əm- (older speakers)	-ən-
Uma Juman Kayan	-əm-	-ən-
	mə-	nə-
Malagasy	-om- (fossilised)	--

Sumatra-Java

Karo Batak	-um-	-in-
Toba Batak	-um-	-in-
Rejang	-əm-	-ən-
	mə-/__l,m,n, #CVC#	nə-/__l,m,n, #CVC#
Old Javanese	-um-	-in-

Sulawesi

Toratán	-um-	-in-
		ni-/__V,l,r
Tondano	-um-	-in-
		ni-/__V
Bolaang Mongondow	-um-	-in-
		i-/h,l,r,y, nasals
Makasarese	-um- (fossilised)	--

Micronesia

Palauan	-m- (?)	-il-
Chamorro	-um- ~ mu-	-in-

Table 6.8 shows that PAN *-um- and *-in- are reflected as surface infixes in about 76% of daughter languages (34 of 43 positionally-determined reflexes of *-um-, and 33 of 45 positionally-determined reflexes of *-in-). Moreover, in nearly every case of phonological conditioning the infixed variant is found in the 'elsewhere' environment, implying that it is underlying and has metathesised into (C)V- position under certain conditions. The environments for prefix-infix allomorphy are sometimes fairly complex, but the one cross-linguistic pattern that emerges clearly from this data is a tendency for prefixal allomorphs to occur before base-initial sonorant consonants, particularly liquids. In Kanakanabu *mu-* occurs before base-initial labials and *ŋ-*, but *-um-* is found elsewhere. On the other hand, *-in-* occurs only before base-initial *t, c* and *s*, and *ni-* elsewhere. The unconditioned allomorph is thus infixal for *-um-*, but prefixal for *ni-*. In other languages prefix/infix allomorphy is absent, but one affix is prefixed and the other infixed, as with Pazeh *mu-*, *-in-* (Blust 1999a), Tagalog *-um-*, *-in-* ~ *ni-* or Chamorro *-um- ~ mu-*, *-in-*. Finally, Mukah Melanau and Uma Juman Kayan of Sarawak, and Rejang of southern Sumatra have prefix/infix allomorphy that depends in part on number of syllables in the base. In Rejang *-əm-* and *-ən-* occur in disyllables that begin with non-labial obstruents, *m-* and *n-* occur before base-initial vowels, zero and *ne-* occur before monosyllabic bases that begin with *m-* or *mb-*, and *mə-* and *ne-* are found before disyllables that begin with a sonorant consonant, or monosyllables that do not begin with *m-* or *mb-* (McGinn 1989:100ff). In both Mukah Melanau and Uma Juman Kayan monosyllabic bases (less than 5% of the total) take only prefixal variants *mə-* and *nə-*, while disyllables take either *mə-/nə-* or *-əm-/-ən-*. In Uma Juman Kayan prefixal and infixal allomorphs in disyllables apparently vary freely, but in Mukah Melanau this is not the case, since speakers born c. 1955 invariably use prefixes, while speakers of their grandparental generation born c. 1910 are said to use only the infixed equivalents (Blust 1988c:184). Finally, a few languages appear to have optionally metathesised *-in-* to *ni-*, and the infixed and prefixed variants of the same base have come to acquire distinct functions. In Kadazan, for example, *-in-* is said to carry 'the idea of past action', while *ni-* refers to 'the object of the action rather than the

action itself', as with *b<in>oos* 'spoke, said', but *ni-boos* 'the thing said' (Kadazan Dusun Cultural Association 1995:xxxviii).

Historically, it is clear that *-um- and *-in- were infixes. Synchronic proposals that *-um-* or *-in-* are underlying prefixes that metathesise under certain conditions cannot readily be separated from the historical facts, and this compels us to conclude that where a language has both variants the movement is most simply explained as from infix to prefix rather than the other way around. Some of the more important considerations that favor this interpretation are:

Cross-linguistic distributional evidence:

If 79% of the positionally-determined reflexes of *-um- and 73%of the positionally-determined reflexes of *-in- in Table 6.6 are infixes, the reconstruction of *mu- and *ni- would require massive convergence in the daughter languages to produce the observed results, but with no clear motivation.

Evidence from child language:

Cross-linguistically infixes are far less common than prefixes or suffixes, and hence are the most marked of the three affix types. The expectation in first-language acquisition is that children will find marked structures more difficult to learn, and will tend to unmark them. Although relatively little work has been done to date on first-language acquisition of languages with infixes, the existing literature conforms to expectation. Galang (1982), for example, has shown that Tagalog-speaking children acquire suffixed verb forms with relative ease, but have greater difficulty with *-um-* and *-in-*. Both children and adults sometimes omit *-um-* altogether, using just the bare base (e.g. *káin* 'eat' instead of *k<um>áin*). In addition, children not uncommonly metathesise *-in-* to *ni-* in environments where adults do not (Galang 1982:12).

Generational differences:

In Mukah Melanau the variation between prefixes and infixes reflecting *-um- and *-in- is correlated with generational differences. As already noted, speakers of the older generation use infixes, while younger speakers use prefixes.

Historical records:

Exactly the same picture emerges where historical records of earlier stages show prefix/infix metathesis, namely earlier infixes have metathesised into prefix position, not vice versa. Lobel (n.d.) notes that Cebuano Bisayan as recorded by Bermejo (1894) had contrasting affixes *mu-* 'future tense', and *-um-*, the latter marking a variety of functions that included imperative, infinitive, and 'actor voice' or 'actor focus'. In modern Cebuano both of these are *mu-* (Wolff 1972:689). Bermejo lived from 1777-1851, so the material for his Cebuano grammar must have been collected during the first half of the nineteenth century. Lobel notes that 'this work is at least to some extent, a re-write by Bermejo of the original grammar by Encina.' Since Francisco Encina lived from 1715-1760, Cebuano probably distinguished *mu-* 'future' from *-um-* 'imperative; actor focus' during the eighteenth and early nineteenth centuries. As Lobel points out, the same distinction is also found in the Waray dialect on northeastern Biliran Island, and in Coastal Miraya Bikol as

spoken in Jovellar and Donsol, where the contrast is seen in pairs such as *um-iliŋ* 'to look' vs. *mu-iliŋ* 'will look', or *l<um>uwas* 'to come out' vs. *mu-luwas* 'will come out'.

Ablaut:

As will be seen in 6.4, many languages in Sarawak have innovated systems of verbal ablaut that derive from *-um- and *-in-. If ablaut in these languages is a product of a single historical change, this change must have occurred at least 2,500 years ago, and *-um- and *-in- must have been infixes at that time. Both the distribution of ablaut systems and differences of detail in their form suggest that this innovation began in the Melanau coastal area of Sarawak and spread northward into the basin of the Baram River. Comparative reconstruction and internal reconstruction indicate an identity between *-um-/-in- and *u/i* ablaut, and this in turn shows that the metathesis of infixes to prefixes is a relatively recent development.

Fossilised forms:

It has long been known that there is a preference for disyllabic word bases throughout much of the AN language family (Blust 2007b). Given this tendency many languages that reflect PAN *kaen 'eat' as a monosyllable through regular sound change have preserved infixed forms of this base after the infixes have otherwise been phonemically altered, or ceased to be productive. An example is Kayan, of central Borneo, which has an active infix *-əm-*, but also has *k<um>an* 'to eat' next to *kan-i* 'eat (imper.)', *ma-kan* 'to feed (people)', and *pa-kan* 'to feed (animals)'. Both a free base *kan* and an infix *-um-* are otherwise unknown in the language. Similarly, in languages such as Malagasy, where the reflex of *-um- is moribund this morpheme appears as an infix.

Pseudo nasal substitution:

The phenomenon of pseudo nasal substitution (PNS) is found in languages reaching from Taiwan to Palauan in western Micronesia. For PNS to take place an affixed word must contain the sequence *b-Vm- or *p-Vm-. A sequence *mV-b- or *mV-p- does not produce the same effect, as witnessed by numerous reflexes of *ma- 'stative' before *b*-initial or *p*-initial bases throughout the AN family (Blust 2004a). The geographical distribution of languages which show PNS effects is testimony to the presence of *-um- as an infix in the history of many languages, some of which no longer make active use of this grammatical morpheme. This is most clearly evident in the PMP prefixes *maŋ- and *maR-, which almost certainly derive from *p<um>aŋ- and *p<um>aR- (Wolff 1973) through CV- truncation.

*PAN *mu- 'movement prefix':*

As noted already, PAN had a prefix *mu- which generally marked verbs of motion (Blust 2003c:451ff, 471). Known reflexes of *mu- are found only in Formosan and central Philippine languages, where they differ from reflexes of *-um-: Thao *tuqris* 'snare trap' : *t<m>uqris* 'catch with a snare trap', but *mu-tuqris* 'blunder into a snare trap', *s<m>apuk* 'to catch (a person, etc.)', but *sazum* 'water' : *mu-sazum* 'enter the water', *fafaw* 'top, upper part' : *mu-fafaw* 'go up, climb up', but *fariw* 'buy', with zero-marked actor focus; Puyuma *dirus* 'bathing' : *d<əm>irus* 'bathe', but *mu-dirus* 'soak with water, as plants', *riwak* 'clearing away' : *mə-riwak* 'clear away', but *ruma* 'house' : *mu-ruma* 'go home,

enter the house', *tamaku* 'tobacco' : *t<ǝm>amaku* 'smoke a cigarette', but *Talu-Talun* 'forest' : *mu-Talun* 'go hunting'. Since PAN also had *pu-* 'causative of motion', it is likely that *mu-* derives from *p<um>u-*. However, the resulting surface form is invariably a prefix, while reflexes of *-um-* are infixal.

Double infixation:

   If *-um-* and *-in-* and their reflexes are treated as prefixes attested forms with double infixation require not one, but two separate metatheses to yield observed sequences such as *-umin-* or *-inum-*. If there is no obvious motivation for a single metathesis there is clearly even less motivation for a double metathesis. In this connection it is worth noting that although *-um-* and *-in-* are reflected positionally or unconditionally as *mu-* and *ni-* in some languages, no case is known of *muni-* or *nimu-*. The same problem arises where reflexes of *-um-* and *-in-* co-occur with other infixes, as with Hiligaynon/Ilonggo *púyoʔ* 'live, dwell, reside': *p<um>
-úyoʔ* 'residents', or Hanunóo *sumá* 'return?' : *s<in><ag>-umá* 'retrogressed'. Double infixation is thus very difficult to reconcile with a hypothesis of historical or underlying prefixes *mu-* and *ni-*.

   Despite this substantial collection of observations in support of an infixal analysis (and the potential additional support provided by the Austric hypothesis, where *-um-* and *-in-* are posited for the common ancestor of both AN and Austroasiatic), some scholars have argued for a prefixal interpretation either for a given language, or for AN languages more generally. Topping (1973:170ff), for example, suggests that Chamorro *-um-* and *-in-* (the only infixes he gives) should be regarded as underlying prefixes *mu-* and *ni-* 'with appropriate rules for metathesis.' The arguments he offers for this analysis are: 1) the canonical form of *-um-* and *-in-* is 'just the opposite of what one would expect in a language that is basically CVCV,' 2) *-um-* and *mu-* are in free variation, especially in the Chamorro spoken on Guam, 3) when 'the infix *-um-* occurs with the prefix *na'-* plus a stem, the resultant form is always *muna'-*,' 4) the form *fanchinemmaʔ* 'forbidden things' is composed of *fan-* 'plural marker', *chommaʔ* 'forbid' and *-in-* (which triggers vowel harmony). However, nasal substitution should operate to produce **fañinemmaʔ*. To explain the absence of nasal substitution in this form Topping suggests a derivation *ni-chommaʔ > ni-chemmaʔ > fan-ni-chemmaʔ > fanchinemmaʔ*. He then concludes 'The above evidence suggests very strongly that infixes in Chamorro (and possibly other Philippine languages?) may actually have prefixes as their underlying forms.'

   There are multiple problems with this analysis. First, the 'appropriate rules for metathesis' are not rules, since there are other CV- prefixes in Chamorro (*gi-*, *ha-*, three distinct types of *ka-* and *ma-*) that do not metathesise. The derivation of *-um-* and *-in-* from underlying *mu-* and *ni-* is thus unmotivated and irregular. Second, it is odd to claim that VC infixes are subject to metathesis because they are canonically deviant. If this is true of infixes it should also hold for suffixes such as *-an* or *-on*, which never metathesise even though -CV suffixes are allowed (*-mu* '2s possessor', *-ña* '3s possessor', *-ña* 'comparative degree'). It is true that Chamorro has few VC syllables, but this is a result of glide epenthesis and fortition in the separate history of this language, not something that is generally true of AN languages which reflect PAN *-um-* and *-in-* as infixes (Blust 2000c). Third, the claim that *mu-* and *-um-* vary freely tells us nothing about direction of change; if anything, the observation that *mu-* is heard mostly on Guam rather than the northern islands points instead to the dialect of Guam as innovative. The last two points that Topping raises are more substantive, but carry little weight. The form *mu-naʔ* for expected **n-um-aʔ* as the actor focus causative could be due to analogy. As Topping

points out, the corresponding object focus form is *ninaʔ*, which is ambiguous for *n\<in>aʔ* or *ni-naʔ*. Such an ambiguity in one member of a set of paired terms that presumably have high text frequency clearly would set the stage for reanalysis of *n\<um>aʔ* as *mu-naʔ*. The absence of nasal substitution in *fanchinemmaʔ* is puzzling, but evidently is confined to this word, and as can be seen in Topping, Ogo and Dungca (1975), nasal substitution is not fully consistent in Chamorro.

Similar proposals have appeared in the general theoretical literature, and these are equally weak. Halle (2001), who provides a useful survey, divides these into 1) the 'traditional' view that reflexes of *-um- and *-in- are infixes, 2) the 'functional' view, expressed in Optimality Theory, that these affixes have the shape VC, but are prefixed both to vowel-initial and consonant-initial bases with subsequent metathesis (although Halle avoids this term), and 3) his own view, which corresponds closely with that of Topping for Tagalog and Chamorro. The traditional (i.e. specialist) view is quickly dismissed in effect by theoretical fiat, and Halle expends his major effort in showing why the 'functional' view of Optimality Theory is ill-founded for Tagalog and Chamorro, but not necessarily for Toba Batak. The observations following Table 6.8, however, show that *any* view which treats the reflexes of *-um- and *-in- as prefixes is ill-founded.[62]

6.3.3 Suffixes

PAN had some clear suffixes, as *-an 'locative voice', *-en 'patient voice', and *-ay 'future tense'. Some other suffixes that are widely shared across the language family may have been prepositions that were phonologically captured during the evolution of the AN languages, as with *i 'generic preposition of location' > *-i 'local transitive'. These will be treated in Chapter 7.

6.3.3.1　*-an 'locative voice'

This probably is the most widespread suffix in AN languages. In Philippine-type languages either singly or in combination with other affixes it has a number of distinct functions, but it is most commonly used to mark the location of an action in relation to the 'focused' NP, as with Ilokano *sagád-an* 'to sweep the floor' (with focused location) next to *sagád-en* 'to sweep dirt' (with focused object). Schachter and Otanes (1972:314) list five actor voice affixes and their locative voice counterparts for Tagalog: 1) *ma-* : *ka-X-an* or *pag-X-an* 2) *maka-* : *ka-X-an*, 3) *-um-* : *pag-X-an*, 4) *mag-* : *pag-X-an*, 5) *maŋ-* : *paŋ-X-an*, as in *ma-matáy* 'die' : *ka-matay-án* 'die in/on', *ma-túlog* 'sleep' : *pag-tulúg-an* 'sleep in/on', *maŋ-isdáʔ* 'go fishing' : *paŋ-isdaʔ-án* 'go fishing in/on'. This reduces to three affix combinations with a common element *-an*. In addition to verbal functions *-an* (often with final stress) creates nouns of location in Philippine-type languages as in Tagalog *hábi* 'texture; woven pattern on fabric' : *habih-án* 'loom', *títis* 'cigar or cigarette ash' : *titis-án* 'ash tray'. In non-Philippine-type languages reflexes of *-an* often function exclusively to derive nouns of location, as in Kelabit *guta* 'wade across a river' : *gəta-an* 'fording place',

---

[62]　In addition, Thao allows *-in-*, which is normally infixed before the first base vowel, to be prefixed as *in-* (or its allomorphs) in some forms, as in *ma-bric* 'heavy' : *im-bric-an* 'weighed' (**b\<in>ric-an), *fari* 'wind' : *min-fari* 'to blow, of the wind' : *in-fari-n* 'was blown by the wind' (**f\<in>ari-n), or *shuqish* 'return' : *in-shuqish-ik* (**sh\<in>uqish-ik) 'I returned (it)'. Conditions remain unclear, but this allomorphy cannot arise from segmental metathesis, and the movement from infix to prefix is difficult to reconcile with any of the views proposed in the generalist literature.

irup 'drink' : *rup-an* 'watering hole (where animals drink)', *tələn* 'to swallow' : *tələn-an* 'throat', or Makasarese *-aŋ* 'formative for locative nouns', as in *əntəŋ* 'stand' : *əntəŋ-aŋ* 'place where one stands'. In a few languages a reflex of *-an that derives locative nouns is fossilised or rare, as with Roviana *huve* 'bathe' : *hu-huve-ana* 'bathing place, bath'.

The semantics of *-an presents an interesting continuum from non-controversially locative senses through senses in which a body part or manufactured implement is the location. Kelabit *tələn-an* 'throat' uses a reflex of *-an to derive the name of a body part (the place where swallowing occurs), and Tagalog *tahíp* 'winnowing; shaking grain to remove husks or chaff' : *táhíp-an* 'flat basket used in winnowing' uses a reflex of *-an to derive the name of a manufactured implement (fossilised in Kayan and other languages of central and western Borneo as *tapan* 'winnowing basket'). More extreme departures from a transparently locative sense are seen in PMP *tian 'abdomen' : *tian-an 'pregnant' (lit. 'in the belly'), and most remarkably in reflexes of *waNiS-an 'wild boar' and *RiNaS-an 'Swinhoe's blue pheasant' in Formosan languages, from bases *waNiS 'boar's tusk' and *RiNaS 'long tail feathers of a pheasant', evidently reflecting a conceptualisation of these animals as the sources (locations) of animal products that were highly valued in the traditional cultures of Taiwan (Blust 1996c).

6.3.3.2 *-en 'patient voice'

This suffix, which Wolff (1973:73) called the 'direct passive' (as opposed to 'local passive' and 'instrumental passive') generally marks the patient. It plays a central syntactic role in the verb systems of Philippine-type languages, and like *Si- and *-an, is often used in nominal derivation as well. As already noted, in its verbal uses the contrast between *-en and *-an is nicely illustrated by Ilokano *sagád-en* 'sweep dirt' vs. *sagád-an* 'sweep floors' (Rubino 2000), and similar examples could be cited from many other languages. As nominalisers, reflexes of *-en and *-an also contrast in that the former are used to derive patient nominals and the latter nouns of location. The geographical distribution of nominals derived by suffixation with *-an or *-en also shows interesting differences. As seen above, in some languages that lack a Philippine-type verb system reflexes of *-an are used productively to derive locative nouns. In general, however, nominalisations formed with a reflex of *-en are far less common in languages that lack a Philippine-type verb system. There is one notable exception: the affixational paradigm for PAN *kaen 'eat' includes *kaen-en 'be eaten', and a noun reflecting this affixed form is found both in Philippine-type languages (Thao *kan-in* 'be eaten by someone', *(ka)kan-in* 'food', Yami *kanən*, Ilokano *kanén* 'food', Casiguran Dumagat *kanən* 'eat something; food, cooked rice', Botolan Sambal, Kalagan *kanən*, Kalamian Tagbanwa *anən* 'cooked rice', Tausug *kaun-un* 'be eaten by someone; cooked rice', Malagasy *hánina* 'be eaten; food'), and in languages that lack a Philippine-type verb system (Mukah Melanau *uaʔ kanən* 'any special food, as one's favorite food', Kayan *kanən* 'cooked rice, food', Palauan *kall* 'food', Tongan *kano* 'flesh or substance', Rennellese *kano* 'kernel, as of a nut; flesh, as of a coconut, fish or bivalve', Nukuoro *gano* 'flesh of, meat of'). Apart from this one form patient nominals derived by a reflex of *-en in languages that do not have a Philippine-type verb system are rare.

6.3.3.3 *-ay 'future'

Clear reflexes of *-ay are confined to Formosan and Philippine languages, where they mark future tense or related notions. In Pazeh stative verbs may be marked for future by

suffixation of *-ay* alone, as in *m-azih* 'ripe, cooked' : *ma-azih-ay* 'will ripen, will be cooked', *ma-busuk* 'drunk', *ka-busuk-an* 'be affected by drunkenness' : *ka-busuk-ay* 'will be intoxicated', or *hakəzəŋ* 'old, of people' : *hakəzəŋ-ay* 'will grow old'. The future of dynamic verbs, on the other hand, requires both CV reduplication and suffixation with *-ay*, as with *mu-bizu* 'to write' : *bi-bizu-ay* 'will write', *mi-kiliw* 'to call' : *ki-kiliw-ay* 'will call', *mu-luzuk* 'to comb' : *lu-luzuk-ay* 'will comb', or *mu-tahan* 'grow wealthy' : *ta-tahan-ay* 'will grow wealthy'. Since CV reduplication marks the future in Tagalog, it is possible that both morphological processes played a role in marking the future tense of at least certain types of verbs in PAN.

In Rukai future tense is marked by *prefixation* of *ay-*, as in *ʔacay* 'die' : *ay-ʔacay* 'will die' (Li 1973:262), and in some Philippine languages *-ay* marks what Wolff (1973:73) describes as 'future-general action dependent subjunctive'.

6.3.3.4 *Single consonant suffixes*

Inherited single consonant suffixes are rare in AN languages, although historically secondary single consonant suffixes occur. Many Oceanic languages in the eastern Admiralty Islands, Vanuatu and Micronesia have lost earlier final vowels, including those on -CV suffixes. This process is clearest in the possessive pronouns reflecting POC *-gu '1sg' ([ŋgu]), *-mu '2sg' and *-ña '3sg', as in Lindrou (Admiralty Islands) *mada-k* 'my eye', *mada-m* 'your eye', *mada-n* 'his/her eye', Sye (southern Vanuatu) *ntelgu-g* 'my ear', *ntelgo-m* 'your ear', *ntelgo-n* 'his/her ear', or Kosraean (Micronesia) *siyuh-k* 'my belly', *siyo-m* 'your belly', *siyac-l* 'his/her belly'. In addition to historically reduced possessive suffixes, many Nuclear Micronesian languages have cliticised the genitive marker *ni to a preceding noun and then lost the final vowel. Dyen (1965b:33) called this the 'construct suffix' in Chuukese, while Goodenough and Sugita (1980:xxv) call it a 'relational particle'. What is noteworthy is that whereas PAN *ni could only occur between two nouns linked in a genitive relationship (or when marking the agent of a non-actor voice verb), in languages such as Chuukese the reflex of *ni has given rise to a generic form of obligatorily possession: *nii* 'tooth' : *nii-y* 'my tooth', *nii-m^w* 'your tooth', *nii-n* 'his/her tooth', *nii-n* 'tooth of' (< POC *nipon ni), *chcha* 'blood' : *chchaa-y* 'my blood', *chchaa-m^w* 'your blood', *chchaa-n* 'his/her blood', *chchaa-n* 'blood of' (< POC *raRaq ni), etc. Apart from such historically shortened morphemes suffixes normally contain a vowel in AN languages. The most noteworthy exceptions to this generalisation are: 1) vocative markers, and 2) nominal suffixes, often of uncertain function, in a number of the languages of eastern Indonesia.

The morphology of vocatives in AN languages stands apart from other types of lexical derivation, as it often involves stress shift and even phoneme subtraction. A number of languages show vocative forms that are distinguished by the addition of *-ŋ*, with concomitant stress shift, as in the following reference : address terms for 'father' (*ama) and 'mother' (*ina): Casiguran Dumagat *áma* : *amə́ŋ*, *ína* : *inə́ŋ*, Hanunóo *ʔáma* : *ʔamá-ŋ*, *ʔína* : *ʔiná-ŋ*, Toba Batak *áma* : *amá-ŋ*, *ína* : *iná-ŋ*. Another set of languages distinguishes vocative from reference forms by final glottal stop: Bario Kelabit *tə-taməh* : *tamaʔ*, *təsinəh* : *sinaʔ*, Long Wat *tamən* : *amaʔ*, *tinən* : *inaʔ*, Long Merigam *tamən* : *maʔ*, *tinən* : *naʔ*, Batu Belah *tamah* : *amaʔ*, *tinah* : *inaʔ*, Bolaang Mongondow *ompu* 'lord, ruler; ancestor' : *ompuʔ* 'formulaic opening of an invocation to the higher powers, as in swearing an oath, healing, etc.', Tae' *adi* 'younger sibling (ref.)' : *adiʔ* 'younger sibling (add.)'. Finally, still another set of languages contains vocative forms of kin terms that are distinguished by the addition of final *-y*: Long Atip Kayan *tama-n* : *ama-y*, *hina-n* : *ina-y*, Long Dunin Kenyah

tamə-n : *ama-y, sinə-n* : *ina-y*, Casiguran Dumagat *ápo* 'grandfather (ref.)' : *bóboy* 'grandparent/grandchild (voc.)' (< *bubu-y).

Atypical single consonant suffixes are also found in a number of the languages of eastern Indonesia, and here their function is sometimes obscure. In Atoni (Dawan) of western Timor nouns that are obligatorily possessed are often suffixed with *-f* when a specific possessor is not indicated. With body parts this appears to mark the noun as a general, or non-individuated category: *mata-f* 'eye (in general); people's eyes, everyone's eyes' (cp. *au mata-k* 'my eye', *hɔ mata-m* 'your eye', *in mata-n* 'his/her eye' < PMP *mata), *nima-f* 'hand' (PMP *lima), *fufu-f* 'fontanel' (PMP *bubun), *ma-f* 'tongue' (PCEMP *maya), *siku-f* 'elbow (PMP *siku), *tu-f* 'knee (PMP *tuhud). With kin terms generic possession requires full reduplication of the base: *aina-f aina-f* 'mothers' (cp. *au ainaʔ* 'my mother'; PMP *ina), *ama-f ama-f* 'fathers' (PMP *ama), *oli-f oli-f* 'younger siblings' (PMP *huaji). Nouns that are not obligatorily possessed cannot take *-f*, hence *ikaʔ* 'fish' (**ika-f), *fafi* 'pig' (**fafi-f), *afu* 'ashes; ground' (**afu-f). The contrast is seen nicely in reflexes of PMP *asu 'dog', which has split semantically into *asu* 'dog' and *asu-f* 'slave', since traditionally slaves were important property (the suffixed form is never used to refer to dogs). Similarly, in *hau nɔ-f* 'leaves in general' (lit. 'leaves of a tree < PMP *dahun kahiw), with the 'reversed' genitive common in eastern Indonesia, *-f* marks non-specific possession in a part-whole relationship. The suffix *-f* in Atoni and other languages of Timor can be viewed in either of two ways. On the one hand, it might be seen as a genitive marker that is structurally comparable to the 'construct suffix' of Chuukese. On the other, it might be seen as a kind of anti-genitive marker, since it is used only with obligatorily possessed nouns, but marks the unpossessed form of the noun, which in most other AN languages would be unmarked. Whichever view is adopted, there are important differences between the construct suffix of Nuclear Micronesian languages and suffixes such as Atoni *-f* in the languages of eastern Indonesia. First, the construct suffix derives from a historical genitive marker *ni, while Atoni *-f* has no known etymology, and evidently was innovated as a single consonant, since final vowels were not lost in this language. Second, some body-part terms or names of substances that take the construct suffix in Nuclear Micronesian languages evidently cannot take *-f* in Atoni, as with *naʔ* 'blood'. Third, in some possessive paradigms *-f* represents all persons except 1sg and 1p incl., as in *au nima-k* 'my hands', *hɔ nima-f* 'your hands', *in naʔ nima-f* 'his/her hands', *hitaʔ nima-k* 'our (incl.) hands', *haiʔ nima-f* 'our (excl.) hands', *hiʔ nima-f* 'your (pl.) hands', *sin naʔ nima-f* 'their hands'.

Tetun, spoken in central Timor, has suffixes *-k*, *-n*, and sometimes *-t* on instrumental nouns formed by Ca- reduplication: *huu* 'blow on' : *ha-huu-k* 'blowpipe', *keke* 'to scratch' : *ka-keke-k* 'a rake', *leno* 'show, be visible' : *la-leno-k* 'mirror' , *firu* 'cast away' : *fa-firu-n* 'a sling', *kore* 'unlock, release' : *ka-kore-n* 'corkscrew', *sui* 'to comb' : *sa-sui-t* 'a comb', and Asilulu of the central Moluccas shows a similar pattern with *-t*, *-l*, and occasionally *-n*: *haʔu* 'strike' : *ha-haʔu* 'beating tool' : *ha- haʔu-t* 'a blow', *kahi* 'to hook' : *ka-kahi-t* 'fruit-hooking pole', *soʔo* 'to cover' : *sa-soʔo-t* 'a cover', *nunu* 'prop up with logs' : *na-nunu-l* 'the upright logs for drydocking boats', *saʔu* 'chop sago pith' : *sa-saʔu-l* 'tool for scraping sago', *tati* 'to lower down' : *ta-tati-l* 'basket in which cloves, etc. are lowered by rope', *hiti* 'lift something held closely to the body' : *ha-hiti-n* 'gleeful leaps', *poro* 'yellow' : *pa-poro-n* 'roe, fish eggs'. As seen in these examples, the single consonant suffixes of Central-Malayo-Polynesian languages contrast, and their semantic contributions are often difficult to determine. For this reason these types of morphemes have been a source of puzzlement at least since Jonker (1906). Some recent writers have given glosses to them in

individual languages, as Grimes (1997), who glosses Buru *-t* 'nominative', but this does not explain why these affixes appear in only some citation forms, nor why different affixes (as Tetun *-k*, *-n*, *-t*, or Asilulu *-t*, *-l*, *-n*) appear in the citation forms of different bases.

6.3.4 Paradigmatic alternations

A number of prefixes and suffixes are in paradigmatic alternation with others or with zero. It was noted above that *ka-* alternates with *ma-* 'stative' in some Formosan and Philippine languages, as seen in Amis *ma-fanaʔ kako* (know I) 'I know' vs. *caay ka-fanaʔ kako* (**neg** know I) 'I don't know', or Bontok *l-om-oto* (cook-AF) 'to cook' vs. *daan ka-loto* (not-yet **stat**-cook) 'not yet cooked'. Although they arrive at their conclusions through rather different lines of argumentation, Zeitoun and Huang (2000) and Blust (2003d) agree that PAN *ka-* 'stative' must have occurred in negative constructions, future constructions and imperatives (hence in 'irrealis' clauses). A second pattern of alternation is seen in Bikol of southern Luzon, where Mintz (1971:141) notes alternate command forms for each of three different verb forms. The first of these sets of commands is marked by *-on*, *i-*, and *-an* and occurs when a pronoun is explicitly stated. The second is marked by *-a*, *-an*, *-i*, and occurs when an implied 2sg pronoun is absent:

19a. sabíh-on mo 'Say it.'
     say-**imp 2s**

20a. i-abót mo an asín 'Pass the salt'
     imp-pass **2sg art** salt

21a. namít-an mo an máŋga 'Taste the mango'
     taste-**imp 2sg art** mango

19b. sabíh-a 'Say it'
     say-**imp**

20b. abut-án an asín 'Pass the salt'
     pass-**imp art** salt

21b. namít-i an máŋga 'Taste the mango'
     taste-**imp art** mango

This pattern is important, as it appears to be quite old in AN. Wolff (1973:73) posited a PAN verb system in which the commonly recognised voice affixes *-um-* 'active', *-en* 'direct passive', *-an* 'local passive' and *i-* (now written *Si-*) 'instrumental passive' alternate with zero, *-a*, *-i* and *-an* respectively (the latter qualified with a question mark). He called the former 'independent' and the latter 'dependent' modes for the four voices of the verb. Bikol and a few other languages preserve the dependent forms of the verb as a system. In many other languages this system has disintegrated, and individual components have expanded their original functions. In some languages of Borneo, for example, where the PAN four-voice system has been reduced to an active/passive voice contrast marked by reflexes of *-um-* and *-in-*, imperative constructions are zero-marked (hence reflecting the dependent mode of *-um-* verbs in Wolff's reconstruction), while in Thao of central Taiwan, which has reduced the PAN four-voice system to three contrasts

marked by reflexes of *-um-, *-en and *-an, all imperatives are marked by -i (equivalent to the 22b pattern in Bikol). Finally, one of the best-known examples of paradigmatic alternation of verbal affixes is the PAN patient voice suffix *-en, which alternates with zero in perfective constructions, as discussed elsewhere in this chapter.

6.4 Circumfixes

Circumfixes (also 'confixes') are prefix-suffix units that attach to a base to form a new word. Since many AN languages, especially Philippine-type languages, permit multiple affixes to occur on a single base in word-formation, the question arises whether these affixes are added to the base in a fixed order, or whether some are added simultaneously. It is not always easy to find evidence that can answer this question unambiguously, but some circumfixes appear to be well-established. One of these is *ka-X-an 'adversative passive' in Taiwan, the Philippines and western Indonesia, as in Pazeh *akux* 'heat' : *ka-akux-an* 'get heatstroke', *lamik* 'cold' : *ka-lamik-an* 'catch a cold', *udan* 'rain' : *ka-udan-an* 'be caught in the rain', Mapun *matay* 'die' : *ka-matay-an* 'for a person or household to experience having a close friend or relative die', *paddi* 'pain, soreness, ache' : *ka-paddih-an* 'experience or be overcome with pain', *saŋom* 'night' : *ka-saŋom-an* 'be overtaken by night (i.e. for nightfall to come while one is doing s.t.)', Javanese *ilaŋ* 'gone, lost' : *ka-ilaŋ-an* 'lose possession of something inadvertently', *turu* 'sleep' : *kə-turu-n* 'drop off to sleep accidentally, doze', *udan* 'rain' : *k-odan-an* 'get rained on, be caught in the rain', Tae' *mate* 'dead' : *ka-mate-an* 'affected by someone's death', *uai* 'water' : *ka-uai-an* 'inundated', *uran* 'rain' : *ka-uran-an* 'caught in the rain'. The use of *ka-X-an in these affixed forms (some of which are cognate, and so support PAN or PMP reconstructions) does not involve any obvious ordering, since bases of related meaning which take either *ka- or *-an alone generally are unknown. It is worth noting that this pattern contrasts with a superficially similar pattern reflecting *ka-X-an 'formative for nouns of location, or abstract qualities', as in Saisiyat *t-om-alək* 'cook' : *ka-talək-an* 'kitchen', Thao *kalhus* 'sleep' : *ka-kalhus-an* 'sleeping place', Ifugaw *kayu* 'tree, wood' : *ka-kayu-an* 'place where there is firewood', Bikol *ma-asgad* 'salty' : *ka-asgad-an* 'saltiness', Mapun *batu* 'stone' : *ka-batu-an* 'a rocky area', Toba Batak *mate* 'die; dead' : *ka-mate-an* 'death; place where one has died', or Javanese *lurah* 'top village official' : *ka-lurah-an* 'residence of the top village official'. In many languages reflexes of *ka- alone function as formatives of abstract nouns, and reflexes of *-an form locative nouns. The use of *ka-X-an to form abstract nouns, many of which carry a locative sense, thus appears to result from cyclical affixation. Although this probably was true historically, in many of the modern languages the reflex of *ka-X-an is just as much a circumfix in forming abstract nouns as it is in forming adversative passives (Blust 2003c).

A second widely distributed circumfix reflects PMP *paR-X-an. Schachter and Otanes (1972:291) describe Tagalog *pag-* … *-an* as an object focus counterpart of actor focus verbs with *mag-*: *mag-áral* 'to study' : *pag-arál-an* 'study', *mag-tiʔís*, *pag-tiʔis-án* 'suffer, endure'. Ramos (1971:63) maintains that Tagalog *pag-* … *-an* 'focuses more on the place or the object where the action takes place rather than on the person,' but many examples cannot easily be reconciled with this definition, and it must be concluded that the semantic contribution of *pag-* ... *-an* to the meaning of Tagalog bases is often vague. Malay/Indonesian has a cognate circumfix *pər* … *an*, which Macdonald and Soenjono (1967:100) describe as a nominaliser that replaces the verb prefix *bər-*, 'and combines with the suffix *-an* to form nouns referring to the process of action referred to by the verb.

Occasionally the noun refers to the place where the action is performed.' There are two allomorphs, *pəl … an* if there is a rhotic in the base, and *pər … an* elsewhere. Some forms show an exact correspondence to Tagalog usage, as with *bəl-ajar* 'to study' : *pəl-ajar-an* 'studying, lesson'. Others are quite different, as with *satu* 'one' : *pər-satu-an* 'state of being a unit, unity, union'. As with Tagalog *pag-* … *-an*, it is often difficult to characterise the contribution of Malay/Indonesian *pər-* … *-an* to the meaning of the base. Where there is full cognation of the affixed form between the two languages there is reason to suspect contact influence, since Tagalog *áral* is generally believed to be a borrowing of Malay *ajar*, even though the affixation process in *pag-arál-an* is native.

6.5   Ablaut

True ablaut is rare in AN languages. Both Egerod (1965:258) and Li (1980a:371) claim that Atayal of northern Taiwan shows ablaut in *m-blaq* : *liq-an* 'good, do it well', *h<m>op* : *hab-an* 'stab', *m-ziup* : *iop-an* 'enter', *m-qes* : *qas-un* 'happy'. However, since all reported examples of this alternation require coexistent bound morphemes, the differences in base shape appear to be conditioned by stress or affixation. Unlike the alternation in English *sing* : *sang* : *sung*, which has morphological value in itself, then, these alternations are best treated as phonological. By contrast, a number of languages in northern Sarawak use vocalic alternations to signal differences of voice, and the typological interest of these cases is increased by internal and historical evidence that this ablaut pattern derives from earlier infixation with *-um- and *-in-.

A very simple ablaut pattern is seen in the Lun Bawang-Kelabit dialects spoken near the Sarawak/Kalimantan border. In Lun Bawang as spoken at Long Semado, which has a nearly defunct Philippine-type verb system with three voices, most disyllabic bases form the actor voice (AV) with *ŋ-*. The corresponding patient voice (PV), however, varies with the phonemic shape of the base. If the first base vowel is *a*, *i*, or *u* the PV is *in-* in vowel-initial bases, and *-in-* in consonant-initial bases, as with *anit* 'bark of a tree' : *ŋ-anit* 'remove bark from a tree' : *in-anit* 'was removed, of bark', or *kubil* 'skin' : *ŋubil* 'to skin, as a pig' : *k<in>ubil* 'was skinned'. However, if the first base vowel is schwa the PF is marked by *i*-ablaut, and is obligatorily perfective: *bəli-ən* 'buy it (imperative)' : *məlih* 'to buy' : *bilih* 'was bought by s.o.', *dərut* 'way or manner of sewing' : *nərut* 'sew' : *dirut* 'was sewed by s.o.', *ədʰuk* 'a command, a request' : *ŋ-ədʰuk* 'ask s.o. to do s.t.' : *idʰuk* 'was asked to do s.t.', *təbʰəŋ* 'way or manner of felling a tree' : *nəbʰəŋ* 'to fell a tree' : *tibʰəŋ* 'was felled by s.o.'. Although *ə* : *i* ablaut is common, in a small number of forms there is a three-way pattern with *ə* : *u* : *i*, as in *ədʰaŋ* 'wall hook' : *ŋ-ədʰaŋ* 'hang something up on a wall hook' : *udʰaŋ* 'be hanging up (without reference to agency)' : *idʰaŋ* 'was hung up by someone', or *gəta* 'ford of a river' : *guta* 'ford a river' : *gita* 'was forded, of a river'. In seeking clues to the origin of this pattern what is immediately striking is the pormanteau function of *i*-ablaut, which is essentially identical to that of the infix *-in- . Moreover, some verbs have historically double layers of infixation, as with *bəbʰat* 'a share of something' : *məbʰat* 'to share' : *bibʰat* and *b<in>ibʰat* 'was shared by s.o.'. Informant reaction suggests that *bibʰat* and *b<in>ibʰat*, and parallel morphological doublets in other verb stems are essentially synonymous, further suggesting that the longer form has been affixed twice with the same infix.

The fullest expression of ablaut can be illustrated by Mukah Melanau, where the system is exceptionally complex (Blust 1997b). In this language the Philippine-type voice system has been lost, and verb bases are marked either for active voice, with a reflex of *maŋ- or

*-um-, or for passive voice, with a reflex of *-in-. This results in six surface patterns as shown in Table 6.9:

Table 6.9 Surface patterns of voice marking in Mukah Melanau

	Base	Active	Passive
(1)	mə- (active) vs. nə- (passive)		
	balas 'revenge'	mə-balas	nə-balas
	biləm 'black'	mə-biləm	nə-biləm
	dipih 'hidden, stashed away'	mə-dipih	nə-dipih
	duga 'measure'	mə-duga	nə-duga
	gaduŋ 'green'	mə-gaduŋ	nə-gaduŋ
	gutiŋ 'scissors'	mə-gutiŋ	nə-gutiŋ
(2)	m- (active) vs. n- (passive)		
	aŋit 'anger'	m-aŋit	n-aŋit
	ituŋ 'count'	m-ituŋ	n-ituŋ
	ulin 'rudder'	m-ulin	n-ulin
(3)	u-ablaut (active) vs. i-ablaut (passive)		
	gəga 'chase away'	guga	giga
	gəgət 'gnaw; moth'	gugət	gigət
	kəkay 'rake'	kukay	kikay
	kəkut 'excavated'	kukut	kikut
	ləpək 'a fold'	lupək	lipək
	ləpəw 'pick'	lupəw	lipəw
	ñəña? 'chew'	ñuña?	ñiña?
	ŋəŋət 'gnaw'	ŋuŋət	ŋiŋət
	səbət 'make'	subət	sibət
	səkəl 'strangle'	sukəl	sikəl
	sələg 'burn'	suləg	siləg
	səpəd 'hack, chop'	supəd	sipəd
	səput 'blowpipe'	suput	siput
	səsaŋ 'pay'	susaŋ	sisaŋ
	səsəp 'sip, suck'	susəp	sisəp
	təbək 'stab'	tubək	tibək
	təbəŋ 'fell a tree'	tubəŋ	tibəŋ
	tətək 'cut'	tutək	titək
	tətəŋ 'drink'	tutəŋ	titəŋ
(4)	məŋ- (active) vs. n- (passive)		
	adək 'sniff, smell'	məŋ-adək	n-adək
	añit 'sharp'	məŋ-añit	n-añit
	apu? 'white'	məŋ-apu?	n-apu?
	ukur 'measure'	məŋ-ukur	n-ukur

	Base	**Active**	**Passive**
(5)	mə- + nasal substitution (active) vs. n- or nə- (passive)		
	kiap 'hand fan'	mə-ŋiap	nə-kiap
	paləy 'a taboo'	mə-maləy	nə-paləy
	sapəw 'broom'	mə-ñapəw	nə-sapəw
(6)	-əm- plus u-ablaut (active) vs. i-ablaut (passive)		
	bəbah 'split (stative)'	mubah	bibah
	bəbəd 'tie around'	mubəd	bibəd
	bənuʔ 'kill'	munuʔ	binuʔ
	pəpah 'hit, whip'	mupah	pipah
	pəpək 'a whip'	mupək	pipək

All passive forms are passive/perfective. The semantic values of the unglossed columns are thus to be read as: *mə-balas* 'take revenge on s.o.' : *nə-balas* 'was the target of s.o's revenge', *mə-biləm* 'blacken s.t.' : *nə-biləm* 'was blackened by s.o.', *məŋ-ukur* 'to measure' : *n-ukur* 'was measured by s.o.', etc. Pattern (1) applies to bases that begin with a voiced obstruent and have a first vowel other than schwa. Pattern (2) applies to bases that begin with a vowel, and pattern (3) to bases in which the first vowel is schwa (since word-initial schwa does not occur in Mukah all of these begin with a consonant). Pattern (4) is like pattern (2) but contains *məŋ-* rather than *-əm-*. Pattern (5) is like pattern (1), but applies to bases that begin with a voiceless obstruent. Pattern (6) applies only to bases that begin with a labial stop and contain schwa as the first vowel. It is the most problematic, and will be discussed at some length below. Underlying forms of active and passive affixes are shown in Figure 6.3:

Pattern	**Active**	**Passive**
(1)	məŋ-	-ən-
(2)	-əm-	-ən-
(3)	-u-	-i-
(4)	məŋ-	-ən-
(5)	məŋ-	-ən-
(6)	ʔ + -u-	-i-

Figure 6.3: Underlying forms of active and passive affixes in Mukah Melanau

Since Mukah Melanau lacks consonant clusters earlier *mam-bilem, *man-deket, etc. would have lost the preconsonantal nasal. Indications that pattern (1) contains underlying *məŋ-* rather than *-um-* are seen in the parallelism of *mə-biləm* 'blacken' with *məŋ-apuʔ* 'whiten', and in the apparent exceptionality of *dəkət* : *mə-dəkət* : *nə-dəkət* as compared with other bases in which the first vowel is schwa (the ablaut paradigm shown as pattern 3). Once *məŋ-* and *-əm-* conjugations are separated the complementation of patterns (2) and (3) becomes transparent. In addition, both *-ən-* and *-i-* show the same semantic peculiarity: they mark a passive voice that is *obligatorily* perfective (attempts to elicit non-perfective passives result in hesitation, circumlocutions, etc.). It follows, then, that *-əm-* and *-u-* < *-um- and that *-ən-* and *-i-* < *-in-. This development is striking for two reasons. First, it

has come about through a separation of allomorphs that is nearly complete: one preserved the consonant but neutralised the vowel with schwa, while the other preserved the vowel but lost the consonant. Second, it resulted in the emergence of a system of ablaut only in certain phonetic environments. Since ablaut arose through conditioned change it can be derived from underlying infixes *-um-* and *-in-* without reference to comparative data.

Mukah ablaut derives from *-um- and *-in- through three ordered changes: 1) *e > Ø/VC__CV, 2) C > Ø/__C, and 3) V > ə/__CV(C)V(C). The first of these, which can be called 'schwa syncope' is found in many AN languages. The second is evident from Mukah morpheme structure, which disallows consonant clusters, and the third is seen in such etymologies as *balabaw > *bəlabaw* 'rat, mouse', *bituka > *bətuka* 'intestines', *sali-matek > *sələmatək* 'large forest leech', and *taliŋa (> təliŋa > tliŋa) > *liŋa* 'ear', as well as in the pattern of CV- reduplication seen in *bəbulan* 'cataract of the eye', *dədian* 'candle', or *ləlaŋaw* 'housefly'.

*bəbah 'split'	*b<um>əbah	*b<in>əbah	Change
	mu-bəbah		IM
	mubbah	binbah	SS
	mubah	bibah	CR
*pəpək 'a whip'	p<um>əpək	*p<in>əpək	
	mu-pəpək		IM
	muppək	pinpək	SS
	mupək	pipək	CR
*biləm 'black'	b<um>iləm	*b<in>iləm	
	mu-biləm		IM
	mə-biləm	b<ən>iləm	PN
		nə-biləm	IM

Figure 6.4: Sample historical derivations of compound ablaut in Mukah Melanau

Pattern (6) is typologically unusual and analytically challenging. For lack of a better term it can be called 'compound ablaut', as it combines ablaut alternations with purely phonological alternations in a single pattern. Since patterns (3) and (6) are in complementary distribution it is likely that they had the same origin, yet pattern (6) shows apparent nasal substitution on top of ablaut. There are two reasons to rule out true nasal substitution in these forms. First, nasal substitution in Mukah active verbs is found with *məŋ-*, and if this affix had been added to e.g. *bəbah* the expected outcome would be **mə-məbah, a double discrepancy in having *mə-*, and providing no account of ablaut. On the other hand, if these bases had been infixed with *-um- and *-in-, the general ablaut process would have produced *bubah* : *bibah*, *bubəd* : *bibəd*, etc. To derive the attested forms we need to recall that pseudo nasal substitution is motivated by strong dissociative tendencies between dissimilar labial consonants in successive syllables. Figure 6.4 provides sample derivations to show how compound ablaut probably arose in Mukah Melanau (IM = infix metathesis, SS = schwa syncope CR = cluster reduction).

In many AN languages the disfavored sequences *pVm and *bVm that arise through infixation with *-um- are avoided by CV- truncation, resulting in pseudo nasal subsitution.

Compound ablaut in Mukah Melanau apparently arose out of a similar motivation, but through use of a different strategy of avoidance, namely metathesis of the first two consonants of a *C<um>VCVC word, where the initial consonant was labial. Given the frequent occurrence of such prefixes as *ma- 'stative' with bases of any shape, the sequence *ma-b- or *ma-p- would have been rather common, and to eliminate it by CV-truncation would have deleted the affix without a trace. For this reason sequences of *mVb- and *mVp- were tolerated, while sequences of *b<um>V- or *p<um>V- were at much greater risk of elimination (Blust 2004a). Once infix metathesis had taken place the three ordered changes that gave rise to other cases of ablaut (schwa syncope, cluster reduction, merger of prepenultimate vowels as schwa) completed the process. Although this explanation assumes a historical recurrence of infix metathesis, the morphological asymmetry in *mə-biləm* : *b<ən>iləm*, etc. could have set the stage for the more general transformation of infixes to prefixes in the recent history of this language.

More recently Lobel (2013:183-188) has shown that similar systems of verbal ablaut have developed independently in Central Subanen, Southern Subanen, Maranao and Iranun of the southern Philippines. In both areas the infixes *-um- and *-in- were transformed to an ablaut pattern only in bases that contained a penultimate schwa. However, the steps in the process were fundamentally different: whereas in North Sarawak languages a schwa first deleted in the environment VC__CV, giving rise to medial clusters that were then reduced, schwa syncope never occurred in Subanen or the Danaw languages Maranao and Iranun. Instead, the intervocalic nasal of the infixes *-um- and *-in- was dropped and the resulting cluster of high vowel + schwa assimilated to produce a stressed or long high vowel, as in *sələd > base *saləd* 'enter', *s<um><in>ələd (> *s<um><i>ələd > *s<um><i>iləd) > *s<um>íləd* (AF past), *s<um>ələd (> *s<u>ələd > *s<u>uləd) > *sūləd* (AF non-past), or *s<in>ələd (> *s<i>ələd > *s<i>iləd) > *sīləd* (PF past), next to *sələd-ən > *sələr-ən* (PF non-past). This occurred not only in open syllables, as in the North Sarawak languages, but also in closed penults of the form CəNCV(C).

6.6   Suprasegmental morphology

Some AN languages use suprasegmental features to mark morphological contrasts. Strategies of this kind can be divided into two types: 1) the use of morphological stress, and 2) the use of morphological tone.

A number of Philippine languages not only have lexical stress, but also use stress in word derivation. In Tagalog, for example, *pátid* 'tripping another's foot' : *patíd* 'cut off' and *túlog* 'sleep' : *tulóg* 'asleep' are both minimal pairs, but the first is lexical, while the second is morphological. The available dictionaries do not always distinguish these consistently, as where Panganiban (1966) gives *túlay*, *pag-túlay* 'effort to balance oneself on narrow or tiny foothold' : *tuláy* 'bridge; (fig.) go-between' as separate lexical entries, despite a clear semantic relationship between these segmentally identical but prosodically distinguished forms. Some languages outside the Philippine group that lack lexical stress contrasts also use stress shift for lexical derivation, as in Chamorro *aságwa* 'spouse' : *ásagwa* 'to wed, marry'. The most productive use of morphological stress probably is found in Toba Batak, where lexical stress is penultimate, but shifts rightward to derive adjectives, stative verbs, or descriptive nouns: *gogo* 'push hard (imp.)' : *gogó* 'strong', *arga* 'bargain (imp.)' : *argá* 'expensive', *hojot* 'go quickly (imp.) : *hojót* 'quick', *dila* 'tongue' : *dilá* 'a big talker, person who brags or exaggerates'.

Since tone is rare in AN languages the use of morphological tone is also unusual. Van der Leeden (1997) has described Ma'ya, a South Halmahera-West New Guinea language of extreme northeast Indonesia, as having four contrasting tones, plus a tonal replacement pattern which in his analysis has morphological value. This analysis, however, is questioned by Remijsen (2001:51ff), who proposes a system of three tonemes plus contrastive stress, and no morphological value for tone.

6.7   Zero morphology: bases as imperatives

The precise form of imperative marking in PAN is unknown. Although *-i 'imperative of locative voice verbs' is fairly well-attested, the imperative forms of other voices are less clearly reflected in cognate morphology. In many of the modern languages the bare stem forms the imperative, as in Mukah Melanau *siən mə-tud kayəw* (**3sg   av**-bend stick) 'he is bending a stick' vs. *tud kayəw iən* (bend stick that) 'bend that stick!', *kain iən n-upuk siən* (clothes those **pv**.**perf**-wash **3sg**) 'she washed those clothes' vs. *upuk kain itəw* (wash clothes these) 'wash these clothes!', or Malay *dia məm-baca buku* (**3sg   av**-read book) 'he is reading a book' vs. *baca buku itu* (read book that) 'read that book!'. Since the absence of an affix on the verb base and of an expressed agent in such constructions signals the imperative mood, it can be argued that the imperative is marked by zero morphology.

6.8   Subtractive morphology

Stevens (1994) has drawn attention to what he calls 'truncation phenomena' in Bahasa Indonesia, by which he means the shortening of a base as a morphological process. He illustrates this in 1) hypocoristics, or nicknames, 2) acronyms, and 3) Prokem, a secret language of Jakarta youth (see Chapter 3). Examples include *Sukarno > Karno* (or *Bung Karno*), *bapak > pak* 'father' (often used as a term of address for older or respected males), *administrasi > min* 'administration', *pəmbinaan > bin* 'development, fostering' (common in governmental discourse), and such Prokem forms as *bokap* (b-ok-ap) = *bapak* 'father', or *tokau* (t-ok-au) = *tahu* 'know'. The most common use of subtractive morphology in AN languages, however, is seen in kin terms, where the dropping of an initial consonant marks the vocative form (Blust 1979):

Table 6.10 Derivation by subtraction in the vocative forms of kinship terms

Language	Reference/address	Vocative	Gloss
Bikol	apoʔ	poʔ	grandparent, elder
Cebuano	nánay	nay	mother
Cebuano	tátay	tay	father, uncle
Malay	adek	dek	younger sibling
Malay	kakak	kak	elder sibling
Banjarese	adiŋ	diŋ	younger sibling
PMP	laki	aki	grandfather
PMP	kaka	aka	elder same sex sibling

In some of the languages of Indonesia the vocative of personal names is formed in a similar way, as in Timugon Murut of Sabah, where subtraction of an initial consonant or

the entire first syllable can be used quite productively, as with *Ohn*, vocative of 'John' (D.J. Prentice, p.c.).

6.9   Reduplication

Reduplication is globally distributed, but its role is far more prominent in some language families than in others. Most Indo-European languages, for example, make relatively little use of this device, whereas it is heavily exploited in many AN languages. No general survey of reduplication in AN languages exists. Among early writers Blake (1917) provided one of the surprisingly rare accounts of reduplication in Philippine languages, and Gonda (1950) examined the functions of reduplication in the languages of Indonesia. More recently some theory-driven work has been done on reduplication in individual AN languages, and two doctoral dissertations have appeared that treat reduplication in a number of AN languages within a coherent descriptive and analytic system (Spaelti 1997, Kennedy 2003). Comparative research has only begun to reveal the number of reduplicative patterns found in this language family and the range of functions they serve. This section surveys some of the major types and functions of reduplication, organising the material by type of copying pattern. It does not pretend to be descriptively exhaustive, nor to provide complete analyses of the data.

Because reduplicative morphology in AN languages often co-occurs with non-reduplicative affixes, the number of reduplicative patterns that can be distinguished is very high, and it would be impractical to try to enumerate them exhaustively. In Thao of central Taiwan, for example, full reduplication (with coda dropping) may occur alone, as in *fariw* 'buy' : *fari-fariw* 'go shopping' or *kaush* 'scoop' : *kau-kaush* 'scoop repeatedly, as water'. However, the same copying pattern co-occurs with many non-reduplicative affixes, as in *acan* 'type, variety' : *mia-aca-acan* 'be replete, have everything', *m-acay* 'die' : *an-m-aca-acay* 'be on the verge of death', *apuy* 'fire' : *pin-apu-apuy-an* 'be exposed to fire repeatedly', or *ian* 'refuge, shelter' : *ia-ian-an* 'lived in, as a house'. In counting reduplicative patterns the additional factor of non-reduplicative affixation will therefore be ignored, although in illustrating these patterns examples of reduplication will be given both with and without other types of affixes.

Reduplication usually is regarded as a form of affixation, and hence is treated here as part of morphology. However, since they arise from copying the base, reduplication patterns occupy an ambivalent position between morphology and phonology, and the analysis of these patterns has led to some major claims in phonological theory in recent years. Because AN languages present such a wealth of examples of this phenomenon they are uniquely situated to test these claims. For this reason some of the principal theoretical implications of reduplication in AN languages will be considered briefly before proceeding to a survey of reduplication patterns. The major topics covered are: 1) reduplicative pattern and reduplicative structure, 2) base-1 and base-2, 3) restrictions on the shape of the reduplicant, 4) restrictions on the content of the reduplicant, 5) patterns of reduplication (full reduplication, prefixal foot reduplication, heavy syllable reduplication, CV-reduplication, fixed segmentism, infixal reduplication, suffixal foot reduplication, suffixal syllable reduplication, other patterns).

6.9.1 Reduplicative pattern and reduplicative structure

Reduplicative patterns are surface phenomena that derive from more abstract and uniform structures which underlie them. Structures thus give rise to patterns under given conditions of canonical shape or prosodic contrast. In listing types of reduplication patterns will be cited, since these are the visible phenomena that appear in most descriptive accounts. However, divergent surface patterns may be variants of a single reduplicative structure. Rehg (1981:73-85), for example, documents eleven surface patterns of reduplication to mark durative aspect in Pohnpeian of central Micronesia, noting that 'it may be possible to combine some of these patterns,' although he does not explicitly do so. Spaelti (1997) carries the logic of this observation further by showing that reduplication patterns which perform similar functions and occur in prosodically or canonically complementary environments may be considered divergent surface realisations of a single underlying structure. He calls surface forms that share an underlying functional unity 'alloduples' of the same reduplicant, in parallel with traditional notions of non-reduplicative allomorphy. In Thao of central Taiwan, for example, where stress is penultimate in about 98% of all forms (and final in the rest) three patterns of reduplication contribute a semantic overlay of repetitive action to the meaning of the base. These are 1) full reduplication, 2) suffixal foot reduplication, and 3) reduplication of the rightmost CCV(C) of a base (Chang 1998). Where two identical morphemes or morpheme fragments are juxtaposed the coda of the first iteration is invariably deleted. Morpheme boundaries are marked by hyphen (angled brackets for infixes), syllable boundaries by period, and reduplicants are bolded:

Table 6.11 Alloduples of a reduplicative template in Thao

A.	'Full' reduplication

ka.ri 'dig up or out'   k<m>a.ri.-**ka.ri** 'dig up repeatedly/habitually'
lha.ri 'flash, of lightning'   kun.-lha.ri.-**lha.ri** 'flash repeatedly, of lightning'
mi.-qi.lha 'drink'   mi.-qi.lha.-**qi.lha** 'drink frequently or repeatedly'
mi.-ta.lha 'wait'   mi.-ta.lha.-**ta.lha** 'wait and wait'
m.-za.i 'say'   m<in.>za.i.-**za.i** 'said repeatedly'
cpiq 'whip, beat'   cpi.-**cpiq** 'whip repeatedly'
fi.lhaq 'saliva, sputum'   ma.-fi.lha.-**fi.lhaq** 'will spit repeatedly'
ki.rac 'light, luminosity'   pish.-ki.ra.-**ki.rac** 'send out sparks'
lhun 'wet nasal mucus'   mak.-lhu.-**lhun** 'blow the nose repeatedly'
qbit 'share, portion'   mi.-qbi.-**qbit** 'portion out, divide into shares'

B.	Suffixal foot reduplication

i.-su.huy 'there'   pi-su.hu.-**huy** 'be put there repeatedly'
ki.ka.lhi 'ask'   ma.-ki.ka.lhi.-**ka.lhi** 'ask around'
pa.ti.haul 'spell, curse'   ma.ti.hau.-**haul** 'cast a spell on someone'
shna.ra 'ignite, catch fire'   pa.-shna.ra.-**na.ra** 'burn something repeatedly'
qri.u? 'steal'   q<un.>ri.u-**ri.u?** 'steal habitually or repeatedly'

C.	Suffixal CCV(C) reduplication

aŋ.qtu 'contemplate'   m-aŋ.qtu.-**qtu** 'think about, mull over'
ma.-par.fu 'wrestle'   ma.-par.fu-**r.fu** 'wrestle repeatedly'
m-ar.faz 'to fly'   m-ar.fa-**r.faz** 'keep flying around'
pa.tqal/pat.qal 'a mark'   pa.tqa.-**tqa.l**-an/pat.qa-t.qa.l-an 'put marks on'
tap.?an 'patch on clothing' t<i.n>ap.?a-**p.? a-p.? an** 'was patched repeatedly'

Since patterns A-C have essentially the same function and occur under predictable but differing canonical conditions, they exemplify what Spaelti has called 'alloduples' of the same reduplicant, a structure that can most plausibly be considered suffixal foot reduplication. By appealing to complementation as an explanation for divergent surface patterns of a single reduplicative structure Spaelti has stretched the meaning of this term beyond its traditional use, where it refers to physical variation in the speech signal that is predictable from phonetic context. In reduplication, by contrast, surface variation is often predictable from the canonical shape of the base or from prosodic factors rather than from the immediate segmental environment. This difference suggests that complementation is best seen as a cover term for two distinct types of phonological relationship: 1) *contextual complementation* (the traditional notion of variation determined by phonetic context, as found in allophony or phonologically conditioned allomorphy), and 2) *canonical complementation* (the patterning which led Spaelti to coin the term 'alloduple'). As will be seen below, distinctions between reduplicative structures and reduplicative patterns are almost invariably products of canonical complementation. However, in rare cases alloduples of the same reduplicant are correlated not with the canonical shape of the base, but with immediate phonetic context, and are therefore products of the more familiar notion of contextual complementation.

6.9.2 Base-1 and Base-2

A second issue that arises in dealing with patterns of reduplication in AN languages is how to distinguish the base from the reduplicant. This problem is particularly acute with full reduplication, as with Malay *oraŋ* 'person' : *oraŋ-oraŋ* 'people', or *kəlapa* 'coconut' : *kəlapa-kəlapa* 'coconuts'.[63]   If full reduplication copies the entire base regardless of how many syllables it contains it may be genuinely impossible to distinguish base from reduplicant.

One possibility for distinguishing base from reduplicant in languages like Malay is to compare patterns of reduplication with patterns of non-reduplicative affixation. Malay has a pattern of full reduplication in which the second iteration takes *məŋ-*, with homorganic nasal substitution of the base-initial consonant, as in *tawar-mənawar* 'to bargain, haggle'. Since affixes normally attach to bases rather than to other affixes, it might be concluded that *tawar* is the reduplicant and *mənawar* the affixed base. What muddles this analysis is a competing pattern in which the *first* half of a reduplicated word takes a non-reduplicative affix. With some bases both patterns occur, sometimes with a difference of meaning and sometimes with none, as with 1) *ganti* 'substitute' : *ganti-bərganti*, *bərganti-ganti* 'alternate with each other', or 2) *masak* 'cook' : *məmasak-masak* 'do the cooking' : *masak-məmasak* 'cookery'. If reduplication is regarded as a type of affixation, and only bases may be affixed, then patterns such as that seen in *tawar-mənawar*, or *masak-məmasak* must be interpreted as R+B, while patterns such as that seen in *məmasak-masak* or *bərganti-ganti* must be interpreted as B+R. However, clear semantic correlates of these patterns remain to be established, and in reduplicated words with non-reduplicative affixes one part of the reduplication must be an affix which itself can be affixed, as in *kənal* 'know a person' : *bər-kənal-kənal-an* 'be acquainted with each other', or *takut* 'afraid' : *mənakut-nakut-i* 'to intimidate'. Examples like these may favor the view that reduplicative affixes that fully copy the base are simultaneously affix-like and base-like in their contribution to word-formation.

Where 'full' reduplication turns out to be a variant of suffixal foot reduplication, as illustrated above for Thao, the issue of base-reduplicant identity can be resolved: pairs such as *patihaul* 'spell, curse': *matihau-haul* 'cast a spell on someone', *shnara* 'ignite, catch fire' : *pa-shnara-nara* 'burn something repeatedly', or *qriuʔ* 'steal' : *q-un-riu-riuʔ* 'steal habitually or repeatedly' show an unambiguous pattern of base + reduplicant in the longer word. Given this pattern there is no reason to posit a different order in bases such as *fariw* 'buy' : *fari-fariw* 'go shopping', *kaush* 'scoop' : *kau-kaush* 'scoop repeatedly, as water', or *kirac* 'light, luminosity' : *pish-kira-kirac* 'shoot out sparks', where the functional unity of these divergent surface patterns is clear.[64]

In cross-linguistic perspective this conclusion is surprising for two reasons. First, McCarthy and Prince (1994) have claimed that reduplicants are characterised by 'the emergence of the unmarked,' where the maximally unmarked syllable is CV, and any addition or deletion of consonants is an increase in markedness. By this they mean that the

---

[63]   Fully reduplicated trisyllables are rare, but it is clear that trisyllabic nominal bases are never pluralized by suffixal foot reduplication. According to Uli Kozok (p.c.) fully reduplicated bases of more than two syllables are fully acceptable in Bahasa Indonesia (even *pərpustakaan-pərpustakaan* 'libraries' reportedly is normal). However, such expressions seem awkward, and one wonders whether most native speakers would not normally omit reduplication in bases of more than two syllables when the information it carries is clear from context.

[64]   A hypothesis of infixal reduplication (pish-kira<kira>c) is possible, but unconvincing, since true infixal reduplication appears to be found in a few forms such as *itiza* 'arrive' : *i-ti-tiza* 'arrive, return, come back', and here it copies the penultimate syllable.

reduplicant will tend to be less marked than the base from which it is copied, and in any case will never contain marked features not present in the base. Yet the data in Table 6.11 show coda deletion (hence unmarking) in the *base*, not the reduplicant. Second, there is a widely shared assumption that a reduplicant may not exceed the size of the base that it 'copies'. Both of these anomalies can be reconciled with theory if it is recognised that the term 'base' has been used in the general phonological literature in two distinct senses: 1) the base is independent (Base-1), or 2) the base is the morpheme to which affixes, including reduplicants, attach (Base-2). In reduplication both Base-2 and the reduplicant are 'copied' from Base-1, as in *fari-fariw* 'go shopping', where each half of this word is copied from *fariw* 'buy', with automatic coda reduction in Base-2. Once these two senses of the term 'base' are distinguished the apparent contradiction of a reduplicant being longer or more highly marked than the base (base-2, not base-1) disappears. However, as will be seen, problems with 'the emergence of the unmarked' are more pervasive and recalcitrant than this in AN languages.

6.9.3 Restrictions on the shape of the reduplicant

Another widely shared theoretical position, sometimes called the 'Prosodic Morphology Hypothesis' (PMH), holds that the shape of reduplicants is constrained by considerations of prosody (cf. McCarthy and Prince 1990:209 and subsequent publications, where reduplication is stated in terms of the affixation of a template to a base, and templates are "defined in terms of authentic units of prosody," namely the mora, syllable, foot, or prosodic word). Although most reduplicants are prosodic units it is clear that others are not, and while they have not provided the only counterevidence, AN languages have played and continue to play a major part in challenging this view.

There is an ambiguity in the McCarthy-Prince formulation that requires clarification, namely whether a template is defined in terms of the base or the reduplicant. Figure 6.5 outlines the four logical possibilities:

Type	Base	Reduplicant
1)	+PU	+PU
2)	-PU	+PU
3)	+PU	-PU
4)	-PU	-PU

Figure 6.5 Relations between prosodic units in base and reduplicant

Type 1) is the structurally simplest, typologically most common and theoretically least problematic of the four. In it a prosodic unit in the base is copied as a prosodic unit in the reduplicant. This is the type seen in full reduplication, CV- reduplication and the like, and will be surveyed briefly below.

Type 2), seen in e.g. Ilokano *ba.két* 'old woman' : *ag.-bak.-ba.két* 'grow old, of women', or *pú.sa* 'cat' : *pus.-pú.sa* 'cats' is somewhat more problematic in that the copied portion is what can be called a 'prosodic chimera' (a syllable plus onset, or σ+o). Hayes and Abad (1989) called this 'heavy syllable reduplication', and it has been incorporated into phonological theory as a conforming case, since the reduplicant is a possible syllable in the language.

Types 3) and 4) are the most problematic. For Type 3), Blevins (2003, 2005) has drawn attention to several instances of prefixal foot reduplication in Oceanic languages that allow phonological contraction of the reduplicant, as with Bugotu *ka.lu* : *kau.ka.lu* 'to stir, knead', *li.ko* : *li.o.li.ko* 'to be crooked', or Hoava *ɣa.sa* : *ɣa.sa. ɣa.sa* (slow/careful speech) : *ɣas. ɣa.sa* (rapid speech) 'to jump, jumping'. Since contraction does not occur with non-reduplicative morphology under similar prosodic conditions, she argues that the exceptional behavior of reduplicants in such forms reflects the general principle that redundantly specified information allows contextual deletion (what is predictable need not be fully specified). Blevins shows that there are intractable problems in stating these patterns in templatic terms, since the form of the reduplicant varies depending upon the sonority of the first vowel (bimoraic syllable with lower-higher, but disyllabic foot with higher-lower). In addition, since the reduplicant in forms such as Bugotu *kau.ka.lu* or *li.o.li.ko* contains an onsetless syllable that is not present in the base, there is a clear violation of the 'emergence of the unmarked.'

Type 4) presents perhaps the most flagrant violations of theoretical expectation. There are two known subtypes: 1) the reduplicant is a single consonant, and 2) the reduplicant has the form c+σ (a syllable preceded by a syllable coda).

6.9.3.1 *The reduplicant is a single consonant*

Nivens (1993) has documented a pattern of single-consonant reduplication in several dialects of West Tarangan, spoken in the Aru Islands of eastern Indonesia. In the North dialect bases with initial stress take what he calls 'a CVC prefix' (= CV, VC or CVC), as in *ke* : *keke* 'wood', *ɔn* : *ɔnɔn* 'shoot', *tun* : *tuntun* 'mosquito', *ɔta* : *ɔtɔta* 'fold', or *lɔpay* : *lɔplɔpay* 'cold'. Bases with non-initial stress, however, reduplicate a single consonant if the syllable immediately preceding the stressed syllable is open. Under these highly specific conditions which are sensitive to both stress and syllable shape the posttonic consonant is copied in front of the pretonic consonant, as in *tapúran* : *tarpúran* 'middle', *gasíra-na* 'old-3sg' : *garsíra*, *dubém-na* : *dumbém* 'seven', or *ga* 'relative' + *let* 'male' : *gatlet* 'bachelor'.

A second example of single-consonant reduplication is seen in the Batad dialect of Ifugaw in the northern Philippines. Newell (1993:6) points out that all Ifugaw dialects have phonemically geminated consonants, and that gemination is a grammatical device used to indicate reciprocal action, but that geminate glides -*ww*- and -*yy*- have undergone fortition to g^w and d^y in the Batad dialect. Many verb pairs exist in which the distinction between singular and plural actors is marked by doubling the consonant that immediately follows the first vowel of a base, as in *ābak* 'defeat someone' : *abbak* 'for two to compete with each other', *awit* 'wrestle' : *ag^wit* 'engage in a wrestling contest' (< *awwit*), *bāliw* 'for someone to protect or rescue s.o. or s.t.' : *him-ba-ballíw-an* 'protect or rescue one another', or *patoy* 'for someone to kill something or s.o. with a weapon' : *pattoy* 'for two people or animals to fight one another in physical combat'. Some verb bases instead use CV-reduplication to mark the same semantic distinction, as with *gubat* 'war' : *muŋ-gu-gubat* 'for a nation or group of soldiers to battle against another nation or group of soldiers' (focus is on three or more nations or groups of soldiers fighting against each other)', or *hāpit* 'language, speech, voice' : *mun-ha-hapit* 'talk to each other (focus is on three or more who talk to each other)'. This suggests that gemination marking reciprocal action may be a recent development from an earlier CV- pattern. However, consonant gemination has been lexicalised in many forms that have an inherently reciprocal or collective meaning, implying instead that single consonant gemination/reduplication has a long

history in Ifugaw: *addum* 'for two people or living things to come together', *ahhiw* 'for two to place or carry a heavy load on the shoulders by center suspension', *ahhud* 'for two people to pound cereal grains in pairs with pestles in a mortar', *ammid* 'for two things to stick to each other', *delloh* 'for two people or things to come beside each other', *dihhul* 'for two people or things to simultaneously do something', etc. Moreover, CV- reduplication with syncope of the prefixal vowel would not produce medial consonant gemination, or any type of gemination in vowel-initial stems (cf. *hapit* : *mun-ha-hapit* above, but *happit* (**hhapit) 'a crafty scheme devised between two or among several for the advantage of the schemers and the disadvantage of someone else'). It appears, then, that Ifugaw innovated morphological consonant gemination to add an overlay of reciprocity or collective action to the meaning of the base, and that gemination is thus another form of single-consonant reduplication.

The use of consonant gemination to signal grammatical information is also found in Ilokano, where doubling of the consonant which follows the first base vowel encodes plurality in some [+human] nouns: *ádi* > *addí* 'younger siblings', *amá* > *ammá* 'fathers', *asáwa* > *assáwa* 'spouses', *babái* > *babbái* 'girls', *iná* > *inná* 'mothers', *laláki* > *lalláki* 'boys' (Rubino 2000:xlvi-xlvii). Again, there appears to be no simple way in which geminating reduplication can derive historically from an earlier pattern of CV- copying.

6.9.3.2 *The reduplicant is a prosodic chimera*

Gafos (1998) has used the term 'a-templatic' reduplication for patterns in which the copying domain is not a prosodic unit. The examples that he gives, like those of Nivens (1993), involve the iteration of single consonants, and so can be called 'subtemplatic reduplication'. By contrast, the following types of non-prosodic reduplicants can be called instances of 'supertemplatic reduplication.' Like so-called 'heavy syllable reduplication' as described under Type 2) supertemplatic reduplication is a prosodic chimera in its source (the base), since it copies a syllable followed by a syllable onset. However, with heavy syllable reduplication (σ+o) the reduplicant is a possible syllable in the language. The cases considered here, which copy a coda plus a following syllable (c+σ), are different in that they are prosodic chimeras both in the base and in the reduplicant.

The Thao examples in Table 6.11, part C illustrate a pattern of c+σ reduplication. Unlike most AN languages Thao permits word-initial consonant clusters, as in *qnuan* 'carabao' or *tqir* 'take offense'. However, such consonant clusters may not begin with a sonorant, nor end with a glottal stop. Although the syllabification of *patqal* 'a mark' could in principle be either *pa.tqal* or *pat.qal*, then, the syllable boundary in e.g. *m-ar.faz* 'to fly': *m-ar.fa-r.faz* 'keep flying around, as a fly over food' or *tap.ʔan* 'patch on clothing' : *t<i.n>ap.ʔa-p.ʔa-p.ʔan* 'was patched repeatedly' is unambiguous. Historically all reduplication patterns in Table 6.11 derive from a process of suffixal foot reduplication that is transparent in trisyllables, or in disyllables with an initial consonant cluster. With CVCVC disyllables foot and base are coterminous, so that foot reduplication and full reduplication are indistinguishable. With bases that contain a medial consonant cluster the equivalent of a suffixal foot reduplicant surfaces as c+σ (-CCVC), a synchronic residue of the historical loss of schwa. Typologically surprising reduplications such as *m-ar.fa-r.faz* thus derive from earlier *m-arəfa-rəfaz. A similar pattern of c+σ reduplication is also known from Central Amis and Southern Paiwan in eastern Taiwan, as seen in Table 6.12.

All three of these languages have penultimate stress, with minor qualifications (about 2% of Thao forms are oxytone, and in Amis stress is final in citation forms, but penultimate in phrasal context). Consonant clustering patterns, however, differ among the

languages. While Thao allows a variety of consonant clusters in initial and medial position, Central Amis and Paiwan have only medial clusters. Furthermore, although Central Amis allows a wide range of medial clusters, Southern Paiwan allows clusters only in historically reduplicated monosyllables of the form $C_1V_1C_2C_1V_1C_2$.

Table 6.12 c+σ reduplication in Central Amis and Southern Paiwan

Central Amis (Affixed) base	Suffixal c+σ reduplication
aŋ.rər 'bitter'	aŋ.rə-***ŋ.rər*** 'very bitter'
aŋ.saw 'smell of smoke' aŋ.sa-***ŋ.saw***	'strong smoke odor (as in clothing)'
faq.loh 'new'	faq.lo-***q.loh*** 'everything is new'
in.tər 'hate, despise'	ma.in.tə-***n.tər*** 'everyone hates'
kaq.soq 'tasty'	kaq.so-***q.soq*** 'everything is tasty'
maq.cak 'cooked'	maq.ca-***q.cak*** 'everything is cooked'
kar.təŋ 'heavy'	kar.tə-***r.təŋ*** 'everything is heavy'
maŋ.taq 'raw'	maŋ.ta-***ŋ.taq*** 'everything is raw'
siq.naw 'cold'	sa.-siq.na-***q.naw*** 'everything is cold'
tam.ɬaw 'person'	tam.ɬa-***m.ɬaw*** 'everyone'

Southern Paiwan

Stem	Gloss	Dual	Plural
A. Suffixal foot reduplication			
panaq	shoot	ma.-pa.-pa.naq	ma.-pa.-pa.na.-***pa.naq***
gətsəl	pinch	ma.-ga.-gə.tsəl	ma.-ga.-gə.tsə.-***gə.tsəl***
kakəlʸaŋ	know	mar.-ʔa.-ka.kə.lʸaŋ	mar.-ʔa.-ka.kə.lʸa.-***kə.lʸaŋ***
bulay	good	mar.-ʔa.-bu.lay	mar.-ʔa.-bu.la.-***bu.lay***
ləva	happy	mar.-ʔa.-lə.va	mar.-ʔa.-lə.va-***lə.va***
tjəŋəlay	love	mar.-ʔa.-tjə.ŋə.lay	mar.-ʔa.-tjə.ŋə.la.-***ŋə.lay***
B. Suffixal c+σ reduplication			
galəmgəm	hate	mar.-ʔa.-ga.ləm.gəm	mar.-ʔa.-ga.ləm.gə-***m.gəm***
kinəmnəm	think	mar.-ʔa.-ki.nəm.nəm	mar.-ʔa.-ki.nəm.nə-***m.nəm***
ḍawḍaw	forget	mar.-ʔa.-ḍaw.ḍaw	mar.-ʔa.-ḍaw.ḍa-***w.ḍaw***
gutsguts	scratch	ma.-ga.-guts.guts	ma.-ga.-guts.gu-***ts.guts***
tsəktsək	pierce	ma.-tsa.-tsək.tsək	ma.-tsa.-tsək.tsə-***k.tsək***
duqduq	shake s.t.	ma.-da-.duq.duq	ma.-da-.duq.du-***q.duq***

Zeitoun (n.d.) describes reciprocal constructions in several Formosan languages, including Southern Paiwan, Puyuma and Rukai. Her material for Southern Paiwan shows a distinction of dual and plural number that is marked by reduplication. Southern Paiwan bases that lack a medial consonant cluster form the plural of reciprocal verbs by copying the rightmost foot (Table 6.12, Part A). As in Thao, and many other AN languages the copied base (Base-2) omits the coda. If this pattern were followed in forms that have a medial consonant cluster the expected reciprocal plurals of *galəmgəm* and *kinəmnəm* would be **galəmgə-ləmgəm and **kinəmnə-nəmnəm. Where the base coda is deleted, then, the only possible segmentation of *marʔgaləmgəmgəm* or *marʔakinəmnəmnəm* is one with reduplicants *-m.gəm* and *-m.nəm*. Like Thao, Southern Paiwan suffixal reduplication

thus copies a foot in bases that lack a medial consonant cluster (part A), but copies c+σ in bases that have a medial consonant cluster (part B). However, Zeitoun observes that some speakers break up consonant clusters through schwa epenthesis. Those born c. 1940 use *mar.-ʔa.-ga.ləm.gə.-m.gəm* 'to hate one another', for example, while those born c. 1970 tend to use the alternative form *mar.-ʔa.-ga.ləm.gə.-mə.gəm*. Younger speakers of Southern Paiwan have thus reinstated the requirement that the reduplicant be a prosodic unit, although it is by no means clear that this was anything other than an accidental by-product of the elimination of surface consonant clusters.

In Amis a similar pattern of c+σ reduplication appears in intensive or all-inclusive forms. The occurrence of a similar typologically unusual pattern of reduplication in three Formosan languages raises the question whether this pattern might have existed in Proto Austronesian. It is reasonably clear, however, that it did not. Since PAN consonant clusters were restricted to historically reduplicated monosyllables ($C_1V_1C_2C_1V_1C_2$), the pattern of copying c+σ in suffixal reduplications could reflect a PAN pattern only in Paiwan. Moreover, since Thao, Paiwan and Amis belong to three different primary branches of the AN language family, c+σ reduplication could not be inherited from an immediate common ancestor. Both Thao and Amis have lost earlier unstressed vowels, usually the reflex of PAN *e (schwa) in the environment VC__CV, producing many of the heterorganic clusters in these languages. This observation suggests that at least some instances of c+σ reduplication in these two languages may have arisen from a pattern of suffixal foot reduplication which copied *-Ce.CV(C). However, this is cannot be true of Paiwan, which reflects *e as schwa in the environment VC__CV, and never had a medial vowel in forms such as *g-ə.m-əm.gəm* 'grasp in the fist' : *g-ə.m-əm.gə-m.gəm* 'keep grasping in the fist' (PAN *gemgem 'fist; hold in the closed hand').

Crowley (1998:143) has noted a similar pattern of c+σ suffixal reduplication in Sye, spoken by about 1,400 people on the island of Erromango in southern Vanuatu, as in *om.ti* 'break' : *om.ti-m.ti* 'dilapidated', or *al.ni* 'fold' : *al.ni-l.ni* 'fold'. The history of this development is unknown, but presumably involves cluster-producing syncope.

6.9.3.3 The reduplicant is a doubly-marked syllable

Healey (1960:10) states that in Central Cagayan Agta, spoken by a foraging Negrito population in northern Luzon, "the first -VC of some words is reduplicated, if the first vowel is *i* or *u*. The vowel of the reduplication is changed from *i* to *e*, or from *u* to *o*, and the reduplication is infixed after the first syllable, thus: CiC-eC-VC or CuC-oC-VC. There is a possible element of 'diminutive' meaning in this reduplication, but the meaning of the word is usually very much changed, and the change is not predictable." Central Cagayan Agta infixal reduplication resembles heavy syllable reduplication (σ+o) in concatenating syllable plus following onset. However in Agta it is only the syllable nucleus that is joined to an onset (σn+o), and reduplication is infixal rather than prefixal.

Healey's description of the insertion algorithm for this infix, however, is inaccurate: as seen in the data reproduced below, for bases in which the first vowel is high the infix is invariably inserted before the last vowel, not 'after the first syllable,' as she states. Undoubtedly the most problematic feature of this pattern is the difficulty of correlating form and meaning, and hence of establishing the reality of a reduplicative morpheme. However, given the unusual form of the reduplicant, which involves both infixation and vowel lowering, it is not unreasonable to supose that this is a single morphological process in the language. This is assumed by Healey, and her position will be adopted here.

Examples, with syllable boundaries supplied and reduplicative infixes set off by angled brackets and bolded, are given in Table 6.13:

Table 6.13 The reduplicant as a doubly-marked syllable in Central Cagayan Agta

Base	Reduplicated word
bi.lág 'sun'	ma.-mi.l<**e.l.**>ág 'bask in the sun'
u.muk 'nest'	ma.g-u.m<**o.m**>uk 'wrap up against the wind'
u.dán 'rain'	u.d<**o.d**>án 'a lot of rain'
ma.g-u.yuŋ 'mad'	ma.g-u.y<**o.y**>uŋ 'mad'
gi.lát 'steel arrowhead'	gi.l<**e.l**>át 'small bamboo arrowhead'
u.lag 'snake'	u.l<**o.l**>ag 'insect'
hu.tug 'bow'	hu.t<**o.t**>ug 'small bamboo bow'
la.vú.n-an 'guess'	ma.ki.-l<**e.l**>a.vún 'be ignorant'
ta.lun 'forest'	i.-t<**e.t**>a.lu.n-an 'forest dweller'

Although Healey is silent on this point, the last two examples suggest that infixal reduplication in Agta is expressed through canonically complementary alloduples. If the first vowel of the base is high (represented in Figure 6.6 by *i*) the reduplicant is formed by 1) copying V_1C (C = C_1 for V-initial stems, C_2 for C-initial stems), 2) lowering the copied vowel from high to mid, and 3. infixation before the last base vowel. If the first vowel of the base is low (represented in Figure 6.6 by *a*), the reduplicant apparently is formed by 1. copying C_1V_1, 2. raising the copied vowel from low to mid, 3) metathesising the copied C_1V_1, and 4) infixation before the first base vowel. In neither case is the reduplicant a plausible template, since 1) VC syllables in Agta are never prevocalic, and 2) the insertion algorithm evidently varies according to the sonority of the copied vowel. Likewise, the reduplicant violates the 'emergence of the unmarked,' since 1) in words such as *ma.-mi.l<e.l.>ág* or *u.d<o.d>án* it lacks an onset even when it is copied from a base without onsetless syllables, and 2) the reduplicant contains vowels that are more marked that those from which they are derived. Agta infixal reduplicants are thus doubly-marked in relation to their source in the base.

First base vowel high	First base vowel low
(C_1)i.$C_2V_2C_3$	(C_1)a.$C_2V_2C_3$
(C_1)i.C_2<**e.C**>V_2C_3	(C_1)<**e.C$_1$**>a.$C_2V_2C_3$

Figure 6.6 Canonically conditioned allomorphy in Agta infixal reduplication

6.9.4 Restrictions on the content of the reduplicant

In addition to the claim that a reduplicant must be an 'authentic unit of prosody,' and cannot be more marked than the base, there is a widely accepted view that requires the reduplicant to share at least some phonemic material with the base. There are obvious circumstances under which a copying process results in *non-identity* between reduplicant and base, as with partial reduplication, fixed segmentism, etc. However, no theory allows reduplication as *null identity*, that is, a copying process in which base and reduplicant share no phonemic material. At least two AN languages nonetheless have such a process.

Sangir and Bolaang Mongondow (BM) of northern Sulawesi have inherited a process of Ca- reduplication to form instrumental nouns that will be described in greater detail later, but have altered it in sometimes striking ways through sound change. In BM two phonological innovations are important to understanding the transformation of Ca-reduplication from a transparent process of partial reduplication to one that presents special complications: 1) prepenultimate *a became schwa, and then schwa from any source became *o*, 2) *t became *s* before *i, and then borrowing of words with *ti-* gave rise to a marginal contrast of *s* and *t* before a high front vowel. As a result of change 1) Ca-reduplication has become Co-reduplication: *mo-dagum* 'to sew' : *do-dagum* 'needle', *mo-liŋkop* 'to close, as a door' : *lo-liŋkop* 'door', or *dupaʔ* 'beat with a hammer' : *do-dupaʔ* 'hammer'. This change had no structural consequences, but as a result of change 2) bases that began with *ti now begin with *si-* and the reduplicant is *to-*, as in *mo-silad* 'to split betel nuts' : *to-silad* 'betel knife', *mo-simbaŋ* 'to weigh' : *to-simbaŋ* 'a scale', *mo-simpat* 'to sweep' : *to-simpat/so-simpat* 'broom', or *mo-siug* 'to sleep' : *to-siug-an/so-siug-an* 'bed'. At a slightly earlier period in the history of BM before the acquisition of *ti-* loans, *silad* was underlying *tilad*, but contained the phonetic sequence [si]-. At this stage Co-reduplication preserved partial identity between base and reduplicant on the phonemic level, but had begun to develop instances of reduplicative null identity on the phonetic level. Once phonetic null identity between base and reduplicant began to approach phonemic null identity through the acquisition of loans with *ti-*, variant pronunciations of *to-simpat* and *to-siug-an* (namely *so-simpat* and *so-siug-an*) began to develop.

The BM variants *so-simpat* and *so-siug-an* cannot be products of primary sound change, which gave rise to an [s] allophone of *t* only before a high front vowel. Rather, they result from analogical leveling based on the predominant pattern of Co-reduplication used to form instrumental nouns. From the standpoint of general linguistic theory it is also worth noting that the classic alternative to analogical leveling, namely rule loss, is not available here, since the rule *t* > *s*/__*i* was not lost, but was rephonemicised and the new phoneme extended to the reduplicant. In other words, rule loss would yield not *so-simpat* and *so-siug-an*, but rather **to-timpat and **to-tiug-an. The upshot of these observations is interesting: since paradigmatic leveling by analogy implies a recognition of patterns, the *so-* variants of these derived instrumental nouns provide evidence for the psychological reality of reduplication as null identity. Put differently, if structural pressure operates on the level of the phoneme rather than the allophone, Mongondow speakers must have considered phonemic forms such as *to-simpat* as instances of Co-reduplication, or there would be no basis for the phonetically unmotivated change of the initial consonant in the *so-* variants. Sangir presents problems of even greater complexity, as seen in Table 6.14.

The most transparent evidence that instrumental nouns in Sangir reflect a process of Ca-reduplication comes from stems that begin with a voiceless obstruent *k, p, s,* or *t*. These behave very much like formally similar stems in canonical languages such as Thao, Puyuma, or Tetun, having only the added complication of homorganic nasal substitution in the paired active verb. With the voiced obstruents we find ourselves one step further removed from this level of morphological transparency, since word-initial *b, d,* and *g* alternate with the corresponding continuants *w, r* and *gh* ([ɣ]).

Table 6.14 Ca-reduplication as base-reduplicant null identity in Sangir

Base	Verb	Instrumental noun
aki	maŋ-aki 'add cordage'	la-aki 'extension to fishline'
baŋgo	ma-maŋgo 'to beat'	ba-waŋgo 'cudgel'
biŋkuŋ	ma-wiŋkuŋ 'to adze'	ba-wiŋkuŋ 'an adze'
dosa	mən-dosa 'to pound'	da-rosa 'wooden mallet'
əmmuʔ	maŋ-əmmuʔ 'to wipe off'	la-əmmuʔ 'dustcloth'
gataʔ	meŋ-gataʔ 'carry under arm'	ga-ghataʔ 'bamboo tongs'
himadəʔ	mə-himadəʔ 'to gouge'	la- himadəʔ 'gouging tool'
ikiʔ	maŋ-ikiʔ 'to tie'	la-ikiʔ 'anything used for tying'
kətuŋ	ma-ŋətuŋ 'seize with pincers'	ka-kətuŋ 'pincers'
lədaŋ	mə-lədaŋ 'file the teeth'	da-lədaŋ 'tooth file'
pədasəʔ	ma-mədasəʔ 'to whet'	pa-pədasəʔ 'whetstone'
sapu	ma-napu 'to sweep'	sa-sapu 'broom'
tubuŋ	ma-nubuŋ 'knock down fruit'	ta-tubuŋ 'fruiting pole'
uhasəʔ	maŋ-uhasəʔ 'to wash'	la-uhasəʔ 'washing water'
lauʔ	mə-lauʔ 'to mix'	da-lauʔ 'mixing tool'
limasəʔ	limasəʔ 'bail out water'	da-limasəʔ 'bailer in boat'

Somewhat more opaque are Ca-reduplications that begin with *l*, as the copied portion of the base dissimilates to *d*. Historically, this dissimilation apparently was not regular, since Sangir has many stems that contain the sequence *lVl*, in which the initial liquid derives from *l or *y. Most puzzling of all are stems that begin with a vowel, as these are reduplicated as *la-*. As just noted, there are two regular historical sources for this intrusive liquid: *l and *y. Assuming that the liquid of *la-* derives from *y might seem to mitigate the anomalous character of this reduplicative allomorph, since glide epenthesis is fairly common in other languages. But this hardly helps, as the presumptive glide was not inserted to break up a vowel cluster, and no such epenthesis took place before unreduplicated vowel-initial stems. Finally, *h*-initial stems are treated as though they begin with a vowel, even though the *h* is historically secondary, and corresponds to *r* in the closely related Sangil of the southern Philippines.

To summarise, Sangir shows the following alloduples for Ca-reduplication. Where relevant, base-initial segments are given in two forms—that of the independent base (Base-1), and that of the reduplicated base (Base-2), hence x/y:

Table 6.15 Ca- alloduples in Sangir showing the extent of base-reduplicant null identity

No.	Base initial	Reduplicant	No.	Base initial	Reduplicant
01.	a	la-	21.	kə	ka-
02.	ba/wa	ba-	22.	ki	ka-
03.	bə/wə	ba-	23.	ko	ka-
04.	bi/wi	ba-	24.	la	da-
05.	bo/wo	ba-	25.	lə	da-
06.	bu/wu	ba-	26.	li	da-
07.	da/ra	da-	27.	lu	da-
08.	də/rə	da-	28.	pa	pa-
09.	di/ri	da-	29.	pə	pa-
10.	do/ro	da-	30.	pi	pa-

No.	Base initial	Reduplicant	No.	Base initial	Reduplicant
11.	du/ru	da-	31.	pu	pa-
12.	ə	la-	32.	sa	sa-
13.	ga/gha	ga-	33.	sə	sa-
14.	ga/ghə	ga-	34.	si	sa-
15.	gu/ghu	ga-	35.	su	sa-
16.	hə	la-	36.	ta	ta-
17.	hi	la-	37.	tə	ta-
18.	ho	la-	38.	ti	ta-
19.	hu	la-	39.	tu	ta-
20.	i	la-	40.	u	la-

Despite forms such as *ma-wiŋkuŋ* 'to adze' : *ba-wiŋkuŋ* 'an adze', it might be argued that partial BR identity is preserved in bases that begin with a voiced obstruent (*biŋkuŋ* 'adzing'), but there is no obvious way that any current theory of reduplication can recognise the prefixal element of Sangir *la-həpiŋ* 'door, window' (*mə-həpiŋ* 'to close a door or window'), *la-inum-aŋ* 'cup' (*maŋ-inuŋ* 'to drink'), or *da-limasəʔ* 'canoe bailer' (*mə-limasəʔ* 'bail out a canoe') as a copy of the base. The easiest escape from this dilemma would be to exclude such cases from the process of Ca-reduplication in Sangir. The problem with this solution is that it fails to recognise the obvious functional unity of these reduplicative allomorphs with the straightforward instances of Ca-reduplication in stems that begin with a voiceless obstruent, the somewhat less straightforward instances of Ca-reduplication in stems that begin with a voiced stop, and so on. Instrumental nouns formed by Ca-reduplication that have been recognised to date exemplify some 40 of the 76 logically possible types of bases defined by the initial CV sequence. Of these 40 types BR null identity is found in 11 types with Base-1, and in 22 types with Base-2.

In both Sangir and Bolaang Mongondow some bases exemplify a straightforward pattern of Ca- (or Co-) reduplication. Without these it would no longer be possible to consider the morphologically more opaque cases as instances of a copying process. Since the realisation of this process shows a more-or-less continuous gradation of phonological transparency, and since there is clear evidence of pattern congruity which connects the worse cases to the best, there seems to be no non-arbitrary way in which to exclude *da-* as the reduplicant of *lu*, or *la-* as the reduplicant of *hu* or *i*. In effect, then, reduplication as null identity is undeniable, but appears to be possible only on the level of the allomorph, and from this it follows that the identity requirement between base and reduplicant must apply on the level of the morpheme. All discussions of the identity requirement appear to illustrate the BR correspondence with examples on the level of the allomorph/alloduple, thus avoiding the issue of whether *some* allomorphs could show null identity with the base. Although the matter is too involved to pursue here, this practice probably results from the fact that the representation of reduplication on the level of the morpheme is formulaic to a far greater degree than is true of e.g. the morphophoneme.

6.9.5 Patterns of reduplication

Some of the issues that reduplication in AN languages raises for general linguistic theory have been addressed in the preceding section. This section briefly surveys the wider range of reduplication patterns that are attested in these languages.

6.9.5.1 *Full reduplication*

The most transparent type of reduplication is complete copying of a base morpheme. This has been reported with a variety of functions in a number of languages. It is noteworthy, however, that full reduplication apparently is never used to mark grammatical functions such as tense, although partial reduplication may carry this type of information. The following examples illustrate:

Botolan Sambal (Antworth 1979:11): A fully reduplicated base indicates a diminutive or make-believe object: *anak* 'child' : *anak-anak* 'doll', *tawo* 'person' : *tawo-tawo* 'scarecrow', *bali* 'house' : *bali-bali* 'playhouse'.

Malay/Indonesian (various sources): 1) full reduplication of nouns usually signals plurality, as in *anak* 'child' : *anak-anak* 'children', *rumah* 'house' : *rumah-rumah* 'houses', or *kəlapa* 'coconut' : *kəlapa-kəlapa* 'coconuts', 2) in some cases, however, it may signal similitude, as in *bantal* 'pillow' : *bantal-bantal* 'railroad tie', or *jala* 'fishing net' : *jala-jala* 'small net such as a hair net', 3) full reduplication of intransitive verbal bases sometimes indicates a leisurely or undirected execution of the action of the verb: *(bər-)jalan* 'walk, go' : *(bər-)jalan-jalan* 'stroll around', *makan* 'eat' : *makan-makan* 'eat not out of hunger, but as in sampling food at a party; go through the motions of eating', *tidur* 'to sleep' : *tidur-tidur* 'lie down to rest (with no intent to sleep)', 4) with other verb bases full reduplication signals repetitive or durative action, as with *marah* 'angry' : *marah-marah* 'get angry at s.o. again and again', *(bər-)cakap* 'to talk' : *(bər-)cakap-cakap* 'chatter on and on', 5) with adjectival bases full reduplication indicates intensity: *kəras* 'hard, of material; intense, of effort' : *kəras-kəras* 'intensely', *tiŋgi* 'high' : *tiŋgi-tiŋgi* 'very high', 6) with some other bases the meaning is unpredictable, as with *apa* 'what?' : *apa-apa* 'whatever'.

Karo Batak (Woollams 1996:92ff): 1) full reduplication of some nouns signals plurality, as in *tulan* 'bone' : *tulan-tulan* 'bones', *sinuan* 'plant' : *sinuan-sinuan* 'plants', or *kəjadin* 'event' : *kəjadin-kəjadin* 'events', 2) with other nouns full reduplication signals imitation or similitude, as in *nahe* 'leg' : *nahe-nahe* 'stilts', *nipe* 'snake' : *nipe-nipe* 'grub, caterpillar', *bərku* 'coconut shell' : *bərku-bərku* 'skull', or *kacaŋ* 'peanut' : *kacaŋ-kacaŋ* 'clitoris'; with colour terms reduplication has an approximative meaning, as with *məgara* 'red' : *məgara-məgara* 'reddish', or *mbiriŋ* 'black' : *mbiriŋ mbiriŋ* 'blackish'; some verbs also acquire an imitative meaning through full reduplication, as with *mədəm* 'sleep' : *mədəm-mədəm* 'lie down, rest'.

6.9.5.2 *Full reduplication plus affixation*

In many languages fully reduplicated bases occur with non-reduplicative affixation. Since some examples of this pattern have already been given, only a few more will be added here: Thao *k<m>an* 'eat' : *k<m>a-kan* 'eat something often', *patash* 'write' : *matash* 'to write' : *mata-tash* 'write and write, keep on writing', Tagalog *l<um>akad* 'walk' : *mag-lakad-lakad* 'walk a little', Malay *bər-jalan* 'walk' : *bər-jalan-jalan* 'stroll about', Malay *mə-lihat* 'see' : *mə-lihat-lihat* 'take one's time viewing something (as when visiting a museum)', or Karo Batak *suŋkun* 'ask' : *nuŋkun-nuŋkun* 'keep on asking', *tatap* 'look' : *natap-natap* 'look around, view', *apus* 'wipe' : *ŋ-apus-ŋ-apus-i* 'wipe repeatedly'. With regard to prefixation there are three patterns of interaction between reduplication and non-reduplicative affixation: 1) the first half of the word is prefixed (Malay *məmasak-masak* 'do the cooking'), 2) the second half of the word is prefixed (Malay *masak-məmasak* 'cookery'), 3) both halves of the word are prefixed (Karo Batak *nuŋkun-nuŋkun* 'keep on asking', base *suŋkun*). Although Malay has examples of the first two types it

apparently has no examples of type 3. Moreover, no examples are known of pattern 1 in conjunction with nasal substitution: *potoŋ* 'cut' : *potoŋ-məmotoŋ* 'cut one another' (but **məmotoŋ-potoŋ), *surat* 'write; letter' : *surat-məñurat* 'write to one another' (but **məñurat-surat), *tari* 'dance' : *tari-mənari* 'dancing' (but **mə-nari-tari). Such a pattern is attested in Palauan, however, where it is well disguised by sound change (final vowels are unknown in some reconstructions and are written as *a* for visual ease): *tub* 'spittle, saliva', *məlub* 'spit on' : *mə-ləb-tub* 'keep spitting', (< *suba : *ma-ñúba : *ma-ñuba-súba), *kimd* 'cut' : *məŋímd* 'cut hair, trim' : *mə-ŋəm-kímd* 'keep trimming s.o.'s hair' (< *kimat : *ma-ŋímat : *ma-ŋima-kímat), *báləʔ* 'slingshot' : *o-máləʔ* 'shoot with a slingshot' : *o-mələ-báləʔ* 'play around with a slingshot' (< *banaq : *pa-mánaq : *pa-mana-bánaq).

In both Thao and Karo Batak active verbs with fully reduplicated bases usually signal frequentative or repetitive action. In some cases a pattern of full reduplication plus affixation is widely shared, as with Kapampangan, Tagalog *anák-anák-an* 'adopted child', Malay/Indonesian *anak-anak-an* 'doll', anak-*anak-an məntimun* (lit. 'cucumber doll') 'child adopted for one's wife', Old Javanese *anak-anak-an* 'doll, or anything which is treated and fondled like a baby; pupil of the eye' (also cf. Malay *oraŋ* 'person' : *oraŋ-oraŋ-an* 'scarecrow', *rumah* 'house' : *rumah-rumah-an* 'playhouse', and various other words that use the pattern base-base-*an* to form simulative nouns).

Various reduplication patterns may require vowel assimilations or other types of phonological change. Since these do not change the morphological composition of the word, however, they will be treated like the similar reduplication patterns without vocalic changes. Paamese of central Vanuatu, for example, has a recurrent pattern of frontness assimilation in full reduplications such as *muni* : *munu-munu* 'drink', *luhi* : *luhu-luhu* 'plant', and *uhi* : *uhu-uhu* 'blow' (Crowley 1982:48), and in Chamorro deverbal nouns derived by CV- reduplication show automatic fronting of back vowels, as in *gupu* 'to fly' : *gi-gipu* 'flyer' (Topping 1973:181-182). The second of these patterns apparently results from a general fronting rule for vowels following the article *i* (*gumaʔ* 'house' : *i-gimaʔ* 'the house'), but the cause of the Paamese assimilation is unclear, since a high front vowel between syllables with *u* is found in lexical forms such as *musinuni* 'put on, wear shirt or hat' (Crowley 1992).

6.9.5.3 *Full reduplication minus the coda*

This pattern appears to have been present in PAN, as it is found both in Formosan and non-Formosan languages. In it the base is copied without a coda, producing a typologically unusual situation in which the base (Base-2) lacks material that is present in the reduplicant. Examples from Thao are seen in Table 6.11, where it is clear that this is actually suffixal foot reduplication minus the base coda. In most languages examples of reduplicated bases longer than two syllables cannot easily be found, leaving the distinction between full reduplication and suffixal foot reduplication ambiguous.

Itbayaten (Yamada 1976): *hapin* 'mat' : *h<in>api-hapin* 'woven one', *koxat* 'heat, warmth' : *ma-ŋoxa-ŋoxat* 'boil repeatedly', *oxas* 'cleanness' : *ma-wxa-wxas* 'cleaner, neater (comparative)', or *takəy* 'field, farm' : *mi-takə-takəy* 'farmer, do farming'.

Ngaju Dayak (Hardeland 1859): *abas* 'strong' : *aba-abas* 'rather strong', *humoŋ* 'stupid' : *humo-humoŋ* 'rather stupid', *tiroh* 'sleep' : *ba-tiroh* 'to sleep' : *ba-tiro-tiroh* 'to sleep a little, or for a short time'.

Bolaang Mongondow (Dunnebier 1951): *bayag* 'light, bright' : *ko-baya-bayag-an* 'become completely bright or clear', *kuntuŋ* 'carry piggyback' : *ko-kuntu-kuntuŋ* 'being carried piggyback', *posad* 'communal work' : *ko-posa-posad* 'doing communal work'.

6.9.5.4 Full reduplication minus the last vowel

Many Oceanic languages that have lost final vowels show reduplication patterns of this type. These arose from CVCV bases which were fully reduplicated with subsequent loss of the final vowel of the word. Where etymological information is available, the medial vowel of such reduplications almost always reflects the lost historical final (POC *kani 'to eat', *roŋoR 'to hear', *taŋis 'to cry', *inum/unum 'to drink', *paŋan 'to feed', *manipis 'thin', etc.):

Seimat (Admiralty Islands; Blust n.d. b), Smythe n.d.): *aŋ* 'eat' : *aŋi-aŋ* 'eating', *hõŋ* 'hear' : *hõŋo-hõŋ* 'hearing', *pak* 'sing' : *paku-pak* 'singing', *put* 'fall' : *puta-put* 'falling', *taŋ* 'cry' : *taŋitaŋ* 'crying', *tele-i* 'kill (something)' : *tele-tel* 'killing', *un* 'drink' : *unu-un* 'drinking'.

Loniu (Admiralty Islands; Hamel 1994:81ff): *cim* 'buy' : *cimi-cim* 'buying', *haŋ* 'feed' : *haŋahaŋ* 'adoptive', *nɔh* 'to fear' : *nɔhɔnɔh* 'fear (N)', *iw* 'call' : *iwiʔiw* 'calling'.

Woleaian (Micronesia; Sohn 1975:102): *bis* 'brother' : *bisibis* 'be in a brother relation', *roŋ* 'hearsay, tradition' : *roŋoroŋ* 'to hear', *masow* 'hard' : *masowe-sow* 'strong', *malif* 'thin' : *malifi-lif* 'thin'.

6.9.5.5 Full reduplication with vocalic or consonantal change, or both

Copying patterns of this type are similar to English 'flip-flop' or 'zig-zag' (with vowel change), or 'hocus-pocus' and 'helter-skelter' (with consonant change). They are known from very few AN languages. The best-known examples are found in Malay/Indonesian, where (Macdonald and Soenjono 1967:53) call them 'imitative reduplications'. These reduplications are more variable than the comparable examples in English, as they may show differences in both vowels of a disyllabic base, or (rarely) variation in both vowels and consonants. Words formed in this way usually imply a multiplicity of referents or actions, and often some sort of chaos:

Malay/Indonesian (Wilkinson 1959; Macdonald and Soenjono 1967:54; Moeliono 1989): *bolak-balik* 'lying this way and that, of bottles or bamboos laid side by side but alternately in different directions; also of a story that is never the same' (= *bolak* 'go back on one's word' + *balik* 'return'), *dəsas-dəsus* 'rumors, whispering' (*dəsus* = 'sound of whispering'), *joŋkat-jaŋkit* 'bob up and down' (*joŋkat* 'stand on one's toes'), *kucar-kacir* 'scattered helter-skelter' (*kacir* = 'depart without taking one's leave, out of embarrassment, fear, etc.), *umbaŋ-ambiŋ* 'drift to and fro' (*umbaŋ* = 'to float'); *cərai-bərai* 'disperse' (*cərai* 'sever, cut ties'), *coreŋ-moreŋ* 'full of scratches' (*coreŋ* = 'scratch'), *pəcah-bəlah* 'shattered' (*pəcah* 'broken into several large pieces', *bəlah* 'split'), *sayur-mayur* 'various sorts of vegetables' (*sayur* = 'vegetable'); *eraŋ-erot* 'zigzag' (*erot* = 'crooked'), *sabur-limbur* 'confused, dusky' (*sabur* = 'vague, dim').

Examples such as *pəcah-bəlah* raise the question of where and how to draw the line between reduplication and compounding. Both *pəcah* and *bəlah* are independent words, and so the combination fits the usual definition of a compound. With words like *sayur-mayur* the situation is different: here only *sayur* may occur alone, and the two halves of the word are so similar that some sort of copying process must be assumed. Historically, imitative reduplication may already have existed, and then words like *pəcah-bəlah* were added to an existing pattern by combining independent words of matching vocalism and similar meaning. In any event the synchronic result is a collection of bipartite words that sometimes appear to be reduplications and sometimes compounds. Unlike the similar expressions in English, where neither half of the word ever occurs independently (*hocus*,

pocus, *helter*, *skelter*, etc.), at least one half of the word in Malay/Indonesian 'imitative reduplications' normally has a meaning.

Hamel (1994:81) cites a few reduplicated forms from Loniu in the Admiralty Islands, which show unpredictable vocalic changes in the reduplicant, as with *kah* 'hunt for' : *kɛhɛkah* 'hunt', or *sah* 'carve (trans.)' : *sɛhisah* 'carve (intrans.)'. Unlike the Malay examples, which almost certainly are deliberate creations based on existing lexical bases, Loniu examples such as these appear to be accidental by-products of sound change (*sahi > sahi-sahi > sahi-sah > *sɛhisah*, etc.).

6.9.5.6 Full reduplication with four consecutive identical syllables

It has been claimed that reduplication will be avoided if it would produce four consecutive identical syllables. This tendency is seen in some AN languages, but not in others. In Puyuma, for example, *taina* 'mother' reduplicates as *mar-taina-ina* 'mother and children', but *tamama* 'father' reduplicates as *mar-tamama-ma* (**mar-tamama-mama) 'father and children'. However, Amis has no problem with *wawa* 'child' : *wawa-wawa* 'children', or *aɬaɬa* 'pain, sickness' : *aɬaɬa-ɬaɬa* 'pretend to be sick'.

6.9.5.7 Prefixal foot reduplication/leftward reduplication

Foot reduplication in AN languages tends to be suffixal. As a result reduplicative patterns which copy CVCV- are relatively hard to find, but the following two cases can be cited.

Araki, a moribund language spoken off the south coast of Espiritu Santo in north-central Vanuatu (François 2002:31): *märahu* : **mära**-*mäarahu* 'fear, be afraid', *mäcihi* : **mäci**-*mäcihi* 'colour', *veculu* : **vecu**-*veculu* 'whistle', *hudara* : **huda**-*hudara* 'dirt'.

Palauan exhibits an unusually complex pattern of prefixal foot reduplication that Josephs (1975:234ff) symbolises as $C_1eC_1V(C_2)$-: *mə-saul* 'tired' : *mə-**sesu**-saul* 'sort of tired', *mə-dakt* 'afraid' : *mə-**dedək**-dakt* 'sort of afraid', *mə-saik* 'lazy' : *mə-**sesi**-saik* 'sort of lazy', *mə-riŋal* 'difficult' : *mə-**rerəŋə**-riŋəl* 'sort of difficult', *mə-kar* 'awake' : *mə-**kekər**-kar* 'half-awake'. This 'quite productive' reduplication pattern, which adds a semantic overlay of mitigation to the meaning of what Josephs calls a 'state verb', is typologically unusual for two reasons. First, it copies the first base consonant *twice* -- once on each side of the fixed vowel /e/. Second, it appears superficially that with stems of the shape CVVC or CVCC the reduplicant is a foot, but with stems of the shape CVCVC it may be a string of three syllables, as in *mə-**rerəŋə**-riŋəl* 'sort of difficult'. However, the latter complication depends crucially on the phonemic status of schwa.

Palauan often breaks up consonant clusters with an excrescent schwa, and the underlying forms of bases probably are better written without it. In trying to predict the shapes of $C_1eC_1V(C_2)$-reduplicants it becomes clear that Josephs' formula is adequate for the first three segments, but then breaks down: with *mə-saul* the formula generates *ses* + $V(C_2)$, but does not specify which vowel to copy (/a/ or /u/), nor the conditions under which C_2 gets copied (ditto for *mə-saik*); with *mə-dakt* the formula generates *ded* + a(k), hence *deda* or *dedak* for actual *dedək*, with *mə-riŋal* it generates *rer* + i(ŋ), hence *reri* or *reriŋ* for actual *rerəŋə*, and with *mə-kar* it generates *kek* + a(r), hence *keka* or *kekar* for actual *kekər*. Josephs tries to cope with these problems by arguing that the reduplicated syllable 'shows the effects of vowel reduction … and vowel cluster reduction' which are independently motivated in the grammar. However, with this analysis he is still unable to predict which vowel is copied onto the reduplicant, and when C_2 is copied or omitted. By

contrast, once the schwa is treated as excrescent, a copying algorithm emerges that can account for all of the data without ambiguity or residue. This algorithm can be stated as 'copy $C_1eC_1S_3$, where S_3 represents the third segment of the base, whether it is a vowel or a consonant, and insert before the base.' This gives 1) *ses + u*, 2) *ded + k*, 3) *ses + i*, 4) *rer + ŋ*, and 5) *kek + r*, with schwa insertion breaking up the clusters in 2, 4 and 5 both within the reduplicant and between the reduplicant and base. The second odd feature of this copying pattern noted above thus needs to be restated, but the restatement is more surprising than the original. Most theoretical approaches assume that reduplicative templates can be stated in term of an alphabet of three primary symbols, V (vowel), C (consonant) and an open class of prespecified segments (usually a single vowel). To describe $C_1eC_1S_3$ reduplication in Palauan, however, it is necessary to add S (segment) to this alphabet, since the S_3 slot of the template includes both vowels and consonants under complementary phonotactic conditions (with schwa epenthesis to break up derived consonant clusters). A similar pattern of reduplication is used to derive what can be called facilitative verbs from the ergative forms of transitive verb bases: *m-daŋəb* 'be/get covered' : *mə-**dedəŋə**-daŋəb* 'easy to cover', *obuid* 'be/get glued' : *o-**bebi**-buid* 'easy to glue', *mə-luʔəs* 'be/get written' : *mə-**leləʔə**-luʔəs* 'easy to write on'.[65]

6.9.5.8 Suffixal foot reduplication/rightward reduplication

This is a common type of reduplication in AN languages, but requires bases of more than two syllables to distinguish it from full reduplication. For this reason it has generally been reported only in languages for which fairly extensive descriptive data is available. Examples are given here from Paiwan of southeastern Taiwan, Manam of New Guinea, and Hawaiian:

Paiwan (Ferrell 1982): *kulalu* 'flute' : *k<m>ulalu* 'play a flute' : *k<m>ulalu-**lalu*** 'be playing a flute', *ɬimatjək* 'leech' : *ɬimatjə-**matjək**-ən* 'covered with leeches', *qatia* 'salt' : *qatia-**tia*** 'tiny glass beads', *saqətju* 'painful, sick' : *p<n>a-saqətju-**qətju*** 'painful, causing pain', *saviki* 'areca nut' : *saviki-**viki**-n* 'areca plantation'.

Manam (Lichtenberk 1983:599) : *salaga* 'be long' : *salaga-**laga*** 'long (sg.)', *moita* 'knife' : *moita-**ita*** 'cone shell', *ʔarai* 'k.o. ginger' : *ʔarai-**rai*** 'green (sg.)', *sapara* 'branch' : *sapara-**para*** 'having branches; pants', *malipi* 'work' : *malipi-**lipi*** 'be working'.

Lichtenberk calls this 'rightward disyllabic reduplication', and describes it as 'by far the commonest type' in Manam.

Hawaiian (Pukui and Elbert 1971): *aloha* 'love, affection' : *āloha-**loha*** 'express affection', *kiawe* : *kīawe-**awe*** 'stream gracefully, as rain in the wind' : *kūpele* : *kūpele-**pele*** 'to knead, as bread dough', *pohole* : *pōhole-**hole*** 'bruised, skinned, scraped', *pueo* 'rock a child on the foot' : *pūeo-**eo*** 'rocking a child on the foot'. The Hawaiian pattern of suffixal

---

[65] Josephs (1975:237) maintains that some pairs of ergative transitive and facilitative verbs such as *mə-ʔəsimər* 'be/get closed' : *mə-ʔe-ʔəsimər* 'easy to close' (**mə-ʔeʔəs-ʔəsimər*), *mə-ʔəlebəd* 'be/get hit' :*mə-ʔe-ʔəlebəd* 'easy to hit' (**mə-ʔeʔə-ʔəlebəd*), *mə- təkoi* 'be/get talked to' : *mə-te-təkoi* 'easy to talk to' (**mə-**tetk**-təkoi*), and *mə-sesəb* 'be/get burned' : *mə-se-sesəb* (**mə-**sess**-sesəb* 'flammable' show only Ce- reduplication. The first two cases are transparent instances of haplology, the third is evidently due to reduction of the derived cluster -*tkt*- followed by geminate reduction, and the last to reduction of the unique derived cluster -*sss*-. The only example in Josephs (1975) that does not appear to conform to the formula C_1eC_{1sg3} is *mə-kiut* 'be/get cleared' : *mə-keki-kiut* 'easy to clear' (expected **mə-keku-kiut). I am indebted to Kie Zuraw (p.c.) for improvements in presentation that resulted from discussion of some of these issues.

foot reduplication is noteworthy in relation to vowel length, which is phonemic in the language: if the first vowel of Base-1 is short it is lengthened in Base-2.

6.9.5.9 CVC- reduplication

This copying pattern, sometimes called 'heavy syllable reduplication', occurs in two known types. Type 1) is found in Saisiyat of northwest Taiwan, in a number of the languages of northern Luzon, including Central Cagayan Agta, Ilokano, Isneg, Isinay, Bontok, Pangasinan, and in a number of Oceanic languages. Type 2) is known so far only from Palawano. In Bontok Type 1) heavy syllable reduplication may be accompanied by additional processes of glottal stop formation, metathesis and consonant gemination (E. Thurgood 1997):

TYPE 1: Agta (Healey 1960:7ff). In Agta CVC- reduplication has several functions. With some bases it marks plurality or multiplicity of referents, as in *uffu* 'thigh' : **uf**-*uffu* 'thighs', *takki* 'leg' : **tak**-*takki* 'legs', *ulu* 'head' : **ul**-*ulu-da* 'their heads', or *d<um>ataŋ* 'arrive' : **d**<*um*>**at**-*dataŋ* 'keep arriving one after another'. With others it marks intensity, as in *adánuk* 'long' : **ad**-*adánuk* 'very long', *apísi* 'small' : **ap**-*apísi* 'very small', *abíkan* 'near' : **ab**-*abíkan* 'very near', or *ma-baŋí* 'delicious-smelling' : *ma-**baŋ**-baŋí* 'very delicious-smelling'. With other bases it marks diminutive or some other senses, as in *átu* 'dog' : **at**-*átu* 'puppy', *balatáŋ* 'girl' : **bal**-*balatáŋ* 'little girl', etc.

TYPE 2: Palawano (Revel-Macdonald 1979:187). In addition to the rather common pattern of CVC- reduplication illustrated for Central Cagayan Agta, Palawano of the west-central Philippines creates a diminutive prefix by copying the first CV of the base together with the *final* base consonant: *kusiŋ* 'cat' : **kuŋ**-*kusiŋ* 'kitten', *bajuʔ* 'clothing' : **bäʔ**-*bajuʔ* 'child's clothing', *libun* 'woman' : **lin**-*libun* 'girl', *kunit* 'yellow' : **kut**-*kunit* 'yellow flycatcher (bird)', *siak* 'tears' : **sik**-*siak* 'crocodile tears/false tears'.

Again, although the reduplicant corresponds to a unit of prosody, it is not formed from a unit of prosody, or even from segments that are adjacent in the stream of speech. In addition to these Philippine examples, Type 1) CVC- reduplication is reported from Yapese of western Micronesia (Jensen 1977a:110), Tigak of New Ireland (Beaumont 1979:93), and Kosraean of central Micronesia, where it is as an optional pattern in free variation with CV- reduplication used to mark repetitive or distributive action, as in *ku.lus* 'to peel' : **kul**.*ku.lus* or **ku**.*ku.lus* 'peel bit by bit', *pih.srihk* 'to flick' : **pihsr**.*pih.srihk* or **pih**.*pih.srihk* 'flick repeatedly', *i.pihs* 'to roll' : **ip**.*i-pihs* or **i**.*i.pihs* 'roll bit by bit', and *la.kihn* 'to spread' : **lak**.*la.kihn* or **la**.*la.kihn* 'to spread' (Lee 1975:219).

6.9.5.10 CV-reduplication

As in languages generally, this is a common copying pattern. In Bunun it marks durative aspect, collectivity, or intensity, in Tagalog it marks future tense, in Manam it marks plurality of adjectives or the continuative, progressive, or perseverative aspects of some transitive verbs when the direct object is a third person plural non-higher-animal, in North-East Ambae it is an allomorph of full reduplication (function unspecified) and in Pangasinan it marks the plural of nouns. In many languages CV-reduplication copies only full syllables, but in some (as Bunun) it copies only the syllable onset and nucleus, or only the nucleus:

Bunun (Blust n.d. c)): *ma-asik* 'to sweep' : *ma-**a**-asik* 'keep sweeping', *bazbaz* 'talk, speak' : **ba**-*bazbaz* 'non-stop talking', *bicvaq-an* 'to thunder' : **bi**-*bicvaq-an* 'thunder a lot, keep thundering', *buntu* 'often' : *mal-**bu**-buntu* 'keep doing s.t.', *cucu* 'breast' : **cu**-*cucu*

'be nursing, of a baby', *mu-dan* 'walk, go' : *mu-**da**-dan* 'keep walking, keep going', *kitnus* 'fart' : ***ki**-kitnus* 'keep farting', *ma-patas* 'to write' : *ma-**pa**-patas* 'keep writing', *qudan-an* 'be raining' : ***qu**-qudan-an* 'keep on raining', *ma-uktic* 'to cut, as paper' : *ma-**u**-uktic* 'keep cutting, as paper', *ma-bulav* 'ripe, yellow' : *ma-**bu**-bulav* 'all ripening at once', *ma-kuis* 'slender' : *ma-**ku**-kuis* 'very slender'.

Tagalog (Ramos1971): *b<um>ilí* 'to buy' : ***bi**-bilí* 'will buy', *um-iyák* 'to cry' : ***i**-iyák* 'will cry', *l<um>ákad* 'to walk' : ***la**-lákad* 'will walk', *s<um>ulát* 'to write' : ***su**-sulát* 'will write', *s<um>unód* 'to follow, obey' : ***su**-sunód* 'will follow, will obey'.

Manam (Lichtenberk 1983:603): *salaga* 'be long' : ***sa**-salaga* 'long (pl.)', *tumura* 'be cold' : ***tu**-tumura* 'cold (pl.)', *noʔa* 'be ripe' : ***no**-noʔa* 'ripe (pl.)', *gara-s* 'scrape' : ***ga**-gara-s* 'is scraping'.

North-East Ambae (Hyslop 2001:45). The Lolovoli dialect of North-East Ambae in Vanuatu shows an unusual condition for CV- reduplication. Bases of two syllables are usually fully reduplicated, as with *kalo* : ***kalo**-kalo* 'climb', *mʷoso* : ***mʷoso**-mʷoso* 'play', or *tomu* : ***tomu**-tomu* 'tell a story'. With bases of more than two syllables, however, only the first syllable is copied, as in *garea* : ***ga**-garea* 'good', *lague* : ***la**-lague* 'big', or *sogagi* : ***so**-sogagi* 'sell'. The motivation for such canonical complementation is unclear, since a few disyllabic bases use CV- reduplication, as *maŋi* : ***ma**-maŋi* 'wipe', or *tunu* : ***tu**-tunu* 'roast', and some bases allow both CV- and full reduplication, with different semantics, as *garu* : ***ga**-garu* 'swim, bathe', ***garu**-geru* 'swim, bathe (emphasis on continuous action)'. Given these observations the status of CV- reduplication as a distinct reduplicative structure or an alloduple of full reduplication remains in doubt in this language.

6.9.5.11 CV-reduplication plus affixation

As with other forms of reduplication, CV- reduplication may co-occur with non-reduplicative affixation. Examples are given from Thao, where with some motion verbs CV- reduplication together with a non-reduplicative prefix marks repetitive action:

Thao (Blust 2003a:193): *luish* 'short, low' : *mak-**lu**-luish* 'breathing hard, short of breath', *iŋkmir* 'grasping in the hand' : *m-iŋkmir* 'to grasp or knead' : *miŋ<m>iŋkmir* 'grasp or knead repeatedly', *tusi* 'there' : *mu-tusi* 'go over there' : *mu-**tu**-tusi* 'go over there repeatedly or often'. In *miŋ<m>iŋkmir* the copied CV- consists of the prefix *m-* plus base-initial *i*, with contraction of the resulting sequence of like vowels.

6.9.5.12 Ca-reduplication

Ca- reduplication copies the first vowel of a base followed by the fixed vowel *a*, as in Puyuma (Tamalakaw) *kədan* 'whet' : *ka-kədan* 'whetstone', or *Tiŋa* 'food caught between the teeth' : *Ta-Tiŋa* 'toothpick' In vowel-initial bases this fixed vowel is the reduplicant, as in Thao *m-iup* 'blow on s.t.' : *a-iup* 'tube used to blow on the fire'. This pattern is found in several Formosan and Philippine languages, in Balinese and Chamorro, in various parts of eastern Indonesia, and in a limited form in some Oceanic languages such as Motu of southeast New Guinea (Blust 1998f).

Ca- reduplication in Proto Austronesian had at least two functions: 1) it marked a derivative series of [+human] numerals *a-esa* 'one', *da-duSa* 'two', *ta-telu* 'three', etc., and 2) it was used to derive deverbal instrumental nouns, as in the above examples from Puyuma and Thao. In addition to these reconstructed functions a similar pattern of reduplication is used to mark other functions in various daughter languages. Examples are given here from Thao, Ngaju Dayak, and Balinese.

Thao (Blust 2003a:190):   1) with verb bases Ca-reduplication creates instrumental nouns: *cpiq* 'thresh, thrash' : ***ca-cpiq*** 'rattan whip', *duruk* 'stab' : ***da-duruk*** 'skewer', *finshiq* 'sow seed in planting' : ***fa-finshiq*** 'seed for planting', 2) with numeral bases it creates numerals used to count [+human] referents, as in *tusha* 'two' : ***ta-tusha*** 'two (of humans)', *turu* 'three' : ***ta-turu*** 'three (of humans)', *shpat* 'four' : ***sha-shpat*** 'four (of humans)', *rima* 'five' : ***ra-rima*** 'five (of humans)', 3) it also forms the durative aspect of dynamic verbs, and the distributive of statives: *m-ishur* 'pry something up' : *ma-a-ishur* 'keep shaking back and forth on its foundations, as a house in an earthquake', *mi-lhilhi* 'stand up' : *mi-**lha**-lhilhi* 'keep standing', *lhulhuk* 'hiccups' : ***lh<m>a**-lhulhuk* 'keep hiccuping', *c<um>piq* 'swat, beat' : ***ca**-c<um>piq* 'keep beating', *ma-diplhaq* 'muddy' : *ma-**da**-diplhaq* 'muddy all over, covered with mud', *mu-luplup* 'scattered, as beads that fall when a necklace breaks' : *pu-**la**-luplup* 'scatter things helter-skelter', 4) with some bases the semantic effect of Ca-reduplication is unpredictable, as with *apu* 'grandparent' : *min-apu* 'become a grandparent' : *min-**a**-apu* 'become a great-grandparent', or *tutu* 'breast' : ***ta**-tutu* 'flat chested, of a woman'. In one known example a Ca- reduplicant is infixed within a non-reduplicative prefix: *pakin-tutuz* 'stack up' : *p<in>a<**ka**>kin-tutuz* 'stacked something up' (with *-in-* marking perfective aspect, and *-ka-* marking repeated action)'.

Ngaju Dayak (Hardeland 1858:67, 1859): 1) Ca- reduplication applies only to consonant-initial bases, to which it adds an attenuative or qualifying sense: *hai* 'large' : ***ha**-hai* 'rather large', *gila* 'unwise' : ***ga**-gila* 'rather unwise', *ka-puti* 'whiteness' : *ba-puti* 'white' : *ba-**pa**-puti* 'whitish', *ka-hijaw* 'greenness' : *ba-hijaw* 'green' : ***ha**-hijaw* 'greenish', *henda* 'turmeric' : *ba-henda* 'yellow' : ***ha**-henda/henda-henda* 'yellowish', *ma-nipis* 'thin, of materials' : *ma-**na**-nipis/ma-nipi-nipis* 'rather thin'. The last two examples show Ca- reduplication and full reduplication (minus the coda) as alternative strategies for adding an attenuative modification to the meaning of the base. However, the data in Hardeland (1859) suggest that this is not always possible. Given the stative prefix *ba-*, Ca-reduplication with *b*-initial stative bases would result in *ba-ba-bV*. This apparently is avoided by using only full reduplication minus the coda in these cases: *behat* 'weight' : ***ba**-behat* 'heavy' : *beha-behat* 'rather heavy' (**ba-ba-behat), *ka-bilem* 'blackness' : ***ba**-bilem* 'black' : *bile-bilem* 'blackish' (**ba-ba-bilem).

Balinese (Barber 1979): 1) Ca-reduplication is used to derive both instrumental and non-instrumental nouns. Many simple and reduplicated bases evidently are synonymous, suggesting that the morphological function of this derivational process is being lost: *bəsah* 'wash the hands' : ***ba**-bəsah-an* 'things to be washed', *bisik* 'to whisper' : ***ba**-bisik-an* 'whispering', *cunduk* 'put something on the head' : ***ca**-cunduk* 'a flower put in the hair', *gitik/**ga**-gitik* 'rod, stick, staff', *gotol/**ga**-gotol* 'the cock of a gun', *ŋəbat* 'spread something out' : ***ka**-kəbat* 'sago leaf plate', *ŋili* 'clean wax from the ear' : ***ka**-kili* 'earpick', *kuyaŋ/**ka**-kuyaŋ* 'shroud, the cloth wrapping a corpse', *ləkas* 'begin' : ***la**-ləkas-an* 'beginning, origin', *lintah/**la**-lintah* 'leech', *pirit* 'roll something up' : ***pa**-pirit-an* 'a rolled-up object; cigarette', *sulit/**sa**-sulit* 'toothpick', *təkən/**ta**-təkən* 'stick, staff'.

Several languages also show a pattern in which *Ca*- reduplication has become entirely fossilised. This has been noted in Amis with many names of flora and fauna, as ***cacido?*** 'dragonfly', ***cacopi*** 'maggot', ***dadipis*** 'cockroach', ***fafikfik*** 'gecko', ***fafokod*** 'large grasshopper', ***kakitiw*** 'yellow jacket (wasp.)', ***kakonah*** 'ant', or ***rarikah*** 'scorpion' (Blust 1999c), and it is also true of several colour terms in Tanimbar-Kei of the southeast Moluccas, as *ŋaŋiar* 'white', ***babul*** 'red', ***tatom*** 'yellow', ***babir*** 'green', and possibly *ŋametan* 'black' (< *ma-qitem), where the unexpected initial consonant may have been reshaped due to close conceptual association with *ŋaŋiar* 'white' (Blust 2001c:27).

A variant of Ca- reduplication is CaC- reduplication, which Bowden (2001) reports for Taba of southern Halmahera. In this pattern, also used to derive instrumental nouns, a heavy syllable CVC- is copied from the base, but with the fixed vowel *a* in the reduplicant, as in *bulay* 'to wind, coil something' : ***bal-bulay*** 'device for winding rope, cord onto', *tek* 'scoop up water' : ***tak-tek*** 'water scoop', *lewit* 'carry on shoulder pole' : ***law-lewit*** 'carry on shoulder pole'. In some cases the final consonant of the reduplicant assimilates to the first consonant of the base, further reducing the phonemic similarity between base and reduplicant: *pit* 'catch with a noose trap' : ***pap-pit*** 'noose trap', *tubal* 'prod with vertical motion' : ***tat-tubal*** 'fruiting pole'.

6.9.5.13 Extensions of fixed segmentism

Ca- reduplication is by far the most common pattern of reduplication with fixed segmentism in AN languages. In the general linguistics literature fixed segmentism shows little typological variation: almost always it is a form of CV- reduplication in which the vowel is prespecified. Although its evolution in some languages, as Bolaang Mongondow or Sangir raises thorny theoretical questions, Ca- reduplication thus conforms to a well-attested pattern. Since some formula is needed to distinguish a CV- pattern in which the vowel is copied from one in which the vowel is prespecified, the latter can be signaled by Cv-. However, a fuller consideration of fixed segmentism in AN languages suggests that Ca- reduplication and similar Cv- patterns in other language families represent a minimal pattern of prespecification. Other, less common patterns of fixed segmentism depart from the usual typological configurations in introducing multiple prespecified segments. As it happens, all such extended patterns of fixed segmentism discovered to date in AN languages are found in the Philippines, where Ca- reduplication itself is attested only marginally in [+human] numerals.

To take the minimal extension of Cv- reduplication first, Botolan Sambal of west-central Luzon forms nonpersonal noun plurals by reduplicating the first stem consonant plus the fixed sequence *-aw-* (hence Caw- reduplication): *lapis* 'pencil' : ***law-lapis*** 'pencils', *dowih* 'thorn' : ***daw-dowih*** 'thorns', *anak* 'child' : ***aw-anak*** 'children', *otan* 'snake' : ***aw-otan*** 'snakes' (Antworth 1979:9). Whether this should be considered a variant of Ca- reduplication or not is moot. In any case it exemplifies the canonical formula Cvc- (C is copied, vc- are prespecified). Caw- reduplication in Botolan Sambal departs from the typologically well-attested pattern of Cv- reduplication found in many language families, but is still a form of syllable reduplication.

Other languages of both the northern and the central Philippines show patterns of fixed segmentism that depart more radically from the minimal Cv- type. Healey (1960:6ff) has described a pattern of Cala- reduplication marking diminutive nouns and related ideas in Agta of northern Luzon. Before vowel-initial bases the reduplicant is *ala?-* (the glottal stop is automatic between vowels), while before consonant-initial bases it is *Cala-*. Expressed formulaically it is thus a form of Cvcv- reduplication: *abbiŋ* 'child' : ***ala?-abbiŋ*** 'a little child', *báhuy* 'pig' : ***bala-báhuy*** 'a little pig', *pirák* 'money' : ***pala-pirák*** 'a little money', *talobag* 'beetle' : ***tala-talobag*** 'lady-bird', *wer* 'creek' : ***wala-wer*** 'small creek'; *assaŋ* 'small' : ***ala?-assaŋ*** 'very small', *kwá-k* 'mine' : ***kwala-kwá-k*** 'my small thing', *mag-poray* 'angry' : ***mag-pala-poray*** 'a bit angry', *mag-simul* 'take a mouthful' : ***mag-sala-simul*** 'take a nibble'. As noted earlier, Viray (1973) documented a number of similar cases in Bikol and the Bisayan languages of the central Philippines, as with (his orthography): Bikol *kawatan* 'toy' : ***karo-kawatan*** 'small toy', *kandiŋ* 'goat' : ***karo-kandiŋ*** 'small goat', or Hiligaynon *panday* 'carpenter' : ***polo-panday*** 'a carpenter of little experience', *sakayan*

'ship' : ***solo**-sakayan* 'small ship', *maestro* 'teacher' : ***molo**-maestro* 'little teacher', *gantaŋ* 'measuring vessel equal to three liters' : ***golo**-gantaŋ* 'small gantang or object like a gantang', *silhíg* 'broom' : ***solo**-silhíg* 'small broom or object like a broom.' Viray viewed these as instances of CV-infixation, but there are problems with his interpretation, and the agreement of this reduplicative pattern in Agta and languages of the central Philippines can be taken as evidence that a pattern of Cvcv- diminutive reduplication may already have existed in Proto Philippines.[66]

6.9.5.14. Reduplicative infixes

Several examples of infixal reduplication in AN languages have already been discussed. Reduplicative infixes, like infixes in general, are rarer than prefixes or suffixes. There are three known types: 1) single consonants, 2) CV or Cv syllables, or 3) VC syllables. Examples of the first type are seen in the West Tarangan dialects reported by Nivens (1993) and further analyzed by Spaelti (1997:134), and in infixal gemination as seen in Batad Ifugaw or Ilokano. Examples of the second type are seen in Thao of central Taiwan or Xârâcùù of south-central New Caledonia:

Thao (Blust 2003a). Thao infixes both CV and Ca syllables, although the former appears to be rare: *i.ti.za* : *i.<**ti**>.ti.za* 'arrive; return, come back', *pash.ʔu.zu* 'phlegm' : *mash.<**ʔa**>.ʔu.zu.-ʔu.zu* 'cough repeatedly' (Ca copying occurs within a base), *tutuz* : *pa.kin-tu.tuz* 'stack, pile up' : *p<i.n>a.<**ka**>.kin.-tu.tu.z-in* 'was stacked up by someone' (Ca copying occurs within the prefix *pakin-*). In view of a CV- copying pattern with consonant-initial forms such as *lu.ish* 'short, low' : *mak.-**lu**-.lu.ish* 'short of breath', or *ma.-lhi.lhnit* 'smiling' : *pi.a.-**lhi**-.lhi.lhnit* 'put on a smile for someone' the affix in *i.<**ti**>.ti.za* probably is best considered an infixal alloduple of CV- reduplication.

Xârâcùù (Moyse-Faurie 1995:180) describes what she calls "reduplication of the second syllable" in Xârâcùù of south-central New Caledonia, but the insertion algorithm is unclear. If it requires insertion after the first vowel all cases are infixal; if it calls for insertion after the second vowel it produces -CV reduplication with disyllables, but -CV- reduplication with trisyllables. I will arbitrarily assume uniform infixation here: *pù.tù* 'put together' : *pù.<**tù**>.tù* 'at the same level', *xwâ.sé* 'numerous' : *xwâ.<**sé**>.sé* 'very numerous', *a.tî.rî* 'count on' : *a.tî.<**tî**>.rî* 'have confidence in', *ji.ki.è* 'rich' : *ji.<**ki**>.ki.è* 'very rich', *xu.tu.è* 'a long time' : *xu.<**tu**>.tu.è* 'a very long time'. Elbert (1988:204) describes a similar pattern for Rennellese, and assumes without argument that the infixal reduplicant is inserted after the second vowel: *ka.i.ti.ʔi* : *ka.i.<**ti**>.ti.ʔi* 'beg silently', *ma.sa.ki* : *ma.sa.<**sa**>.ki* 'sick, weak', *ma.ta.ku* : *ma.ta.<**ta**>.ku* 'fear'.

In addition, Crowley (1998:143) gives three examples of apparent infixal reduplication with vowel changes in Sye of southern Vanuatu: *e.lwo* 'vomit' : *e.l<**i.l**>wo* 'produce disgusting bodily exusion' (exudation?), *e.tur* 'stand' : *e.t<**e.t**>ur* 'stand in large numbers', and *o.rut* '(of tuber) past its prime' : *o.r<**u.r**>.ut* '(fruit) dried in the sun'. The first and last of these agree in copying V_1C_1, raising the vowel and inserting the resulting string immediately before V_2. If the second example followed this pattern it would be **et<**it**>ur, and in view of the probable low frequency of such reduplicated forms in field materials it is possible that *et<et>ur* is a transcriptional error.

---

[66]   The phonemic shape of this pattern cannot be retrieved from available reflexes. Bikol points to *Caru- and most eastern Bisayan languages to *Ca(lr)u- or *Cu(lr)u-, while Agta indicates *Cala- or possibly *Cara-. Even in the absence of reconstruction, however, the close phonological and functional similarity of these patterns manifesting a rare type of reduplication strongly suggests a historical connection.

6.9.5.15 Suffixal syllable reduplication

CV- reduplication can be distinguished from CVC- reduplication (since the former is a syllable in Base-1, while the latter often is not), but both -CV and -CVC reduplication are forms of suffixal syllable reduplication, and so can be treated together. Although historical examples are visible in derivations such as PMP *bekaŋ 'unbent' > Tagalog *bikaŋkáŋ* 'open (forced) at one end' (Blust 1970a), or Proto Philippines *bujak > Botolan Sambal *bolaklak*, Casiguran Dumagat *bulaklak* 'flower' suffixal syllable reduplication is rare both diachronically and synchronically. Known cases include the following from Chamorro and Yapese of western Micronesia, Manam of northeast New Guinea, and Sye of southern Vanuatu:

Chamorro (Topping 1973:183): *ña.laŋ* 'hungry' : *ñá.la.-laŋ* 'very hungry', *dán.ko.lo* 'big' : *dán.ko.lo.-lo* 'very big', *met.got* 'strong' : *mét.go.-got* 'very strong', *bu.ni.ta* 'pretty' : *bu.ní.ta.-ta* 'very pretty' (with loss of coda in the suffixed base).

Yapese (Jensen 1977a:110-111): *qa.thib* : *qa.thib.-thib* 'sweet', *qa.thuk-* 'to mix' : *ma-q.thuk.-thuk* 'mixed up'.

Manam (Lichtenberk 1983:600-602) : *ra.go.go* 'be warm' : *ra.go.go.-go* 'warm', *ʔo.ʔo* 'be plentiful' : *ʔo.ʔo.-ʔo* 'many, much', *re.re* : *re.re.-re* 'like', *ma.la.boŋ* 'flying fox (generic)' : *ma.la.bom.-boŋ* 'k.o. flying fox', *ʔ.ulan* 'desire' : *ʔu.lan.-laŋ* 'desirable'.

Sye (Crowley 1998:143) : *a.cum.su* 'black' : *a.cum.su.-su* 'pitch black', *o.u.rup* 'lob' : *o.u.ruv.-rup* 'lob', *o.wa.top* 'walk to meet someone coming the other way' : *o.wa.tov.-top* 'walk stealthily in search of prey'

Although the pattern in Manam can be analyzed as suffixal syllable reduplication today, it may have originated as suffixal foot reduplication, since final vowels were lost after nasals, and earlier forms such as *ragogo-gogo may have reduced by haplology to avoid a disfavored sequence of four consecutive identical syllables.

6.9.5.16 Other patterns of reduplication

While theoreticians have spared no effort to find principled limits on the form that reduplication can take, new languages continue to present a seeming endless collection of novel patterns. Examples from AN languages that cannot easily be subsumed under the above types include the following:

Vacuous reduplication

Crowley (1982:49) uses this term to describe a Paamese reduplication process that is undone by morpheme structure constraints, specifically those that prevent copying from producing a sequence of three or more vowels. Since there is an independently required process of 'homorganic vowel deletion' seen in examples such as *sii + itee > siitee* 'juice of it', or *mee + ene > meene* 'urinating', reduplication of VV- or -VV would reduce to the same sequence, leaving no trace of a reduplication process, which would then be detectable only from the semantics of the zero-derived word.

Full reduplication minus the initial vowel

Lynch (2000:82ff) notes that Anejom of southern Vanuatu truncates some vowel-initial bases when these are fully reduplicated: *aces* 'bite' : *ces-ces* 'taste, nibble', *aged* 'write' : *ged-ged* 'scribble, scrawl', *ahedej* 'whistle shrilly' : *hedej-hedej* 'whistle continuously', *ahen* 'roast' : *hen-hen* 'warm up'. This process is unpredictable, since other bases permit the same vowel in reduplicated forms: *acal* 'be crooked' : *acal-acal* 'twist', *adiat* 'be

daylight' : *adiat-adiat* 'be midday, bright and sunny', *avak* 'bend down' : *avak-avak* 'walk in bent position'.

Full reduplication plus an initial glide

Lee (1975:216-217) points out that when Kosraean VC monosyllabic bases are reduplicated 'the glide *y* appears before the second syllable in some words.' Examples include *af* 'rain' : *af-yaf* 'rainy', *ef* 'fade' : *ef-yef* 'faded', *eŋ* 'wind' : *eŋ-yeŋ* 'windy', and *ek* 'change' : *ek-yek* 'keep on changing'. He notes that there is no synchronic basis for predicting when glides will be inserted, since words that begin with the same vowel, and even homophones may differ in this feature: *an* 'lie down' : *an-an* 'sit uncomfortably', *ek* 'rub' : *ek-ek* 'rub repeatedly'. All examples that he gives are in monosyllables that begin with *a* or *e*. This pattern is of general interest in that it suggests that the reduplicant may be longer than the base.

Partial reduplication minus initial glottal stop

Elbert (1988:203) has noted a pattern of reduplication in Rennellese, where bases with an initial glottal stop reduplicate the first vowel but omit the first consonant, as with *ʔaga* : *a-ʔaga* 'to wake up', *ʔagoha* : *a-ʔagoha* 'to pity', or *ʔaua* : *a-ʔaua* 'to float'. All examples that he gives are formed from bases that begin with *ʔa-*. Oddly, he also notes a pattern in which glottal stop is added to reduplicated vowel-initial bases, as with *aku* : *a-ʔaku* 'mine (a-class)', *o-ʔoku* 'mine (o-class)', *ana* : *a-ʔana* 'his, her (a-class)' or *ona* : *o-ʔona* 'his, her (o-class)'.

True CV- reduplication: Pangasinan, spoken in north-central Luzon, presents an interesting twist on the usual pattern of CV- reduplication (Benton 1971:99). In most languages bases with an initial vowel copy only the vowel in the surface realisation of a CV- copying algorithm, as with Bunun ma-a-asik or ma-u-uktic (forms such as Tagalog *i-iyák* contain an automatic glottal stop both word-initially and intervocalically, and it can therefore be argued that the copied portion here is CV-). However, in Pangasinan CV-reduplication copies the first CV- both in consonant-initial and in vowel-initial bases. In the first case the copying pattern is prefixal, while in the latter it is infixal: kúya 'older brother or man of same generation' : ku-kúya 'older brothers or men of same generation', but amígo '(male) friend' : a<mi>mígo '(male) friends'. Since this reduplication pattern can be stated alternatively as 'copy C + stressed V' it could be argued that the template for CV reduplication in Pangasinan includes both segmental and prosodic information.

This case raises questions about what a CV- template means, since it may produce more than one surface copying pattern. In effect, CV- reduplication as it is normally understood allows C to be null, while this is not possible in Pangasinan. What is usually called CV-reduplication is thus perhaps more appropriately called (C)V- reduplication, reserving CV-reduplication for the rarer copying pattern seen in Pangasinan. The same provision applies to e.g. Ca- reduplication and CVC- reduplication, although to date no examples of 'true Ca- reduplication' or 'true CVC- reduplication' have been reported. If such existed the former would be realised as e.g. *itin* : *i<ta>tin* (not **a-itin), and the latter as e.g. *alutap* : *a<lut>lutap* (not **al-alutap).

Manam rightward trisyllabic reduplication

Lichtenberk (1983:600) describes a pattern of suffixal reduplication in Manam that is 'restricted to sources whose last three syllables are Coa(C)V#, i.e., the antepenultimate

vowel must be *o*, and the penultimate vowel must be *a'*: *goaza* 'be clean' : *goaza-goaza* 'clean (sg.)', *boadu* 'suffice, be possible' : *boadu-boadu* 'powerful', *raboaʔa* 'Alstonia tree' : *raboaʔa-boaʔa* 'plumeria'. This is an odd restriction for a phonological process, and the explanation almost certainly lies in the phonetics of the sequence Coa(C)V. Elsewhere (1983:21ff) Lichtenberk notes that when *o* is unstressed in this sequence the result is Cwa(C)V, as in *damoa* > [dámwa] 'forehead'. He recognises five consonant orders for Manam (bilabial, dental, velar, uvular/glottal), of which the last two are dialectally equivalent, but he does not include a labiovelar series. Other writers, such as Ross (1988:128), however, recognise both *b^w* and *m^w* as unit phonemes. Since Manam has a process of suffixal foot reduplication the recognition of labiovelars eliminates the need for this canonically peculiar statement of 'rightward trisyllabic reduplication'.

Double reduplication

Sohn (1975:103) has described a pattern of reduplicative prefix formation for Woleaian that can be called 'double reduplication', seen in forms such as *shal* 'water' : **che-chal** 'to water' *raŋ* 'yellow powder' : **che-chaŋ** 'apply yellow powder', or *liuwanee(-y)* 'think (it)' : **niu-n**iuwan 'to think'.

According to Sohn (1975:103) 'The doubling of *l*, *sh*, *r*, *g*, and *b* results in a change in the quality of the respective consonants, as in *n*, *ch*, *ch*, *k* and *bb* (stop sound *bb* in contrast with the fricative *b*).' Two unrelated processes that must be mentioned to account for the forms of these reduplicated words are coda deletion, a process already familiar from other languages, and low vowel dissimilation, a process whereby the first of two low vowels in successive syllables raises (Lynch 2003a). Since (diachronic) consonant gemination is one form of reduplication that exists independently in Woleaian, as in *bug(-a)* 'boil (it)' : *bbug* 'boiled', forms such as *chechal* appear to result from two historically successive but linked reduplicative processes: 1) initial consonant doubling (which may have begun as CV- reduplication followed by syncope), and 2) reduplication of the first foot of the base (*iu* is a high central rounded vowel). The reduplicants in these three words appear to be *che-ch*, *che-ch* and *niu-n*, but they contain an internal consonant doubling with morphological significance. Since the same process affects both halves of the reduplicated word it seems clear that at least historically consonant doubling preceded copying of the initial foot.

From one point of view Woleaian double reduplication is simply prefixal syllable reduplication. From another it is single consonant reduplication (gemination). In fact, it appears to be both: in words like **che-chal** or **che-chaŋ** there is a historically double layer of reduplicative morphology, much like the historically double layer of non-reduplicative morphology in Dutch *schoen* 'shoe' (< *schoe + plural -en) : *schoen-en* 'shoes' (*schoe + plural -en + plural -en), or the syntactic doubling in English expressions such as 'all alone' (< all + all + one). How this type of pattern should be analyzed synchronically is another question, but if the reduplicant is identified with both prefixation and gemination it does not correspond to an 'authentic unit of prosody'.

6.10 Triplication

Triplication is a rare morphological device. To date it has been reported only in Thao of central Taiwan (Blust 2001b). Thao words formed by triplication use a single pattern of reduplication iteratively, or two different patterns of reduplication in combination. Examples include *m-apa* 'carry' : *apa-apa-apa-n* [apapápan] 'be carried', *shkash* 'fear' :

makit-shka-shka-shkash 'slowly overwhelmed by a sense of apprehension or foreboding' (with full reduplication applied twice), *tapʔan* 'a patch on clothing' : *t<in>apʔa-pʔa-pʔan* 'was patched repeatedly, of tattered clothing' (with suffixal foot reduplication applied twice and automatic coda reduction), *pashʔuzu* 'phlegm' : *mash<ʔa>ʔuzu-ʔuzu* 'cough repeatedly' (with suffixal foot reduplication + infixal Ca- reduplication), *ma-shimzaw* 'cold' : *muk-sha-sha-shimzaw* 'shiver constantly with chills, as in a malarial attack' (with prefixal Ca-reduplication iteratively), *untal* 'to follow' : *m-un-ta-ta-tal* 'follow someone's actions closely, imitate' (with iterative infixal Ca-reduplication), *zumzum* 'hold in one's mouth' : *za-za-zumzum* 'keep holding in one's mouth' (with prefixal Ca-reduplication applied iteratively).

Serial reduplication

With verb bases Thao triplication adds a layer of intensity. In effect, whatever semantic contribution reduplication makes to the base is magnified or intensified by triplication. A similar process is found with numeral bases, but here there is a clear distinction of function: *tusha* 'two' : *ta-tusha* 'two (of people)' : *ta-ta-tusha* 'two at a time (of people)', *turu* 'three' : *ta-turu* 'three (of people)' : *ta-ta-turu* 'three at a time (of people)', *rima* 'five' : *ra-rima* 'five (of people)' : *ra-ra-rima* 'five at a time (of people)'. Although distributive numerals such as *ta-ta-tusha* are structurally similar to triplicated verbs such as *qa-qa-qucquc* 'bind tightly or securely', there are important derivational differences. Triplicated verbs result from a single process that contributes essentially the same semantic content to the reduplicated and triplicated forms, but distributive numerals are products of two quite distinct processes, the first adding [+human] to the meaning of the base, and the second adding [+distributive] to the entire word. Given these differences the process used to form distributive numerals in Thao is better called 'serial reduplication' (Blust 2001b). It should be noted that both triplication and serial reduplication are distinct from 'double reduplication' in Woleaian, where the synchronic complexity of the pattern is almost certainly due to historically successive accretions of reduplication that apparently performed the same function.

6.11 Compounding

Compounding is rarely mentioned in grammars of Formosan or Philippine languages, where the morphological resources available for word-formation are so extensive that compounding probably would add little derivational capability. The languages of eastern Indonesia and the Pacific, however, typically have a less developed morphology, and compounding appears to be fairly common. Fox (1979:34) states that in Big Nambas of north-central Vanuatu 'There is only one type of compound noun; it consists of a noun followed by a verb stem,' as in *pət m'iel* (head-be red) 'red-head (a type of bird, also a term for 'policeman')', or *nep' kris* (fire-scrape) 'matches'. Crowley (1982:87), writing of Paama in central Vanuatu, notes that compounding is 'a very productive noun-deriving process in Paamese.' Compounds are distinguished from similar sequences of free morphemes that do not form a compound in at least three ways: 1) the parts of a nominal compound are syntactically inseparable, 2) it is not possible to left-dislocate the components of a compound, and 3) the meaning of a compound is often not predictable from the meaning of its parts. Elbert and Pukui (1979:123ff) list a number of noun-noun compounds in Hawaiian, as *huaʔōlelo* 'word' (lit. 'verbal fruit'), verb and noun-verb compounds, as *hanu-ā-puaʔa* 'gasp for breath' (lit. 'breathe like a pig'), and compound

proper names, among others. Some languages of western Indonesia have nominal compounds, as seen in Malay *papan* 'board' + *tulis* 'write' = *papan tulis* 'blackboard', or Toba Batak *ujuŋ* 'end, extremity' + *hosa* 'breath, life' = *ujuk-kosa* 'the end of life, hour of death', but in general compounding appears to be less developed in languages with richer morphologies.

6.12 Morphological change

Although sound change will be treated at some length in Chapter 9, a few remarks on morphological change may be appropriate here. Changes in morphology occur primarily through one of three processes: 1) fossilisation of affixes, 2) fossilisation of word boundaries, and 3) reanalysis of function. The first process affects the distribution of morpheme boundaries in words, and so may give rise to alterations in canonical shape. However, it normally has little effect on general features of structure. The second process has potentially more far-reaching consequences, as it is implicated in the familiar 'typological cycle' leading from isolating to agglutinating to inflectional and back. The last process has similarly important repercussions for the structural properties of language. Only the first of these processes will be discussed here, as the last two can more appropriately be discussed in Chapter 7 'Syntax'.

6.12.1 Fossilisation of affixes

The fossilisation of affixes is known from many language families, but relatively little attention has been devoted to the reasons that morpheme boundaries become lost. In cases like Dutch *schoen* 'shoe' (< *schoe 'shoe' + -en 'plural') loss of the morpheme boundary probably was a frequency effect: since shoes normally occur in pairs and hence are most frequently mentioned in the plural, it is likely that the plural form came to be unmarked. In the case of Spanish plural nouns like *sapato-s* 'shoes' or *arco-s* 'arches' that were borrowed into Tagalog as single morphemes (*sapatos* 'shoe', *alakos* 'arch') the motivation for borrowing the plural rather than the singular form is unclear, but loss of the morpheme boundary followed automatically from failure to borrow the corresponding singular forms. As in other language families, little attention has been paid as to why fossilisation of affixes occurs in AN languages. There are, however, some clues in the data that are worth mentioning here.

Perhaps the most important observation is that some affixes show a greater tendency to fossilise than others. This is particularly true of *ma- 'stative', and of the *qali/kali-prefixes. Fossilisation generally refers to unanalyzability. PMP *ma- has become fossilised in Hawaiian words such as *makaʔu* (PMP *ma-takut) 'fearful, afraid', *malino* (PMP *ma-linaw) 'calm, quiet, as the sea', or *maʔi* 'patient, sick person; sickness, disease; sick, ill, menstruating' (PMP *ma-sakit 'sick, painful') because semantically related forms **kaʔu, **lino, and **aʔi do not occur, and similar cases could be cited from many other languages.[67] Cases like these contrast strikingly with the 'voice' affixes *Si-, *-um-, *-an, *-en, or the perfective/nominaliser *-in- which rarely fossilise, and even then only under special conditions.

The most likely explanation for varying frequencies of fossilisation among affixes is that there were earlier differences in the freedom of bases to occur without the affix in

---

[67]   Pukui and Elbert (1971:125) suggest a morphological connection between *makaʔu* and *kaʔukaʔu* 'slow down, delay, procrastinate, hesitate, linger; inhibited, checked; reluctance', but this is marginal at best.

question. Since the voice-marking affixes commute within a syntactic paradigm bases commonly occur both with and without any one of these, thus ensuring their contrastive status and hence separability: *kaen : *k<um>aen, *k<um>in-aen, *kaen-en, *k<in>aen, *kaen-an, etc. 'eat'). Some affixes, however, appear to have been almost obligatory. In Lun Bawang of northern Sarawak stative verbs are usually offered with the prefix *mə-* in citation forms, much like monomorphemic bases of other types. Some bases can occur with other affixes, as when *-um- 'inchoative' is used with colour terms, but in general there appears to be less affixal commutability with stative verbs than with their dynamic counterparts. Where affixal commutability is low the bond between base and affix must be stronger than in affixed words with a higher rate of commutation, since associations of morphemes and weakening of the boundaries between them develop through repeated co-occurrence. This impression is supported by reflexes of *qali/kali- words, which have often lost the morpheme boundary, probably because affixation was all but obligatory in order to render such words canonically distinctive (Blust 2001d).

Although affixes with high commutability are less likely to become fossilised, there are some circumstances under which these can become difficult to analyze within a synchronic system. In Kayan of central Borneo, for example, the paradigm for 'eat' includes the forms 1) *kuman* 'to eat (active)', 2) *kani* 'to eat (imper.)', 3) *makan* 'to feed (people)', and 4) *pakan* 'to feed (animals)'. It is clear from this fossilised paradigm and from comparative evidence that these forms once contained productive affixes on a base *kan*. Kayan still has productive affixes that reflect *-um- 'actor voice' and *pa-, *paka- 'causative', but their shapes are *-əm-* (obligatorily metathesised to *mə-* in monosyllables, and optionally in polysyllables), and *pə-* (before consonants), *pək-* (before vowels). It might be said, then, that these forms show fossilisation of *shape* rather than complete loss of morpheme boundary, although the recognition of morpheme boundaries in *k<um>an*, *kan-i*, *ma-kan*, and *pa-kan* clearly is problematic. Examples such as these suggest that fossilisation is a matter of degree rather than of unambiguous presence or absence of a morpheme boundary. Neither *-i* 'imperative' nor *ma-* 'active verb' appears to have any counterpart in the productive morphology of Kayan, and so these affixes can be said to show an even greater degree of fossilisation than *-um-* and *pa-*, but a lesser degree than affixes which are found on a historical base that has become synchronically invariant, such as Hawaiian *makaʔu*, *malino* or *maʔi*. It is noteworthy that the weakening of morpheme boundaries in the paradigm for 'eat' began with the reduction of PAN *kaen to a monosyllabic base through regular sound change. Since monosyllabic content morphemes are strongly disfavored in many AN languages the affixed forms of *kan* (which were disyllables) came to stand increasingly as independent free forms, thus preserving the earlier phonemic shapes of some affixes that continue to be productive in Kayan, and stranding other affixes that have since disappeared from the language.

Some patterns of fossilised reduplication have been noted in passing, as the residues of Ca-reduplication in Amis words for flora and fauna, and the colour terms of Tanimbar-Kei. In addition, as noted by Blust (2001c), the dictionary entries for basic colour words in many Oceanic languages appear only in reduplicated form, as with Motu *kakakaka* 'red', *gadokagadoka* 'light green, as young leaves; blue', *laboralabora* 'yellow', *vaiurivaiuri* 'blue', *uriuri* 'brown, colour of Motuan's skin', Kosraean *sroalsroal* 'black', *raŋraŋ* 'yellow', or Fijian *karakarawa* 'blue, green', *dromodromoa* 'yellow, dirty in colour'. Although each of these languages is represented by a dictionary (Lister-Turner and Clark 1930, Lee 1976, Capell 1968) none of these terms is given in simplex form. In other cases, however, a reduplicated colour term corresponds either to an unreduplicated noun which

serves as a transparent source of the longer word, or to a simplex base of apparently identical meaning. The first type of pairing is seen in Pohnpeian *pwe:t* 'lime made from coral' : *pʷetepʷet* 'white; grey hair', *nta* 'blood' : *wei-ta:ta* 'red', *ɔ:ŋ* 'turmeric' : *ɔ:ŋ-ɔ:ŋ* 'yellow', *pe:s* 'ashes' : *pe:se:s* 'grey, greyish', or *mpʷul* 'flame' : *mpʷulapʷul* 'pink'. The second type is seen in Hawaiian *ʔele* 'black' (less common) : *ʔeleʔele* 'black' (the usual term), *keʔo/keʔokeʔo* 'white, clear', *ʔula* 'red, scarlet; brown, as skin of Hawaiians' : *ʔulaʔula* 'bay, as a horse; various snappers; variety of taro with red or purple petioles', *ʔōmaʔo/ʔōmaʔomaʔo* 'green', or *mele/melemele* 'yellow'. Comparative evidence from languages that actively use reduplication with colour terms suggests that such monomorphemic reduplicated colour terms may reflect morphological processes that once added a semantic overlay either of intensity or attenuation to the meaning of the base (really red, reddish, etc.) which in time became 'bleached' through overuse, leaving the reduplicated form as unmarked.

A second type of fossilised reduplication that is found throughout the AN language family is seen in reduplicated monosyllables such as Cebuano Bisayan *budbud* 'wind string, wire, strips, etc. around something', *supsup* 'to suck', or *tiktik* 'tap lightly on a hard surface'. These and cognate forms in many other languages reflect PAN disyllabic bases (*bejbej, *sepsep, *tiktik) which evidently were formed by reduplication at some pre-PAN period. The lexicalised reduplicated monosyllables of AN languages will be treated at greater length in discussing reconstruction and change.

Finally, a few synchronically unanalyzable words have clearly been derived by reduplication from semantically related bases. A striking example is seen in reflexes of PMP *saŋa-saŋa 'starfish', a transparent derivative of *saŋa 'bifurcation, fork of a branch' (the animal being conceived as a radial series of bifurcations), but one that evidently already existed as an independent word at the time the Malayo-Polynesian languages began to diverge from a linguistically united speech community.

7 *Syntax*

7.0 Introduction

Anyone attempting to write a broad description of the syntax of AN languages is at once faced with a double dilemma. First, given the enormous number of languages, their wide geographical distribution, and their varied contacts with languages of other families, it should not be surprising to find that there is great variation in syntactic types. In fact, at the very outset of his landmark reconstruction of 'Uraustronesisch' phonology and lexicon, Dempwolff (1934:13) declared that "These languages have no uniform grammatical structure like the Semitic or Bantu languages, but they possess a common vocabulary that includes many hundreds of lexical forms" (my English translation). Second, probably more than any other component of language, syntax is difficult to describe without reference to a theoretical framework. Because there are so many theories of syntax, any attempt to describe the syntax of AN languages in terms of a given theory will require choices that many scholars will not accept. Moreover, the purpose of a survey volume such as this is not to advocate a theoretical position, but rather to present the old and the new, the good and the bad, the right and the wrong, and so familiarise the reader with what s/he will actually encounter in exploring the published literature on this large and diverse language family. This is how I have approached issues of phonological reconstruction, distant genetic relationship and the like, and a similar approach is adopted in this chapter.

The range of topics to be covered will be constrained by limitations of space, and by the ease with which it is possible to show interconnections among them. As a result some topics on which much theory-bound work has been done will hardly be touched, if at all (e.g. complex sentences, relativisation, extraction, and clitics, to name a few). The topics covered are: 1. voice systems (together with case-marking), 2. word order, 3. negation, 4. possession, 5. word classes, 6. directionals, 7. imperatives, and 8. questions. Before proceeding it will be helpful to briefly review the aims of linguistic typology as these relate to issues in syntax.

7.1 Voice systems

AN languages are perhaps best known to the general linguist for their theoretically incorrigible systems of voice marking, or, as it is often called 'focus marking' in so-called 'Philippine-type' languages. Given this variation in names, a brief aside on terminology may be useful to the nonspecialist.

From the earliest attempts to characterise Philippine-type verb systems in their own terms, it has been recognised that one argument can be marked as having a special relationship to the verb. This relationship has been described in a perplexing variety of ways, some of which may now be of only historical interest. For Tagalog, Schachter and Otanes (1972:69) describe 'focus' as 'the feature of a verbal predicate that determines the semantic relationship between a predicate and its topic.' For the structurally similar

442

Malagasy, however, Keenan (1976:249) describes verbs as occurring in four 'voices', and the specially marked argument as a 'subject', although one with 'certain characteristic topic properties, which makes Malagasy more topic prominent than, for example, English.' Schachter (1976) suggested that in Philippine-type languages such as Tagalog subject properties are distributed over several types of sentence constituents (topic, actor, actor-topic), and concluded that these languages do not have true subjects. Kroeger (1993), on the other hand, has argued that Tagalog (and by implication other Philippine-type languages) has true subjects, and that the retreat from this position by some other scholars was due to a confusion of syntactic and semantic criteria for determining subjecthood. For Ilokano, Rubino (2000:lxi) holds that 'focus' refers to 'formal marking that reflects the privileged syntactic status of the absolutive noun phrase,' thus implying an ergative actancy system. Most recently Ross and Teng (2005) have argued that the 'Philippinist' framework for describing voice in syntactically conservative AN languages obscures important generalisations, and isolates the study of these languages from the wider world of linguistic scholarship through an idiosyncratic use of terms such as 'topic' (= subject) and 'focus' (a discourse notion elsewhere in the world, but not in Philippine linguistics). A preliminary, and certainly incomplete survey of terminology in Blust (2002d), shows the vocabulary used to describe this phenomenon in 67 published works from Adriani's Sangir grammar of 1893 to the papers in Wouk and Ross (2002), with the following breakdown: voice (28), focus (25), case (7), case/topicalisation (1), topicalisation (2), theme (1), verb class (1), recentralisation (1), trigger (1). Some individual scholars have changed their practice over time, as Frank Blake, who used 'case' in 1906, 'theme' in 1925, and 'case' again in 1930, Howard McKaughan, who used 'voice' in 1958, but 'case/topicalisation' in 1970, or Paul Kroeger, who used 'focus' in 1988, but 'voice' in 1993. Other writers have used both terms, as Sneddon (1970), who labels sentences in Tondano of northern Sulawesi as illustrating 'actor focus', 'object focus' and the like, but in a footnote describes the affixes that mark these grammatical roles as 'voice affixes.' Historical trends in usage are somewhat difficult to discern, but in publications since 1990 'voice' has been used in at least 12 publications and 'focus' in at least nine (or 10, if we add Rubino (2000), which was not included in the earlier count). This vacillation in terminology is principally due to the difficulty of matching Philippine-type verb systems with those of more familiar languages elsewhere in the world, a situation that Himmelmann (1991) has called 'the Philippine challenge to Universal Grammar'. The term 'voice' raised objections on several grounds, only one of which was the perceived anomaly of having three 'passives'. The term 'focus', however, has raised other kinds of objections, most notably that this term has been preempted in general linguistic discourse for a different purpose, and the use of a single term for two very distinct phenomena can only invite confusion.[68]

   Although there are indications that the preferred term is now 'voice' (Shibatani 1988, Mithun 1994, Wouk and Ross 2002, Arka and Ross 2005), in some ways the term 'focus' is more convenient than 'voice', since it uniquely identifies languages that otherwise must be called by the longer and more cumbersome term 'Philippine-type languages', while 'voice' obviously does not. For this reason the terms 'voice' and 'focus' are used more-or-less interchangeably in this chapter, reflecting the variable usage in the literature. Klamer (2002:374) points out that a number of languages in the Lesser Sundas and possibly southern Sulawesi should be regarded as lacking true passive constructions, and the same

---

[68]   In relation to the generalist literature, Matthews (1997) defines 'focus' as 'An element or part of a sentence given prominence by intonational or other means,' generally where there is contrast or emphasis, or a distinction of new vs. given.

is true of most OC languages outside Polynesia (Lynch, Ross and Crowley 2002). In the absence of an active/passive contrast it can be argued that these languages lack voice systems altogether. However, many other non-Philippine-type languages have voice systems, and the term 'focus' must therefore be restricted to those languages that allow more than one type of 'passive' construction. Wolff (1973) showed convincingly that PAN must be reconstructed with what he described as a four-voice system. The morphological expression of this system has been discussed in Chapter 6, but will be recapitulated here. Table 7.1 presents Wolff's reconstruction of the PAN voice system, based primarily on evidence from Atayal of northern Taiwan and Samar-Leyte Bisayan of the Central Philippines, with supporting evidence from Tsou of south-central Taiwan, and Javanese (AV = active voice, DP = direct passive, LP = local passive, IP = instrumental passive; FGA = future-general action):

Table 7.1 The Proto Austronesian voice system

	Independent		FGA	Dependent	Subjunctive
	Non-past	Past			
AV	-um-	-inum-	?	Ø	-a
DP	-en	-in-	r- -en	-a	?
LP	-an	-in-an	r- -an	-i	-ay
IP	i-	i- -in-(?)	?	-an(?)	?

Wolff (1973:79) cited evidence from both Formosan and Philippine languages that instrumental and benefactive voices were marked the same in PAN. This raises the classic lexicographic issue of homophony vs. polysemy, and it seems likely that the two usages were complementary realisations of a single affix which had an instrumental reading when the focused argument was inanimate, and a benefactive meaning when it was animate. What Wolff proposed, then, was essentially a four-voice system, with past forms (usually called 'perfective' in other sources) marked by *-in-, which functioned as a portmanteau affix in the DP, marking direct passive and perfective as an inseparable union. As seen in the preceding chapter (6.5.), this peculiarity of the voice morphology has persisted in some languages that have reduced the original system to a simple active-passive contrast. Some cells in Table 7.1 could not be filled at all, and others were filled tentatively. In addition, it is now known that the prefix for the IP was *Si-, and that this affix shows irregular *S > Ø (for expected **h) in Tagalog and many other Central Philippine languages.[69] Table 7.2 shows the affix potential and syntax that can be inferred for PAN *kaen 'to eat' in the independent nonpast form (past forms in Wolff's system would be: AV *k<in><um>aen, DP *k<in>aen, LP *k<in>aen-an, IP *Si-k<in>aen, but as noted in Chapter 6, many languages reflect *k<um><in>aen in the AV). The DP differs from other voices in having a zero allomorph in the perfective. Noun phrase markers are only partially reconstructed,

---

[69]   Tausug *hi-* "Indicates that something (semantically a patient) that is being conveyed or conceived as being conveyed is in grammatical focus" (Hassan, Ashley and Ashley 1994:173). Together with complex affixes such as Samar-Leyte Bisayan *mahi-* 'contingent perfective of the durative instrumental voice' (Zorc 1977:118), this suggests PAN *Si- > PMP *hi-, with convergent loss of *h in Central Philippine languages, an inference that is also supported by PAN *Sika- > Tagalog *ika-*, Tausug *hika-* 'ordinal prefix'. Caution is needed, however, since the simple form of the IF throughout the central Philippines is *qi-*. Moreover, it is not clear that the 'conveyance voice' of Tausug reflects *Si- 'IP', since *s sometimes became Tausug *h* (*sa > ha* 'locative NP marker', *sa-ŋa-puluq > haŋ-puʔ '10'), and this voice may thus derive from *si-.

and those that are unknown are represented here by F (= focused) and NF (non-focused), which stand for morphemes that cannot be specified phonologically. Each voice/focus is illustrated with a common noun phrase, a personal noun phrase, and a pronoun in both the singular and the plural. Adan is a hypothesised personal name. Wolff's terminology has been left unchanged:

Table 7.2 Reconstructed partial voice paradigm for PAN *kaen 'to eat'

AV: a) k<um>aen   Semay   Cau
              eat-**av**   NF rice   **F** man
              'The man is eating some rice'

         b) k<um>aen   Semay   si Adan
              eat-**av**   NF rice   **F** Adan
              'Adan is eating some rice'

         c) k<um>aen   Semay   si-á
              eat-**av**   NF rice   **F-3sg**
              'He is eating some rice'

         d) k<um>aen   Semay   si-dá
              eat-**av**   NF rice   **F-3p**
              'They are eating some rice'

DP: a) kaen-en   nu   Cau   Semay
              eat-**dp**   **gen**   man   **F**-rice
              'A/the man is eating the rice'

         b) kaen-en   ni adan   Semay
              eat-**dp**   **gen** Adan   **F**-rice
              'Adan is eating the rice'

         c) kaen-en   ni-á   Semay
              eat-**dp**   **gen-3sg**   **F**-rice
              'He is eating the rice'

         d) kaen-en   na-ida   Semay
              eat-**dp**   **gen-3pl** **F**-rice
              'They are eating the rice'

LP: a) kaen-an   nu Cau   Semay   Rumaq
              eat-**lp**   **gen** man   **NF**-rice   **F**-house
              'The man is eating rice in the house'

         b) kaen-an   ni adan   Semay   Rumaq
              eat-**lp**   **gen** Adan   **NF**-rice   **F**-house
              'Adan is eating rice in the house'

   c)   kaen-an   ni-á   Semay   Rumaq
       eat-**lp**   **gen-3sg**   **NF**-rice   **F**-house
       'He is eating rice in the house'

   d)   kaen-an   Na-ida   Semay   Rumaq
       eat-**lp**   **gen-3p**   **NF**-rice   **F**-house
       'They are eating rice in the house'

IP   a)   Si-kaen   nu Cau   Semay   lima-ni-á
       eat-**ip**   **gen** man   **NF**-rice   **F**-hand-**gen-3sg**
       'The man is eating rice with his hand'

   b)   Si-kaen   ni adan   Semay   lima-ni-á
       eat-**ip**   **gen** Adan   **NF**-rice   **F**-hand-**gen-3sg**
       'Adan is eating rice with his hand'

   c)   Si-kaen   ni-á   Semay   lima-ni-á
       eat-**ip**   **gen-3sg**   **NF**-rice   **F**-hand-**gen-3sg**
       'He is eating rice with his hand'

   d)   Si-kaen   na-ida   Semay   lima-na-ida
       eat-**ip**   **gen-3p**   **NF**-rice   **F**-hand-**gen-3p**
       'They are eating rice with their hands'

Verb systems that derive from structures very similar to this reconstruction are found in some Formosan languages, in nearly all languages of the Philippines, northern Borneo and northern Sulawesi, and in Malagasy and Chamorro. A sample of these systems can be taken to illustrate what is meant by 'Philippine-type language'. av = actor voice, pv = patient voice (also called 'object focus' or 'goal focus'), lv = locative voice (also called 'referent voice/focus'), iv = instrument voice, bv = benefactive voice (sometimes lumped together as 'conveyance voice'. All grammatical terms are bolded; usage varies somewhat with the source:

Mayrinax Atayal (northern Taiwan). The following examples are taken from Huang (2001), who uses the term 'focus', and describes the relationship between affixed verb and marked NP as "a kind of agreement system between the subject (i.e. the focused noun phrase) and the verb, though showing no person, gender, or number agreement between them." In earlier publications on other dialects of Atayal, as Huang (1993) she described the same relationship as a voice system, with *m-/-um-* = active voice, *-un* = culminitative voice, *-an* = transversal voice, and *s-* = circumstantial voice). Huang (2001:61) points out that in addition to the locative role, "the focused argument in an LF construction can be a recipient…a goal…or a source."

Table 7.3 The focus/voice possibilities of Mayrinax Atayal

AV: a)   h\<um>akay   kuʔ   ʔulaqiʔ
       walk-**av**   **nom.ref**   child
       'The child is walking/the child walked'

b) t<um>aquʔ   ckuʔ   nabakis   kuʔ   ʔulaqiʔ
    push:down-**av** **acc.ref** old:man **nom.ref** child
    'The child pushed the old man down'

PV: a) tutiŋ-un=mu   kuʔ   xuil
    beat-**pv**=**1sg.gen**   **nom.ref** dog
    'I beat the dog'

LV: a) qilap-an niʔ   yayaʔ   kuʔ   paɣaʔ=suʔ
    sleep-**lv** **gen** mother **nom.ref** bed=**2sg.gen**
    'Mother slept on your bed'

IV: a) si-culh=miʔ[70]   cuʔ   siyam   kuʔ   batah
    roast-**iv**=**1sg.gen** **acc.nref** pork   **nom.ref** charcoal
    'I roasted pork with the charcoal'

BV: a) si-cabuʔ cuʔ   qulih nkuʔ   nabakis   ʔiʔ   yumin
    wrap-**bv** **acc.nref** fish   **gen.ref** old:man **nom** Yumin
    'The old man wrapped a fish for Yumin'

Tagalog (central Philippines). The following examples (written phonemically) originally appeared in Foley (1976), and were confirmed for me by Hsiu-chuan Liao, who checked them with seven native speakers of Manila Tagalog. Morpheme glosses reflect what appears to be the most common functional interpretation of the *aŋ* phrase. Acc= accusative, nom = nominative (but called 'topic' by some writers), perf = perfective, gen = genitive. Locative voice covers source, goal, and location. Only a source usage is illustrated here:

Table 7.4 The voice/focus possibilities of Tagalog

AV: a) b<um>ilí   naŋ   kotse aŋ   lalake
    buy-**av**   **gen** car   **nom**   man
    'The man bought a car'

   b) b<um>ilí   naŋ   kotse si   Juan
    buy-**av**   **gen** car   **nom**   John
    'John bought a car'

   c) b<um>ilí   siyá   naŋ   kotse
    buy-**av**   **3sg.nom**   **gen** car
    'He bought a car'

   d) b<um>ilí   silá   naŋ   kotse
    buy-**av**   **3pl.nom**   **gen** car
    'They bought a car'

---

[70] For reasons that are not explained in the original source, both =mu and =miʔ mark the 1sg genitive.

PV a) b\<in\>i-bilí   naŋ lalake aŋ   kotse
      red-pv.perf-buy **gen** man   **nom** car
      'A man is buying the car'

    b) b\<in\>i-bilí   ni   Juan aŋ   kotse
      red-pv.perf-buy **gen** John **nom** car
      'John is buying the car'

    c) b\<in\>i-bilí   niyá   aŋ   kotse
      red-pv.perf-buy **3sg.gen** **nom** car
      'He is buying the car'

    d) b\<in\>i-bilí   nilá   aŋ   kotse
      red-pv.perf-buy **3pl.gen** **nom** car
      'They are buying the car'

PV: a) b\<in\>ilí   naŋ lalake aŋ   kotse
      buy-**pv.perf** **gen** man   **nom** car
      'A man bought the car'

    b) b\<in\>ilí   ni   Juan aŋ   kotse
      buy-**pv.perf** **gen** John **nom** car
      'John bought the car'

    c) b\<in\>ilí   niyá   aŋ   kotse
      buy-**pv.perf** **3sg.gen** **nom** car
      'He bought the car'

    d) b\<in\>ilí   nilá   aŋ   kotse
      buy-**pv.perf** **3pl.gen** **nom** car
      'They bought the car'

LV: a) b\<in\>i-bilh-án   naŋ laláke naŋ   isdáʔ aŋ   bátaʔ
      red-perf-buy-**lv** **gen** man   **gen** fish **nom** child
      'A man is buying fish from the child'

    b) b\<in\>i-bilh-án   ni   Juan naŋ   isdáʔ aŋ   bátaʔ
      red-perf-buy-**lv** **gen** John **gen** fish **nom** child
      'John is buying fish from the child'

    c) b\<in\>i-bilh-án   niyá   naŋ   isdáʔ aŋ   bátaʔ
      red-perf-buy-**lv** **3sg.gen** **gen** fish **nom** child
      'He is buying fish from the child'

d) b\<in\>i-bilh-án   nilá   naŋ isdáʔ aŋ   bátaʔ
   red-perf-buy-**lv** **3pl.gen** **gen** fish **nom** child
   'They are buying fish from the child'

BV: a) i-b\<in\>ilí   naŋ   lalake naŋ isdáʔ aŋ   bátaʔ
    bv-buy-**perf** **gen** man   **gen** fish **nom** child
    'A man bought some fish for the child'

    b)   i-b\<in\>ilí   ni   Juan naŋ isdáʔ aŋ   bátaʔ
       bv-buy-**perf** **gen** John **gen** fish **nom** child
       'John bought some fish for the child'

    c) i-b\<in\>ilí   niyá   naŋ isdáʔ aŋ   bátaʔ
      bv-buy-**perf** **3sg.gen** **gen** fish **nom** child
      'He bought some fish for the child'

    d) i-b\<in\>ilí   nilá   naŋ isdáʔ aŋ   bátaʔ
      bv-buy-**perf** **3pl.gen** **gen** fish **nom** child
      'They bought some fish for the child'

IV: a) (i-)p\<in\>am-bilí naŋ lalake naŋ isdáʔ aŋ   peraʔ
    iv-perf-buy   **gen** man   **gen** fish **nom** money
    'A man bought some fish with the money'

    b) (i-)p\<in\>am-bilí ni   Juan naŋ isdáʔ aŋ   peraʔ
      iv-perf-buy   **gen** John **gen** fish **nom** money
      'John bought some fish with the money'

    c) (i-)p\<in\>am-bilí niyá   naŋ isdáʔ aŋ   peraʔ
      iv-perf-buy   **3sg.gen** **gen** fish **nom** money
      'He bought some fish with the money'

    d) (i-)p\<in\>am-bilí nilá   naŋ isdáʔ aŋ   peraʔ
      iv-perf-buy   **gen.3pl** **gen** fish **nom** money
      'They bought some fish with the money'

Speakers of some provincial varieties of Tagalog are said to distinguish benefactive senses marked by *i-b\<in\>ilí* and *b\<in\>ilh-án*. In the first, a historically general benefactive use has been narrowed to meanings in which the actor performs an action for the recipient to save the latter from doing the work. In the second, the recipient receives the benefit in some tangible form. In a sentence such as 'John bought flowers for his wife', then, the first reading is that John's wife needed to buy flowers for someone else and John did it for her, while in the second reading John's wife keeps the flowers. Although the verb form in such benefactive constructions is homophonous with the locative voice the two senses reportedly are distinguished by the order of the arguments: *b\<in\>i-bilh-án naŋ laláke naŋ isdáʔ aŋ bátaʔ* 'The man is buying fish *from* the child', *b\<in\>i-bilh-án naŋ laláke aŋ bátaʔ naŋ isdáʔ* 'The man is buying fish *for* the child'. In the Manila dialect

b<in>ilí is also used as a benefactive, but this is seen as a colloquial abbreviation of *i-b<in>ilí*. The use of *-án* to mark benefactive is said to be increasingly common in Philippine languages south of Tagalog (Hsiu-chuan Liao, p.c.).

Perhaps the most surprising feature of the Tagalog voice system is the use of *–in-* to mark progressive or imperfective aspect. This is the result of a semantic innovation whereby PAN *-in-, which marked perfective aspect, came to mark inceptive aspect in many Central Philippine languages. However, it has this sense only in combination with CV- reduplication. Without reduplication *–in-* is still contrastively perfective: *b<in>i-bilí naŋ lalake aŋ kotse* 'A/the man is buying the car' vs. *b<in>ilí naŋ lalake aŋ kotse* 'A/the man bought the car', *b<in>i-bilh-án naŋ laláke naŋ isdáʔ aŋ bátaʔ* 'A/the man is buying fish from the child' vs. *b<in>ilh-án naŋ laláke naŋ isdáʔ aŋ bátaʔ* ' A/the man bought fish from the child'.

<u>Malagasy</u> (Madagascar). The following examples come from multiple sources, including Garvey (1964), Keenan (1976), Dahl (1978) and Dahl (1986). The last of these describes a five voice (called 'focus') system (Actor, Object, Referent, Instrument, Circumstantial). The circumstantial voice is formally discontinuous (*i … ana*):

Table 7.5 The voice/focus possibilities of Malagasy

AF: a)   (mi)-t<om>ány   ízy
             av-cry-(**av**)   **3sg**
             '(S)he is crying'

       b)   mi-sótro   ny   dite   áho
             av-drink   **art**   tea   **1sg**
             'I'm drinking the tea'

       c)   ni-sótro   ny   dite   áho
             av.past-drink   **art**   tea   **1sg**
             'I drank the tea'

       d)   ma-nótotra   tány   ny   lávaka   ízy
             av-fill   earth   **art**   pit   **3sg**
             'He is filling the hole with earth'

OF: a)   tehen-ína-ko   ny   lákana
             punt-**ov-1sg**   **art**   canoe
             'I am punting the canoe'

RF: a)   totof-ána   ny   tány   ny   lávaka
             fill-**rv-3sg**   **art**   earth   **art**   pit
             'He is filling the hole with the earth'

IF: a)   a-tápaka   ny   tády   ny   ántsy
             iv-cut   **art**   rope   **art**   knife
             'The knife is used to cut the rope'

CF: a) i-vidi-ána-ko   mófo   ny   ankízy
         cv-buy-**cv**-**1sg** bread **art** child
         'I bought bread for the child'

Tondano (northern Sulawesi). Sneddon (1970:13) provides the data for this language. He refers to the indexing of grammatical roles through verb morphology as a 'topic-voice (focus) system,' and recognises four 'focuses': af = actor focus, pf = patient focus, rf = referent focus, if = instrument focus. T = topic, O = object, In = instrument, R = referent, A = actor:

Table 7.6 The focus/voice possibilities of Tondano

'The man will pull the cart with the rope to the market'

AF: a) si tuama   k<um>eoŋ   roda   wo tali   waki pasar
         man:**T**   **af**-will-pull cart:**O** with rope:**In** to market:**R**

PF: b) roda   keoŋ-en   ni tuama   wo tali   waki pasar
         cart:**T** **pf**-will-pull man:**A**   with rope:**In** to market:**R**

RF: c) pasar   keoŋ-an   ni tuama   roda   wo tali
         market:**T** **rf**-will-pull man:**A**   cart:**O** with rope:**In**

IF: d) tali   i-keoŋ   ni tuama   roda   waki pasar
         rope:**T** **if**-will-pull man:**A**   cart:**O** to market:**R**

Chamorro (western Micronesia). The examples presented in Table 7.7 are taken from Topping (1973:243ff), who describes the nominal argument that is singled out for a special relationship to the verb as the 'theme', and recognises five morphologically marked 'focus' constructions in Chamorro, namely actor focus, goal focus, causative focus, referential focus and benefactive focus constructions (nf = non-focused NP):

Table 7.7 The focus/voice possibilities of Chamorro

AF: a) guahu l<um>iʔeʔ i   palaoʔan
         1sg   see-**af**   **art** woman
         '*I* am the one who saw the woman'

GF: a) l<in>iʔeʔ i   palaoʔan an   ni   lahi
         see-**gf**   **art** woman   **an**   **rel** man
         'The man saw *the woman*'

     b) l<in>iʔeʔ si   Maria as   Pedro
         see-**gf**   **art** Maria **art** Pedro

         'Pedro saw *Maria*'

CF: a) i   maŋga ha   naʔ-malaŋu i   patgon
         art mango **3sg** **cf**-sick   **art** hild
         'The *mango* made the child sick'

RF: a) hu   toʔlaʔ-i   hao
        1sg spit-**rf** **2sg**
        'I spit at/on *you*'

    b) hu   faʔtinas-i hao kafe
        1sg make-**rf** **2sg** coffee
        'I made some coffee for *you*'

BF: a) hu   saŋan-iyi si   Pedro ni estoria
        1sg tell-**bf** **art** Pedro **nf** story
        'I told the story for *Pedro*'

Topping contrasts most of these focus types with what he calls a 'non-focus' equivalent. His interpretation identifies focus with emphasis on the focused NP, and so makes it very similar to (but not identical with) topicalisation. Table 7.8 lists the voice affixes attached to the verb in each of the above languages:

Table 7.8 Voice/focus affixes in five 'Philippine-type' languages

	ATY	TAG	MLG	TND	CHM
AV	-um-	-um-	(-um-)	-um-	-um-
		maŋ-	man, mi-		man-
PV	-un	-in	-ina	-ən	-in-
PV(perf)	-in-	-in-	n(i)-	-in-	-in-
LV	-an	-an	-ana	-an	-i
IV	si-	i-	a-	i-	–
BV	si-	i-	i-…-ana	–	-iyi

The following points are noteworthy in connection with Table 7.8. First, nearly all languages with a Philippine type voice system mark the AV with –um-. Most MP languages of this type contrast AV types reflecting *-um- and *-maŋ-, or *-um-, *maŋ-, and *maR-. This is the result of an apparent innovation in the immediate ancestor of the non-Formosan AN languages. As noted in Chapter 6, in some languages reflexes of *maŋ- have completely or almost completely replaced the old AV infix. Dahl (1986) notes that reflexes of *-um- are found in only a handful of Malagasy morphemes, the AV construction generally being marked by *man-*, or *mi-*. Richardson (1885) gives *t<om>ány* and *mi-t<om>ány* 'to cry', showing that *t<om>ány* was beginning to undergo reanalysis as a monomorphemic base in the late nineteenth century. Keenan (1976:267) gives only the longer form, suggesting that roughly a century later reanalysis was already complete. A similar phasing out of *–um-* is fairly widespread in the central Philippines (Lobel 2004).

Second, nearly all languages that reflect *-en 'patient voice' have a zero allomorph in the perfective aspect, leaving the reflex of *-in- to mark PV and perfective aspect as an inseparable unity.[71]

---

[71]   A few languages have lost the portmanteau function of *-in-: Akamine (2002:360) notes that Sinama (Sama-Bajaw) *–in-* marks goal focus without any aspectual implication, and may even be used in future constructions; Costenoble (1940:383) and Topping (1973:245) make a similar claim for Chamorro *-in-*, although Safford (1909:91) described this infix as a marker of past definite or preterite verb tense.

Third, most languages that have a LV construction represent it by a reflex of *-an. Chamorro is exceptional in using –i, apparently a reflex of the generic locative marker *i 'at, on' which has been cliticised to the preceding verb stem, as in some other languages (e.g. Malay, where the verb base *tanam* 'to plant' can be affixed as *mə-nanam-kan* 'to plant (object)', or *mə-nanam-i* 'to plant (in location) with object'. As noted for Tagalog, the LV may represent location, source, or goal, and this is true of many languages. As already noted in 6.3.3.1, some uses of *-an, however, are difficult to reconcile with any notion of voice, or even location, as in PMP *tahep-an 'winnowing basket' (< *tahep 'to winnow'), PAN *RiNaS-an 'male of Swinhoe's blue pheasant' (< *RiNaS 'long tail feathers'), PAN *waNiS-an 'boar' (< *waNiS 'boar's tusk'), PWMP *bulu-an 'a hairy fruit, the rambutan' (< *bulu 'body hair'), and perhaps PMP *tian-an 'pregnant' (< *tian 'belly').

Finally, the IV/BV is marked the same in most languages. However, Tondano and Chamorro mark benefactive and instrumental case relations respectively with prepositions, and Malagasy distinguishes IV and BV with separate affixes that may reflect PAN *Sa- 'marker of deverbal instrumental nouns', and *Si- 'instrumental voice; marker of deverbal instrumental nouns' (Blust 2003c).

One other aspect of the Tagalog voice system requires discussion, namely the syntactic status of the morphemes *aŋ*, *naŋ*, *si*, *ni*, etc. that introduce noun phrases such as *aŋ lalake*, *naŋ kotse*, *si Juan*, or *ni Juan*. Reid (2002) has addressed this point forcefully, noting that writers on Philippine languages over the past century have used more than 25 different names in an attempt to capture the functions of these elements. Reid himself has vacillated considerably on this point, calling these elements 'construction markers' (Reid 1978), nouns (Reid 2002), and phrase markers (Reid 2006). As he points out (2002:297) "Probably the most common description of these forms is one that identifies them as marking the case of the noun phrase of which they are a part." This is equally true of the literature on Formosan languages (Huang, Zeitoun, Yeh, Chang and Wu 1998). However, under at least one common interpretation of 'case' (where case is identified with thematic role) *aŋ* cannot be a case marker, since its only role is to mark one nominal argument as having a special relationship to the verb, and the details of this relationship are coded entirely in the verbal morphology. Himmelmann (1991:15) has gone so far as to claim that *aŋ* does not mark any morphosyntactic function. He uses the term 'predication base' to denote the predicative relation that holds between the *aŋ* phrase and the predicate, and argues that the element which introduces a noun phrase marks neither predication bases nor topics. Despite these problems, it seems best to adopt the most widely accepted terminology, and so call these syntactic elements 'case markers'.

As seen in the Tagalog examples above, with a pronominal NP in focus the argument is case-marked by pronoun selection. Following earlier work by Blust (1977a), Ross (2002a:36) has reconstructed the following PAN personal pronouns, modifying the earlier proposal almost entirely on the basis of Formosan evidence:

Table 7.9 Proto Austronesian personal pronouns (after Ross 2002a)

	Free	Free polite	PIV, GEN1	GEN2	GEN3
1sg	[i-]aku	–	=ku	maku	n-aku
2sg	[i-]Su	[i-]ka-Su	=Su	miSu	ni-Su
3sg	s(i)-ia	–	(=ia)	–	n(i)-ia
1ex.pl	i-ami	[i-]k-ami	=mi	mami	n(i)-ami
1in.pl	([i])ita	[i-]k-ita	=ta	mita	n-ita
2pl	i-amu	[i-]k-amu	=mu	mamu	n(i)-amu
3pl	si-da	–	(=da)	–	ni-da

This system is revised and expanded by Ross (2006) who posits seven pronominal categories for each person, labeled 'neutral', 'nominative 1', 'nominative 2', 'accusative', 'genitive 1', 'genitive 2' and 'genitive 3'. The full system will not be presented here, but to give an idea of the increase in complexity this revision entails, the following 1sg forms are proposed: 1. NEUT *i-aku, 2. NOM1 *aku, 3. NOM2 *=ku, *=[S]aku, 4. ACC *i-ak-ən, 5. GEN1 *=[a]ku, 6. GEN2 *(=)m-aku, 7. GEN3 *n-aku. Ross (2006:532) notes that "It is immediately obvious that too many sets of pronouns are reconstructed," and he attempts to explain this by subgrouping: if there is more hierarchy in the AN family tree than is usually assumed, it is possible that some of these forms were innovated after the break-up of PAN, but still very early in the history of the language family.

As pointed out in Blust (1977a), the paradigmatic relationship of 'long form' pronouns such as *aku to their 'short form' equivalents (*-ku) is generally predictable, but breaks down with the third person singular (a relationship that emerges more clearly in the original formulation than in the revisions proposed by Ross). To account for this departure from regularity it was shown that the widely accepted *-ña '3sg genitive' must originally have been *ni 'genitive of personal nouns' + *a '3sg'. Contrasting forms such as Tagalog *siyá* '3sg nom.' : *niyá* '3sg gen.' thus reflect *si 'nominative of personal nouns' + *a '3sg', and *ni 'genitive of personal nouns' + *a '3sg', where genitive marks possessor and non-subject agent. Since a similar syntactic complementation is found between cognate nominative and genitive pronouns in many AN languages, it follows that pronominal arguments in PAN were marked for case. In languages that have lost the Philippine-type voice system these elements may be bleached of all syntactic value, as with Malay *si*, which marks personal names and epithets regardless of their syntactic position.

The preceding remarks are mostly concerned with the morphological realisation of voice or 'focus' in Philippine-type languages. However, there are other observations that can be made in connection with the examples in Tables 7.2-7.8. First, it is clear that a system similar in general type to those described here must be reconstructed for PAN. From this it follows that languages cannot be subgrouped purely on the basis of sharing a Philippine-type voice system, since this is a retention from PAN rather than an innovation in some later proto language. Second, there are gaps in the data, as many sources give only partial paradigms, generally with common noun agents in the non-AF voices, but no personal noun or pronominal agents. Third, a number of writers have commented that the 'focused' nominal in Tagalog and some other Philippine-type languages must be definite, although Himmelmann (1991:15) has argued that the determining factor is actually referentiality. Fourth, many Formosan and Philippine languages use case markers to distinguish 'focused' from 'non-focused' NPs. Few of these are cognate, and it has proven difficult to reconstruct the PAN system. Nonetheless, given the presence of case markers in many of the attested systems it appears likely that they were also found in PAN. Ross

(2002a:35, 51) proposed the following systems of what he then called 'phrase markers' for PAN and PMP (SPEC = specific, GEN = genitive, NPIV = non-pivot, LOC = location):

Table 7.10: Early Austronesian phrase markers (after Ross 2002a)

PAN	TOPIC	SPEC	GEN	NPIV
common (present)	*a	*ka	*na	*Ca, *sa
common (absent)	*u	*ku	*nu	*Cu, *su
personal	–	(*i, *ti, *si)	*ni	–

PMP	SPEC	GEN	NPIV	LOC
common (default)	*i	*ni	*si	*di, *i
common (present)	*a, (*sa)	*na	*ta, *sa	*da, *ka, *sa
common (absent)	*u, (*su)	*nu	*tu, *su	*du (?)
personal	*si	*ni	–	*ka[n]i

More recently Ross (2006:529) has revised this reconstruction for PAN, now calling the elements 'case markers', as shown in Table 7.11 (C = common nouns, PS:S = singular personal nouns, PS:P = plural personal nouns):

Table 7.11: Proto Austronesian case markers (after Ross 2006)

	C??	C??	PS:S	PS:P
Neutral	*[y]a	*u	*i	–
Nominative	*k-a	*k-u	*k-i	–
Genitive	*n-a	*n-u	*n-i	*n-i-a
Accusative	*C-a	*C-u	*C-i	–
Oblique	*s-a	*s-u	–	–
Locative	*d-a	–	–	–

Although this is clearly the most complete reconstruction of the system of PAN case markers that has been offered to date, it contains a number of features that are still highly provisional. To cite one prominent problem that appears in both Table 7.10 and Table 7.11, the functions assigned to the genitive markers are difficult to reconcile with the agreement between Amis (Huang 1998:33) *nu* 'genitive of common nouns', *ni* 'genitive of singular personal nouns', *na* 'genitive of plural personal nouns', and West and East Miraya Bikol (Jason Lobel, p.c.) *nu* 'genitive of common nouns, + referential ~ + past', *ni* genitive of singular personal nouns', *na* genitive of plural personal nouns'. Since these languages are generally assigned to different primary branches of the AN family, this agreement in a three-way contrast is most simply explained by positing a similar set of semantic/functional distinctions for PAN (Reid 1978, Blust 2005b).[72]

---

[72]   Reid (2006) has challenged this interpretation, suggesting instead that an incipient system of 'vowel grades' in oblique and locative prepositions was analogically extended to other forms, hence complicating reconstruction, since the process may have led to widespread parallel developments. Nonetheless, in a footnote (fn. 4) he points out that in an earlier publication (Reid 1978) he himself suggested 'The use of *na* as a plural morpheme associated with personal markers … may need to be reconstructed for Proto-Philippines, [because] it is also attested outside of the Philippines as a plural Genitive marker in Amis.'

In Malagasy and some other languages word order has usurped the role performed by case markers in most Formosan languages and languages of the Philippines. Here, the last nominal argument is the one 'in focus'. It is worth noting that although the order of nominal arguments in an NP-marked language such as Tagalog could in principle be free, there reportedly is a strong preference among Tagalog speakers to place the *aŋ* phrase last unless there is a *sa*-phrase marking a locative complement (Hsiu-chuan Liao, p.c.). This agrees with Malagasy in putting the 'focused' NP last, and has led some writers to assume a VOS word-order typology for many Philippine-type languages. A number of writers (e.g. Naylor 1986) have noted that focus or voice selection (what determines the choice of focus/voice by a speaker at a given point in a conversation/narrative) is determined by definiteness, specificity or referentiality, and therefore is discourse-sensitive. Many of the same writers have noted that the 'basic' voice as determined by text frequency appears to be the PV, making it more difficult to reconcile the patient voice/focus of Philippine-type languages with the passive voice of languages such as English in terms of markedness, and for this reason and others some writers prefer to avoid using the term 'passive' in relation to Philippine-type languages.

Another widespread trait connected with the voice systems of Philippine-type languages and some non-Philippine type languages of western Indonesia, is that infinitival complements must be expressed passively, both in Philippine-type languages and in western Indonesian languages, as shown in the following examples from Thao of central Taiwan, Tindal Dusun of Sabah, and Malay:

Thao

(7.1) yaku   a   ma-sas   afu   a   kan-in   ihu
       1sg **fut** **fut**-bring rice **fut** eat-**pv** **2sg**

'I will bring rice to feed you' (lit. 'I will bring rice to be eaten by you')

(7.2) haya   wa   fatu   ma-qitan   tamuku-n,   ma-zaŋqaw
       that **lig** stone good/easy lift-**pv**   **stat**-light

'That stone is easy to lift; it's light' (lit. 'That stone is easy to be lifted; it's light'

Tindal Dusun

(7.3) korot-oʔ   loʔ   manuk   a-kan-on   tokoʔ
       cut.throat-**imp** **def** chicken **a**-eat-**pv** **1pl.in**

'Kill the chicken for us to eat' (lit. 'Kill the chicken to be eaten by us')

Malay

(7.4) buku   itu   susah   di-təmu-kan
       book that hard **pass**-find-**tr**

'That book is hard to find' (lit. 'That book is hard to be found')

Philippine-type languages make up no more than 15-18% of all AN languages, and since they are syntactically conservative it follows that most AN languages have altered an original four-voice system in various ways. Blust (2002e:69ff) characterised the breakdown of this system in Taiwan and western Indonesia in terms of a progressive loss of morphologically marked voice distinctions. Some languages that still are 'Philippine-

type' have reduced the original system to three-voices, as in Kavalan of eastern Taiwan, where AV = *-um-*, *mə-*, PV = *-an*, and B/IV = *tə-* (Y.L. Chang 1997:35ff). Similarly, in Thao the IF/BF has been replaced by prepositions, and the system now functions with just three voice possibilities marked by verbal affixation, although the semantics of patient voice (marked by *-in*) and locative voice (marked by *–an*) are sometimes difficult to distinguish, as in *kupur* 'body hair; feathers' : *kupur-an* 'develop feathers, of a growing bird', *kupur-in* 'be hairy':

(7.5) inay   a   rumfaz niwan tu kupur-an
        this **lig** bird   not.yet *tu* feather-**lf**

        'This bird hasn't developed feathers yet'

(7.6) nak   a   rima kupur-in
        1sg.gen **lig** arm   hair-**pf**

    'My arms are hairy'.

In Lun Dayeh of northern Sarawak the locative voice has been lost, leaving the system with an AF marked by *–um-* and *ŋ-*, a PF marked by *–ən*, and a moribund IF/BF, marked by *i-* (Clayre 1988, 1991). The data to hand shows no dominant pattern of collapse: four-voice systems apparently can be reduced to three-voice systems in various ways by replacing a non-AF (NAF) voice with prepositional case-marking, or by conflating two NAF voices into one. Logically it might be expected that when three-voice systems collapse further through loss of one of the earlier voice distinctions the result would be a simple active-passive voice system. This appears to be true of many languages in western Indonesia, as Malay, but the reduction of the original four-voice system to a two-voice system in other languages has left a more complex residue. Casiguran Dumagat of northeast Luzon is said to have a focus system with just two possibilities: 'The verb in the predicate is affixed to indicate either Subject Focus or Object Focus, and correspondingly the subject or object in the clause is given a preposed topic-marking particle' (Headland and Headland 1974:xxxv):

(7.7) mag-buno ək   ta   manok
        sf-kill   **1sg** **obl** chicken

        'I will kill the chicken'

(7.8) bunu-ən ko   tu   manok
        kill-**of** **1sg** **top** chicken)

        'The chicken is what I will kill'

However, focus (regarded as highlighting one of the nominal arguments) is said to co-occur with a second device for giving prominence to one element in the clause, called 'orientation'. Four orientations are described for Casiguran Dumagat, and the third of these, which is said to have as its object 'the location, end point, or recipient of the action' is marked by a reflex of PAN *-an: *mə-ginan-an du anak to baybay* (run **top** children **obl** beach) 'The children are racing along the beach'. The exact characterisation of this system in terms of the number of focus/voice possibilities thus depends critically on how 'orientation' is treated in the grammar.

In several languages of coastal Sarawak the PAN four-voice system has reduced to an active-passive contrast, but one that differs from languages such as English or even Malay in ways indicative of its historical origin. In Mukah Melanau, for example, *-en (PF), *-an (LF) and *Si- (IF/BF) have been lost as productive elements of the grammar, and a two-voice system has been built from the remnants of the original morphologically richer system, with –əm- (< *-um-) marking active and –ən- (< *-in-) marking passive voice. However, the passive voice is inseparable from perfective aspect, as seen in (7.9)—(7.12):

(7.9)   akəw   mə-lasuʔ   nasiʔ
         1sg   **av**-heat   cooked rice
         'I'm warming up the cooked rice'

(7.10)   akəw   mə-lasuʔ   nasiʔ   mabəy
         1sg   **av**-heat   cooked rice   yesterday
         'I warmed up the cooked rice yesterday'

(7.11)   akəw   ŋaʔ   mə-lasuʔ   nasiʔ   səmunih
         1sg   **fut**   **av**-heat   cooked rice   tomorrow
         'I'll warm up the cooked rice tomorrow'

(7.12)   nasiʔ   nə-lasuʔ   kəw
         cooked rice   **pv.perf**-heat   **1sg**
         'I warmed up the cooked rice'

As shown by the grammaticality of (7.10) and (7.11), aspect is unstated and free to vary in the active voice. However, in the passive voice the only possible reading is perfective: any attempt to force a non-perfective reading on sentences such as (7.12) results in the speaker resorting to alternative ways of stating the same propositional content. The same inseparability of voice and aspect is seen in ablaut forms such as *asəw subut ləŋan kəw*/*ləŋan kəw sibut asəw* 'a dog bit my arm', where the active voice marked by *u*-ablaut (< *-um-) has neutral aspect, but the passive voice with *i*-ablaut (< *-in-) has only a perfective reading (Blust 1997b). This differs from western Indonesian languages such as Malay, where both active and passive sentences such as *dia mə-manas-kan nasi* (3sg **av**-heat-**tr** cooked.rice) or *nasi di-panas-kan dia* (cooked.rice **pv**-heat-**tr** 3sg) 'S/he is warming up the cooked rice' can be used with temporal adverbs indicating past or future action (*dia mə-manas-kan nasi kəmarin*/*nasi di-panas-kan dia kəmarin* 'S/he warmed up the cooked rice yesterday', *die akan mə-manas-kan nasi besok*/*nasi akan di-panas-kan dia besok* 'S/he will warm up the cooked rice tomorrow'. While Malay has reduced the original four-voice system to an active/passive contrast reflecting PMP *maŋ- (AF), and *-in- (> ni- > di-), then, Mukah Melanau and other languages of coastal Sarawak have collapsed the system to a superficially similar two-voice system, but have preserved the portmanteau function of PAN *-in- in marking both passive voice and perfective aspect.

Most languages of Sulawesi south of the Gorontalic region have also reduced the original four-voice system to just two, but have retained some structural peculiarities of the earlier system. Speaking of Proto Celebic, the putative ancestor of all non-Philippine type languages of Sulawesi except those of the South Sulawesi group, Van den Berg (1996b:91), for example, notes that 'Proto Celebic had a two-way focus system (actor and goal focus). As in the modern languages that still have these two foci, focus selection was

to a large extent determined by discourse considerations. Specifically, goal focus was used when the agent was known and the goal was definite, often in clauses forming the backbone of a story.'

Many AN languages have a verb system that is rather different from Philippine-type languages. This is especially true of CEMP languages, which evidently began to diverge from the Philippine type quite early. So far as can be determined, Proto Oceanic lost all active traces of the verbal functions of *-um-, *-en, *-an, *Si-, *-in-, although apparent reflexes of the last three affixes are found in some Oceanic languages with nominalising functions. In place of this lost set of verbal affixes POC and many of its descendants use two transitive suffixes: *-i and *aki(ni) to mark case relations such as actor, agent, experiencer, instrument, cause, patient, place and goal. Pawley (1973:119) pointed out that 'subjects' in Philippine languages and direct objects in Oceanic languages such as Fijian and the Polynesian languages share certain properties that suggest a diachronic lineage. First, the 'subject' in Philippine languages is always definite or specific, as are direct objects in many Oceanic languages. Second, the 'subject' slot in Philippine languages can be filled by an NP playing a wide range of semantic roles, and this is equally true of the direct object in many Oceanic languages. Third, according to Pawley the division of labor between affixes is very similar in the two cases, with one affix marking instrument, cause, concomitant, or beneficiary, and another marking patient, experiencer, and goal. 'Somewhere in the history of the Oceanic languages,' Pawley concludes, 'it seems that non-actor subjects became direct objects, and the complex subject-selection system became a complex direct-object selection system.' Since these words were written there have been several other attempts to bridge the gap between Philippine-type verb systems and the verb system reconstructed for POC, most notably by Starosta, Pawley and Reid (1982), and by Ross (2002a). In each of the latter publications a distinction is drawn between the voice-marking morphology of PAN indicative independent clauses on the one hand, and of imperative and dependent clauses on the other, and in both publications central features of the POC verb system are derived from PMP verb morphology that was restricted to dependent clauses and imperatives, although the interpretations differ in detail. Starosta, Pawley, and Reid, for example, derive POC *-i 'close transitive' from a 'captured' locative marker *i which lost its connection to a following prepositional phrase and became part of the preceding verb phrase, while Ross (2002a) sees it as arising from the PMP local passive suffix *-i used in dependent clauses and imperatives.

Most languages of Borneo, Sumatra, Java and mainland Southeast Asia that have reduced the original system of four voice contrasts have also simplified their overall morphological systems. While it is not unusual for Philippine-type languages to have 200-300 distinct affixes or affix combinations, the number of bound morphemes in western Indonesian languages is far smaller.[73] In Sulawesi south of Mongondow the Philippine voice system has also been lost, but unlike the situation further west, a fairly rich innovative morphology has sprung up in its place. A central feature of these restructured systems is what van den Berg (1996b) has called 'conjugated verbs'—the use of proclitic pronouns or affixes on the verb that are marked for person and number. For Muna of southeast Sulawesi van den Berg (1989) lists some 54 affixes, including reduplication

---

[73]   Some 201 affixes or affix combinations have been recorded for Thao of central Taiwan (Blust 2003a), and Rubino (2000:xviii) lists over 400 for Ilokano of the northern Philippines, although some of these clearly are allomorphs. By contrast, Macdonald and Soenjono (1967) give fewer than 20 affixes for Bahasa Indonesia, and much the same appears to hold true for most of the languages of northern Sarawak, as Kiput, for which Blust (2003b) recorded just twelve affixes (including reduplication).

processes. Some of these are retentions, but others are innovative, and point to a transformation of the original voice system that was very different from that in the languages of Borneo. Even greater differences are seen in eastern Indonesia, where Klamer (2002) points out that nearly all languages are verb-medial, the reconstructed voice system has disappeared almost without a trace, full NPs typically are not marked for case, some languages (as Kambera of eastern Sumba) lack a passive voice, and others (as Manggarai of western Flores) have no affixes at all.

Although nearly all other languages have lost a Philippine-type voice system, if this is defined as a system of verb morphology that allows at least two types of passive, active or fossilised portions of this system are widespread. This is most clearly seen in the Batak languages of northern Sumatra, and in Old Javanese, both of which have a number of 'Philippine-type' features. In addition, many other languages preserve reflexes of one or more PAN voice-marking affixes either as fossilised morphemes, or in a non-verbal function, as in Malay, where a reflex of *-um- is fossilzed as *–əm-* in reduplications such as *gilaŋ g<əm>ilaŋ* 'sparkling, glittering', or *guruh g<əm>uruh* 'rumbling, of thunder', or in Kelabit of northern Sarawak, where reflexes of *-um-, *-en and *-in- mark active and passive voice (the last of these passive-perfective), but a reflex of *-an apparently is used only in locative nouns: *dalan* 'path, road' : *dəlan-an* 'path made by walking repeatedly over the same course', *irup* 'what is drunk; way of drinking' : *rup-an* 'watering hole for animals in the jungle', *tələn* 'swallowing' : *tələn-an* 'internal throat'.

Recently Himmelmann (2005) has proposed a novel typological schema in which all 'Western Austronesian' languages (those of Asia and Madagascar) are assigned to one of three categories: 1. symmetrical voice languages, 2. preposed possessor languages, and 3. transitional languages (those that fit comfortably into neither of these categories). He defines symmetrical voice languages (2005:12) as languages with "at least two voice alternations marked on the verb, neither of which is clearly the basic form," preposed possessor languages as languages in which "possessors regularly precede the possessum," and transitional languages as languages that "do not adhere to a common typological core profile." It is clear, then, that 'transitional languages' are a residual category rather than a well-defined class, and it is necessary to comment only on symmetrical voice and preposed possessor languages.

Table 7.12 Characteristic features of symmetrical voice and preposed possessor languages

Symmetrical voice languages	Preposed possessor languages
Symmetrical voice alternations	No or asymmetrical voice alternations
Postposed possessor	Preposed possessor
No alienable/inalienable distinction	Alienable/inalienable distinction
Few or no differences between narrative and equational clauses	Clear-cut differences between narrative and equational clauses
Person marking only sporadically attested	Person marking prefixes or proclitics for S/A arguments
Numerals/quantifiers precede head	Numerals/quantifiers follow head
Negators in pre-predicate position	Clause-final negators
V-initial or SVX	V-second or -final

In support of this classification Himmelmann (2005:175) lists eight typological features that are said to cluster together in symmetrical voice languages and eight contrasting

typological features that are said to cluster together in preposed possessor languages, as shown in Table 7.12.

Although he does not provide an explicit listing of languages that exemplify each type, symmetrical voice languages are said to include Philippine-type languages and most of the languages of western Indonesia apart from Sulawesi. Thus, Malay is said to meet the basic criterion of a language that has 'at least two voice alternations marked on the verb, neither of which is clearly the basic form,' since a verb such as *lihat* (base form) 'see' is prefixed for both active and passive voices: *mə-lihat* (active), *di-lihat* (passive). According to Himmelmann (2005:113) 'The *Philippine-type languages* are a subset of the symmetrical voice languages' which have a) at least two formally and semantically different *undergoer* voices, b) at least one non-local phrase-marking clitic for nominal expressions, and c) pronominal second-position clitics. Under this definition he excludes Malagasy, Chamorro, Palauan, the languages of Brunei and Sarawak, Tomini-Tolitoli, Gorontalo-Mongondic, Sama-Bajau and Bilic from the category 'Philippine-type language'. While this is not problematic for most of these cases, it is clearly problematic for Malagasy, Chamorro, and Mongondow which are grouped with e.g. Malay as non-Philippine-type symmetrical voice languages, although virtually all comparative remarks in the literature stress their Philippine-type grammars.

Although he clearly distinguishes typological from genetic classifications, Himmelmann's tripartite typological division of the AN languages of insular Southeast Asia shows an interesting correlation with certain features of language history. First, it combines those languages which have retained in a relatively intact form the major structural features of the PAN voice system ('Philippine-type languages') and those languages of western Indonesia that have reduced this system to a two-way active/passive voice contrast, perhaps as a consequence of a more general reduction in the total inventory of affixes. In both cases active and passive voice morphology are present, largely as retentions. Western Indonesian languages such as Malay, Javanese, or Toba Batak, then, are essentially Philippine-type languages that have been morphologically 'scaled down', with resulting consequences for some aspects of the syntax, but which generally show little in the way of fundamental restructuring. Preposed possessor languages, on the other hand, correspond closely to the languages of eastern Indonesia that Brandes (1884) classified as showing the 'reversed genitive', and many of the typological traits that they display may well be products of substratum influence. While Western Indonesian languages such as Malay can be understood as fairly drastically 'scaled-down' versions of Philippine-type languages, then, this is not true of preposed possessor languages, many of which have a radically different morphosyntactic groundplan than that of PAN or PMP. Himmelmann's third category, 'transitional languages' is confined mainly to Sulawesi. Many of these languages appear to be 'scaled-down' versions of Philippine-type languages like Malay, Javanese or Toba Batak which have subsequently been rebuilt with a relatively elaborate morphology, much of it innovative, and concerned with what van den Berg (1996) has called 'the spread of conjugated verbs.'

Himmelmann's proposal is original and in many ways insightful, but it is also problematic. There is in fact no tight association between the traits given for either symmetical voice languages or preposed possessor languages. Languages such as Tagalog, for example, are verb-initial, but have preposed possessors, while languages such as Malay are verb-medial, but have postposed possessors, numerals/quantifiers that precede the head, and negators in pre-predicate position. It is moreover, unclear how how many types of criteria are relevant to determining whether the choices in a two-voice alternation are

equally basic, and what weight should be assigned to each of them. One would expect that frequency would play a central role in deciding the issue, yet frequency data permitting such determinations are generally lacking. Finally, it is obscure why a fundamental typological division of genetically related languages would revolve around criteria that have no apparent interrelationship, namely whether the voice system lacks a dominant-subordinate profile vs. whether the possessor is preposed or postposed to its head noun. Himmelmann (2005:113-114) claims that the contrast of symmetrical voice vs. preposed possessor aligns well with other typologically useful features, and notes further that the two categories 'correlate negatively with each other in that languages with symmetrical voice alternations generally show postposed possessors, and languages with preposed possessors either do not show any grammaticised voice alternations at all or the voice alternations are clearly asymmetrical,' but further research is needed to establish confidently that this is true.

7.1.1 Verbs or nouns?

Nominal reflexes of the voice affixes raise another basic issue in connection with voice or 'focus' morphology. In most Philippine-type languages agents and possessors carry identical marking, making it possible for the same affixed word to function either as a verb or as a noun, depending upon the larger syntactic context. The following examples from Thao of central Taiwan and Tindal Dusun of Sabah illustrate:

Thao

> *apa* 'carry on the back'; *in-apa* 'was carried (PF); what was carried; load, burden'

(7.13)   wazish   in-apa   sa   suma
            wild pig   **pv**.**perf**-carry   **sa**   someone
            'Someone carried the pig'

(7.14)   inay   nak   a   in-apa   wa   aniamin
            this   my   **lig**   **pv**.**perf**-carry   **lig**   thing
            'This is the load of things that I carried'

Tindal Dusun

> *ligod* 'throw stones'; *ligod-on* 'be thrown, of a stone; place of throwing stones'

(7.15)   isio   ligod-on   dokow
            3sg   throw-**pv**   **3pl**
            'They threw stones at him'

(7.16)   ligod-on   tulun   ti
            stone.throwing.place-**pv**   person   this
            'This is a place where people throw stones'

Based on the widespread occurrence of nominal readings for word bases that carry voice or 'focus' morphology Starosta, Pawley and Reid (1982) argued that these affixes had *only* nominalising functions in PAN, with verbal functions developing in many daughter languages through syntactic analogy with equational constructions. This view

now seems overstated for several reasons. First, it assumes a drift from noun to verb, although the voice affixes of almost all Philippine-type languages have both verbal and nominalising functions, and the verbal functions are clearly primary (reflexes of *-um-, for example, rarely have nominalising functions).[74]   Second, reflexes of PAN nominative markers for personal nouns and pronouns such as Malay *si* 'marker of personal nouns' show that a voice system in which grammatical morphemes had well-integrated syntactic functions, has broken down, leaving pieces of the system as syntactically functionless residues. It is difficult to argue that languages like Malay represent a more conservative type in comparison with languages like Tagalog, since this would entail an argument that morphemes that were essentially meaningless were innovated for no obvious reason, and only later acquired syntactic functions that were well-integrated with those of the voice morphemes. Third, if PAN *Si- was used only to derive instrumental nouns it is difficult to understand why Ca- reduplication was used for the same purpose. Rather than exclusively nominalising functions, then, the comparative evidence suggests that 'focus' morphology in PAN had both verbal and nominalising functions, much as it does in many attested Philippine-type languages.

7.2   Ergative to accusative or accusative to ergative?

Another issue that has stirred controversy is whether languages that are now ergative have developed from an accusative prototype, or vice versa. This question was first posed for the Polynesian languages, since even within this relatively closeknit group some languages are ergative (e.g. Tongan) and others accusative (e.g. Maori or Hawaiian). Hale (1968) drew attention to parallels between the distribution of accusative and ergative languages in Australia and Polynesia: in Australia it is clear that the typological cleavage does not follow genetic lines, and he suggested that the same holds true for Polynesia. Hohepa (1969) elaborated Hale's programmatic remarks regarding the accusative/ergative split in Polynesian languages, and like him argued for an accusative-to-ergative drift. He characterised languages that have passed through this process (Niue), as 'ergative', languages that have not begun this process (Eastern Polynesian) as 'accusative', and languages that are at an incipient stage in this process (Tongan, Samoan, Pukapukan, etc.) as 'accusative-ergative'. The Hohepa-Hale discussion triggered a number of commentaries on the history of ergative and accusative languages in Polynesia which are notable for the diametrically opposed positions taken by highly qualified scholars. Chief among these are Clark (1973), who argued that the evidence favors an ergative-to-accusative drift, Chung (1978), who favored the position originally expressed by Hale (1968) and Hohepa (1969), and Kikusawa (2002), who defends an ergative-to-accusative development for Central Pacific languages (Rotuman-Fijian-Polynesian) as a whole, basing her argument on the history of pronominal forms rather than the case marking of noun phrases. Most recently, Otsuka (2011) has argued that the traditional typological bifurcation of Polynesian languages into ergative and accusative varieties is misleading, and that Eastern Polynesian

---

[74]   See, for example, Ross (2002a:41), where a useful compendium of verbal and nominal uses for the same voice marking affix is given for thirteen Formosan languages, as follows (exclusively verbal = EV, verbal and nominal = VN, exclusively nominal = EN; some languages reflect only parts of the reconstructed system): *-um-, EV = 10, VN = 2, EN = 0; *-en, EV = 2, VN = 6, EN = 2; *-an, EV = 0, VN = 8, EN = 5, *Si-, EV = 2, VN = 3, EN = 3; *-in-, EV = 3, VN = 6, EN = 1. These data suggest that reflexes of *-um- rarely have nominalising functions, and that reflexes of *-an are often exclusively nominalising.

languages are better characterized, in the words of Foley (2012b:914), as "symmetrical languages, in which transitive verbs can occur freely in two basic clausal configurations of equal unmarkedness." If accepted, this analysis provides a novel perspective that makes many of the earlier arguments for an accusative-to-ergative or ergative-to-accusative drift lose relevance, although the directionality of change from one alignment type to the other would appear to remain a live issue.

Although attempts to explain ergative/accusative differences among related languages began in Polynesian, the same issue subsequently arose in regard to Philippine-type languages. An ergative interpretation of Philippine languages apparently was first proposed by Donna Gerdts in an unpublished paper on Ilokano that was presented in 1980, and further developed in Gerdts (1988). Very similar views are presented in De Guzman (1988). However, this type of analysis was developed most fully by Stanley Starosta, in such publications as Starosta, Pawley and Reid (1982), and Starosta (1986). While the ergative/accusative debate in Polynesian linguistics generally adhered to classical definitions of transitivity to define ergativity, ergative analyses of Philippine-type languages often depend on the scalar view of transitivity advocated by Hopper and Thompson (1980), and on the distinction between valency and transitivity presented by Dixon and Aikhenvald (2000). Perhaps the fullest development of this approach with a wide range of Philippine-type languages is found in Liao (2004).

Liao (2004:8) adopts several concepts from Dixon and Aikhenvald (2000), including the distinction between core and peripheral arguments, and between valency and transitivity. Four core arguments (S, A, O and E) are distinguished, and defined as follows: "A is the more active core argument of a canonical transitive verb; O is the less active core argument of a canonical transitive verb; S is the sole argument of a canonical intransitive verb, or the core argument of a dyadic intransitive verb that has the same morphological marking as the sole argument of a canonical intransitive verb; E (stands for 'extension to core') is the second core argument of a dyadic intransitive verb, and which does not have the same morphological marking as the sole argument of a canonical intransitive verb." Following the same source, Liao further distinguishes valency (the number of core arguments that a verb takes) from transitivity (whether those arguments include S, A, O, and/or E). Intransitive constructions can thus be monadic (canonical, or plain intransitives) or dyadic (extended intransitives). She illustrates the contrast between dyadic transitive and dyadic intransitive constructions with the English example 'Harry hit the ball' (dyadic transitive) vs. 'Harry hit at the ball' (dyadic intransitive). Although a number of qualifications are given that cannot be fully presented here, this is the basic notion that is applied to the analysis of Philippine-type languages. Using a morphological identification test developed by Gibson and Starosta (1990) she then shows that in her analysis Ilokano intransitive verbs, whether monadic or dyadic, take the affixes *-um-*, *ag-* or *maŋ-*, while transitive verbs take the affixes *-en*, *-an*, and *i-*. Under this interpretation the Ilokano dyadic clause (7.17) is intransitive, while the dyadic clause (7.18) is transitive (*-ek* is said to carry 'an agreement feature for the A of the clause'):

(7.17)   um-inúm=ak   ití   danúm
          drink=**nom.1sg** *iti* water
          'I drink water (any kind of water)'

(7.18) inum=ek   ti   danúm
       drink=?**1sg** **art** water
       'I drink the water (not any kind of water)'

As shown in these examples and others in Liao (2004), the ergative analysis of Philippine-type languages interprets the 'Actor voice' as a marker of intransitive verbs, and all other voices as markers of transitive verbs. This interpretation is justified by several considerations, including the higher text frequency of patient voice as compared with actor voice in those Philippine-type languages for which this type of information is available. Differences in the glosses for (7.17) and (7.18) also suggest that the argument of a dyadic transitive clause is more directly, completely or specifically affected than that of a dyadic intransitive clause. A similar difference of definiteness or specificity appears to characterise the active and passive voices of some western Indonesian languages that have lost the Philippine-type voice system. In Sebop, a Kenyah language of northern Sarawak, for example, the definite marker *inah* is optional with active verbs, but obligatory with passive-perfectives:

(7.19) iah   m-ui   saŋəp
       3sg.nom **av**-wash   clothes
       'She is washing clothes'

(7.20) iah   tipo   m-ui   saŋəp
       3sg.nom already   **av**-wash   clothes
       'She washed clothes'

(7.21) saŋəp   inah n-ui   nah
       clothes **def** **pv.perf**-wash **3sg.gen**
       'She washed the/those clothes'

(7.22) **saŋəp n-ui nah
       'She washed clothes'

The ergative analysis of Philippine-type languages provides several descriptive advantages. First, all PAN verbs that can be reconstructed with *-um- are monadic or optionally monadic: *Caŋis 'weeping' : *C<um>aŋis 'to weep, cry', *laŋuy 'swimming' : *l<um>aŋuy 'to swim', *aRi 'come!' : *um-aRi 'to come', *quzaN 'rain' : *q<um>uzaN 'to rain', *utaq 'vomit' : *um-utaq 'to vomit', etc. Minimally, then, *-um- was an intransitive affix in some verb stems. Second, passive constructions are commonly used as imperatives in AN languages, while this is less often true of active constructions, an observation which agrees with the recognition of the patient voice as less marked than the actor voice. Third, in a number of Philippine languages PV constructions imply that the patient NP is more completely affected by the action of the verb than AV constructions of similar propositional content, thus indicating a higher degree of transitivity on the Hopper and Thompson transitivity scale. In some cases the semantic distinction between expressions of roughly the same propositional content in AV and PV is subtle, and easily missed in translation, as with the following sentences from Tindal Dusun of Sabah, both of which use the verb base *kawin* 'to marry' (borrowed from Malay):

(7.23) k<um>awin   i   Wendell om   i   Trixie koniab
        marry-**av**   **def** Wendell and **def** Trixie yesterday
        'Wendell and Trixie got married yesterday' (normal wedding)

(7.24) noko-kawin   i   Wendell om   i   Trixie koniab
        marry-**pv.perf** **def** Wendell and **def** Trixie yesterday
        'Wendell and Trixie got married yesterday' (under pressure, as of premarital pregnancy)

Despite these explanatory successes the ergative analysis of Philippine-type languages raises other questions. In PMP the AV was marked by two morphological patterns: infixation with *-um- and prefixation with *maŋ-, the latter often reduced to simple homorganic nasal substitution (Blust 2004a). In the standard ergative analysis of Philippine-type languages reflexes of both *-um- and *maŋ- are intransitive, but in some languages these affixes apparently differ in transitivity, as seen in the following affixed forms of *turun* 'to descend' in Bario Kelabit of northern Sarawak, where *-um- > -əm-, and *maŋ- > ŋ- (realised as nasal substitution before bases beginning with an obstruent):

(7.25) iəh   t<əm>urun   əd^{h}an
        3sg descend-**av.intr** ladder
        'He is descending the ladder'

(7.26) iəh   nurun   əd^{h}an
        3sg descend-**av.tr** ladder
        'He is lowering the ladder'

A major test that Liao used to establish the transitivity of dyadic verbs is one that Gibson and Starosta (1990:199) called the 'morphological identification' test, which makes the following prediction: 'If a language has three verbal clause patterns (one monadic pattern and two dyadic patterns) and the verbs in the three clauses are ALL MORPHOLOGICALLY COMPLEX, then the dyadic clause pattern that has the same verbal morphology as the intransitive clause pattern counts as intransitive, whereas the other dyadic clause pattern counts as transitive.' Liao (2004:31ff) notes a number of construction types in which a reflex of *-en occurs with monadic verbs, most notably in the widespread pattern X-en 'be afflicted with X' (where X can be a health condition, insect pests, a bad turn of the weather, etc.): PMP *quban 'grey hair' : *quban-en 'get grey hair', *anay 'termite' : *anay-en 'damaged by termites, eaten by termites', *quzan 'rain' : *quzan-en 'get caught in the rain'. She points out that if these adversative passive constructions are taken as the standard of comparison it must be concluded that reflexes of *-um- mark transitive verbs, and reflexes of *-en and other non-AF affixes mark intransitives. Since this conclusion would contradict the other evidence she considers, however, she takes adversative passives as evidence that the morphological identification test is unreliable for determining transitivity.

Linguists have long shown an interest in assigning languages to types. This is understandable, as it follows from the search for generalisations: rather than treat each language as a unique configuration of properties, it seems worthwhile to look for commonalities that will enable populations of languages to be grouped together as members of a common class distinct from others. At an earlier time the classificatory

criteria used to assign languages to types were mostly morphological: languages were 'agglutinative', 'fusional', 'isolating' or 'polysynthetic'. With the tremendous advances in syntactic theory over the past half century it is not surprising that there has been a shift away from an appeal to morphological criteria in assigning languages to types toward one which gives greater weight to syntactic criteria.

In recent years patterns of case alignment have been most often used to form typological categories, as in the dichotomy between ergative-absolutive vs. nominative-accusative languages. While the motivation behind such classifications is clear, the significance of the resulting categories often is not. One can legitimately ask what has been gained, for example, in determining that a given Polynesian language is ergative rather than accusative. Does it mean that such ergative AN languages are structurally more similar to ergative languages in other language families (Australian, Kartvelian, Northeast Caucasian, Basque, etc.) than they are to accusative AN languages with which they have shared a very large part of their history? The answer seems clearly to be negative. In fact Dixon (1994:219) has gone so far as to say that the mere fact that a language is ergative does not necessarily imply *any* other typological property. As a general theory of types, then, the ergative/accusative distinction is clearly unsatisfactory, and it would be more meaningful to seek a typology in which a given criterial property is implicationally related to as many others as possible.

7.3   Word order

One common strategy is to group languages into categories by order of major sentence constituents, namely subject (S), verb (V), and object (O). For at least some scholars, this may be regarded as problematic with respect to languages with ergative case alignment. To avoid this problem as much as possible I will categorise AN languages only as 'verb-initial', 'verb-medial' or 'verb-final'. All three types of structure occur, but their geographical distribution differs markedly. Unless noted otherwise, statements about word order refer to what Schachter and Otanes (1972:60) call 'basic sentences', that is, declarative sentences in the indicative mood:

Table 7.13 Geographical distribution of verb-initial, verb-medial
and verb-final Austronesian languages

Area	Verb-initial	Verb-medial	Verb-final
Taiwan	most languages	one or two	none
Philippines	all or nearly all	none?[75]	none
Borneo	some in north	most elsewhere	none
Madagascar	all dialects	none	none
Mainland SE Asia	none	all	none
Sumatra	a few in north	most elsewhere	none
Java-Bali-Lombok	none	all	none
Sulawesi	a few in north	most elsewhere	none
Lesser Sundas	a few	nearly all	none
Moluccas	few or none	all or nearly all	none
New Guinea	none	many	many

---

[75]   SVO order occurs in Tagalog and some other Philippine languages as a free option in some types of constructions, but unmarked word-order is verb-initial.

Area	Verb-initial	Verb-medial	Verb-final
Bismarcks	none	all	none
Solomons	a few	most	a few
Vanuatu	none	all	none
New Caledonia	none	all	none
Micronesia	one or two	most	none
Fiji-Polynesia	all	none	none

7.3.1. Verb-initial languages. Verb-initial languages occupy two generally solid blocks, with a scattering of examples elsewhere. The first of these blocks corresponds very closely to the distribution of Philippine-type languages, and so covers Taiwan, the Philippines and northern portions of Borneo, but not all of northern Sulawesi, where languages such as Tondano have a Philippine-type four-voice verb system, but a verb-medial word order typology. Virtually all other Philippine-type languages are verb-initial, or more accurately, predicate-initial, the major exception perhaps being Chamorro. The following data from Bunun and Amis in Taiwan, Ilokano and Maranao in the Philippines, Timugon Murut in northern Borneo, and Malagasy serve to illustrate (morphemes that could not be glossed are in italics):

Bunun

(7.27) uka   an ca puaq   di   is-duli
        not.have   *an ca* flower this **iv**-thorn
        'This flower has no thorns'

(7.28) ma-asik   lumaq   azak
        av-sweep house   **1sg**
        'I'm sweeping the house'

(7.29) muŋ-qanu   ca   lukic
        av-drift *ca* wood
        'The wood is drifting away'

(7.30) pataz-un   bunun   acu
        kill-**pv** man   dog
        'The man killed a dog'

Amis

(7.31) ma-fanaq   kaku   cima   ciŋra
        av-know **1sg** who   **3sg**
        'I know who he is'

(7.32) ci-sołaq   to   kasiʔnaw-an
        have-snow   already   winter
        'In the winter it snows'

(7.33) koh-te-tiŋ   kina   paliłin   maʔmin
        black-**red** those   car   all
        'All those cars are black'

(7.34) fətək-un   no   mako   ko   mata
close-**pv**   **gen**   **1sg**   **conn**   eye
'I closed my eyes'

Ilokano

(7.35) nag-sáŋit   ti   ubíŋ
act.**past**-cry   **art**   child
'The child cried'

(7.36) saán-ko   a   ma-awát-an
neg-1sg.erg   **lig**   understand
'I don't understand it'

(7.37) kukuá-da   ti   baláy
goods-**3pl**   **art**   house
'The house is theirs'

(7.38) Insík   da   Wei   ken   Yi
Chinese   **3pl**   Wei   and   Yi
'Wei and Yi are Chinese'

Maranao

(7.39) t<om>abas   so   bəbai   sa   dinis   ko   gəlat
cut-**av**   **foc**   woman   **obj**   cloth   **ref**   knife
'The woman will cut cloth with the knife'

(7.40) ma-dakəl   a   tao   sa   masgit
stat-many   **lig**   person   **loc**   mosque
'There were many people at the mosque'

(7.41) m<i>anik   si   Anak
av-perf-climb   **foc**   Anak
'Anak climbed up'

(7.42) mia-ilai   i   Anak   so   kambiŋ
pv-see   **pm**   Anak   **foc**   goat
'Anak saw the goat'

Timugon Murut

(7.43) tataŋ-on   mu   korojo-mu-no
leave-**ov**   **2sg.gen**   work-**2sg.gen-art**
'You will leave your work'

(7.44) ma-riuq   io   ra   suŋoy
av-bathe   **3sg**   **loc**   river
'He bathes in the river'

(7.45)   nag-anak   karabaw   raitio
        av.past-child   water buffalo   this
        'This water buffalo has calved'

(7.46)   na-kito   min   aku
        av.past-see   **2sg**   **1sg**
        'You saw me'

Malagasy

(7.47)   lavitra   ny   lalana
        long   **art**   way/road
        'The way is long'

(7.48)   ni-vidy   maŋga   telo   aho
        av.past-buy   mango   three   **1sg**
        'I bought three mangos'

(7.49)   manana   boky   aho
        have   book   **1sg**
        'I have a book'

(7.50)   mi-tady   raharaha   i   Koto
        av-search   work   **pm**   Koto
        'Koto is looking for work'

To say that the above languages and others from the same regions are verb-initial is, of course an oversimplification. In most active declarative sentences the predicate is the first element, whether this element is an active verb, a stative verb/adjective, a noun or some other part of speech. Within the predicate, however, other elements may precede the verb, and in some cases this may trigger movement. In Timugon Murut (Prentice 1971:218) the negative marker precedes the initial verb, but does not affect word order (7.52). Verbal auxiliaries and sentential adverbs, including at least some temporal expressions, on the other hand, invert the order of actor and verb (7.54, 7.56):

(7.51)   inum-on   takaw   gitio
        drink-**of.fut**   **1pl.in**   this
        'We will drink this'

(7.52)   kalo   inum-on   takaw   gitio
        neg   drink-**of**.**fut**   **1pl.in**   this
        'We will not drink this'

(7.53)   tataŋ-on   mu   korojo-mu-no
        leave-**of**   **2sg.gen**   work-**2sg.gen-art**
        'You will leave your work'

(7.54)   ma-buli   mu   tataŋ-on   korojo-mu-no
        aux   **2sg.gen**   **of**-leave   work-**2sg.gen-art**

'You may leave your work'

(7.55) ma-riuq   io   ra   suŋoy
        af-bathe   **3sg**   **loc**   river
        'He bathes in the river'

(7.56) monsoŋ-orow   io   ma-riuq   ra   suŋoy
        all day   **3sg**   **af**-bathe   **loc**   river
        'He bathes in the river all day'

Similarly, nearly all languages allow verb-medial constructions if there is a nominal predicate, as in the following Ilokano sentences:

(7.57) na-pán-ak   idiáy   Tagudin
        pot.**perf**-go-**1sg**   there   Tagudin
        'I went to Tagudin'

(7.58) si-ák   ti   na-pán   idiáy   Tagudin
        pm-**1sg**   **art**   **pot**.**perf**-go   there   Tagudin
        'I was the one who went to Tagudin'

In some languages word order appears to be undergoing change from verb-initial to verb-medial, but is not affecting all types of constructions at once. Bario Kelabit of northern Sarawak is spoken in the border area between the verb-initial Philippine-type languages of Sabah, and verb-medial western Indonesian languages such as Malay. As might be expected, it shows features typical of both areas. In particular, the four-voice system of PAN has been simplified to an active-passive voice distinction somewhat like that of many western Indonesian languages, and active transitive verbs are now sentence-medial, but passive and intransitive verbs remain sentence-initial:

(7.59) iəh   ŋə-lanit   bərək   inəh
        3sg   **av**-skin   pig   **def**
        'He is skinning the pig'

(7.60) l<in>anit   iəh   bərək   inəh
        skin-**pv**.**perf**   **3sg**   pig   **def**
        'He skinned the pig'

(7.61) ŋi   iəh   k<um>an   kərid
        emph   **3sg**   Eat-**av**   vegetable
        'He is eating vegetables'

(7.62) k<in>an   iəh   kərid   inəh
        eat- **pv**.**perf**   **3sg**   vegetable   **def**
        'He ate the vegetables'

(7.63)   ŋi   iəh   ŋə-linuh idih
          emph **3sg** **av**-think **def**
          'He is thinking of it'

(7.64)   l\<in\>inuh   iəh   idih
          think-**pv**.**perf** **3sg** **def**
          'He thought of it (already)'

(7.65)   riər   bataŋ
          turn   log
          'The log is turning (by itself)'

(7.66)   iəh   ŋə-riər   bataŋ
          3sg **av**-turn log
          'He is turning the log'

Similar syntactic innovations are found in languages that are either undergoing the transition from verb-initial to verb-medial typology, or that have already become verb-medial as a result of this transition. In Bintulu, a language of northern Sarawak that subgroups with Kelabit, but not closely, both active and intransitive constructions have become verb-medial, leaving only the passive verb in initial position:

(7.67)   batəw   iəʔ   pə-laləg buʔay mətid
          stone   **dem** **av**-roll down   hill
          'This stone is rolling down the hill'

(7.68)   isa   mə-lakaw
          3sg **av**-walk
          'He is walking'

(7.69)   isa   m-itip   pəñaʔ   njen   inəh
          3sg **av**-nibble eat   fish   **dem**
          'He is eating the fish bit-by-bit'

(7.70)   n-itip   ña   pəñaʔ njen   inəh
          pv.**perf**-nibble **3sg** eat   fish   **dem**
          'He ate the fish bit-by-bit'

(7.71)   isa   lupək   bajəw   inəh
          3sg fold-**av** shirt   **dem**
          'He is folding the shirt'

(7.72)   lipək   ña   bajəw inəh
          fold-**pv**.**perf** **3sg** shirt   **dem**
          'He folded the shirt'

(7.73)   n-atəb   ña   mata anak   inəh
          pv.**perf**-close **3sg** eye   child **dem**
          'She closed the child's eyes'

However, other sentences in Bintulu show passive verbs in medial position:

(7.74)   agəm-ña   sibut   ñipa
         hand-**3sg.gen**   bite-**pv.perf**   snake
         'His hand was bitten by a snake'

(7.75)   isa   n-upuʔ   tama-ña
         3sg   **pv.perf**-embrace   father-**3sg.gen**
         'His father embraced him'

Bintulu, which is spoken further south than Kelabit, and hence farther from the 'boundary' between Philippine-type and western Indonesian-type languages, thus appears to be closer to losing verb-initial constructions entirely. Huang and Tanangkingsing (2005) report a similar pattern for Saisiyat of northwest Taiwan, and a generally similar pattern of word order change in which the passive verb is the last to abandon word-initial position is found in comparing Old Javanese, known from written documents dating from the ninth to the fifteenth century with the modern language (Poedjosoedarmo 2002).

Cumming (1991) has demonstrated a similar change from verb-initial to verb-medial constituent order between the Classical Malay texts of the seventeenth century and modern Malay and Bahasa Indonesia. She argues that both orders were available in the classical language and remain available in the modern language, but that there has been a change in the frequency of clause types, or the distribution of functions over forms. She proposes a rather different explanation for this development in the history of Malay, but the examples she cites show a high correlation between passive voice (her 'patient trigger') and verb-initial position, as opposed to active voice (her 'actor trigger') and verb-medial position. Although this clearly oversimplifies the facts, and is based on evidence that is still rather fragmentary, these agreements suggest a drift from (1) invariant verb-initial to (2) verb-initial (patient voice)/verb-medial (actor voice), to (3) invariant verb-medial. Why active verbs have recurrently preceded passive verbs in the change from verb-initial to verb-medial is unknown. Blust (2002d:72) raises the possibility that since the actor voice appears to be statistically less frequent than the passive voice in most Philippine-type languages, word order change may have affected less frequent, and hence more marginal types of constructions before affecting the more central constructions in the grammar.

So far as is known, nearly all other verb-initial AN languages are Oceanic. There are relatively few of these, including a few in the western Solomons, Fijian, and most of the Polynesian languages. Although it is universally agreed that Proto Polynesian was verb-initial, a number of Polynesian Outliers have become verb-medial, apparently through reinterpretation of topicalised constructions as unmarked sentence types, a change that in turn probably was stimulated through contact with SVO languages in Melanesia. Polynesian languages that are verb-initial appear to be typologically consistent in all clause types, but this is not necessarily true of verb-initial languages elsewhere in the Pacific. In Roviana of the western Solomons, for example, many clauses are verb-initial, while others are verb-medial (Corston-Oliver 2002):

(7.76)   kote   sage   la   si   goi
         fut   ascend   go   **abs**   **2s**
         'You will go up'

(7.77)   ke   turu   mo   sari   ka-ŋeta
         so   stand **downtoner 3pl card**-three
         'So the three of them just stood there'

(7.78)   raro-a   gami   sa   talo
         cook-**3sg 1pl.ex def** taro
         'We cooked the taro'

(7.79)   kote   la   sa   igana   gan-i-u
         fut   perf   def fish   eat-**tr-1sg**
         'The fish will eat me'

7.3.2 Verb-medial languages

   South of Sabah nearly all AN languages of Indonesia-Malaysia and mainland Southeast
Asia are verb-medial. This is also true of most Oceanic languages, apart from several
dozen languages of the New Guinea area. Around 80% of all AN languages are thus verb-
medial. Sentences 7.80-7.83 illustrate this for the Long Anap dialect of Kenyah in northern
Sarawak, sentences 7.84-7.87 show it for Wolio of southeast Sulawesi, sentences 7.88-7.91
for the Fehan dialect of Tetun in central Timor, sentences 7.92-7.95 for Titan of the
Admiralty Islands in western Melanesia, and sentences 7.96-7.99 for Naman of Vanuatu:

   Long Anap Kenyah (Blust, n.d. a)

(7.80)   akeʔ   ŋə-lidəp   sapay   ia   kaʔ   ndiŋ
         1sg   av-hang   shirt   **3sg**   on   wall
         'I hung his shirt on the wall'

(7.81)   ləto   ina   ñəbutiŋ   kain
         woman   that   **av**-cut.with scissors   cloth
         'That woman is cutting cloth with scissors'

(7.82)   laki   ina   pə-tiraʔ   ma   meʔ
         man   that   **caus**-talk   **ben   1pl.incl**
         'That man is talking for us (on our behalf)'

(7.83)   anak   ina   ŋə-təmbu   tabat   ina
         child that   **av**-spit.out medicine   that
         'The child spit out the medicine'

   Wolio (Anceaux 1952)

(7.84)   maŋa   mia   i   Wolio   a-kande   talu   mpearo   sa-eo
         pl   person   in   Wolio   **3pl**-eat   three   times   one-
                                                                 day
   'People in Wolio eat three times a day'

(7.85)   o   malo-malo   a-kande   jepe
         art   morning   **3pl**-eat   rice porridge
   'In the morning they eat rice porridge'

(7.86)   maŋa   mia   a-aso   sagala   giu
         pl   person **3pl**-sell various   kind, type
   'People sell all kinds (of things)'

(7.87)   a-poili-mo   i   taliku-na
         3sg-look.back-**pst**   **to**   back-his
   'He looked back'

   Fehan Tetun (van Klinken 1999)

(7.88)   ita   tau   musan   há-hát   lima-lima
         1pl.incl   put   seed   **redup-**four   redup-five
   'We plant four or five seeds at a time'

(7.89)   ita   ruas   bá   harís   lai
         1pl.incl   two   go   bathe   first
   'Let's we two go and bathe now'

(7.90)   sira   bá   hotu   toʔos
         3pl   go   all   garden
   'They all went to the garden'

(7.91)   ó   m-alo   sá
         2sg   **2sg-**do   what
   'What are you doing?'

   Titan (Blust n.d. b)

(7.92)   John   pa-ki-ani   ni
         John   **fut-3sg-**eat   fish
   'John will eat the fish'

(7.93)   yo   lis-i   John   **ti**   leŋ
         1sg   see-**tr**   John   **loc**   beach
   'I saw John at the beach'

(7.94)   yo   pa-ku-caliti-i   key
         1sg   **fut-1sg** cut-**tr**   wood
   'I will cut the wood'

(7.95)   o-po   ni   ceh
         2sg-catch   fish   how.many
   'How many fish did you catch?'

Naman (Crowley 2006c)

(7.96)　në-luolu　　　　usër　　khën　melëkh
　　　　1sg.real-vomit　**cause**　**obl**　*kava*
　　'I vomited because of the *kava*'

(7.97)　ai　　Ø-leg　　　　raŋan　nevet
　　　　3sg　**3sg.real**-sit　**loc**　　rock
　　'She sat on the rock'

(7.98)　Ø-ve　　　　　nelag　　nakha-n
　　　　3sg.real-make　pudding　**ben-3sg**
　　'She made pudding for him'

(7.99)　get　　　ne　　attët-khan　　　　　net
　　　　1pl.incl　only　**top.1pl.incl.real**-eat　**dem**
　　'It's just us who eat that

7.3.3 Verb-final languages

As observed by Capell (1971), a number of AN languages of the New Guinea area are SOV, a typology that agrees with most of the more than 700 Papuan languages. Capell distinguished between what he called 'AN₁' (SVO) and 'AN₂' (SOV), noting that the latter type is found in most of southeast New Guinea and the Massim region, as well as in pockets on the north coast of New Guinea, and in Torau-Uruava, spoken on the east coast of Bougainville Island near the western end of the Solomons chain. He cites lexical data from fourteen languages identified as AN₂ (Mekeo, Motu, Sinaugoro, Suau, Dobu, Misima, Wedau, Mukawa, Laukanu, Labuʔ, Kawaʔ, Yabem, Wampar and Adzera), but does not indicate the total number of verb-final AN languages. Ross (1988) is more explicit, listing 48 languages in the 'Papuan Tip Cluster', all of which apparently have inherited an SOV typology from their immediate common ancestor. While this represents the bulk of verb-final AN languages, the total number may be in the range of 60-70. As might be expected on general typological grounds, SOV word order in these languages is accompanied by implicationally-related structural features such as postpositions and preposed relative clauses. Capell (1971:243) illustrates these features with the sentence 'The man planted the tree in the middle of the garden' from Motu:

(7.100)　tau　ese　au-na　　imea　　bogaragi-na-i　vada　e　　hado
　　　　　man　**def**　tree-the　garden　middle-its-in　**perf**　**Vp**　plant

Although it is not apparent in this sentence, many Motu sentences have obligatory object-marking on the verb, as in

(7.101)　hahine　ese　natu-na　　e　　ubu-a
　　　　　woman　**def**　child-her　**vp**　feed-**3sg**
　　　　'The woman fed her child'

(7.102)   hahine   ese   natu-na   e   ubu-dia
        woman   **def**   child-her   **vp**   feed-**3pl**
        'The woman fed her children'

What Capell writes as 'vp' is described by Lister-Turner and Clark (1930) as a 'verbal particle of the third person used with present and past tenses.' In effect, then, the verb complex found in final position contains an SVO order that is arguably reflective of the order of major sentence constituents before the change to a verb-final typology.

Although virtually all verb-final AN languages have been characterised as SOV, there is one noteworthy exception. Donohue (2002) has reported that Tobati, an Oceanic language inYotefa Bay, near the city of Jayapura in Indonesian New Guinea, has unmarked OSV word order, as in:

(7.103)   hony-o   for-o   rom-i
        dog-**o**   pig-**o**   see-**3sg**
        'The pig saw the dog'

Donohue claims that the alternative reading 'The dog saw the pig' is impossible, and that OSV basic word order is a recent innovation, since SOV order was recorded half a century earlier. This case is unique, and raises questions about the motivations for such change. As will be seen below, the SOV typology in a number of New Guinea AN languages almost certainly is due to Papuan contact influence, and while no Papuan language is known to be OSV, some type of contact influence cannot be completely ruled out as a factor in this unusual development.

The transitions between typological categories based on order of major sentence constituents can be summarised as follows. PAN almost certainly was verb-initial, an order that is retained in most languages with a Philippine-type voice system (the major exceptions being Tondano and some other languages in northern Sulawesi). In the languages of Borneo that have begun to lose the original voice system there has been concomitant word-order change which appears to have taken place in two steps, with patient voice and intransitive constructions remaining verb-initial after actor voice constructions have become verb-medial. A similar pattern of change conditioned by distinctions of voice or transitivity also appears in the history of other languages, as in the transition from Old Javanese to modern Javanese (Poedjosoedarmo 2002), and from Classical to modern Malay (Cumming 1991). In Sulawesi, eastern Indonesia, and Proto Oceanic the verb-initial typology of PAN also developed into a verb-medial typology, but here the steps leading from one dominant constituent order to another are less clear. In the New Guinea area contact with SOV Papuan languages led to a number of Oceanic languages becoming verb-final, and to the acquisition of other structural features universally associated with verb-final typology (postpositions, etc.). The full range of verb-final AN languages remains to be determined. Further east a verb-initial typology developed independently in parts of the western Solomons, and in Proto Central Pacific from a Proto Oceanic interstage that was presumably verb-medial. The motivations for this change and the manner in which it was implemented remain obscure.

7.4   Negation

Although the distinctions may be difficult to draw in some languages, negation may be encoded by single lexical items, as in English *no, not, never, nothing*, and *none*, or by collocations, as in *not have, is/was not, not want* or *not yet*. Cross-linguistically the codification of negation varies among these semantic categories. Where English has (at least synchronically) a single morpheme, 'never', for example, Bahasa Indonesia uses a negated adverb: *tidak pərnah* (**neg** ever), and where Bahasa Indonesia has a single morpheme, *bəlum*, English uses a negated adverb 'not yet'.

Even a cursory glance at the grammars of AN languages suggests wide variation in the coding of negation. The challenge in a survey is therefore to find common threads that reveal a unity beneath the surface diversity of patterns. To date the most ambitious attempt to do this is that of Hovdhaugen and Mosel (1999). This volume, which grew out of a conference devoted to a broader range of topics, contains in-depth studies of negation in seven Oceanic languages, along with passing comments on other languages in the Oceanic group. It is thus concerned only with languages belonging to one well-defined subgroup of AN. To some extent these seven studies can be supplemented by the sketches in Lynch, Ross, and Crowley (2002), although these are often very brief. Both Mosel (1999) and Lynch, Ross, and Crowley (2002:51-52) provide generalising statements about the codification of negation in Oceanic languages, and these preliminary attempts at generalisation are perhaps the best place to begin a survey.

In less than a page, Lynch, Ross, and Crowley (2002) state some interesting general properties of negative marking in Oceanic languages that do not emerge in the longer formulation of Mosel (1999), no doubt because the latter statement is founded on a much smaller database. First, it is maintained that in languages that express tense, aspect and mood (TAM) with free forms the negative marker also tends to be a free form, but in languages with extensive inflectional prefixation, negation tends also to be marked with a prefix. The morphological expression of negative marking thus shows a strong tendency to co-vary with the morphological expression of TAM marking. Second, according to Lynch, Ross, and Crowley (2002:51) "There is a recurring tendency in Oceanic languages for negation to be expressed discontinuously," with the two negative elements bracketing the verb. Since bipartite negative markers generally show little cross-linguistic similarity it is most reasonable to assume that they have arisen through parallel development. This is a striking observation in view of the fact that discontinuous negative marking, although found, is relatively rare in AN languages outside the Oceanic group. One language, Lewo of central Vanuatu, is even reported as having tripartite negative marking (Lynch, Ross and Crowley 2002:52):

(7.104)   Pe   wii   re   poli
          neg₁ water **neg₂** **neg₃**
          'There is no water'

Mosel (1999) stresses the diversity of negation patterns in the languages she has examined, but tries to find some generalising statements about them. The core of her generalisations can be expressed through reference to the patterns in Table 7.14:

Table 7.14 Patterns of negation in twelve Oceanic languages

	1 NO!	2 NEG- EXIST	3 NEG- NON-V.	4 NEG- V.	5 NEG- IMPER
Loniu	pʷa	?	pʷa	pʷa	topu
Manam	tágo	tágo	tágo	tágo	moaʔi
Saliba	nige(le)	nige	nige	nige	tapu
Tolai	pata	pata, vakir	pata, vakir	pa, vakir	koko
Teop	ahiki	ahiki	saka … haa	saka … haa	(goe)
Nêlêmwa	ai, ayai	kia	kio	kio	a, axo
Fijian	seŋa	seŋa	seŋa	seŋa	ʔua, waaʔua
Tongan	ʔikai	ʔikai	ʔikai	ʔikai	ʔoua
E. Futunan	(l)eʔai	leʔai	leʔaise	leʔaise leʔaiʔaise leʔaiʔokise	auana
		leʔe	leʔese, se	leʔese, se	auase
Tokelauan	hēai	hēai	hē	hē	nahe, nā, iā, inā, einā
Samoan	leai	leai	lē	lē, leai	ʔaua
Tahitian	ʔaita	ʔaita	ʔaita	ʔaita	
	ASP-ʔore	ASP-ʔore	ASP-ʔore	ASP-ʔore	
	ʔeita, eʔore		eʔere	ʔeita, eʔore	
			ʔeiaha		ʔeiaha

Mosel calls the negatives of category 1) 'prosentences'; they form a subset of the negative answers to what are traditionally called 'yes-no questions' or 'polar questions'. Category 2) is the negative existential 'there is/are none'; categories 3) and 4) are the negations of non-verbal and verbal constituents, and 5) is the negative imperative or vetative. Although she stresses that her sample is too small to draw definite conclusions, four general tendencies are outlined: 1) Oceanic negatives tend to distinguish three functions: negation of existential constructions, predicates and imperatives, 2) if a language has negative verbs and particles, it will use the verb for existential constructions and the particle for predicates, 3) the negative prosentence tends to have the same form as the existential negative, and 4) when focusing of NPs is expressed by clefting, the cleft NP can be negated by the negative used for the negation of predicates.

A somewhat different set of patterns emerges from the comparison of negative marking in the non-Oceanic AN languages. Although Mosel mentions that some Oceanic languages negate verbal and non-verbal clauses differently, the differences are very slight (Tolai *pata*, *vakir* vs. *pa*, *vakir*, East Futunan *leʔaise*, *leʔese*, *se* vs. *leʔaise*, *leʔese*, *se*, *leʔaiʔaise*, *leʔaiʔokise*, Samoan *leai* vs, *lē*, *leai*, Tahitian *ʔaita*, ASP-*ʔore*, *eʔere* vs. *ʔaita*, ASP-*ʔore*, *ʔeita*, *eʔore*). In every case at least one option allows both types of clause to be marked in the same way, and even where they are not the same the differences show a large degree of similarity in terms of phonemic shape or morphological composition. Patterns of negative marking in the AN languages of insular Southeast Asia diverge sharply from this pattern, since in many languages nominal and verbal constituents are negated by morphologically unrelated forms.

In Bahasa Indonesia the negative answer to a 'yes-no' question depends upon whether the interrogated predicate is nominal or verbal, hence *apa dia guru*? 'Is he a teacher?' is answered negatively with the one-word reply *bukan*, or with the extended reply *bukan, dia bukan guru*, but *apa dia pərgi* 'Did he go?' is answered negatively with *tidak*, or less commonly with the extended reply *tidak, dia tidak pərgi*. It might be argued that a similar distinction is partially encoded in English *no* and *not*, since either of these can be used to negate a nominal expression (the first a bare noun, the second a determiner plus noun), but only *not* can negate a verb. However, the division of labor between negators of nominal and non-nominal expressions is generally more complete in non-Oceanic AN languages, and often involves lexical markers that have no apparent historical connection to one another. McFarland (1977) provides a particularly useful overview of negators in the northern Philippines, showing that these fall into four patterns: 1. noun vs. non-noun, 2. verb vs. non-verb, 3. no contrast, and 4. noun vs. future verb vs. present verb. Table 7.15 provides a sample of AN languages that use distinct markers to negate nouns on the one hand, and verbs on the other

Table 7.15 Nominal and verbal negators in selected Austronesian languages

Language	Nominal negator	Verbal negator
Atayal	(i)yat	iniʔ
Ilokano	saán, (di)	saán, di
Bontok	bakən	adi, əgʔay
Gaddang	bəkkən	əmme
Kalinga	bokon	adi
Botolan Sambal	alwa	aheʔ, ag-
Central Tagbanwa	bèlagiŋ/bèlahiŋ	data
Tboli	sundu, (laʔ)	laʔ
Yakan	dumaʔin	gaʔ, gaʔ-i
Malay/Indonesian	bukan	tidak
Lampung	lain	maʔ
Muna	suano	miina, pa/pae/paise, pata/tapa

Tables such as this oversimplify the facts, and so are in need of discussion. While the basic distinction of nominal vs. verbal negators appears to run through a number of AN languages, the details of how this distinction is manifested differ from one language to the next. In Wulai Atayal as described by Rau (1992:169ff), for example, *iniʔ* negates verbal predicates, and *(i)yat* negates nominal predicates:

(7.105) iniʔ   ku   nbuw   qwaw
         neg **1sg** **act.dep**.drink   wine
         'I did not drink wine'

(7.106) yat   libuʔ naʔ   ŋtaʔ   sa,   libuʔ naʔ   yuŋay   sa
         neg cage   **gen** chicken that cage   **gen** monkey   that
         'That's not an enclosure for chickens, it's an enclosure for monkeys'

The categorial contrast between *iniʔ* and *yat*, then, appears to correspond closely to that seen in Malay/Indonesian *tidak* vs. *bukan*. In Bahasa Indonesia, however, *tidak* also negates adjectives (*tidak bəsar* 'not big'), adverbs (*tidak besok* 'not tomorrow', *tidak di-*

sini 'not here'), numerals (*tidak dua* 'not two'), and some other parts of speech (*apa* 'what', *tidak apa* 'it's O.K., never mind'). Macdonald and Soenjono (1967:159) describe *tidak* as 'the negative for predicatives,' but this is not true of equational sentences, where the predicate may be nominal. For Ilokano Rubino (2000:lxxix) states that *saán* and *di* may be used to negate either verbs or nouns, but that '*saán* is preferred to *di* when negating noun phrases.' A similar statement is given by Porter (1977:33) for Tboli of the southern Philippines, where *sundu* is said to negate "nouns or noun substitutes only," while *laʔ* "is most commonly used with verbs and statives, but may negate a noun." Categories that are mutually exclusive in languages such as Atayal or Malay/Indonesian, then, appear to overlap in languages such as Ilokano or Tboli. Other differences that can be seen as superimposed on the pattern of nominal vs. verbal negators, include Bontok *bakən*, which Reid (1976) describes as 'not; negative of nouns', *adi* 'no; not; negative of verbs and adjectives', and *əgʔay* 'no; not; completive aspect negative', where the nominal/verbal distinction is clear, but the category of verbal negators is subdivided by aspect, and Muna, for which van den Berg (1989:203ff) lists five negators: 1. *miina*, which negates verbal clauses that refer to past or present events, 2. *pa/pae/paise*, which negates verbal clauses that refer to future events, 3. *pata/tapa*, which negates "active and passive participles, and *ka-/-ha-* reason clauses," 4. *suano*, which negates NPs, and 5. *ko/koe/koise*, used in negative imperatives. Finally, while most languages that have a lexicalised distinction between nominal and verbal negators appear to contrast the categories noun : non-noun, a few appear to contrast the categories verb : non-verb. Antworth (1979:50-52), for example, reports four negative markers in Botolan Sambal: *ahəʔ*, *ag-*, *alwa*, and *ayin*. The first two, which are in syntactic complementation, are used to negate verbal sentences. By contrast, *alwa* is used to negate most nonverbal sentences, and *ayin* negates existential sentences, and is the negative form of the locative adjectives *anti* and *anto* (roughly 'here' and 'there'):

(7.107) ahəʔ   p\<in\>ati   nin   tawo   ya   domowag   ko
        neg   kill-**pv.perf** **gen** person **nom**   carabao   my
        'The person didn't kill my carabao'

(7.108) ag-ko   naka-ka-toloy   na-yabi
        neg-1sg   **apt.perf**-*ka*-sleep   last night
        'I couldn't sleep last night'

(7.109) alwa-n   ma-hipəg   ya   tatay   ko
        neg-lig   **stat-**ambitious   **nom** father   my
        'My father isn't ambitious'

(7.110) alwa-n   hiko   ya   naŋ-gawaʔ   nin   habayto
        neg-lig   **1sg**   **nom** make/do-**perf**   **gen** that
        'I'm not the one who made/did that'

Another type of distinction among negative markers that is found in a several AN languages in island Southeast Asia has already been noted in passing: the fusion of tense with negation. Negative markers which differ in tense sometimes show partial phonemic similarity that suggests a historical morpheme boundary, as withYakan *gaʔ* 'negative marker for past events' : *gaʔi* 'negative marker for future events' (Brainard and Behrens 2002:123). In most languages, however, these markers show no morphological relationship

to one another, and so presumably did not arise historically from the fusion of tense and negation morphemes, as with the Bontok markers *adi*, and *əgʔay* cited above. A particularly clear-cut example of morphologically unrelated tensed negative markers is seen in Sarangani Manobo of the southern Philippines, which lacks a nominal/verbal distinction in its negative markers, but distinguishes tense in *wədaʔ* 'did not' and *əkəd* 'will not' (Dubois 1976:20):

(7.111)   t-im-ədogi   sə   bayi
          sleep-**sf.perf** **foc** woman
          'The woman slept'

(7.112)   t-om-ədogi sə   bayi
          sleep-**sf**   **foc** woman
          'The woman will sleep'

(7.113)   wədaʔ   tədogi sə   bayi
          neg.**past** sleep   **foc** woman
          'The woman did not sleep'

(7.114)   əkəd   tədogi sə   bayi
          neg.**fut**   sleep   **foc** woman
          'The woman will not go to sleep'

Dubois (1976:132) recognises four tenses in Sarangani Manobo, which he calls 'past', 'future', 'neutral' and 'iterative'. In 7.11) and 7.112 tense is expressed by verbal affixation marking past and future forms of the subject focus. In 7.113 and 7.114 the verb is temporally neutral, and the expression of tense is carried by contrasting negation markers. It is unclear from the available data how present tense forms would be expressed. A similar use of tensed negatives is found in Cebuano Bisayan, where *díliʔ* negates future action, and *waláʔ* past action (Wolff 1966:43), and in Saisiyat of northern Taiwan (Yeh, Huang, Zeitoun, Chang and Wu 1998).

In order to show more explicitly how patterns of negation in the AN languages of insular Southeast Asia differ from those in Oceanic languages, Table 7.16 adopts the format in Mosel (1999), and applies it to thirteen languages of Taiwan, the Philippines and western Indonesia (Sambal = Botolan Sambal):

Table 7.16 Patterns of negation in thirteen non-Oceanic Austronesian languages

	1 NO!	2 NEG- EXIST	3 NEG- NON-V.	4 NEG- V.	5 NEG- IMPER.
Atayal	(i)yat, iniʔ	iŋat	(i)yat	iniʔ	laxi
Thao	ani	uka	ani, antu	ani, antu	ata
Ifugaw	adí, bokon, ugge	maid	bokon	adí, ugge	???
Sambal	ahəʔ, alwa	ayin	alwa	ahəʔ, ag-	ag-mo
Tagalog	hindíʔ	waláʔ	hindíʔ	hindíʔ	huwág
Cebuano	díliʔ, waláʔ	waláy	díliʔ	díliʔ, waláʔ	ayáw
Yakan	gaʔ/gāʔ	gaʔ	dumaʔin	gaʔ, gaʔi	daʔa
Urak Lawoi'	hoy, bukat tet	hoy	bukat	tet	jaŋan
Indonesian	tidak, bukan	tidak ada	bukan	tidak	jaŋan
Muna	miina, paise	miina	suano	miina, pa, pae, paise	ko, koe, koise
Kambera	nda	nda niŋu	nda	nda	àmbu
Taba	te	te	te	te	oik
Palauan	diak	diak	diak	diak	lak

Tables 7.14 and 7.16 show a clear difference of patterning. While more than half of the languages in Table 7.14 have a configuration A : A : A : A : B (only the vetative differing), in Table 7.16 this pattern is rare. Two factors are mainly responsible for this difference. First, while only one Oceanic language (Nêlêmwa) is reported with a distinct negative existential, it is fairly common in the AN languages of Taiwan, the Philippines and western Indonesia to represent the negative existential by a distinct form, often reflecting *wada, which carries the contradictory senses 'be, have' and 'not be, not have' in different languages (e.g. Ilokano *wadá* 'have, be; there is, there are', Tagalog *waláʔ* 'none, nothing, absent'). Second, as already noted, although some Oceanic languages negate nominal and verbal constituents differently, every language that shows this pattern also has an option in which nominal and verbal constituents are negated in the same way. Moreover, in Oceanic languages the differences in nominal and verbal negation appear to be morphological. By contrast, half of the languages in Table 7.16 show different negation patterns for nominal and verbal constituents, and these usually are expressed lexically. It is possible that the description of these patterns of negation as associated with nominal or verbal constituents is misguided. Sentence examples in the published sources suggest that negators of nominal constituents are often contrastive negatives. Brainard and Behrens (2002:121), for example, illustrate Yakan *dumaʔin* with sentences such as 'It is not Sadda'alun who is going to Isabela, but Toto', or 'It is not fish that she will buy, but vegetables,' and similar examples can be found in other languages.

There is currently no known basis for positing a PAN distinction of negatives marking nominal and verbal constituents. The history of most negative markers is quite obscure, but there are two exceptions. Lampung *lain* 'negator of nominals' clearly is cognate with Malay *lain* 'other, different', and similar non-negative forms in other languages. Similarly, *beken is reflected in a number of languages in the meaning 'other, different' (both Bontok *bakən* and Malay/Indonesian *bukan* appear to continue this form, with unexplained irregularities in the first vowel): Isneg *bak-bakkán* 'another kind of; different', Aborlan

Tagbanwa, Kelabit *bəkən*, Kapuas *beken* 'other, different', Ngaju Dayak *beken* 'different, be distinct from; another'), and in others either as a generic negative, or as a negator of nominals (Gaddang *bəkkən* 'generic negative', Isneg *bakkán* 'no, not', Ifugaw *bokón* 'restrictive negation; as a verb it means 'refuse', 'not want', 'reject', etc', Hanunóo *bukún* 'an emphatic negative…most frequently used in statements of denial or contradiction', Tiruray *bəkən* 'not', Tausug *bukun* 'negator of nominals'). Here the evolution seems clearly to have been from a morpheme that originally meant 'other, different' to a negator of nominal expressions. The same semantic history appears to apply to the etymologically more obscure Central Tagbanwa *bèlagiŋ/bèlahiŋ* since, according to Scebold (2003:82) "The word *bèlagiŋ* is actually the combination of the words *bèlag* and *iŋ*. The precise meaning of *bèlag* isn't clear. But it seems to mean 'different from' or 'not'." Given the rarity of a nominal/verbal split in the negation patterns of CEMP languages, it seems likely that the negative meaning of *beken* in the languages of the Philippines and western Indonesia is a relatively recent development that may be the outcome of recurrent parallel change (drift). What remains unclear is why there has been such a strong historical tendency in the more westerly part of insular Southeast Asia to innovate nominal negators from words meaning 'other, different', since there is very little evidence of such a tendency in the Oceanic languages, or (so far as the available evidence shows) in the languages of eastern Indonesia.

7.4.1 Double negatives

The use of double, or bipartite negatives is not common in AN languages, but appears in a few languages. Klamer (1998:143) reports two types of double negatives in Kambera of eastern Sumba:

(7.115) àmbu   bobar   ndoku   -ma   -ya
        neg.irr preach   **neg**   **emp**   **3sg.acc**
        'Do ***not*** talk about it!'

(7.116) nda   niŋu   ndoku
        neg   be   **neg**
        'There are none/I have none'

Klamer describes *nda* as a general marker of negation, and *àmbu* as marking an irrealis negative (glossed 'won't') or a vetative (glossed 'don't'). Onvlee (1984) glosses *ndoku*, as 'mistake, error; wrong, mistaken'. In addition, Mosel and Spriggs (1999:46) note that Teop, an Oceanic language spoken on the island of Bougainville in the western Solomons, has a double negative *saka … haa*, which they describe as a "double particle, negating verbal and non-verbal predicates and cleft noun phrases." The history of this construction is obscure. To date, then, double negatives have been reported for AN languages in eastern Indonesia and the Pacific, but not for those in Taiwan, the Philippines or western Indonesia.

7.4.2 Emphatic negatives

The gloss of the Kambera sentence 'Do ***not*** talk about it!' might be taken as evidence that double negation is emphatic, but this does not appear to be the case, since the

emphasis placed on the negative marker evidently is coded in the postclitic emphatic marker *-ma*. Similar types of emphatic negation are reported for some other languages, as Sye of southern Vanuatu, where the contrast between sentences glossed as 'I did not walk' and 'I did not walk at all' or 'I did not know you' and 'I did not know you at all' is carried by an emphatic suffix *-hai* (Crowley 1998:106), or in Pohnpeian, where the contrast between sentences glossed as 'That man is not a teacher' (using *kaide:n* 'negative') and 'That man is not a teacher!' (using *kaide:nte* 'emphatic negative') is carried by the sentential adverb *-te* (Rehg 1981:326). In a few languages, however, a contrast of simple and emphatic negatives is coded in the negative marker itself, as in Bontok of northern Luzon, where Reid (1976) gives *adí* 'no; not; negative of verbs and adjectives', but *adʔí* 'emphatic negative of verbs and adjectives'. Although the available sources are silent on this point, it is likely that the apparent synonyms in such languages as Isneg *adí*, *addí* 'not' express (or once expressed) a similar contrast.

7.4.3 Negative verbs

In a number of AN languages the negative marker may be inflected as a verb**.** This appears to be particularly common in Philippine-type languages. Thao, of central Taiwan, for example, has general negative markers *ani* and *antu* which appear in many cases to overlap in function:

(7.117)   ani/antu   yaku   tu   Caw
        neg   **1sg** *tu* Thao
        'I'm not Thao'

(7.118)   ani   yaku   sa   Shput
        neg **1sg** *sa* Chinese
        'I don't want Chinese people (e.g. as company)'

(7.119)   antu   yaku   Shput
        neg **1sg** Chinese
        'I'm not Chinese'

Despite this overlap in distribution, *ani* takes a number of different affixes to form negative verbs, while this reportedly is impossible with *antu*, which is never affixed:

(7.120)   ani-wak   tu   a   m-untal
        neg-1sg *tu* **fut** **av**-follow
        'I don't want to come with you' (**antu-wak)

(7.121)   maka-ani   cicu,   numa   m-usha
        maka-**neg** **3sg**   hence   **av**-go
        'He didn't like it, so he left' (**maka-antu)

(7.122)   minu   cicu   pish-ani
        why   **3sg** *pish*-**neg**
        'Why did she deny it?' (**pish-antu)

Similar morphological elaborations of negative markers can be found in many Philippine languages and some of the languages of western Indonesia, as with Isneg *adí*, *addí* 'no, not' : *max-adí* 'to separate, divorce' : *um-addí* 'to refuse, not to like', Itawis *awán* 'no!; not exist, not have' : *m-aw-áwan* (< ma-awa-awan) 'get lost' : *maŋ-aw-áwan* 'lose something', Tiruray *ʔəndaʔ* 'none, not any, not' : *fə-ʔəndaʔ-əndaʔ-ən* 'discount completely, give up all hope for something', or Tae' (south-central Sulawesi) *taeʔ* 'no, not; not be or have' : *maʔ-taeʔ* 'tell someone that there is none' : *ka-tae-ran* 'the one that is missing or lacking'.

7.4.4 Negative personal pronouns

Mosel (1999:4-5) notes that negative determiners, prepositions and conjunctions are found in Oceanic languages, and proposes that these be added to the lexical categories that are recognised in the general typological literature as taking negation (verbs, auxiliaries, particles, affixes, nouns, quantifiers, and adverbs). Several Philippine languages also have what are described as negative personal pronouns. These are reported for Pangasinan of central Luzon, and for Central Tagbanwa of Palawan Island in the central Philippines. In Pangasinan according to Benton (1971) a sentence may be negated by prefixing *ag-* to the verb, or to a subject or attributive pronoun standing before the verb. Note the following positive sentences (a) and their negative counterparts (b):

(7.123) a. antá   nən   Pedro   ya   wadiá   ka
             know   **gen** Pedro **lig** here   **2sg**
             'Pedro knows that you are here'

         b. ag-antá   nən   Pedro   ya   wadiá   ka
             neg-know **gen** Pedro **lig** here   **2sg**
             'Pedro doesn't know that you are here'

(7.124) a. táwag-ən   ko   ra
             call-**pv** **1sg** **3p**
             'I'll call them'

         b. ag-ko   ra   táwag-ən
             neg-1sg **3pl** call-**pv**
             'I won't call them'

(7.125) a. maŋ-asawá   ak   la
             av-marry **1sg** already
             'I'll be getting married (already)'

         b. ag-ák   ni   maŋ-asawá
             neg-1sg yet   **av**-marry
             'I'm not getting married yet'

Ifugaw of northern Luzon distinguishes negators of verbs (*adí, ugge*) from negators of nouns and pronouns (*bokon*); use of the latter is sometimes called 'exclusive' or

'restrictive' negation. In this language both restrictive and non-restrictive negative markers may be inflected with a clitic pronoun, representing the actor with verbal negators, and the patient with nominal negators:

(7.126) adí-m   i-ad-ʔadí
         neg-2s   i-red-neg
         'Do not (always) forbid it'

(7.127) ugge-ak   im-m-ali
         neg.past-1sg   perf-av-come
         'I did not come'

(7.128) bokón-ak
         neg-1sg
         'It isn't me'

In Central Tagbanwa the general negation marker is *data*. When followed by *na* 'recently completed action' these two morphemes fuse into *dana* 'no longer'. When *data* is followed by *ako* '1sg' the negative marker and pronoun fuse to form *dako*, and the negative-pronoun unit is then followed by *na* (Scebold 2003):

(7.129)   dako   na   man-luak   kaito   ka   patag
          neg-1sg   now   **av**-farm   here   **obl**   plain
          'I no longer farm here in the plain'

7.4.5 Responses to polar questions

Responses to polar questions are discussed in a few grammars, and in most cases these appear to be the opposite of responses in languages such as English, where the reply to a negative question requires the affirmation of its contrary. In English, for example, 'Are you hungry?' and 'Aren't you hungry?' both receive the affirmative reply 'Yes, I'm hungry', whereas in languages such as Indonesian or Pohnpeian the affirmative interrogative must be answered with 'yes' and the negative interrogative with 'no' when the proposition is being affirmed, and the affirmative interrogative must be answered with 'no' and the negative interrogative with 'yes' when the proposition is being denied. This is schematised in Table 7.17:

Table 7.17 Responses to polar questions in English and some Austronesian languages

	Positive Q	**Negative Q**
English	+ = yes	+ = yes
	− = no	− = no
Austronesian	+ = yes	+ = no
	− = no	− = yes

As seen in Table 7.17, for positive interrogatives AN languages show a response pattern like that of English, but for negative interrogatives it is the opposite: English speakers affirm a negative question with 'yes' and deny it with 'no', while speakers of at least some

AN languages affirm a negative question with 'no' and deny it with 'yes'. Macdonald and Soenjono (1967:251) give the following examples for Bahasa Indonesia:

(7.130) Q: Siti tidak pulaŋ? (Siti **neg** go.home) 'Didn't Siti go home?'
        A: Ia 'yes' = 'No, she didn't' (i.e., 'Yes, it is true that Siti did not go home')
        Tidak 'no' = 'Yes, she did' (i.e., 'No, it is not true that Siti did not go home)

For Pohnpeian, Rehg (1981:329) gives the following examples, noting that both responses indicate that the speaker is hungry. However, the positive question receives a positive response and the negative question a negative response. As in Indonesian, then, the negation of a negative is used to express a positive, whereas in English the reply to a negative question requires the affirmation of its contrary:

(7.131) Q: Ke menmʷeŋe? 'Are you hungry?'
        A: Ei, i menmʷeŋe 'Yes, I am hungry'
        Q: Ke sou menmʷeŋe? 'Aren't you hungry?'
        A: Sou, i menmʷeŋe 'No, I am hungry'

7.4.6 Negative affirmatives

One last feature of negative constructions can be mentioned here, not because it is typical, but because it is exceptional, and so shows the range of phenomena that can be found associated with negation in AN languages. Durie (1985:269) reports that in Acehnese of northern Sumatra "It is common … to use a negative exclamatory sentence to imply a positive meaning," as in:

(7.132) kön   bit   baŋay=keuh   that
        neg really stupid=**2s**   very
        'You are really stupid!' (lit. 'You really aren't stupid!')

(7.133) bôh   h'an   ka=pumuntah   dilee
        do!   **neg**   **2s**=undercook   now
        'You'll undercook it—so don't!' (lit. 'You won't undercook it, so do it!')

Durie calls these 'negative affirmatives', but they could just as well be called 'ironic negatives', since they appear to exploit irony to achieve greater effect.

7.5   Possessive constructions

In the AN languages of Taiwan, the Philippines and western Indonesia possessive relationships generally are simple and uninteresting, but in Oceanic languages the situation is quite different, since a fundamental distinction between obligatorily or inalienably possessed nouns, and alienably possessed nouns is commonplace. Morever, the category of alienable possession often carries with it an implicit notion of intended use. Standard (Bauan) Fijian as described by Schütz (1985) illustrates the contrast. Possessive constructions 1)—4) are inalienable, 5)—9) are alienable, and 10) may be either:

Table 7.18 Possession marking in Bauan Fijian

1)	na	tama-na		2)	na	luve-na	
	art	father-**3sg**			**art**	offspring-**3sg**	
	'his/her father'				'his/her offspring'		

3)	na	ulu-na		4)	na	boto-na	
	art	head-**3sg**			**art**	bottom-**3sg**	
	'his/her head'				'its bottom'		

5)	na	no-na	vale	6)	na	ke-na	dalo
	art	**poss.gnr-3sg**	house		**art**	**poss.ed-3sg**	taro
	'his/her house'				'his/her taro'		

7)	na	me-na	moli	8)	na	vale	ne-i	Jone
	art	**poss.dr-3sg**	citrus		**art**	house	**poss.gnr-pm**	John
	'his/her citrus fruit'				'John's house'			

9)	na	dalo	ke-i	Jone	10)	na	vale	ni	kana[76]
	art	taro	**poss.ed-pm**	John		**art**	house	**gen**	eat
	'John's taro'					'restaurant' (lit. 'house of eating')			

The most thoroughgoing treatment of possessive constructions in Oceanic languages is that of Lichtenberk (1985), who notes (94) that "A possessive construction may or may not express true possession." In addition to encoding ownership ('My car') the same type of construction may express part-whole or kinship relations ('My hand', 'My father'), or even involvement in an event ('John's arrival'). Lichtenberk is careful to distinguish the formal criteria used to identify possessive constructions from the semantic criteria used to describe the type of relationship between possessor and possessed. In this classification there are three types of possessive constructions in Oceanic languages: direct possession, indirect possession, and prepositional possessive constructions.

7.5.1 Direct possession

In most Oceanic languages direct possession applies to body part and kin terms, the words for 'name' and 'shadow/spirit' (seen as parts of one's personal identity), and to almost anything viewed as having a part to whole relationship in either a physical sense (body parts, leaf, branch or fruit of a tree, etc.), a social sense (kin terms), or a general relational sense (the flight of an arrow). Nouns in these semantic domains are often difficult to elicit in isolation, and are commonly offered with an attached possessive pronoun (usually the third person singular). In some languages native speaker reaction suggests that attempts to elicit terms like 'eye' or 'hand' as a bare stem will be conceptualised as though the referent is disembodied. The same resistance to decontextualisation evidently is true of kin terms, but here the visual imagery that so

---

[76]   Lichtenberk (1985) calls this an 'associative construction', a type of configuration in which the alienable/inalienable possessive distinction is neutralised.

clearly distinguishes possessed from non-possessed body part terms is less transparent. Because the relationship expressed by direct possession is one in which the possessor is so intimately bound together with the possessed that a conceptual separation is difficult, it is often called 'inalienable possession'.

   Although an alienable : inalienable distinction is nearly universal in Oceanic languages, the membership of these contrasting classes varies cross-linguistically. Lichtenberk (1983:278ff) states that in Manam inalienable possession (*matá-gu* 'my eye', etc.) encompasses body parts (including fluids and discharges), parts of wholes (branch of a tree, mango juice, etc.), kinship terms, including terms for friends and trading partners, psychological states ('I am afraid' = 'My fear is bad', 'I am angry' = 'My inner feeling is bad'), verbal nouns expressing events or states, and (typically verbal) nouns which express the properties or characteristics of objects, including the customary manners of performing events. In other languages, however, the class of inalienably possessed nouns may only partially overlap with this. Thus, for Loniu of the Admiralty Islands, Hamel (1994:29) gives *ŋah* 'lime' : *ŋaha* + *w* = *ŋoho* 'my lime', *ŋaha-m* 'your lime', *ŋaha-n* 'his/her lime', p^w*ahacan* 'route' : p^w*ahacala* + w = p^{w}ahacɔlɔ 'my route', p^w*ahacala-m* 'your route', p^w*ahacala-n* 'his/her route', and *cim* 'purchase' : *cima* +*w* = *cimɔ* 'my purchase', *cima-m* 'your purchase', *cima-n* 'his/her purchase', nouns that would not be expected *a priori* to be inalienably possessed. In some cases an explanation emerges from a fuller understanding of the cultural context (betel chewing is a frequent activity in the Admiralty Islands, and the lime that forms an integral part of the betel chew is habitually carried by the side in a gourd container), while in others it may be more elusive, as with 'route' or 'purchase'. In Kosraean of Micronesia virtually all body parts are inalienably possessed. Lee (1976), who is thorough in indicating the possessive category of nouns, however, marks *fohk* 'feces' as inalienably possessed, but *kof* 'urine', *acni* 'spit (n.)', *fiyoh* 'sweat', and *uswacnwen* 'pus' as alienably possessed. This suggests that in Kosraean bodily secretions generally take a different form of possessive marking than body parts. In other languages obligatorily possessed nouns and inalienably possessed nouns that allow the base to be expressed without a possessive suffix must be distinguished. In some languages of the eastern Admiralty Islands, for example, most body part terms must occur with a possessive suffix, but this does not appear to be necessary for body fluids. Table 7.19 gives the free and singular possessed forms of nine nouns in Lou of Lou Island, and Lenkau of Rambutjo Island:

Table 7.19 Free and singular possessed forms of nine nouns in Lou and Lenkau

Free base	Lou	
(none)	moro-ŋ : moro-m : mara-n	'eye'
(none)	tio-ŋ : tio-m : tia-n	'belly'
(none)	tino-ŋ : tino-m : tina-n	'mother'
(none)	noru-ŋ :noru-m : noru-n	'child'
tur	ture-ŋ : turɪ-m : turɪ-n	'blood'
mimiya	?	'urine'
te	?	'feces'
porak	?	'pus'
roŋus	?	'snot'

Free base	Lenkau	
mara-n	moro-ŋ : moro-m : mara-ni	'eye'
tria-n	trio-ŋ : trio-m : tria-ni	'belly'
?	trino-ŋ : trino-m : trina-ni	'mother'
notr	notra-ŋ : notro : notri	'child'
troh	troh heno-ŋ : troh heno, troh heni	'blood'
mimiya	?	'urine'
tre	?	'feces'
pohoan	?	'pus'
trow	?	'snot'

   Although the material in Table 7.19 is based on fieldnotes of limited scope, and contains some gaps, certain patterns emerge with reasonable clarity. In Lou, body part and kin terms could not be elicited in isolation, but body fluids could. Where possessive data is available, as with the word for 'blood', it is clear that body fluids also take the suffixed pronouns marking inalienable possession, but they appear to be free from the conceptual impediment of decontexualised body parts or kin terms, since body fluids may be observed in isolation. In Lenkau the situation is different, since some body-part terms were given as 'free forms', but these correspond historically to the third person singular possessed form (*mara-n* < *mata-ña, etc.), which has been replaced in the modern language by an innovative suffix *-ni*. Moreover, some kin terms, including 'child' (*notr*), 'grandfather' (*pʷapʷaw*), 'grandmother' (*pʷepʷew*), and 'mother's brother' (*caca*) were offered without a historical third person singular suffix, and two words (*troh : troh heno-ŋ : troh heno, troh heni, caca : caca raŋ : caca ro : caca ri*) were offered as free forms that take inalienable possessive markers, but these are attached to a separate word which follows the possessed noun. No information is available as to what these separate words mean, or what their function in the possessive system might be. Finally, as noted above, some languages of the Admiralty Islands allow direct possession with a number of alienable nouns, as in Loniu *puriya-n* 'his/her work' (Hamel 1994:94), a situation that is distinct from the patterns observed in most other Oceanic languages.
   The fundamental notion of inalienable possession is that of an inseparable union of part and whole, where it makes little pragmatic sense to refer to or conceive of the part in isolation from the physical, psychological or social context that gives it meaning. Although inalienable possession is almost always marked by direct suffixation to the possessed noun, a few languages prepose the possessive pronoun, at least in some persons and numbers, as

in Wayan (Western Fijian) *ŋgu-ulu* 'my head', *mu-ulu* 'your head', *ulu-ya* 'his/her/its head', *dra-lima* 'their hands', *o tama-dra* 'their fathers'. In many Micronesian languages obligatorily possessed nouns may be cited without a possessive pronoun, but in this case they take what is commonly called the 'construct' suffix, a reflex of POC *ni* 'genitive', as seen in Chuukese *masa-n* 'his eye' (< POC *mata-ña), but *mese-n* 'eye of' (< POC *mata ni). The use of construct forms thus underlines the point that direct or inalienable possession is conceived as a part to whole relation, since where no explicit indication of the relationship of part to whole is given, a generic genitive serves to mark nouns as being part of something larger.

7.5.2 Indirect possession

As seen at the beginning of this section, the Fijian possessive forms for 'his/her house', 'his/her taro' and 'his/her citrus' differ both from directly possessed nouns, and from one another. First, while inalienable possession is signaled by suffixing possessive pronouns directly to the possessed noun, alienable possession is signaled by the suffixation of possessive pronouns to a possessive relation marker that precedes the possessed noun. In the literature these relation markers go by a variety of names. For Fijian Milner (1967) called the set of possessive relationships a 'gender' system, with separate marking for neutral, edible, drinkable, and familiar 'gender'. Schütz (1985:446) refers to the entire system as one of 'possession', calling forms such as *no-na*, *ke-na*, and *me-na* 'attribute possessors,' and with regard to the Wayan language of western Fiji Pawley and Sayaba (2003) refer to these elements as 'prenominal particles' or 'possessive markers'. For most languages of Micronesia these preposed elements have been called 'possessive classifiers', a term that was first proposed by Lichtenberk (1985), who drew attention to parallels between them and the typologically better-known numeral classifiers found in many language families. Perhaps the most striking feature of this system in Standard Fijian is that alienable possession must be distinguished as neutral/general, edible or drinkable. Since most possessed nouns are neither edible nor drinkable, the neutral possessive marker *no-* occurs with higher list frequency than either *ke-* 'marker of edible possession' or *me-* 'marker of drinkable possession'. Whether higher list frequency agrees with higher text frequency is unknown, but it is clear that edible and drinkable possessive relationships are very common in spoken Fijian.

Many languages permit some nouns to be marked for more than one type of possession. In Loniu of eastern Manus, for example, Hamel (1994:48) notes that the base *pihin, pihine-* 'female, woman' may be inalienably possessed as *pihine-n* (female **3sg)** 'its female (of species)', or alienably possessed as in *hetow pihin a yo* (**3paucal** woman **poss 1sg**) 'my women', and she notes that 'the more mutable type of possessive relationship may be indicated by the use of the alienable possessive phrase, and the relationship which is not likely to change is expressed by the inalienable possessive.' On the other hand, in Fijian and many other Oceanic languages alienably possessed nouns take neutral, edible, or drinkable possessive markers to signal nuances of relationship between possessor and possessed: *na no-dra dalo* 'their taro (e.g. for selling)' : *na ke-dra dalo* 'their taro (to eat)', *na no-daru dovu* 'our (dual) sugarcane (e.g. for selling)' : *na me-daru dovu* 'our sugarcane for drinking' (since one sucks the juice from the stalk and ejects the fiber). Only a neutral vs. edible distinction is common to most of western Melanesia, but Lynch (1996:109), reconstructs six possessive classifiers for Proto Oceanic, namely *ka-* 'food', *ma-* 'drink', *na/a* 'general (definite?)', and *ta/sa* 'general (indefinite?)', with the suggestion that they

have developed from articles.[77]   No language is known to reflect more than half of these forms, and some languages have only a single possessive classifier, as Kwaio of the southeast Solomons, which contrasts inalienable and alienable possession, but marks the latter by suffixing a possessive pronoun to a possessive classifier *a-*: *nima-na* 'his hand', *susu-na* 'her breast' vs. *ʔifi a-na* 'his house', *susu a-na* 'its breast (i.e. the breast an infant suckles at)'.

While some languages have reduced the system of possessive classifiers to just one, others have expanded it. In the Lolovoli dialect of the Northeast Ambae language in Vanuatu, for example, the classifiers *ga-* 'food possession', *me-* 'drink possession' and *no-* 'general possession' are functionally similar to, and clearly cognate with Fijian *ke-*, *me-* and *no-*. However, this set of forms has been expanded by the addition of *bula-* 'natural or valued object possession', which Hyslop (2001:178) describes as relating primarily to ownership of animals and plants. Other languages of northern Vanuatu also have an innovative classifier marking prized possessions (pigs, chickens, and more recently cars and radios). Nowhere has the expansion of the inherited set of possessive classifiers gone so far as in Micronesia. Here a great many independent nouns may also be used as classifiers in possessive constructions. Harrison (1976:130) lists fourteen of "the most common possessive classifiers" of Mokilese in their 3sg forms, as follows: *ah* 'his thing', *nah* 'his child, pet, valuable', *kanah* 'his food', *nimah* 'his drink', *ŋidah* 'his chaw', *warah* 'his vehicle', *imʷah* 'his house', *mʷarah* 'his garland', *dapah* 'his ear decoration', *siah* 'his earring', *kiah* 'his mat', *japʷah* 'his land', *upah* 'his sheet' and *wiliŋah* 'his pillow'. One of the distinctions that this system encodes is reminiscent of the 'prized possession' marking of northern Vanuatu:

(7.134)   oai   wusso
          1sg.poss   banana.tree
          'My banana tree'

(7.135)   noai   wusso
          1sg.poss   banana tree
          'My banana tree that is particularly valuable to me'

Many other Micronesian languages have a similarly large set of possessive classifiers, although the categories they encode show considerable variation apart from the core notions of 'food', 'drink', 'house', 'vehicle' and 'general'.

7.5.3 Prepositional possessive constructions

The third type of possessive construction that Lichtenberk (1985) recognises expresses the possessive relationship with a preposition, as in the following example from Babatana in the western Solomons:

---

[77]   Palmer and Brown (2007) argue instead that the possessive classifiers of Oceanic languages are directly possessed nouns that are the heads of indirect possessive constructions. Lichtenberk (2009) presents convincing arguments against this analysis, holding instead that these syntactic elements "form a category of their own" (Lichtenberk 2009: 379).

(7.136) pade   ta   mamalata
          house   of   uncle
          (My) uncle's house

He notes that there is a strong tendency for such constructions to express alienable possession, although they may express either type of possessive relationship. It is unclear whether this type of construction expressed possession in POC, although it clearly expressed part-whole relationships, as in *raqan ni kayu 'branch of a tree'.

7.5.4 Proto Polynesian innovations

In Proto Polynesian the inherited Oceanic possessive system underwent a basic restructuring. The properties of this system are described in many individual grammars, but the most complete account of the history and typology of possessive marking in Polynesian languages is provided by Wilson (1982). Wilson (1982:35ff) notes that possession is indicated in some Polynesian languages by direct suffixation in a small number of kin terms. This pattern is interesting for several reasons. First, it occurs only in the Polynesian Outlier languages Mae, Rennellese, Pileni, Mele-Fila, Tikopia, and West Futunan. Second, it is restricted to singular possessive suffixes. Third, it is found in about half a dozen kin terms, but is unknown in body part terms, or other parts of wholes. Fourth, Rennellese has only two suffixed possessive markers: *-u* '1sg, 2s', and *-na* '3sg'. Fifth, in most Rennellese kin terms *-na* is obligatorily attached to the independent form of most kin terms, as with *te tama-na* (**art** father-**3sg**) 'the father' (*te tama-u* 'my father/your father'), or *te tina-na* (**art** mother-**3sg**) 'the mother' (*te tina-u* 'my mother/your mother'). However, the word for 'mother's brother' (built from *tuʔaa* + *tina* 'mother') has an independent form without *-na*: *te tuʔaatina* 'uncle' : *te tuʔaatina-u* 'my uncle/your uncle', *te tuʔaatina-na* 'his uncle' : uncle'. Given this data Pawley (1967:262) and Wilson (1982:35) concluded that Proto Polynesian made limited use of possessive suffixes in some kin terms, but that these suffixes have fossilised in the Polynesian Outliers of Micronesia, and in all languages of triangle Polynesia: *tina-na > Kapingamarangi *dinana* 'mother', *tama-na > *damana* 'father, uncle', *tahi-na > Hawaiian *kaina* 'younger same sex sibling', *tuaka-na > *kuaʔana* 'older same sex sibling', *makupu-na > *moʔopuna* 'grandchild', *tupu-na > *kupuna* 'grandparent; ancestor'. This distribution shows an ongoing pattern of reduction of the inherited POC system of possessive marking by direct suffixation, and it may not be accidental that the only Polynesian languages that preserve traces of this system are those spoken in Melanesia, where direct possessive marking on kin terms has generally been retained.

Everywhere else in Polynesia, and everywhere else in the lexicon of Polynesian Outlier languages such as Rennellese, the inherited POC system of possessive marking was replaced by an innovative system that is variously described as one of 'dominant' vs. 'subordinate' possession, or of 'A' vs. 'O' possession. Possessive pronouns in the 'A' class mark relationships in which the possessor is dominant/in control, or acts as an initiator, while those in the 'O' class mark relationships in which the possessor is subordinate/not in control, or has not acted as an initiator. Elbert and Pukui (1979:139-140) illustrate this contrast in Hawaiian with examples such as the following (**dp** = dominant possession. **sp** = subordinate possession):

(7.137) ka   leo   a   Pua
      art sound **dp** Pua
      'The tune composed by Pua; Pua's command'

(7.138) a   leo   o   Pua
      art sound **sp** Pua
      'Pua's voice'

(7.139) ka   iʔa   a   kākou
      art fish **dp** **1in.pl**
      'Our fish'

(7.140) ka   iʔa   o   keia   wahi
      art fish **sp** this   place
      'The fish of this place'

(7.141) ka   wahine   a   ke   aliʔi
      art woman   **dp** **art** chief
      'The wife of the chief'

(7.142) ka   wahine   o   ka   lua
      art woman   **sp** **art** pit
      'The woman of the pit' (Pele)

The history of the *a/o* type possessive distinction is not entirely clear. Although it is most fully developed in Polynesian languages, a semantically parallel distinction is found in some other OC languages. Thus in Kove of west New Britain nouns may be possessed in any of three ways: 1. directly with a possessive pronoun, 2. by suffixing the possessive classifier *a* with a possessive pronoun and preposing this whole word to the possessum, or 3. by suffixing the possessive classifier *le* with a possessive pronoun and preposing this whole word to the possessum. Sato (n.d.) describes the semantic contrast between the latter two categories as one of undergoer (*a*-class possessor) vs. agent (*le*-class possessor):

(7.143) a-ghu   ninipuŋa
      clas-**1sg** story
      'My story' (a story about me)

(7.144) le-ghu   ninipuŋa
      clas-**1sg** story
      'My story (a story which I tell)

There is some evidence of similar distinctions of possessive marking in other AN languages, but these are much less fully developed than the systems of possessive marking in typical Oceanic languages. Most CMP languages distinguish what is sometimes called 'inalienable' from 'alienable' possession. However, as noted by Laidig (1993), many nouns may be possessed in either way, making the distinction less rigid than is true of most Oceanic languages. Typically, the same possessive morpheme is used to mark both types of possessive relationship, and the distinction is maintained positionally, as in Paulohi

nife-u 'my tooth' : *nife-mu* 'your tooth' : *nife-ni* 'his/her tooth' vs. *u-utu* 'my louse' : *mu-utu* 'your louse' : *ni-utu* 'his/her louse' (Laidig 1993:317), or Kaitetu *mata* 'eye' : *au mata-u* 'my eye' : *ale mata-m* 'your eye' : *ini mata-ñ* 'his/her eye' vs. *au luma* 'my house' : *ale-m luma* 'your house' : *ini-ñ luma* 'his/her house' (Collins 1983a:28). In a number of the languages of central and western Borneo only one possessive paradigm is used, but body part and kin terms are often obligatorily marked with a suffix *–n* when they are not explicitly possessed, as in Kayan *tama-n* 'father of' : *tama-k* 'my father' : *tama-m* 'your father', *tama-n naʔ* 'his/her father', or *bulu-n* 'feathers, scales of fish, body hair' : *bulu*-k 'my body hair' : *bulu-m* 'your body hair' : *bulun naʔ* 'his/her body hair'.

Historically, most Kayan dialects lost final glottal stop, and added glottal stop after final vowel (Blust 2002a). The possessive pattern illustrated here appears only on bases that originally ended in a vowel, and a different possessive pattern is found on bases that originally ended in a consonant. Superficially, there is no alienable/inalienable difference: **zelaq > jəla kuy* 'my tongue', **bulu > bulu-k* 'my body hair', **buaq niuR > bua ñuh kuy* 'my coconut', **asu > asu-k* 'my dog'. What marks the alienable/inalienable distinction is the freedom of alienably possessed nouns to occur without *–n*: *asuʔ* 'dog' is common, whereas *buluʔ* 'body hair' is unusual, and far more likely to be heard as *bulu-n* if it is not suffixed for specific possession. There are two further indications that body part terms and kin terms have been historically marked with a near-obligatory *–n* in these languages. First, in the Uma Juman dialect of Kayan (Blust 1977c) some words that originally ended in **-n* have been reanalyzed as ending in *-ʔ* (reflecting earlier final vowel): **qutin > uti-k* 'my penis' (not ***utin kuy*), **ipen > ipə-k* 'my tooth' (not ***ipən kuy*). These are most simply explained as analogical back formations from a pattern in which body part terms that originally ended in a vowel required a suffixed **-n*. Second, in other Bornean languages kin terms are reflected without affixation, but exist alongside historically related non-kin terms with fossilised **-n*, as in Bario Kelabit, where **t-ama > tə-taməh* 'father (ref.)', *tamaʔ* 'father (add.)', but *taman* 'leader of a pack of animals', or **t-ina > tə-sinəh* 'mother (ref.)', *sinaʔ* 'mother (add.)', but *sinan* 'female, of animals'. Unlike typical Oceanic languages, then, in which alienable vs. inalienable is marked by the contrast of an entire possessive paradigm, in central and western Borneo the only evidence of an inalienable possessive relationship is the suffix *–n*, reminiscent of the 'construct' suffix of Nuclear Micronesian languages such as Chuukese.

Distinctions within the category of alienable possession are unknown in Borneo or in CMP languages, but do appear in a few SHWNG languages, as Buli of southeast Halmahera, which marks general possession with the possessive classifier *ni-*, and edible, or alimentary possession with *na-*. It is not known whether the same alienable noun (e.g. *pira* 'sago bread') can take both general and alimentary possessive classifiers:

(7.145)   ya-   ŋahñ-k
           1sg   name-**1sg**
           'My name'

(7.146)   ya-   ni-k   ebai
           1sg   **pc-1sg**   house
           'My house'

(7.147) ya-   na-k   pira
       1sg **pc-1sg** sago bread
       'My sago bread'

7.6   Word classes

Word classes have traditionally been problematic in many AN languages. Dempwolff (1934:28) claimed that unaffixed word bases in Tagalog, Toba Batak, and Javanese are predominantly nouns. Similarly, Schachter and Otanes (1972:62) hold that "Tagalog verbs and verb phrases are … much more noun-like than their English counterparts," and Himmelmann and Wolff (1999:17) note for Toratán of northern Sulawesi, that "The distinction between common nouns and verbs in Toratán … is less clearly developed (less clearly grammaticised) than in the European languages. That is, all open class items can appear in almost every morphosyntactic slot for open class items." Two reasons are usually given for the difficulty of distinguishing nouns from verbs in languages like Tagalog. First, as observed by Himmelmann (1991:17) in Tagalog (and by implication many other Philippine-type languages) every full word can occur in each major morphosyntactic function. On distributional grounds, then, there is no firm basis for distinguishing word classes, and for this reason some writers, as Lemaréchal (1982) have claimed that Tagalog lacks word classes altogether. Second, in languages that have an overt copula nouns and verbs can be more easily distinguished, but most AN languages lack a copula. In addition, non-subject agents and possessors are marked alike in many AN languages, so that a sequence such as Tagalog *inum-ín ni Juan* can mean either 'was drunk by John', or 'what John drank'. It is this difficulty of clearly separating nouns from verbs that made it possible for Starosta, Pawley and Reid (1982) to claim that the PAN voice system had only nominal functions.

In Philippine languages the difficulty of distinguishing nouns from verbs is encountered in both affixed and unaffixed bases. In some of the non-Philippine-type languages of western Indonesia, where word classes appear to be more clearly defined, many lexical bases that might be regarded *a priori* as verbal, turn out on distributional grounds to be nominal, but affixation generally produces unambiguous verbs. In several languages of northern Sarawak, including Kelabit and Kenyah, for example, unaffixed verb stems often appear to be nouns that acquire their verbal properties from the affix system. The following syntactic frames can be used to show that a number of Kelabit verbs are derived from nominal bases by affixation even though their inherent semantics might suggest that the bases themselves are verbal:

(7.148) bəkən   təh   siʔər   laʔih inəh
       different **emph** way.of.looking man   that
       'That man's way of looking is different'

(7.149) bəkən   təh   uit   laʔih inəh laʔal
       different **emph** way.of.bringing man   that chicken
       'That man's way of bringing a chicken is different'

Although some monovalent verbs and abrupt commands have no affix, active bivalent verbs in Kelabit carry the prefix *ŋ-*, with allomorphs *ŋ-* before vowel-initial stems, *ŋə-* before liquids and nasals, and homorganic nasal substitution before stems that begin with

an obstruent. Passive verbs are suffixed with *–ən* (generally only in questions of reason) or infixed with *–in- ~ -ən-*:

(7.150)   ŋi   iəh   niʔər   uih
         ŋi   **3sg**   **av**-see   **1sg**
         'He is looking at me'

(7.151)   sir-ən   muh   kənun   ukuʔ   nəh   inəh
         see-**pv**   **2sg**   why   dog   already   that
         'Why are you looking at that dog?'

(7.152)   s<ən>iʔər   iəh   uih
         see-**pv.perf**   **3sg**   **1sg**
         'He looked at me'

The bases *siʔər* and *uit* in the above sentences cannot be verbs 1. because they carry no verbal affix, and 2. because they are possessed by the NP *laʔih inəh*. Many other examples of possessed nominal bases representing what may appear to be inherently verbal concepts can be cited in similar syntactic frames, including *bukut* 'punching' (**av**: *mukut*), *diŋər* 'listening' (**av**: *niŋər*), *linuh* 'thinking' (**av**: *ŋə-linuh*), *pəpag* 'slapping' (**av**: *məpag*), *pudik* 'swimming' (**av**: *mudik*), or *pupuʔ* 'washing' (**av**: *mupuʔ*), *pupuʔ* 'hitting' (**av**: *mupuʔ*). In a few cases a possessed nominal may be suffixed with *–ən*, as in *bəkən təh ligət-ən iəh* (different **emph** way.of.turning **3sg**) 'His way of turning is different'; cp. *ligət-ligət təh iəh* (turn-**red emph 3sg**) 'He is turning this way and that'.

In the Highland Kenyah dialect of Long Anap possessive constructions also mark the bases of affixed verbs as being inherently nominal:

(7.153)   akeʔ   tay   m-asat
         1sg   go   **av**-walk
         'I'm walking'

(7.154)   salun   asat-ia
         slow   walking-**3sg**
         'Her walking is slow'

Klamer (1998:91-144) provides an extensive discussion of the problem of word classes in one of the few full grammars for any language of eastern Indonesia. Although she shows that Kambera nouns and verbs differ in distribution, and so can be clearly separated, she devotes a large part of her analysis to illustrating significant areas of overlap in nominal and verbal morphology. In short, the difficulty of distinguishing nouns from verbs on distributional grounds probably was a property of PAN, and this characteristic has survived most clearly in morphosyntactically conservative Philippine-type languages that permit nouns and verbs to fill most of the same morphosyntactic positions, and that lack an overt copula or diagnostic negative markers. In languages such as Kambera, or in western Indonesian or Philippine languages that use distinctive negative markers for nouns and non-nouns or verbs and non-verbs, the noun-verb distinction is clearer, although perhaps less clear than in typical European languages.

Given problems in establishing what is arguably the most fundamental word class distinction (noun vs. verb), it should not be surprising that other word class distinctions are also problematic. Ross (1998a) has drawn attention to the problem of identifying a category 'adjective' in many Oceanic languages, noting that all known posibilities are attested, namely languages with 1. an open class of adjectives, 2. no adjectives, and 3. a small closed class of adjectives. In languages of the last two types the functions that would otherwise be performed by an open class of adjectives are represented by verbs or nouns. In AN languages generally adjective-like words appear to be stative verbs, but in some languages of northwest Melanesia attributive adjectives are often affixed like possessed nouns, as seen in the following examples from Tolai of New Ireland and Tawala of southeast New Guinea (Ross 1998b):

Tolai

(7.155)   a   mapi   na   davai
          art   leaf   **lig**   tree
          'leaves of a tree'

(7.156)   a   mamat   na   vat
          art   heavy   **lig**   stone
          'a heavy stone' (= 'a heavy one of a stone')

Tawala

(7.157)   koida   poha-na
          yam   basket-**3sg**
          'basket of yams'

(7.158)   tahaya   bigabiga-na
          path   muddy-**3sg**
          'a muddy path' (= 'a muddy one of a path')

There are many other problems connected with the establishment of word classes in AN languages. As noted in Chapter 6, locative prepositions in many AN languages are morphologically complex, containing a generic locative marker (usually reflecting *i or *di), followed by a specifier, which is often an independent noun, as in Bahasa Indonesia *di-atas* 'above', *di-bawah* 'below', *di-muka* 'in front' (*muka* 'face, front') *di-bəlakaŋ* 'behind' (*bəlakaŋ* 'back; rear; hind portion'), *di-dalam* 'inside' (*dalam* 'deep; depth'), or *di-luar* 'outside'. In such languages the entire complex may be analyzed synchronically as a preposition, but historically this does not appear to have been the case, and in many of the modern languages it may prove best to recognise only the initial generic marker of location as prepositional and the following element as nominal.

Even when the same word class is recognised cross-linguistically, its membership may differ between languages. Ross (1998a), for example, reports nineteen languages that have a small class of adjectives for which full documentation is available. He divides the adjectives in 'small class' languages into categories of dimension (big, small, long/tall, short, thin, far), age (new, old, ripe), value (good, bad, true/real, beautiful), and 'other' (strong). Eight of these fourteen meanings are represented by adjectives in four or more languages: big (16), small (16), new (9), long/tall (6), old (5), good (5), short (4) and bad

(4). In most Philippine-type languages adjective-like words are classified as stative verbs, and are marked with a reflex of PAN *ma- 'stative', as in Lun Dayeh (northern Sarawak) *mə-lauʔ* 'hot', *mə-tənəb* 'cold', *mə-kəriŋ* 'dry', *mə-baaʔ* 'wet', *mə-səlud* 'smooth', *mə-bərat* 'heavy', or *mə-raan* 'light in weight'. However, a small class of adjective-like words occurs as a bare stem, as with Lun Dayeh *rayəh* 'big', *suut* 'small', *dooʔ* 'good', *daat* 'bad'. Table 7.20 maps the semantic categories represented as adjectives in Ross's 'small class' languages onto Lun Dayeh. Numbers in parentheses give number of languages in Ross's sample for which the semantic category is morphosyntactically adjectival:

Table 7.20 The marking of 'small class' adjectives in Lun Dayeh

big	(16)	rayəh
small	(16)	suut
new	(9)	mə-bəruh
long/tall	(6)	mə-kadaŋ, mə-ditaʔ
good	(5)	dooʔ
old	(5)	mə-ŋərəd (people)
bad	(4)	daat
short	(4)	mə-bənəh (height), mə-kəməʔ (length)
true/real	(2)	mə-tuʔuh
beautiful	(1)	mə-taga
far	(1)	mado (< *ma-zauq)
ripe	(1)	mə-laak
strong	(1)	mə-tuəh
thin	(1)	mə-ruguʔ (people), mə-lipi (materials)

Many other languages that mark stative verbs or adjectives with a reflex of PAN *ma- also exempt certain semantic categories from the normal pattern of affixation. This can readily be seen for a large number of languages in Reid (1971). To choose just one example, Western Bukidnon Manobo marks most stative verbs with *mə-*: *mə-ʔitəm* 'black', *mə-rigaʔ* 'red', *mə-ʔiləm* 'green/blue', *mə-laŋkəw* 'tall (people)', *mə-vəgat* 'heavy', *mə-guraŋ* 'old (people)', *mə-layat* 'long', *mə-ʔupia* 'good', *mə-zaʔat* 'bad', *mə-vavaʔ* 'short'. However, some statives are zero-marked, as with *dəkəl-aʔ* 'big' (but *mə-zakəl* 'many', both from a base *dakəl*), *dəʔisək* 'small', *bəgu* 'new', *daʔan* 'old; before', *nipis* 'thin (materials)', and *putiʔ* 'white'. Although there are unexplained idiosyncracies in the membership of these categories (why do all colour terms except 'white' take mə- in Western Bukidnon Manobo?), there appears to be a clear tendency for stative verbs or adjectives that are distributionally deviant to represent the semantic categories in Table 7.20, and particularly those near the top of the list. Since these show a high degree of correspondence with what are sometimes called the 'baby adjectives' of French and other Romance languages (big, small, good, bad, etc.) some universal determinant appears to be at work (Dixon 1977, Croft 2003).

There is a need to recognise noun classes in many AN languages. Unlike some language families, where an animate/inanimate distinction figures prominently in the grammar, the most pervasive distinction between grammatically marked noun classes in AN languages separates personal nouns from common nouns. In Tagalog this distinction for focused arguments is signaled by the contrast of *si* (singular) or *sina* (plural) + focused nominal as against *aŋ* + focused nominal: *k<um>ain si Maria* 'Mary is eating' vs. *k<um>ain aŋ*

babaʔe 'The woman is eating'. For non-focused nominals the corresponding contrast is marked by *ni* (singular) or *nina* (plural) vs. *naŋ* (marking actor, goal, or instrument complements to the verb), *kay* (singular) or *kina* (plural) vs. *sa* (marking singular locative complements), and by *para kay* (singular), *para kina* (plural) vs. *para sa* (marking benefactive complements). Table 7.21 shows the personal/common noun contrast for focused nominals in twelve Philippine languages. A few languages, as Ilokano, mark a singular/plural distinction in common nouns, but this appears to be rare (data from Yamada and Tsuchida 1983, with some additions):

Table 7.21 Markers of personal and common focused nominals
in fifteen Philippine languages

	Personal		Common
	sg	**pl**	
Ivatan	si	sira	o
Ibanag	si	ra	in
Ilokano	ni	da	ti (sg), dagití (pl)
Tingguian	si		din
Kankanaey	si		nan
Gaddang	i	da	na/yo
Pangasinan	Ø	irá di	so
Botolan Sambal	hi	hili	hay/ya
Kapampangan	i	reŋ	iŋ
Tagalog	si	sina	aŋ
Bikol	si	sa	an
Maranao	si		so

A similar personal/common noun distinction is found in a number of Oceanic languages, where reflexes of *na function as a common noun article, and reflexes of *qa as a personal article (Pawley 1972:58). Although the personal/common distinction probably is more widespread than any other for subcategorising nouns, the numerals of some languages distinguish either human vs. non-human referents (Thao *tusha wa fafuy* 'two pigs' vs. *ta-tusha wa azazak* 'two children'), or animate from inanimate referents (early twentieth century Chamorro *lima* 'five, of inanimates', *la-lima* 'five, of animates').

Given the complex and only partly regular morphology of most Philippine-type languages it should not be surprising that some writers speak of verb classes in terms of differing affix potential. One of the most thorough studies of this type is McFarland (1976). Many languages of Sulawesi which have complex morphological systems of a non-Philippine-type have also been described as having multiple verb classes based on affix potential.

7.7 Directionals

A prominent feature of the verb systems of many Oceanic languages is the use of directional morphemes as obligatory components of the VP, indicating motion toward or away from the speaker. In some languages these form part of a larger set that also includes elements meaning 'up' and 'down', as with Hawaiian *mai* 'to me; near or toward the speaker', *iho* 'downward, self; reflexive, near future', *aʔe* 'up, nearby, adjacent, adjoining,

next in space or time', and *aku* 'away, future' (Elbert and Pukui 1979:91). In expressions such as *hele mai* 'Come!', or *hele aku* 'Go away!' the verb is identical, and the difference of meaning that would be encoded in the inherent semantics of the verb in most languages is associated instead with the directional element. Similar examples can be found in widely separated Oceanic languages, as with Mussau *kasu-a mai* 'to bring' : *kasu-a laa* 'to take' (*mai* = 'hither, toward the speaker', *laa* = 'go, walk'), Rennellese *tauʔi mai* 'buy' : *tauʔi atu* 'sell', or the following examples from Seimat of the Admiralty Islands, where the venitive suffix -(V)*ma* and the allative suffix –(V)*wa* indicate directionality, but also have a benefactive or recipient sense when suffixed to the prepositional verb *hani* 'to' (Wozna and Wilson 2005:50ff):

(7.159)   i   nahi   sohot-uma   leil-i   iŋ
          3sg   walk   go.out-**ven**   inside-**gen**   house
          'He came out of the house' (speaker is outside the house)

(7.160)   i   nahi   sohot-ua   leil-i   iŋ
          3sg   walk   **o**.out-**all**   inside-**gen**   house
          'He went out of the house' (speaker is in the house)

What is striking is how deeply the use of directionals translatable as 'hither' and 'thither' penetrates the grammar of many Oceanic languages, usurping grammatical functions that are expressed by different parts of speech in most languages of the world. Elbert and Pukui (1979:91-95) note that the directionals, like pronouns, possessives and demonstratives, express relative distance between speaker and addressee in place and/or time, and this assumes a number of different forms: 1) they convey degrees of temporal distance; when followed by *nei* these are related to past time: *aku nei* 'distant past', *aʔe nei, iho nei* 'recent past', *aʔe* 'adjoining the present', *iho* 'near future', *aku* 'distant future', 2) with verbs of saying they indicate the directionality of spoken messages (**ts** = toward speaker, **as** = away from speaker) : *ʔ ī mai-la ʔoia* (say **ts**-there **3sg**) 'He said to me', *ʔ ī aʔe-la ʔoia* (say up-there **3sg**) 'He said to someone nearby', *ʔī aku-la ʔoia* (say **as**-there **3sg**) 'He said to someone else', 3) in connection with direct quotations *ʔ ī aku* means 'say to someone else', while *ʔ ī mai* appears to mean 'say back to the first speaker', 4) *aʔe* is also used to express the comparative degree, as with *maikaʔi* 'good' : *maikaʔi aʔe* 'better', 5) *mai* may function like a main verb, but without any verb markers, as in *Mai! Mai e ʔai* 'Come! Come and eat!', 6) *mai* may follow the indefinite article *he*, yet clearly function as a verb, as in *He mai! E kipa i kauhale* (**art** come **imp** visit **io** house.complex) 'Come! Visit the house!', 7) *iho* also follows such verbs as *ʔai* 'eat', *aloha* 'love', *inu* 'drink', *makaʔu* 'fear', and *manaʔo* and *noʔonoʔo* 'think', as in *ʔai iho-la ʔoia i ka puaka* (eat down-there **3sg   io art** pork) 'He ate the pork', 8) as a noun and after pronouns and locatives *iho* may mean 'self', 9) directionals with opposite meanings may occur with place names, as in *mai Honolulu mai* 'from Honolulu this way', or *mai Honolulu aku* 'away from Honolulu', 10) the directionals occur with many types of verbs, including statives, as in *Aloha mai!* 'May you be welcome here! Greetings!' or *nahā aku-la ka hale* (broken **as**-there **art** house) 'The house is broken', 11) in verbless sentences they carry meanings of coming or going, as in *I Maui aku nei au* (**loc** Maui **as past 1sg**) 'I was on Maui'. Besnier (2000:525) reports an even larger number of functions for the directionals *mai* 'hither', *atu* 'thither', *aka* 'up' and *ifo* 'down' in Tuvaluan. Among other uses, he notes 1) that the first two elements may "appear in discourse in which neither first-person

nor second-person entities are participants. In such cases the speaker chooses the participant with whom he or she identifies most closely as the deictic frame of reference," 2) that changes from sleeping to waking, from childhood to adulthood, from nonbeing or death to life, from darkness to light, from poor to good health, and from less to more desirable states are marked by *mai*, and their opposites by *atu*, 3) that *aka* 'up' and *ifo* 'down' are also used to denote 'landward' and 'seaward' respectively, 4) that *ifo* can indicate a decrease in size, intensity, or prominence, while *aka* marks the reverse, and 5) that *aka*, like *mai*, can denote changes from darkness to light or childhood to adulthood, but where there is overlap in function *mai* and *atu* are used 'to provide a more affectively charged connotation to the utterance than *aka* and *ifo* could convey.' Besnier illustrates this statement with the following contrast:

(7.161) Te   mataŋi   koo   tuku   mai
        art   wind   **inc**   let   **ts**
        'The wind is increasing (and is going to affect us)'

(7.162) Te   mataŋi   koo   tuku   aka
        art   wind   **inc**   let   up
        'The wind is increasing (and may or may not affect us)'

Even this brief survey shows how pervasive directionals are in the grammar of Oceanic languages. Moreover, their freedom of occurrence raises questions about how they are to be categorised in terms of word classes, a problem that is clearly reflected in the range of terminology used to describe them in the published literature. Table 7.22 summarises the categorisation of these elements in a number of published grammars:

Table 7.22 Descriptive terms for the directional particles in Oceanic languages

Language	Source	Term
Woleaian	Sohn (1975)	directionals
Mokilese	Harrison (1976)	directional suffixes
Hawaiian	Elbert and Pukui (1979)	directionals
Fijian	Milner (1967)	particles
Fijian	Schütz (1985)	directionals
Sye	Crowley (1998)	directional suffixes
Tuvaluan	Besnier (2000)	deictics
NE Ambae	Hyslop (2001)	directionals
Araki	François (2002)	directional adverbs
Hoava	Davis (2003)	directional verbs

What is most distinctive about this list is the general avoidance of established word class terminology to categorise these elements, which are most commonly called 'directionals', or in languages in which they are phonologically attached to a preceding verb, 'directional suffixes'. Where a commitment is made to an established word class there is little agreement (deictics, directional adverbs, directional verbs). Despite the dangers of a highly generalised cross-linguistic comparison of this kind (where even cognate forms may have rather different distributions from one language to the next), it seems clear that for the most part systems of directionals in Oceanic language share many

features in common, and one of these is the difficulty of associating these elements with any traditional part of speech.

Although most or possibly all systems of directionals in Oceanic languages may be traceable ultimately to a common ancestral type, some languages diverge widely from the better-known Polynesian type. Tinrin of New Caledonia (Osumi 1995:133ff), for example, has directional suffixes that indicates motion upwards, a bit higher, downwards, on the same level, apart or away, and dispersed hither and thither. The equivalents of *aku* and *mai* in Hawaiian are reported to be 'action verbs' *fi* 'to go away—motion from speaker' and *mê* 'to come, approach—motion to speaker.' These verbs are said to occur independently 'but they often combine with other verbs in serialisation … to clarify the direction or the place of an action with reference to the speaker' (1995:76). Similarly, Crowley (1982:157ff) reports that in Paamese of central Vanuatu, *mai* 'come', *maa* 'come up', *miitaa* 'come down', *haa* 'go', *hinaa* 'go up' and *hiitaa* 'go down' are full verbs.

Systems like this raise questions about the history of directionals in Oceanic languages. It is known that the ubiquitous *mai* in Oceanic directional systems reflects PMP *um-aRi* 'to come'. Similarly, PMP *sakay* 'to mount, ascend' is reflected in a number of Oceanic languages as the directional of upward motion. The history of the other directionals is obscure, but for at least those two for which pre-POC etymologies are known the source is an independent verb. The way in which such independent verbs of motion became subordinated to other verbs is unclear, although some type of earlier serialisation process is possible. Topping (1973:115) reports a similar set of 'motion locatives' in Chamorro, and it is noteworthy that the morphemes used to express these meanings are cognate with Proto Polynesian *mai* and *atu: Chamorro *magi* 'here—in direction of speaker', *g^watu* 'there—away from speaker'. While the Chamorro system of directionals does not appear to be as richly developed as the systems found in most Oceanic languages, the fact that the relevant morphemes for 'hither' and 'thither' are cognate suggests a history for these systems that predates Proto Oceanic. Alternatively, the Chamorro motion locatives could be a product of early contact with an Oceanic language or languages, even if the term *magi* itself is native (reflexes of *atu have no known meaning other than 'away from speaker', and their ultimate source in an independent verb remains purely speculative).

7.8   Imperatives

Imperative verbs are formed in a number of different ways in AN languages, and many languages have multiple imperative forms. In some cases there are indications that these correlate with degrees of abruptness or politeness, but in others the differences appear to be syntactic.

A number of AN languages mark the imperative with a reflex of *-i. In Thao of central Taiwan this is the most frequent form of the imperative:

(7.163)   kaiza   ihu   pa-kan   ranaw
        when   **2sg**   **caus**-eat   chicken
        'When did you feed the chickens?'

(7.164)   pa-kan-i   uan   ranaw
        caus-eat-**imp**   please   chicken
        'Please feed the chickens'

Several widely separated languages, including Atayal of northern Taiwan and Bikol of the central Philippines, reflect an alternation of *-an in the indicative mood, and *-i in the imperative mood. These affixes appear to be identical to what Wolff (1973:73) called the 'independent local passive' and the 'dependent local passive' voice markers. Table 7.23 shows the affixation pattern for declarative and imperative verbs in PAN as reconstructed by Wolff (1973) and adopted by Ross (2002a:49), in Atayal as described by Rau (1992), and in Bikol as described by Mintz and Britanico (1985). Rau discusses these in terms of the contrast of independent, subjunctive and dependent modes, where the dependent mode is used in negative, emphatic and imperative sentences. Mintz and Britanico discuss them in terms of the contrast of 'the regular series of commands which are identical in form to the infinitive of the verb,' and the 'alternative command forms.' The division of labor between *mag-* in the regular command forms of Bikol, and of *–um-* in the alternative forms is striking, since both affixes mark the actor voice of declarative sentences in languages such as Tagalog, but *mag-* has almost completely replaced *–um-* in this function in modern Bikol, leaving *–um-* only as an imperative, a typological oddity, since reflexes of *-um- rarely have this function in AN languages.

Table 7.23 Morphological correspondences between declarative and imperative verb forms in PAN, Atayal and Bikol

	Declarative	**Imperative**	
PAN			
Actor	*-um-	Ø	
Patient	*-en	*-a	
Locative	*-an	*-i	
Instrumental	*Si-	*-an	
Atayal			
Actor	m-/-m-	Ø	
Patient	-un	-i	
Locative	-an	-i	
Instrumental	s-	s-	
Bikol		regular	alternative
Actor	mag-/-um-	mag-	-um-
Patient	-on	-on	-a
Locative	-an	-an	-i
Instrumental	i-	i-	-an

The data in this table show why many AN languages have multiple forms of imperative marking: the imperative mode, like the declarative mode, must encode a particular voice, and since Philippine-type languages typically have four voices it can be expected that imperative marking in languages of this type is more complex than is true in many languages of the world. Despite important areas of overlap, the imperative systems of Atayal and Bikol differ in a number of respects. First, Mintz and Britanico (1985:41) distinguish a 'regular series of commands' (*mag-, -on-, -an, i-*) that are identical to the declarative form of the verb, from a set of 'alternative command forms.' The alternative command forms *-a*, *-an* and *-i* are said to differ from regular commands in lacking an overt pronoun that indexes the one commanded to perform an action, as partially illustrated in

the following contrastive structures (but note that 7.147 *does* contain an overt pronoun that indexes the one commanded to perform the action):

(7.165) isíp-isíp-on   mo   pa
         think-**red-imp   2s**   still
         'Just think it over a little more'

(7.166) hapot-á   man   tábiʔ   siyá
         ask-**imp**   also   please   **3sg**
         'Please ask him'

(7.167) hugás-hugás-an   mo   na   laŋ   an   máŋa   pláto
         wash-**red-imp   2s**   already   just   **top   pl**   plate
         'Just give the dishes a quick wash'

(7.168) bayáʔ-i   na   ŋáni   an   pig-gi-gíbo   mo
         stop-**imp**   already   please   **top**   thing done   **2s**
         'Just stop what you are doing'

Just as many languages preserve fragments of the PAN voice system, sometimes bleached of syntactic function (e.g. Malay *si* 'marker of personal names'), fragments of the PAN system of imperative marking are found in many languages. Zero-marked imperatives are common in the languages of central and western Borneo and in some other parts of western Indonesia, as in Mukah Melanau *upuk kain itəw* (wash clothes these) 'Wash these clothes' (base: *upuk*, **av**: *m-upuk*). In Javanese the imperative of intransitive verbs and active transitive verbs that are unsuffixed in their non-imperative forms is usually marked by *-a*, while the imperative of passive transitive verbs is marked by *-ən* (Robson 2002:82). As already seen, the most common imperative marker in Thao is *-i*, with no clear correspondence to voice. In Kayan the imperative of active verbs is normally zero-marked, as in *(im) jat ue anih* (**2sg** pull rattan this) 'Pull this rattan' (**av**: *mə-jat*), but the verb 'to eat' (base: *kan*), which often contains fossilised morphology in the languages of Borneo, has two imperative forms, one corresponding to the active verb (*k<um>an*), and the other to the passive verb (*kan-i*). In some languages with multiple types of imperatives zero-marked imperatives are considered abrupt unless tempered with an ameliorative particle, as in Thao *k<m>an* (**av**) 'to eat' : *kan afu* (eat rice) 'Eat your rice' (said, e.g. to a child who won't eat) : *kan uan afu* (eat please rice) 'Please eat your rice'. In Bahasa Indonesia transitive verbs with the active prefix *məŋ-* drop the prefix in imperative constructions, but intransitive verbs retain it (Macdonald and Soenjono 1967:261): *baca buku ini* (read book this) 'Read this book' (**av**: *məm-baca*), but *məm-buŋkuk-lah* (**av-**bend.down-**emph**) 'Bend down!'. Although use of the bare verb stem is not considered abrupt or impolite in itself, the verbs *toloŋ* 'help', *coba* 'try' and *silahkan* 'do it for your own benefit' may be used with bare imperatives to convey a sense of empathy or amelioration: *coba baca buku ini* (try read book this) 'Read this book'.

7.8.1 Presence or absence of a pronoun

In many languages pronouns are not overtly expressed in the imperative mode. AN languages vary considerably in this respect: some languages do not allow overt pronouns,

others require them, and still others distinguish constructions which have an overt pronoun from those that do not.

Lee (1975:333) reports that imperative sentences in Kosraean never have an overt subject (signalled by the subject marker *el*), and that this is the surest way to distinguish them from declarative sentences. In Anejom of southern Vanuatu, on the other hand, subject pronouns are rarely deleted (Lynch 2000:136):

(7.169) lep   elad-se-sjak   ajourau
         again   look-down-**pol**   **2d**
         'Both of you look down again!'

In Pohnpeian of central Micronesia according to Rehg (1981:304ff) positive imperatives do not have overt pronouns, but their negative equivalents do:

(7.170) mʷo:ndi
         sit.down
         'Sit down!'

(7.171) ke   de:r   mʷo:ndi
         2s   **vet**   sit.down
         'Don't sit down!'

In Sarangani Manobo of the southern Philippines the persons addressed in imperatives are represented by the pronouns *ka* '2s' and *kaw* '2pl'. Both pronouns appear on the surface in positive commands expressed in the subject focus. However, in non-subject focuses the singular pronoun *ka* is omitted, and in negative commands the singular pronoun *ka* is omitted and the plural pronoun becomes *niyo* (Dubois 1976:88). The correlation of negation with presence of an overt pronoun in imperative constructions thus appears to vary across languages.

There are other variations, but space will permit mention of only one more. Woollams (1996:234) notes that in Karo Batak imperatives may or may not express the addressee overtly, but the choice carries with it a difference of emotional tone: "inclusion of the addressee in an imperative normally implies a moderation in the directness of the command and conveys a more persuasive tone to the appeal being made." This contrasts with English, where the use of a second person pronoun in an imperative normally would be accompanied by forceful intonation, and so be interpreted as increasing the strength of the command ('Come here' vs. '*You* come here!').

7.8.2 The illocutionary force of imperatives

The examples from the preceding section touch on another aspect of this type of construction. Because imperatives contain an inherent tension between the need for physical action and the need for social cohesion some means must be available for softening the force of a command. Most languages consequently have both brusque and polite forms of commands, the former generally used with children or social inferiors, and the latter in 'normal' social settings.

AN languages use a number of strategies to render a command more polite. One strategy was just seen in Karo Batak—the inclusion of an overt pronoun representing the

addressee mitigates the force of a command. Another way to accomplish this end is to add mediating words which imply that the addressee has greater control in initiating the requested action, as in Bahasa Indonesia, where commands may be accompanied by *coba* 'try' or *toloŋ* 'help'. Whereas *aŋkat batu* 'Lift the stone!' might be appropriate if a prison guard were commanding a prisoner to work, one construction worker asking another to perform the same action would more likely use *coba aŋkat batu* for an action in which he himself was not involved, or *toloŋ aŋkat batu* for a jointly performed task. A similar means of softening commands is found in Sye of southern Vanuatu (Crowley 1998:89):

(7.172) tapmi   m-etehep
        try   **s.es**-sit
        'Please sit!'

Still another form of 'polite' command is one in which the addressee is asked to perform an action that is ostensibly for his or her own benefit. This construction type is found in other Asian languages (as Japanese), and within AN appears to be confined to insular Southeast Asia, and even more specifically to those languages that have been subjected to centuries of Indian contact influence. In Malay/Indonesian, for example, the equivalent of the Sye command illustrated above is *silahkan duduk*, not *coba duduk*; both are polite, but the former would be used, e.g. in inviting a guest to be seated in one's home, and the latter in a situation where the addressee's sitting down benefits the person making the command/request, as in a movie theatre where a person who is standing is requested to sit so as not to block the view for those behind.

Lynch (2000:137) notes that Anejom has strategies for mitigating and intensifying the force of a command by preposing either the hortative particle *mu* or the intensifying particle *fi* to the simple imperative:

(7.173) adia   aak
        go.away   **2s**
        'Go away!'

(7.174) Mu   adia   aak
        hort   go.away   **2s**
        'Please go away!/Would you mind leaving?'

(7.175) fi   adia   aak
        intens go.away   **2s**
        'Piss off!'

In some languages passive verb forms are used to form polite imperatives, as in the following examples from Malay/Indonesian:

(7.176) jaŋan   gaŋgu   guru-mu
        vet   disturb   teacher-**2s**
        'Don't disturb your teacher!' (said e.g. in annoyance)

(7.177) jaŋan di-gaŋgu   guru-mu
         vet   pass-disturb   teacher-**2s**
         'Don't disturb your teacher!' (said e.g. as gentle words of advice)

In other languages, however, the use of a passive form is regarded by native speakers as a more direct form of command, as in the following contrastive constructions from Lun Dayeh, a marginally Philippine-type language of northern Sarawak:

(7.178) məlih   kuyuʔ inəh
         av-buy   shirt   **dem**
         'Buy that shirt' (request)

(7.179) bəli-ən   kuyuʔ inəh
         buy-**pv**   shirt   **dem**
         'Buy that shirt' (felt to be mandatory)

This difference of emotional tone is consistent with the view that the actor voice in Philippine-type languages marks intransitive verbs, and that the patient voice encodes actions which more thoroughly or significant affect the patient.

In Kambera of eastern Sumba differences of politeness in imperatives are expressed through the case marking of the addressee (Klamer 1998:164). Specifically, the use of accusative case for the addressee is said to be more blunt or direct, while the use of nominative case conveys something closer to the sense of a request. Klamer suggests that the use of the nominative case is more respectful because it implies that the addressee has greater control over the event (and hence greater volition in performing it). Although Kambera lacks a true passive, then, the nominative/accusative contrast in this language evidently conveys some of the same emotional overtones as the active/passive distinction in Philippine-type languages, where the patient voice implies that the patient is more drastically affected (and less in control) than the actor voice.

Finally, in many languages the emotional tone of an imperative is conveyed most directly by intonation. Although Javanese has several options for marking imperatives, for example, Robson (2002:81) notes that 'in spoken Javanese the imperative is often not marked by any particular affixes, but rather by one's tone of voice.'

7.8.3 Direct and indirect imperatives

Although moderated commands of the type just noted may be called indirect, the direct/indirect distinction must be reserved for another type of imperative. Lee (1975:334) has observed that Kosraean distinguishes direct imperatives, in which the addressee is expected to carry out the action demanded or requested, from indirect imperatives, in which the addressee mediates between the person giving the command and the person for whom it is intended. Alternatively, indirect imperatives might be called 'third person imperatives':

Kosraean

(7.180) orek   ma   lututεŋ
         work   *ma*   early.morning
         'Work tomorrow morning!' (direct imperative)

(7.181) ɛl-an   uniyɛ pik   soko   æ
         3sg-allow   kill   pig   one   **det**
         'Let (allow, make, tell) him to kill the pig' (indirect imperative)

7.8.4 Singular and plural imperatives

Although most imperatives contain or imply a second person singular referent, plural imperatives also occur. These can be expressed either as a command to two or more persons to perform an action, or as an invitation (exhortation) to one or more persons to join the speaker in performing an action. In Loniu of the Admiralty Islands, hortative commands are marked with a first person dual inclusive pronoun:

Loniu

(7.182) tɔʔu   kɛyɛni   ɛ
         1du.incl   **pot.ns.**eat   **emph**
         'Let's eat now!'

While dual inclusive pronouns are a natural means to express hortatives, they may occur in commands directed toward a single referent. Muna of southeast Sulawesi, for example, uses the pronominal suffix *–kaeta* in commands that are ostensibly for the benefit of both speaker and hearer. While this may take the form of a hortative, as in *fumaa-kaeta* (eat-**imp**) 'Let us eat', it can equally well be used in commands that are directed to an addressee who is expected to carry out the action alone, as in the following sentence from van den Berg (1989:68):

Muna

(7.183) me-gholi-kaeta   kenta   naewine
         imp-buy-for.us   fish   tomorrow
         'Buy some fish tomorrow' (for us, so that we can eat)

The Muna imperative with *–kaeta* thus appears to stand somewhere between a true hortative, and invitational commands such as Malay/Indonesian *silahkan*.

7.8.5 Imperatives of coming and going

Verbs of coming and going, or their historical continuations, are used in some AN languages to create particular types of imperative. In Tetun, for example, *mai ita* (come **1pl.in**) is preposed to the predicate to form a hortative command. This is structurally parallel to, and cognate with Malay/Indonesian *mari kita*, as seen in the following examples:

Tetun

(7.184) mai   ita   bá   ne-bá
         come   **1pl.in**   go   there
         'Let's go over there!'

Bahasa Indonesia

(7.185) mari   kita   pərgi kə-sana
        let's   **1pl.in**   go   to-there
        'Let's go over there!'

The difference is that although Malay/Indonesian *mari* reflects PMP *um-aRi 'to come', its only function in the modern language is to signal the hortative (cf. *dataŋ* 'to come' < PMP *dateŋ 'arrive'). Historically, however, the Malay and Tetun patterns appear to be continuations of a PMP *um-aRi kita X (come 1pl.in X) 'Let's X!', where X probably was an unaffixed verb stem.

Tetun distinguishes hortatives, marked by *mai ita* from commands in which the speaker is not included. The latter are signalled by imperative *bá*, which is functionally distinct, but clearly derived from *bá* 'go' (van Klinken 1999:244):

(7.186) em   bá   bá-n   té   haʔu ha   ulu-n   moras
        2p   go   **imp-imm**   since   **1sg** **1sg** head-**gen** sick
        'You go (without me) because I have a headache'

In Tetun, then, historical verbs of coming and going are used to distinguish inclusive or hortative, from exclusive commands. If exclusive commands are regarded as unmarked, in Tetun the verb 'to come' is used in marked, and the verb 'to go' in unmarked commands. In some other AN languages, however, the verb 'to go' is used in what appear to be marked imperatives.

Wozna and Wilson (2005:76) describe positive imperatives in Seimat of the Admiralty Islands as obligatorily marked with the venitive suffix -(V)*ma* 'action directed towards the speaker', or -(V)*wa* 'action directed away from the speaker'. They note that a second person pronoun is optional; when such a pronoun is present the imperative and declarative clause structures are identical and the distinction is carried by intonation:

(7.187) ke-ma
        pass-**imp.ven**
        'Pass it (to me)'

(7.188) ke-wa
        pass-**imp.all**
        'Pass it (to him, her, them, etc.)'

These imperatives are thus marked for directionality of the action, in accordance with general properties of the verb system of this and many other Oceanic languages. In some languages, however, the coding of imperative messages in relation to directionality is meant to distinguish commands that literally require the addressee to change location to accomplish the desired action from those that don't. The term 'ambulatory imperative' was used in Blust (2003e) to distinguish two types of imperative construction in Selau, spoken on the island of Buka in the western Solomons:

(7.189) ase-i   moni
        count-**imp**   money
        'Count the money!'

(7.190)   (na)   ase-ia   moni
          go   count-**imp**   money
          'Go and count the money!'

(7.191)   ss-i
          nurse-**imp**
          'nurse it ! (e.g. a crying baby)'

(7.192)   na   ss-ia   aksə
          go   nurse-**imp**   child
          'Go and nurse the child!'

Examples (189) and (191) are unmarked imperatives, while (190) and (192) imply that the person who is ordered to perform an action cannot do it at the location where the order is given. Since the full expression of the ambulatory imperative is redundantly marked by *na* 'go (and do something)', and the suffix *–ia*, the verb is dropped in some recorded examples (hence: *ase-ia moni*). Unlike the parallel between Malay/Indonesian *mari kita* and Tetun *mai ita*, that between Selau and Tetun imperatives marked by the verb 'to go' has no demonstrable historical connection. Nonetheless, the parallelism suggests that verbs of coming and going may commonly be recruited to distinguish types of imperatives, whether the distinction is between hortative/non-hortative, or between action to be performed at the site of the command or at a distance from it.

7.8.6 Tense/aspect in imperatives

In languages generally imperative verbs usually are tenseless, and this is also true of most AN languages. However, in Palauan, Chamorro, and many languages in Melanesia, imperatives are expressed by future, hypothetical, or irrealis verb forms, a pattern that is almost unknown elsewhere. Loniu of the Admiralty Islands, for example, uses the potential form of the verb to mark the imperative mode (Hamel 1994:147), Manam of New Guinea uses the definite irrealis mood to express any type of obligation, necessity, demand, command, or exhortation (Lichtenberk 1983:417), and Araki of central Vanuatu uses irrealis modality for all imperative sentences "by definition, since they refer to virtual events" (François 2002:168). In Palauan the imperative of imperfective verbs is derived by substituting the second person hypothetical pronoun prefix *mo-* for the prefix *mə-* of the corresponding imperfective verb (Josephs 1975:394): *mə-lim* 'to drink' : *mo-lim a kərum* (**2p.hyp a** medicine) 'Drink your medicine!'. In Chamorro imperatives are marked with future tense (Topping 1973:264):

(7.193)   para   bai   u   falagu
          fut   **fut**   **1sg**   run
          'I will run'

(7.194)   falagu 'Run!' (cf. *malagu* 'to run')

The distribution of this feature clearly has an areal character, but how it might have originated through contact is unclear. If an imperative verb form is to carry tense or aspect

marking of any kind it might be considered natural, as François suggests in relation to Araki, that it should be marked as irrealis or future, since at the time of utterance the desired action exists only in the will of the speaker. Surprisingly, however, several AN languages use perfective verb forms to mark certain types of imperatives. Although Palauan imperative constructions are basically expressed through the second person hypothetical pronoun *ʔomo-*, which is shortened to *mo-*, Josephs (1975:110) notes that such imperatives can be perfective (ordering that an action be completed), in which case the prefix is shortened to *m-*:

(7.195)   m-ŋilmii   a   iməl-əm
          av-drink   *a*   drink-**2s**
          'Finish up your drink!'

While the use of perfective imperatives is one of several options open to speakers of Palauan, Lee (1975:335) notes that in Kosraean of central Micronesia "verbs in imperative sentences usually appear in the complete aspect," as in:

(7.196)   ise-ɛk   tutpes   æ
          squeeze-**perf**   toothpaste   **dem**
          'Squeeze out the toothpaste'

His explanation for this peculiarity is that in its primary usage *-ɛk* is a directional suffix. In declarative sentences directional suffixes disambiguate otherwise identical utterances such as 'John is digging up a box' vs. 'John is burying a box.' Moreover, they denote "that a certain action has come to an end and a certain result has been achieved." In imperative sentences that are otherwise ambiguous ('Dig up a box' or 'Bury a box'), the directional suffixes do not indicate perfectivity, but specify the result of a certain action. The fact that Kosraean imperatives are usually perfective is thus an accidental by-product of the multifunctionality of the directional suffixes.

One of the most intriguing uses of perfective imperative verb forms is found in Mukah Melanau of coastal Sarawak. In this language reflexes of PAN *-um- and *-in- have become the markers of active and passive verbs respectively. There are several surface realisations of these affixes, one of which is an ablaut pattern whereby bases that contain penultimate schwa alternate with active forms in *–u-* and passive forms in *-i-*, as with *ləpək* 'a fold' : *lupək* 'to fold' : *lipək* 'was folded by someone'. As in a number of other languages in this area, passive verbs are obligatorily perfective. Imperatives can be expressed in various ways, one of which is by use of a passive/perfective verb:

(7.197)   iən   dudut   kayəw   iən
          3sg   pull.up-**av**   tree   **dem**
          'He is pulling up the tree'

(7.198)   ayəw   iən   didut   siən
          tree   **dem**   pull.up-**pv.perf**   **3sg**
          'He pulled up the tree'

(7.199)   didut   kayəw   iən
   pull.up-**pv.perf**   tree   **dem**
   'Pull up that tree!'

In the vetative, however, the verb form must be active/imperfective:

(7.200)   kaʔ   dudut   kayəw   iən
   vet   pull.up-**av**   tree   **dem**
   'Don't pull up that tree!'

The same relationship holds for verbs in which voice distinctions are expressed by prefixes, as with *siən mə-biləm kain iən* (**3sg** **av**-black cloth **dem**), *kain iən nə-biləm siən* (cloth **dem pv.perf**-black **3sg**) 'She blackened the cloth' as against the imperative sentences *nə-biləm kain iən* 'Blacken the cloth!', *kaʔ mə-biləm kain iən* 'Don't blacken the cloth!', where *biləm kain iən* is acceptable but not preferred, and ***mə-biləm kain iən* is impossible. Whereas most languages that distinguish tense or aspect in imperatives use a future or irrealis verb in commands, then, Mukah Melanau prefers a perfective verb unless the command is negative. Unlike perfective imperatives in Palauan, which signal a contrast between ordering that an action be begun and ordering that it be completed, imperatives in Mukah Melanau appear to be inherently perfective unless they are negated. This typologically odd situation may have arisen from an earlier preference for passive imperatives, and since the Mukah passive is obligatorily perfective the incongruous aspect was imported with the preferred voice. In the vetative, however, the perfective verb form would have been doubly incongruous, and so is strictly avoided.

7.8.7 Vetative stress shift

Thao has an unusual pattern of alternation in the vetative of verbs suffixed with imperative *-i*. Although stress is normally penultimate, objectless vetative constructions with *-i* undergo rightward stress shift:

(7.201)   ata   (tu)   karí   (base: kari, av: k<m>ari)
   don't   *tu*   dig-**imp**
   'Don't dig!'

(7.202)   ata   (tu)   fariw-í   (base: fariw, av: fariw)
   don't   *tu*   buy-**imp**
   'Don't buy (it)!'

Where the imperative suffix *-i* marks a negative command with an expressed object, stress shift does not occur:

(7.203)   ata   (tu)   cpiq-i   sa   shaqish
   don't   **pol**   slap-**imp**   *sa*   face
   'Don't slap (his) face!'

7.9   Questions

There are a number of strategies for forming 'yes-no', or polar questions in AN languages. One common strategy is to impose a different intonation pattern on a sentence that is structurally identical to its declarative counterpart. In Palauan this is the common method for transforming statements into questions: *kə mle smeʔər* (low even pitch, falling slightly at the end) 'You were sick' : *kə mle smeʔər* (steadily rising intonation, which remains high at the end) 'Were you sick?' (Josephs 1975:409). A similar strategy is found in many other languages, including Tsou (Zeitoun 2005:282), Bikol (Mintz 1971:104), Sarangani Manobo (Dubois 1976:13), Hawu (Walker 1982:40), Seimat (Wozna and Wilson (2005:77), and in most languages of Vanuatu (Lynch, Ross and Crowley 2002). In virtually all cases polar questions are distinguished from declarative statements by a pattern of final rising intonation.

Another method of question formation is to place a general question word at the beginning of a declarative sentence, as with Malay *apa* 'what; marker of questions', or Chamorro *kao* 'general interrogation marker': Malay *dia bəraŋkat* (**3sg**   depart) 'He left' : *apa dia bəraŋkat* (**QM 3sg**   depart) 'Did he leave?', Chamorro *g<um>u-gupu i páharu* (**red-av**-fly **def** bird) 'The bird is flying' : *kao g<um>u-gupu i páharu* (**QM av**-fly **def** bird) 'Is the bird flying?'. Although question formation in these languages is structurally similar it is noteworthy that Malay *apa* also serves as the general interrogative 'what?', whereas Chamorro *kao* has no independent function. In many Oceanic languages, including Sobei, Takia, Yabem, and Gapapaiwa of New Guinea, Bali-Vitu, Kaulong and Siar of the Bismarck Archipelago, Nggela, Longgu, and Arosi of the Solomon Islands, and Lamenu, Ifira-Mele (Mele-Fila) and Anejom of Vanuatu, polar questions are indicated by a clause-final question tag or general interrogation marker (Lynch, Ross and Crowley 2002).

In most Philippine languages the general interrogative for polar questions is inserted immediately after the predicate, as in Tagalog *Amerikano si Jorge* (American **top** George) 'George is an American' : *Amerikano ba si Jorge* (American **QM top** George) 'Is George an American?', or Central Tagbanwa *ma-intidi-an mo layan* (understand **2sg** that) 'You understand that' : *ma-intidi-an mo va layan* (understand **2sg QM** that) 'Do you understand that?'. According to Ramos (1971:118) the Tagalog question marker *ba* 'usually follows the first full word of a sentence. However, when the topic is the pronoun *ka*, then *ba* follows': *Amerikano ka* (American **2sg)**   'You are an American' : *Amerikano ka ba* (American **2sg QM**) 'Are you an American?'. The single example from Central Tagbanwa suggests a similar pattern.

Some languages show other complications in polar questions. Healey (1960:92), for example, gives Central Cagayan Agta 'post-adverbs' *hud* and *de* both of which have other functions, but may mark yes-no questions when the expected reply is negative (*hud*) or positive (*de*). These uses of interrogative markers are unusual in that the form of the question presupposes the polarity of the answer.

(7.204)   ittá   hud   ya   danum
          exist   **QM**   **lig**   water
          'Is there any water?   (expected answer negative)

(7.205)   ittá   de   ya   danum
          exist   **QM**   **lig**   water
          'Is there any water?   (expected answer positive)

Constituent questions are formed with a small number of interrogative words some of which show clear historical if not synchronic evidence of morphological complexity. One widespread feature of constituent questions is especially noteworthy. In many languages throughout the family the question 'What is your name?' does not use the general interrogative morpheme 'what?', but rather the personal interrogative 'who?': Pazeh *asay* 'what?', *ima* 'who?' : *ima laŋat pai siw* (who name **QM 2sg**), Mansaka *nana* 'what?', *sini* 'who?' : *sini-ŋ ŋaran-mo* (who-**lig** name-**2sg**), Ngaju Dayak *naray* 'what?', *eweh* 'who?' : *eweh ara-m* (who name-**2sg**), Bahasa Indonesia *apa* 'what?' : *siapa* 'who?' : *siapa nama anda* (who name **2sg**), Manggarai *apa* 'what?', *cei* 'who?' : *cei ŋasaŋ de hau* (who name of **2sg**), Chamorro *hafa* 'what?', *hayi* 'who?' : *hayi naʔan-mu* (who name-**2sg**), Roviana *sa* 'what?', *esei* 'who?' : *esei poza-mu si agoi* (who name-**2sg foc 2sg.foc**) 'What is your name?', Kosraean *meœ* 'what?' : *sə* 'who?' : *sə ine-l an* (who name-**3sg   dem**) 'What is his/her name?'. In a smaller number of languages the question word appropriate for inquiring about personal names is the general interrogative morpheme, as in Tagalog *anó* 'what?', *síno* 'who?' : *anó aŋ paŋalan-mo* (what **top** name-**2sg**) 'What is your name?'. The motivation for 'who name **2s**?' almost certainly is structural, and the following observations appear relevant to understanding how it might have developed. First, the general and personal interrogatives in PAN and PMP appear in Figure 7.1:

	what?	who?
PAN	*anu	*ima
PMP	*apa	*i-sai

Figure 7.1 General and personal interrogatives in PAN and PMP

The reconstruction of PMP interrogative markers includes complications not represented in Figure 7.1. Several widely scattered languages, for example, reflect *anu 'what?', but the larger comparative picture suggests that this was an indefinite interrogative, best rendered in English as 'whatchamacallit': Bontok *ano-ka* 'an empty form substituting in any form class when the normal form is inappropriate or when it cannot be recalled; what's-it', Aborlan Tagbanwa, Sama *anu* 'whatchamacallit', Bintulu *anəw* 'thing mentioned, whatchamacallit', Ngaju Dayak *anu* 'a certain (person, etc.)', Karo Batak *anu* 'indefinite pronoun indicating a person whose name one doesn't know, or doesn't want to say', Javanese *anu* 'substitution for a word that has slipped the mind'. A few languages in the Philippines and Borneo also reflect *inu 'what?', although the comparative evidence suggests that this meant 'when?' in PAN, and probably in PMP.

What is striking in comparing these reconstructed terms with the range of terms found in the modern languages is the heavy concentration of replacement innovations for 'who?' that contain a reflex of *si 'focus/subject marker for personal nouns':

Table 7.24 The captured focus/subject marker of personal nouns in words for 'who?'

Language	Base	Source
Thao	tima	*si ima (= si + who?)
Bunun	sima	*si ima
Amis	cima	*si ima
Paiwan	tima	*si ima
Yami	sinu	*si inu (= si + what?)
Tagalog	sino	*si inu
Subanen	sinu	*si inu
Itneg	siʔanu	*si anu (= si + what?)
Pangasinan	siopá	*si + Pangasinan opá 'what?'
Ibaloy	sipa	
Kalagan	siŋan	
Palawan Batak	siʔu	
Mansaka	sini	
Sarangani Bilaan	sinto	
Malay	siapa	*si apa (= si + what?)
Balinese	sira	

For at least those forms with a demonstrable base *ima*, *inu*, *anu*, *opá* or *apa*, it is clear that the personal interrogative contains a reflex of *si. The common occurrence of other words for 'who?' that begin with *si-* implies a similar history that cannot yet be traced. Even where the forms are cognate, however, as with Formosan reflexes of *si ima, the non-correspondence with established subgroup boundaries indicates a drift.

Linguistic drifts are motivated by the continued operation of inherited structural pressures after language split. Since *si marks third person pronouns (PAN *i-aku '1sg', *i-kaSu '2s', but *si-ia '3sg'), the capture of *si in the personal interrogative presumably began with third-person reference, and was subsequently generalised to all persons, through the following steps:

Stage 1)
Q: ima ia 'Who is he?'
A: si Adan '(He is) Adan'
Q: anu ŋajan ni-á 'What is his name?'
A: si Adan '(He is) Adan'

Stage 2)
Q: si ima ia 'Who is he?' (anticipates the person marker of the answer)
A: si Adan '(He is) Adan'
Q: si anu ŋajan ni-á 'What is his name?' (anticipates the person marker of the answer)
A: si Adan '(He is) Adan'

Although the answer to 'Who is he?' could be supplied by a common noun, the probability is high that personal nouns would be more frequent. If so, the transition between stage 1) and stage 2) would be motivated by pattern pressure; since the answer to the question 'Who is he?' would generally begin with a reflex of *si, over time this element became incorporated into the question itself by anticipating the person marker of the answer. With the question 'What is his name?' the general interrogative 'what?' would

have acquired a reflex of *si for the same reason (anticipating the person marker of the answer). Over time this would have led to the creation of many new personal interrogatives that began with a reflex of *si, and initially these would have had a transparent morphological relationship to the general interrogative 'what?'. Apart from Subanen *sima* 'who?', which may be a convergent development that has no relationship to similar Formosan forms, there are no clear reflexes of *ima outside Taiwan, and so stages 1) and 2) cannot correspond to PAN and PMP. They do, however, suggest a general direction of development which serves to explain both why the personal interrogative 'who?' so often contains a reflex of *si, and why in so many widely distributed AN languages an inquiry about personal identity must be couched in the somewhat surprising form 'Who is your name?'.

8   *Reconstruction*

8.0   Introduction

In previous chapters forms have been cited from PAN, PMP and sometimes other proto languages, and occasional reference has been made to issues of reconstruction. However, no systematic presentation of the basis for reconstruction has yet been given. This chapter is intended to fill that gap. For reasons of space the discussion of reconstruction will be limited mainly to phonology. A brief review of morphosyntactic reconstruction in AN is given at the end of Chapter 7. By way of introduction it will be useful to have an overview of the history of comparative scholarship on the AN languages. This has the advantage of showing the cumulative nature of the growth of knowledge in this field from limited beginnings to ever more inclusive generalisations.

8.1   History of scholarship

Before reconstruction is possible in any language family it is necessary to establish the genetic relationship of the languages and to recognise the full range of relevant sound correspondences between them. As in other parts of the world, progress in both of these areas was piecemeal in the study of the AN languages. Table 8.1 lists dates marking major events in 1) the discovery of the scope of the AN language family, 2) the discovery of recurrent sound correspondences, and 3) the reconstruction of the phonology, lexicon and aspects of the grammar of PAN and other early proto languages. Important publications concerned with subgrouping are also included if they have a bearing on the reconstruction of higher-level proto languages:

TABLE 8.1: Important dates in the comparative study of the Austronesian languages

Year	Event
1521	Pigafetta collects vocabularies from the Philippines and Indonesia
1603	de Houtman recognizes the Malay-Malagasy connection
1615	Le Maire collects vocabularies from western Polynesia
1708	Reland recognizes a 'common language' from Madagascar to western Polynesia
1768-1779	Cook's voyages in the Pacific: Főrster (1778) sees Polynesian as a unity, but the languages of Melanesia as unrelated to these, or to one another
1784	Hervas y Panduro recognizes a 'common language' from Madagascar to eastern Polynesia
1836-1839	von Humboldt's *Die Kawi-Sprache* appears posthumously
1841	Bopp calls the language family 'maleisch-polynesisch'

1861-1873	von der Gabelentz shows that 'Melanesian' and Polynesian languages share a common grammar
1865	van der Tuuk identifies three important 'sound laws'
1884	Brandes establishes the 'Brandes Line', and coins the terms 'van der Tuuk's first law' and 'van der Tuuk's Second Law'
1885	Codrington publishes *The Melanesian languages*
1886	Kern publishes *De Fidji-taal*
1889	Kern places the Austronesian homeland in Indochina
1897-1912	van der Tuuk compiles the *Kawi-Balineesch woordenboek*
1899	Schmidt calls the language family 'Austronesian'
1906	Schmidt proposes the Austric hypothesis Brandstetter compiles the *Prodromus*
1910	Brandstetter recognizes widespread monosyllabic 'roots'
1911	Brandstetter begins to reconstruct 'Original Indonesian' morphology
1914	Jonker questions the 'Brandes Line'
1915	Brandstetter reconstructs the 'Original Indonesian' phonetic system
1920	Dempwolff publishes his "Lippenlaute"
1924-1925	Dempwolff publishes his "l-, r- und d-Laute"
1926	Ray publishes *A comparative study of the Melanesian island languages*
1927	Stresemann publishes *Die Lauterscheinungen in den ambonischen Sprachen*
1934-1938	Dempwolff publishes his *Vergleichende Lautlehre des austronesischen Wortschatzes*, a reconstruction of the 'Uraustronesisch' sound system with 2,215 base forms, and recognizes a 'melanesisch' (= Oceanic) subgroup
1935	Ogawa and Asai publish *The myths and traditions of the Formosan native tribes* (in Japanese), noting two phonological distinctions not found in non-Formosan Austronesian languages
1942	Benedict publishes 'Thai, Kadai and Indonesian: a new alignment in Southeastern Asia'
1943	Capell publishes *The linguistic position of South-Eastern Papua*
1946	Leenhardt publishes *Langues et dialects de l'Austro-Mélanésie*
1949	Dyen publishes 'On the history of the Trukese vowels'
1951	Dahl publishes *Malgache et Maanjan: une comparaison linguistique*
1953	Dyen publishes *The Proto-Malayo-Polynesian laryngeals* and 'Dempwolff's *R'
1955	Grace publishes 'Subgrouping of Malayo-Polynesian: a report of tentative findings'
1956	Dyen publishes 'The Ngaju Dayak 'Old speech stratum', and 'Language distribution and migration theory'
1959	Grace publishes *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family*
1961	Milke publishes 'Beiträge zur ozeanischen Linguistik'
1963	Dyen publishes 'The position of the Malayopolynesian languages of Formosa'
1965	Dyen publishes *A lexicostatistical classification of the Austronesian languages* and 'Formosan evidence for some new Proto-Austronesian phonemes' Chrétien publishes 'The statistical structure of the Proto-Austronesian morph'

Haudricourt publishes 'Problems of Austronesian comparative philology'

Biggs publishes 'Direct and indirect inheritance in Rotuman'

1966   Walsh and Biggs publish *Proto-Polynesian word list I*, subsequently expanded as the ongoing online Proto-Polynesian lexicon (POLLEX)

Pawley publishes 'Polynesian languages: a subgrouping based upon shared innovations in morphology'

Green publishes 'Linguistic subgrouping within Polynesia: the implications for prehistoric settlement'

Grace publishes 'Austronesian lexicostatical classification: a review article'

1967   Pawley publishes 'The relationships of Polynesian Outlier languages'

Hudson publishes *The Barito isolects of Borneo: a classification based on comparative reconstruction and lexicostatistics*

1968   Milke publishes 'Proto-Oceanic addenda'

1969   Ferrell publishes *Taiwan aboriginal groups*: *problems in cultural and linguistic classification*

Grace publishes 'A Proto-Oceanic finder list'

Blust publishes 'Some new Proto-Austronesian trisyllables'

1970-1989   Blust publishes over 2,800 new lexical reconstructions for PAN, PMP, and PWMP, and shows that several hundred of Dempwolff's reconstructions must be considered late innovations in western Indonesia

1971   Reid publishes *Philippine minor languages: word lists and phonologies*

Haudricourt publishes 'New Caledonia and the Loyalty Islands'

1972   Pawley publishes *On the internal relationships of Eastern Oceanic languages*

1973   Pawley publishes 'Some problems in Proto-Oceanic grammar'

Wolff publishes 'Verbal inflection in Proto-Austronesian'

Dahl publishes *Proto-Austronesian* (2nd, rev. edition 1976)

1975   Benedict publishes *Austro-Thai language and culture*

Mills publishes *Proto-South Sulawesi and Proto-Austronesian phonology*

Nothofer publishes *The reconstruction of Proto-Malayo-Javanic*

1976   Clark publishes *Aspects of Proto-Polynesian syntax*

Tsuchida publishes *Reconstruction of Proto-Tsouic phonology*

Tryon publishes *New Hebrides languages: an internal classification*

Blust publishes 'A third palatal reflex in Polynesian languages' and 'Austronesian culture history: some linguistic inferences and their relations to the archaeological record'

1977   Zorc publishes *The Bisayan dialects of the Philippines: subgrouping and reconstruction*

Blust publishes 'The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report'

1977-1984   Blust proposes a qualitative subgrouping of the Austronesian languages that differs fundamentally from that of Dyen (1965)

1978   Chung publishes *Case marking and grammatical relations in Polynesian*

Blust publishes *The Proto-Oceanic palatals*

Sneddon publishes *Proto-Minahasan: phonology, morphology and wordlist*

Zorc publishes 'Proto-Philippine accent: innovation or Proto-Hesperonesian retention?'

Reid publishes 'Problems in the reconstruction of Proto-Philippine construction markers'

Lynch publishes 'Proto-South Hebridean and Proto-Oceanic'

Lincoln publishes 'Reef-Santa Cruz as Austronesian'

1980   Blust publishes 'Early Austronesian social organisation: the evidence of language'

1981   Dahl publishes *Early phonetic and phonemic changes in Austronesian*

Pawley publishes 'Melanesian diversity and Polynesian homogeneity: a unified explanation for language'

Lynch publishes 'Melanesian diversity and Polynesian homogeneity: the other side of the coin'

Blust publishes 'Some remarks on labiovelar correspondences in Oceanic languages', and 'Linguistic evidence for some early Austronesian taboos'

1982   Starosta, Pawley and Reid publish 'The evolution of focus in Austronesian'

Haudricourt and Ozanne-Rivierre publish *Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie)*

Zorc publishes 'Where, o where have the laryngeals gone? Austronesian laryngeals re-examined'

Pawley publishes 'Rubbish-man commoner, big man chief? Linguistic evidence for hereditary chieftainship in Proto-Oceanic society'

Simons publishes 'Word taboo and comparative Austronesian linguistics'

Blust publishes 'The linguistic value of the Wallace Line'

1983   Collins publishes *The historical relationships of the languages of central Maluku, Indonesia*

Geraghty publishes *The history of the Fijian languages*

Tryon and Hackman publish *Solomon Island languages: an internal classification*

1984   Sneddon publishes *Proto-Sangiric and the Sangiric languages*

Verheijen publishes *Plant names in Austronesian linguistics*

1985   Lichtenberk publishes 'Possessive constructions in Oceanic languages and in Proto-Oceanic'

1986   Zorc publishes 'The genetic relationships of Philippine languages'

Lichtenberk publishes 'Leadership in Proto-Oceanic society: linguistic evidence'

Marck publishes 'Micronesian dialects and the overnight voyage'

1987   Reid publishes 'The early switch hypothesis: linguistic evidence for contact between Negritos and Austronesians'

Blust publishes 'Lexical reconstruction and semantic reconstruction: the case of Austronesian 'house' words

1988   Ross publishes *Proto-Oceanic and the Austronesian languages of western Melanesia*

Mahdi publishes *Morphophonologische Besonderheiten und historische phonologie des Malagasy*

Blust publishes *Austronesian root theory*

1989   Adelaar publishes 'Malay influence on Malagasy: linguistic and culture-historical implications'

1990   Grace publishes 'The "aberrant" (vs. "exemplary") Melanesian languages'

Geraghty publishes 'Proto-Eastern Oceanic *R and its reflexes'

1990-1995   Blust begins *An Austronesian comparative dictionary* (available online)

1991	Dahl publishes *Migration from Kalimantan to Madagascar*
	Blust publishes 'The Greater Central Philippines hypothesis'
	Rehg publishes 'Final vowel lenition in Micronesian languages: an exploration of the dynamics of drift'
1992	Adelaar publishes *Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*
	Reid publishes 'On the development of the aspect system in some Philippine languages'
	Ross publishes 'The sound of Proto-Austronesian: an outsider's view of the Formosan evidence'
1993	Sagart publishes 'Chinese and Austronesian: evidence for a genetic relationship'
	Blust publishes 'Central and Central-Eastern Malayo-Polynesian', and 'Austronesian sibling terms and culture history'
	Rehg publishes 'Proto-Micronesian prosody'
1994	Reid publishes 'Morphological evidence for Austric' and 'Possible non-Austronesian lexical elements in Philippine Negrito languages'
1995	Tryon publishes the *Comparative Austronesian dictionary*
	Ross publishes 'Reconstructing Proto-Austronesian verb morphology: evidence from Taiwan'
1996	Lynch publishes 'Proto-Oceanic possessive marking'
	Grace publishes 'Regularity of change in what?'
	Ross publishes 'Is Yapese Oceanic?'
	Blust publishes 'Some remarks on the linguistic position of Thao', and 'The Neogrammarian hypothesis and pandemic irregularity'
1997	Ross publishes 'Social networks and kinds of speech community event'
	Blust publishes 'Ablaut in northwest Borneo'
1998	Ross, Pawley and Osmond publish *The lexicon of Proto Oceanic, vol. 1: Material culture*
	Hage publishes 'Was Proto-Oceanic society matrilineal?'
	Blust publishes 'Ca- reduplication and Proto-Austronesian grammar'
1999	Thurgood publishes *From ancient Cham to modern dialects*
	Blust publishes 'Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics'
	Hage publishes 'Reconstructing ancestral Oceanic society'
2000	Marck publishes *Topics in Polynesian language and culture history*
	Blust publishes 'Low vowel fronting in northern Sarawak'
	Zeitoun and Huang publish 'Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages'
2001	Lynch publishes *The linguistic history of southern Vanuatu*
2002	Wouk and Ross publish *The history and typology of western Austronesian voice systems*
	Lynch, Ross and Crowley publish *The Oceanic languages*
	Reid publishes 'Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages'
	Ross publishes 'The history and transitivity of western Austronesian voice and voice-marking'
	Lynch publishes 'The Proto-Oceanic labiovelars: some new observations'
	Kikusawa publishes *Proto-Central Pacific Ergativity: Its Reconstruction*

	and Development in the Fijian, Rotuman and Polynesian Languages
2003	Ross, Pawley and Osmond publish *The lexicon of Proto Oceanic, vol. 2: The physical enironment*
	Bender, Goodenough, Jackson, Marck, Rehg, Sohn, Trussel and Wang publish 'Proto-Micronesian reconstructions – 1' and 'Proto-Micronesian reconstructions – 2'
	Hage and Marck publish 'Matrilineality and the Melanesian origin of Polynesian Y chromosomes'
	Blust publishes 'Three notes on Early Austronesian morphology'
2004	Li publishes *Selected papers on Formosan languages*
	Sagart publishes 'The higher phylogeny of Austronesian and the position of Tai-Kadai'
	François publishes 'Reconstructing the geocentric system of Proto-Oceanic'
2005	Greenhill and Gray make the Austronesian Basic Vocabulary Database (ABVD) publicly accessible through the AN-LANG mailing list
	Adelaar and Himmelmann publish *The Austronesian languages of Asia and Madagascar*
	Sagart publishes 'Sino-Tibetan and Austronesian: an updated and improved argument'
	François publishes 'Unravelling the history of the vowels of seventeen Northern Vanuatu languages'
	Ross and Teng publish 'Formosan languages and linguistic typology'
	Scaglion publishes '*Kumara* in the Ecuadorian Gulf of Guayaquil?'
2006	Ross publishes 'Reconstructing the case-marking and personal pronoun systems of Proto-Austronesian'
2007	Blust publishes 'The linguistic position of Sama-Bajaw'
	Ross and Næss publish 'An Oceanic origin for Äiwoo, the language of the Reef Islands?'
	Blevins publishes 'A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands'
2008	Ross, Pawley and Osmond publish *The lexicon of Proto Oceanic, vol. 3: Plants*
	Lichtenberk publishes *A grammar of Toqabaqita*, and *A dictionary of Toqabaqita* (Solomon Islands)
	Liao publishes 'A typology of first person dual pronouns and their reconstructibility in Philippine languages'
	Greenhill, Blust, and Gray publish 'The Austronesian Basic Vocabulary Database: from bioinformatics to lexomics'
2009	Lobel and Riwarung publish 'Maranao revisited: an overlooked consonant contrast and its implications for lexicography and grammar'
	Clark publishes *Leo Tuai: a comparative lexical study of North and Central Vanuatu languages*
	Gray, Drummond and Greenhill publish 'Language phylogenies reveal expansion pulses and pauses in Pacific settlement'
2010	Zeitoun, Teng, and Ferrell publish 'Reconstruction of '2' in PAN and related issues'
	Lobel and Hall publish 'Southern Subanen aspiration'
	Blust publishes 'The Greater North Borneo hypothesis'
	Lobel publishes 'Manide: an undescribed Philippine language'

	Wolff publishes *Proto-Austronesian phonology with glossary*
2011	Ross, Pawley and Osmond publish *The lexicon of Proto Oceanic, vol. 4: Animals*
	François publishes 'Where *R they all? The geography and history of *R-loss in Southern Oceanic languages'
	Blust publishes 'The problem of doubletting in Austronesian languages'

8.1.1 The age of discovery

Although it is hard to know where to begin a history of comparative scholarship on the AN languages, it probably is safe to say that anything written before van der Tuuk (or perhaps von der Gabelentz) is 'prescientific' in the sense that term is normally understood today. Those who collected data (usually sea captains or their crews) had no professional training in the study of language, and although de Houtman, who commented that "there is much of Malay in Malagasy", drew valid conclusions about the relationship of Malagasy to Malay, and Cook readily recognised the unity of Polynesian, these insights were based on such extensive lexical similarity that a historical connection would be difficult to miss even without special training. Reland and Hervas y Panduro were classical scholars, but neither had expertise in the Austronesian area. Despite this Reland did compile short comparative vocabularies of Malagasy and Malay, and even drew attention to the correspondence MLG *v* : MAL *b*. His description of a *v* : *b* sound correspondence between Malagasy and Malay was, however, an isolated instance of precision in an era dominated by vague observations of 'similarity', and it was not until the second half of the nineteenth century that a serious attempt was made to describe some of the phonetically more variable sound correspondences holding between AN languages in terms of what, using the parlance of the day, came to be known as 'sound laws'.

8.1.2 Von Humboldt and von der Gabelentz

With von Humboldt and von der Gabelentz there is a clear sense of passing from the age of discovery to the early stages of a more analytic period, even if it was still dominated by armchair scholarship. Von Humboldt occupies a transitional position: the novel insights and approach that he introduced make it impossible to place him in the same category as Reland or Hervas y Panduro, but at the same time they do not qualify him for inclusion with scholars in the succeeding period. His colossal work of scholarship, *Über die Kawi-Sprache auf der Insel Java, nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* (1836-1839) was a major contribution to the study of Old Javanese with an excursus into comparative Austronesian linguistics, and a considerably more famous excursus into the relationship of language to human intellectual development. In over 1,800 pages of text and tables he laid out the most complete synthesis of descriptive and comparative knowledge about the AN languages available in his time. In comparison with the Dutch work that was soon to follow, however, it was superficial and derivative. Von Humboldt's sweeping consideration of the world's languages made reference not only to Kawi (Old Javanese), but also to Sanskrit, Greek, Chinese, Burmese, various American Indian languages (Cora, Nahuatl, Mayan, Arawakan, Delaware), and to a number of other

'Malayan' (Malayischen) languages.[78]   For Malay and Javanese he drew heavily on the work of the English scholars Raffles, Marsden, and above all Crawfurd. At the same time he made extensive use of the available material on Malagasy, Tagalog, and several of the Polynesian languages. Hundreds of pages are dedicated to describing the 'alphabets' and 'particles' (= grammatical morphemes) of various of these languages, to descriptions of verbal and nominal morphology, numeral systems, etc., and to numerous comparative remarks. Many of his comparative remarks are valid and insightful, but many others are not, and show that in comparing morphemes cross-linguistically von Humboldt (like his Indo-Europeanist contemporaries) was not yet sensitive to the notion of regularity in sound change. He cited many etymologies where he clearly recognised recurrence, as with the correspondence of Malagasy *h* to *k* in most other languages, or Malagasy *f* to *p* in most other languages, but at the same time he was willing to resort to vague phonetic similarity in comparing forms that looked like they might be historically related. To cite one of many possible examples, he noted the valid cognate relationship in Malay *laki-laki*, Javanese *laki*, Malagasy *lahy*, Tagalog *lalaki* 'male, man', but adds to these the unrelated forms Tongan *lahi*, Tahitian, Maori *rahi* 'big, large, great' which are not just semantically deviant, but also are not linked to the non-Polynesian languages by recurrent sound correspondences (1838:219). Examples such as this are instructive: since Malagasy *h* corresponds to *k* in most 'Malayan' languages, von Humboldt evidently assumed that a *k*: *h* correspondence was plausible whether it was recurrent or not (if the Polynesian forms were cognate we would expect Tongan *k* : Tahitian *ʔ* : Maori *k*). From the standpoint of lexical comparison one of the highlights of von Humboldt's work is a fold-out table of 131 words in nine languages (Malay, Javanese, Buginese, Malagasy, Tagalog, Tongan, Maori, Tahitian, Hawaiian). Since these clearly are presented as evidence of genetic relationship it is easy to fall into the trap of believing that they are treated as cognate sets, but in fact they simply contain the most common equivalent for the German semantic categories in question (sky, earth, water, sea, salt, etc.). The era of assembling cognate forms and clearly distinguishing them from non-cognates had not yet arrived.

There is no question that von Humboldt's treatise was a landmark of scholarship in several areas. However, with regard to the comparative study of the AN languages it seems fair to say that he stood on the very brink of the scientific era, but had not yet crossed into it. His philosophical treatise on the relationship of language to thought was far ahead of its time, but his approach to comparative issues in linguistics was in many ways no more advanced than that of Reland 130 years before him.

A little over twenty years after von Humboldt's *Über die Kawi-Sprache* another German linguist, Hans Conon von der Gabelentz (1861-1873), published an impressively resourceful study of a number of the languages of Melanesia. Drawing for the most part on privately circulated school primers, translations of scripture and religious tracts that had appeared during the previous two decades, von der Gabelentz constructed well-organised, informative sketches of the phonology and grammar of ten languages. These included not only such relatively 'easy' languages as Fijian and Bauro (southeast Solomons), but also a number of 'difficult' languages such as Anejom (Aneityum) of the southern New Hebrides

---

[78]   Ross (1996a) has pointed out that although von Humboldt is often credited with coining the term 'Malayo-Polynesian' this is untrue. Rather von Humboldt used four terms to refer to AN languages: 1) 'Malayischen' (= AN), 2) 'westlichen Malayischen' (= AN languages of insular Southeast Asia), 3) 'Südsee-Sprachen' (= AN languages of the Pacific in general), and 4) 'Polynesischen'. Although he did not explicitly comment on subgrouping, von Humboldt (1838:x) clearly took the unity of Polynesian ('den polynesischen Sprachzweig') as a given.

(Vanuatu), and Nengone and Dehu (called 'Maré' and 'Lifu' respectively) of the Loyalty Islands. The principal aim of this study was to determine whether the 'Melanesian' languages are genetically related, or whether they form a number of isolates or disparate groups as had been generally believed since the days of Förster. As noted earlier, von der Gabelentz firmly supported the view that most languages of Melanesia are AN. He based this conclusion primarily on grammatical resemblances, especially in the pronouns and genitive construction, observing that lexical similarities also occur, but are less common in the 'difficult' languages. Although von der Gabelentz tended to downplay comparative phonology, he did note (1861:69) that Anejom differs from Fijian in having an initial vowel corresponding to zero in some apparent cognates, as in *epeg* : *bogi* (*boŋi*) 'night', *etmai* : *tama* 'father', *ateuc* : *tico* (*tiðo*) 'sit', or *ero* : *rua* 'two'. However, he progressed no further in establishing regular sound correspondences (and included erroneous etymologies such as *ateuc* : *tico* alongside valid ones). Finally, von der Gabelentz also cited 'Malayan' and Polynesian forms for comparative purposes, as in comparing Fijian *laŋi* with 'Polynesian' *laŋi*, 'Malayan' *laŋit* 'sky'. Despite his recognition of such valid cognate sets he did not unite these under a common 'Malayo-Polynesian' prototype, as was subsequently the case in the wide-reaching comparisons of Kern.

8.1.3 The observational period: van der Tuuk to Kern

In describing the transition from a prescientific to an early scientific period in the comparative study of the AN languages it is necessary to take into account the wider social and scientific context in which research on these languages was embedded. In 1822 Jakob Grimm had presented his celebrated statement of the First Germanic Consonant Shift, setting the comparative study of the Indo-European languages into full motion, in 1859 Charles Darwin published his revolutionary ideas about the origin of species through natural selection, in 1863 August Schleicher had published the first family tree diagram, and although it was more than a decade before the Neogrammarians would issue their famous manifesto, by the early 1860s Dutch and German scholarship on the languages of Indonesia had begun to reach new standards of precision. It was within this intellectual climate that H.N. van der Tuuk (1824-1894), a maverick Eurasian employee of the Netherlands Bible Society (he was an outspoken atheist), initiated what might be called the 'observational period' in the comparative phonology of the AN languages.

In his 'Note on the relation of the Kawi to the Javanese' (1865) van der Tuuk pointed out three non-trivial sound correspondences holding between a few of the then better-known languages of western Indonesia and the Philippines, as follows (MAL = Malay, BTK = Batak, TAG = Tagalog, BIS = Bisayan, OJ = Old Javanese (Kawi), JAV = modern Javanese, BAL = Balinese, LPG = Lampung, MLG = Malagasy):[79]

---

[79] ě = schwa. Differences of meaning are not indicated in these or other tabulated cognate sets that will be discussed. I have modified minor details of van der Tuuk's orthography, but have intentionally left certain of his errors intact so as to better convey a sense of the pioneering character of his work, which by modern standards is flawed in various respects. Where he does not cite a cognate I have left the entry blank, even though a related form may occur (e.g. Tagalog *túlog* 'sleep').

Table 8.2 Sound correspondences first recognised by H.N. van der Tuuk (1865)

1. MAL *r* : BTK *r* : TAG *g* : BIS *g* : OJ = Kawi Ø : JAV Ø

	hear	**bathe**	**squeeze**	**belch**
MAL	dĕŋar	dirus	pĕrah	—
BTK	—	dirus	poro	—
TAG	—	—	pigá	tigáb
BIS	duŋúg	digus	pogá	togáb
OJ	rĕŋĕ	dyus	pwah	twab
JAV	ruŋu	a-dus	poh	a-tob

2. MAL *d* : BAL *d* : TAG *l* : BIS *l* : OJ = Kawi *r* : JAV *r*

	nose	**sleep**	**leaf**
MAL	hiduŋ	tidur	daun
BAL	—	—	don
TAG	ilóŋ	—	—
BIS	—	tulug	—
OJ	hiruŋ	turū	ron
JAV	iruŋ	turu	ron

3. MAL j : BAL j : BTK d : OJ d : JAV d

	lick	**road**	**far**	**rain**
MAL	jilat	jalan	jauh	hujan
BAL	—	jalan	joh	hujan
BTK	dilat	dalan	dao	udan
OJ	dilat	dalan	ma-doh	hudan
JAV	dilat	dalan	a-doh	udan

Shortly thereafter, in a dialect study of the Lampung language of south Sumatra with comparative notes, van der Tuuk (1872) included a seven-page section (141-148) devoted to 'sound laws.' Among other, less important sound correspondences, he drew attention to the following:

Table 8.3 A putative fourth sound correspondence recognised by van der Tuuk (1872)

4. LPG *r* : MAL *d/t* : BAL *d* : BTK *g/k* : JAV *d/r*:

	two	**wall**	**nose**	**navel**	**maggot**
LPG	rua	rindiŋ	iruŋ	pusor	hulor
MAL	dua	dindiŋ	hiduŋ	pusat	ulat
BAL	—	—	—	—	ulĕd
BTK	—	—	iguŋ	pusok	—
JAV	-do, -ro	—	iruŋ	pusĕr	ulĕr

In modern terms correspondence (1) is assigned to *R, and correspondence (3) to *z or *Z. However, we now know that correspondences (2) and (4) represent the same two proto phonemes (written by most scholars today as *d and *j). Curiously, van der Tuuk did not include Batak evidence in (2) which would have indicated, at least provisionally, that he

had confounded distinct correspondences (Toba Batak *iguŋ* 'nose', but *daon* 'leaf'). Moreover, in (4) he omitted Toba Batak *dua* 'two' and *diŋdiŋ* 'wall' which would have suggested either that distinct sound correspondences had been conflated, or that the correspondences in this and (2) are in complementary distribution. Since he did not consider the latter possibility, it is not clear how van der Tuuk would have proposed to distinguish correspondences (2) and (4), or to reconcile the internal contradictions in each of them.

Finally, van der Tuuk reiterated the *r* : *g* : Ø correspondence that he had first stated in 1865, and expanded it to include Lampung *y* : Malay *r* : Batak *r* : Tagalog *g* : Bisayan *g* : Balinese *h* : Ngaju Dayak *h* : Javanese Ø. In none of these cases did he propose reconstructions; the correspondences were observed, but not explained as the outcome of divergent descent from common ancestral forms. Unfortunately for comparative AN linguistics most of the remainder of this remarkable scholar's life was devoted to the compilation of his massive *Kawi-Balineesch-Nederlands woordenboek* (1897-1912), an oddly constructed guide to the translation of Old Javanese literature which, though it contains an abundance of comparative linguistic material, is extremely difficult to use for comparative purposes.

Somewhat after van der Tuuk had put his interest in comparative linguistics at the service of Old Javanese philology, J.L.A. Brandes (1857-1905), in a dissertation defended at the University of Leiden in 1884 again took up the question of sound correspondences in the AN languages. Brandes expanded the comparison initiated by van der Tuuk to Sundanese, Madurese, and several of the languages of Sulawesi (Tombulu, Buginese, Makasarese) which were then becoming known through the labors of B.F. Matthes and others. In addition he included some material identified only as 'Formosan' that was taken from the Siraya gospels of Gravius.[80]

Brandes' principal contribution to comparative AN phonology is the proposal (1884:19-20) that the correspondence Malagasy *i* : Tagalog *i* : Batak *o* : Bisayan *o* (= *u*) : Minangkabau *a* : Ngaju Dayak *a* : Makasarese *a*, together with schwa in a number of other languages, derives from an original mid-central vowel which he called after the Javanese name 'pĕpĕt'. Brandes also discussed palatal : dental correspondences in the languages of western Indonesia and the Philippines (but could not decide whether the palatals were original or secondary), gave the first inkling of the problem of doubleting in AN languages, and labeled the *y* : *r* : *g* : *h* : Ø and *d* : *l* : *r* : *g* correspondences 'van der Tuuk's First and Second Laws' respectively in honor of their discoverer (1884:139).[81]

Shortly after he defended his dissertation Brandes secured a post in the Dutch colonial civil service. As a result he spent his later years working at the Batavia museum where he was responsible for the decipherment of Old Javanese inscriptions and the translation of texts. During the remainder of his life he published almost nothing on AN linguistics. By

---

[80] Van der Tuuk (1897-1912) cites languages of Indonesia and the Philippines that are not mentioned in his earlier publications, but the Formosan material is never used, apparently because it was unknown to him. In an obituary for Brandes, Kern (1906b:302) credits his former student with rediscovering the Gravius volume which had 'slumbered forgotten in the University of Leiden for over two centuries.'

[81] Blagden (1902) added the correspondence Javanese *d* : Batak *d* : Malay *j* : Balinese *j* : Malagasy *r* as 'van der Tuuk's third rule', though this correspondence generally was not assigned a distinct proto source until Dyen (1951). Following the lead of van der Tuuk himself, Brandes confused at least two correspondences under the name 'van der Tuuk's second law.' This confusion was recognised by Conant (1911) on the basis of Philippine reflexes, and by Lafeber (1922) on the basis of reflexes in Nias, but not by Brandstetter. The relevant distinction was finally documented *in extenso* by Dempwolff (1924-1925).

contrast, his thesis advisor at the University of Leiden was a prolific writer on a variety of linguistic subjects. Hendrik Kern (1833-1917) began his academic career with the study of Dutch and Sanskrit, and did not publish on AN linguistics until he was 47. Kern's major contributions to the comparative phonology of the AN languages are his 1886 monograph *De Fidji-taal vergeleken met hare verwanten in Indonesië en Polynesië* (Fijian compared with its relatives in Indonesia and Polynesia), and his 1906 paper 'Taalvergelijkende verhandeling over het Aneityumsch' (A comparative study of Aneityum). In both of these publications Kern tried to identify lexical material shared by Fijian or Anejom (the modern spelling) with other AN languages, especially those of Indonesia. In doing so he performed the invaluable service of collecting many widely distributed cognate sets, and of demonstrating the AN affinity of some languages whose genetic affiliations were until then unclear. This is true not only of Anejom, which the missionary-lexicographer John Inglis in 1882 had declared to be Papuan (despite the earlier Anejom-Fijian comparisons of von der Gabelentz), but also of the 'language' (there is more than one) of the Philippine Negritos, and of several languages in the New Guinea area (Numfor, Yotafa) which were given comparative treatment in shorter studies.

   Kern was exceptional among Dutch scholars in including the entire AN language family within his purview, thus expanding his comparisons far beyond what van der Tuuk or Brandes had attempted. Although anticipated in some ways by von der Gabelentz, he was the first to employ rough approximations of lexical reconstructions capable of explaining the phonological development of AN languages both in island Southeast Asia and the Pacific, as with Fijian *walu* < 'M.P.' (Malayo-Polynesian) ualu, uwalu 'eight'. Quite frequently, however, Kern's 'reconstructions' are simply copies of the reflex in Old Javanese, which he evidently believed to be more conservative than other AN languages. He derives Fijian *vanua* 'land, region, place', *vatu* 'stone', and *uvi* 'yam', for example, from 'M.P.' wanua, watu, (h)uwi (cp. OJ *wanua/wanwa* 'continent; district, region, village', *watu* 'stone', (h)uwi 'a tuber'), despite such cognates as Malay *bənua, batu, (h)ubi*, Toba Batak *banua, batu, ubi*, or Tagalog *banwá, bató, úbi*, which unambiguously support prototypes with *b. Where a comparison shows greater phonetic variation, as with Fijian *ulo*, OJ *ulər*, Tagalog *uwód*, Malay (h)ulat, Malagasy *ulitra* 'maggot' Kern cited cognate forms but did not venture an etymon. Moreover, Kern never attempted a reconstruction of Proto Malayo-Polynesian phonology. Probably as a consequence of this procedure ('reconstruction' of lexical items but no phonological system) he also adopted a more casual approach to sound correspondences than is characteristic of van der Tuuk or even Brandes.

8.1.4 The early explanatory period: Brandstetter

   Although Kern proposed a number of rough and ready 'reconstructions' and was followed in this by Jonker (1914) no scholar discussed so far attempted to raise such a practice above the level of an *ad hoc* convenience. To do this it was necessary to relate lexical reconstructions to a system of original phonological contrasts. Because the languages of Indonesia and the Philippines are both phonologically and lexically more conservative than most Oceanic languages it is understandable that the earliest progress toward the reconstruction of the phonology of their common ancestor was made through comparison of the western languages. A first approximation to such a phonology was proposed at last in 1910, and refinements of it in 1911 and 1915 by a scholar who had an

excellent comparative grasp of Malagasy and the AN languages of insular Southeast Asia, but who excluded the languages of the Pacific from his investigations.

Renward Brandstetter (1860-1942) was born and received his early schooling in Lucerne, where he later returned to serve until retirement as a language teacher in the cantonal school. Like Kern, Brandstetter's interests were remarkably broad. The Dutch missionary-linguist S.J. Esser (1930) mentions Swiss folklore and dialectology, legal and musical history, the botany of the Lucerne canton and the study of Rhaeto-Romansch among his scholarly pursuits. Through a chance meeting with the Dutch Indologist G.K. Niemann who frequently vacationed in Switzerland, he also developed an intense interest in the languages of island Southeast Asia.[82]

Brandstetter was in many respects the epitome of the armchair scholar. Although he never visited any part of the AN-speaking world, he devoured nearly every grammar, dictionary and wordlist in print, and where these were not available he combed through texts to extract lexical and grammatical information relevant to his purposes. The results were impressive. In a number of well-organised, clearly written and neatly interlocking essays and monographs Brandstetter produced a systematic overview of the then-current comparative knowledge of the languages of island Southeast Asia that was conspicuously lacking in the rather fragmented Dutch work on the languages of Indonesia. No unbiased reader perusing the work of Kern or Brandes and Brandstetter could fail to conclude that, despite the important achievements of the early Dutch scholars, with Brandstetter AN comparative linguistics had reached a new level of factual and theoretical integration.

Brandstetter's account of the comparative phonology of 'Indonesian' languages appeals to certain key concepts that require prior discussion. Chief among these is the distinction between 'Common Indonesian' and 'Original Indonesian.' Although he proposed no subgrouping, Brandstetter divided his region of study into geographical areas that were regarded as corresponding in some degree with important linguistic breaks. In 1911 he recognised 'seven great insular regions' and 'three border districts' as follows: 1) Philippines, 2) Celebes, 3) Borneo, 4) Java-Madura-Bali, 5) Sumatra, 6) the Malay Peninsula with the adjacent islands, 7) Madagascar, 8) northern border (Batanes Islands and Formosa), 9) eastern border district (the islands from Lombok towards New Guinea), 10) southwestern border district (the Barrier Islands west of Sumatra, including Simalur, Nias, and Mentawai). A linguistic feature found in at least seven of these regions was unreservedly regarded as 'Common Indonesian' (CIN). Features found in fewer regions were regarded as CIN only with reservations. To illustrate, Brandstetter presented the following comparison for 'sky' (geographical regions are in parentheses):

---

[82] For a recent appraisal of the work of Brandstetter and its role in subsequent Austronesian studies see Blust and Schneider (2012), which represents the proceedings of a conference held in Lucerne on the sesquicentennial of Brandstetter's birth.

(1)	Tagalog	laŋit	(6)	Malay	laŋit
(2)	Tontemboan	laŋit	(7)	Malagasy	lanitră
(3)	Ngaju Dayak	laŋit	(8)	Ivatan	gañit
(4)	Javanese	laŋit	(9)	Bimanese	laŋi
(5)	Gayō	laŋit	(10)	Mentawai	laŋit

Figure 8.1 Data illustrating Brandstetter's concept of 'Common Indonesian'

Since the form *laŋit* appears in seven of his ten regions, Brandstetter concluded that this was the CIN term for 'sky'. Apparent cognates that diverge from the shape of the CIN norm were explained through the operation of general 'sound laws,' or where the material was insufficient for this, by citing parallel cases (e.g. Ivatan *añin* next to CIN *aŋin* 'wind'). In effect, these CIN forms were linguistic reconstructions, and Brandstetter acknowledges this in part II of his essay 'Common Indonesian and Original Indonesian' (1916:128): "We saw in sect. 1 that the word *laŋit*, either unchanged or modified only in accordance with strict phonetic laws, runs through a number of IN languages. How do we account for this fact? By the assumption that there was once a uniform original IN language, which possessed the word *laŋit*, and from which its offshoots, when they parted from it, took the word with them."

Having proposed an explanatory mechanism for the observed similarities and differences between languages Brandstetter then suggested that the term 'Original Indonesian' (OIN) be applied 'to all linguistic phenomena which in Part I have been pronounced to be Common IN.' In retrospect it appears that the concept of 'Common Indonesian' served mainly as a control on the assignment of phonological and lexical features to Original Indonesian in the absence of a well-established subgrouping theory. As Dyen (1971a:22) points out, however, Brandstetter's OIN differs from his CIN in having two distinct *r* sounds. While the concept of a CIN acted as a subgrouping control then, it did not impose limits beyond which OIN could not go.

Brandstetter proposed successive approximations of the OIN phonological (called 'phonetic') system in 1910, 1911, and 1915. His final statement (1916:248) shows the following contrasts:

Table 8.4 The 'Original Indonesian phonetic system' (Brandstetter 1916)

Vowels	a	i	u	e	o	ě
Semivowels	y	w				
Liquids	r_1	r_2	l			
Laryngeal	q					
Velars	k	g	ŋ			
Palatals	c	j	ñ			
Dentals	t	d	n			
Labials	p	b	m			
Sibilant	s					
Aspirate	h					

Seven comments are appended to this table, of which the follow merit mention: 1) the vowels *e* and *o* are generally of secondary origin, being reconstrutible as CIN in only two words, 2) r_1 is lingual (= van der Tuuk's second law), while r_2 is uvular (= van der Tuuk's first law/the RGH law), 3) *q*, the hamzah or glottal stop, is "almost always secondary" in the modern languages, there being only one good case for assignment to OIN. In addition Brandstetter defends his reconstruction of the palatal series and **b*, which he says are rejected by 'some scholars' (presumably Brandes and Kern respectively). The evolution of these segments is traced through a number of languages and 'laws' (Laws of the vowels, Laws of the semi-vowels, Laws of the liquids, etc.) are established for their development.

Brandstetter singles out four sound laws as particularly important: 1) the Pĕpĕt-Law, 2) the RGH-Law, 3) the Hamzah-Law, and 4) the law of the Mediae. Of these only the first two merit discussion today. In 1) he follows Brandes (1884) in recognising an original mid-central vowel that has changed in many languages, and like Brandes he calls it by its Javanese name, *pĕpĕt*. In 2) the name 'RGH-Law' (Malay *r* : Tagalog *g* : Ngaju Dayak *h*) is given as a description of r_2, and is contrasted with the 'RLD-Law,' a name identified variously with 'Van der Tuuk's Second Law' (Brandstetter 1906:61, where he illustrates it with the word for 'nose'), and with r_1 (Brandstetter 1916:280, where he illustrates it with the word for 'thousand'). As already noted, these identifications refer to different correspondences, and Brandstetter's 'RLD Law' is thus a confusing label:

Table 8.5 Correspondence of van der Tuuk's First and Second Laws
with Brandstetter's RGH and RLD Laws

Brandes	Brandstetter	Malay	TB	Tagalog		NgD
vdT 1	r_2 = RGH	urat	urat	ugat	uhat	vein
vdT 2	RLD	hiduŋ	iguŋ	iloŋ	uroŋ	nose
—	r_1	ribu	ribu	libo	ribu	thousand

Brandstetter's scholarly influence is perhaps most clearly visible in the work of the American Philippinist Carlos Everett Conant (1911, 1912), and his essays form the basis of the important comparative study of the languages of Melanesia by the English linguist Sidney Herbert Ray (1926). By contrast, his work was all but totally ignored in Holland. Thus Lafeber (1922), writing a decade after most of Brandstetter's major essays had appeared, referred to van der Tuuk in describing the phonological correspondences holding between Nias and other languages, with no mention of Brandstetter. Similarly, Esser (1927) in a 40-page discussion of the phonological history of Mori (east-central Sulawesi) never mentions Brandstetter by name, referring to the 'R-G-H' and 'R-L-D' laws only through the publications of Conant. Finally, Jonker (1932) contains a 27-page discussion of phonological changes in Leti of the Lesser Sunda Islands, with no reference of any kind to the rather clear advances that Brandstetter had succeeded in bringing to the comparative study of the languages of Indonesia. He proposes lexical reconstructions only where these can be made with minimal assumptions of change (MP ama > Leti *ama* 'father', MP anak > Leti *ana* 'child'). Where the correspondences are more complex he reverts to direct citation of cognate sets (e.g. Leti *asna*, Malay *asaŋ* 'gills', Leti *talla*, Malay *jalan* 'path, road'), very much as Kern had done nearly half a century earlier.

Brandstetter complained bitterly of the Dutch neglect of his work, and there were some belated efforts to smooth relations. In the 1920s he was made an honorary member of the Batavian Society of Arts and Sciences, and Esser (1930) offered him a generous encomium on his seventieth birthday that is remarkable for its admission of shortcomings in the Dutch

approach to historical linguistics in Indonesia (with the notable exception of van der Tuuk). But these concessions apparently came too late. On his death Brandstetter left the manuscript of an unfinished comparative dictionary in Paris with specific instructions that no Dutchman ever be given access to it. The Dutch attitude toward Brandstetter seems to have been motivated in part by the feeling that he was an 'outsider'— in particular someone who had no firsthand field experience with the languages of Indonesia –, and in part by a feeling that he had neglected the pioneering work of Dutch scholars, especially that of van der Tuuk, and had done little more than restate in more basic terms what others had said before him.[83] It would thus be well to review Brandstetter's achievements before proceeding to a discussion of his successors.

To begin with, it should not be forgotten that Brandstetter's orientation was primarily comparative. Since he had never visited the AN world his opportunities for fieldwork, and hence for original contributions to the description of AN languages were almost nonexistent. In this respect his experience differed fundamentally from that of his Dutch contemporaries and predecessors, nearly all of whom had lived in Indonesia, some for extended periods. Brandstetter's principal research goals in AN linguistics can be stated as follows: 1) to provide a typological overview of the languages of island Southeast Asia, including not only Indonesia and Malaysia, but also the Philippines, Madagascar and (insofar as the published data permitted) Formosa, 2) to reconstruct the sound system and some features of the morphology of the common parent of these languages, 3) to describe the changes (especially sound changes) that have taken place in the daughter languages in well-documented general statements ('sound laws'), 4) to provide a deeper morphological analysis of these languages than had hitherto been attempted, and 5) to compile a comparative dictionary of the languages.

With regard to 1) any fair reader must conclude that Brandstetter was eminently successful. Although they were originally published over 90 years ago, such essays as 'Root and word in the Indonesian languages' (1910), and 'The Indonesian verb: a delineation based upon an analysis of the best texts in twenty-four languages' (1912) are still among the most readable and informative general discussions of the typology of word formation in the entire literature of AN linguistics. As for 2), Brandstetter's was the first attempt to reconstruct the full system of 'Original Indonesian' phonological contrasts. Van der Tuuk had correctly identified several phonetically variable sound correspondences, including the 'RGH Law,' but these identifications a) left open the question of the direction of change, and b) were not placed within the context of a reconstructed system. Brandstetter further corrected Kern's erroneous 'reconstruction' of *w for *b in words such as *batu 'stone', and demonstrated convincingly—contra Brandes—that the palatal series in certain of the languages of Indonesia is the reflex of an ancient series of similar type, not a secondary development. As for 3) Brandstetter, because of his attention to explicit phonological reconstruction, was the first to describe the *direction* of sound changes in AN languages, a question that had befuddled all his predecessors. Although certain of the details would now be treated differently, his 'Phonetic phenomena in the Indonesian languages' (1915) remains a model of lucid exposition, and a source of useful comparative information. With respect to 4) Brandstetter, following a few programmatic remarks by earlier scholars, pioneered the analysis of disyllabic word bases into smaller

---

[83] Apart from some playing down of van der Tuuk's work there appears to be little basis for this attitude. Brandstetter's papers in fact contain abundant references to the Dutch literature on the languages of Indonesia—especially to the contributions of Kern and Jonker, which he praises in various passages— but also to Brandes, van der Tuuk, Juynboll, Niemann, Adriani and others.

recurrent partials, a matter that has already been discussed in Chapter 6. It is only with 5) that Brandstetter's efforts might be regarded as genuinely disappointing. His *Ein Prodromus zu einem vergleichenden Wörterbuch der malaio-polynesischen Sprachen* (1906) is a very provisional comparative study of the lexicons of a score of AN languages which makes little if any advance over the work of Kern and other Dutch scholars.

On the debit side Brandstetter failed to recognise the ambiguity in 'Van der Tuuk's Second Law,' and in fact compounded the problem by associating the same label with his *r₁. His treatment of *q and *h were inadequate, and he failed to grasp the significance of certain phonological contrasts that are preserved in the Formosan languages (including Siraya, which he consulted), but which are weakened or lost outside Taiwan. Moreover, as already stated, he chose to exclude the languages of the Pacific from his studies despite Kern's pathbreaking papers which had clearly demonstrated the connection not only of Fijian, but also of more problematic languages such as Anejom, Numfor and Yotafa, with the better-known languages of Indonesia and the Philippines.

On balance, then, Brandstetter left some major problems unsolved and did not attempt to confront others (e.g. subgrouping), but in several respects he advanced the comparative study of the AN languages well beyond the level attained by van der Tuuk, Brandes, and Kern. By the 1920s no serious scholar in this field could afford to ignore his work.

8.1.5 The developed explanatory period: Dempwolff

What might be called 'the developed explanatory period' begins with the appearance of truly systematic work in phonological reconstruction. While Brandstetter was well-informed about the languages he described, and gave a reasonably good overview of typology for his time, his work in phonological and lexical reconstruction lacked the discipline and precision of later work. One of those who did not ignore Brandstetter's work was Otto Dempwolff (1871-1938).

The initial period

Dempwolff began his career as a medical doctor, developing an interest in linguistics as a by-product of his experiences in the German colonial possessions of the Pacific and east Africa (Blust 1988b). His first experience of the tropics was in Papua (then German New Guinea), where he served as a medical doctor from 1895 to 1897. During this period, while professionally engaged in malaria research on the local population, Dempwolff studied Siar, a dialect of the AN language Gedaged, with the missionary Bergmann. During a second stay in Papua from 1901 to 1903 he studied two other AN languages: Tami, with the missionary Bamler, and Yabem, with the missionary Pfalzer. At some time before 1901 he had also learned Pidgin English and Malay for use as *lingua franca*s. Although all of these languages appear to have been acquired initially as tools to facilitate the execution of his medical duties, Dempwolff soon turned them into objects of study. Moreover, through them he was able to collect pioneering information on many other languages for which he claimed no spoken fluency.

In his first publication Dempwolff (1905) presented brief vocabularies, a few sentences and short texts for some 28 languages spoken on the north coast of New Guinea, and on other islands in the Bismarck Archipelago. The languages are divided into 'melanesische Sprachen' (15 languages), and 'Papuasprachen' (13 languages), although it is evident that some of the latter (as Bunu) contain some basic AN lexical items. The material is written in accordance with a system for the transcription of unwritten languages that was

developed by the German Egyptologist and palaeographer Karl Richard Lepsius in the mid-nineteenth century, and which Dempwolff would continue to use throughout his career. Perhaps the most distinctive feature of this publication is Dempwolff's precise identification of his sources. For each language he noted where, and in which month and year the material was collected. In addition, he often indicates the language of elicitation and the names and ages of his informants. In most other respects, however, there is little to distinguish this from contemporary articles by other colonial civil servants of the era. Contact with speakers of the languages clearly had been brief, the materials are limited, and the orthography bristles with not always helpful diacritics. Finally, the paper was descriptive; a few comparative remarks appear in connection with Wuvulu (Admiralty Islands), but Dempwolff's interest in comparative problems was yet to awaken, or at least to take definite form.

Following his early service in New Guinea Dempwolff worked in German East Africa, where he met Carl Meinhof, sometimes called the 'father of comparative Bantu linguistics,' and carried out fieldwork on several previously undescribed languages. During this period his concern with phonetic problems deepened, and he acquired a sound understanding of the comparative method. For reasons of health Dempwolff eventually was forced to leave the tropics. In 1911 he began to work at the phonetics laboratory of the Kolonial Institut (later to become the University of Hamburg), where he was concerned primarily with the analysis of his African data. From 1919, when the University of Hamburg emerged from a collection of earlier institutes, he taught African and AN linguistics as an assistant to Meinhof, and in 1931 he founded his own department, the Seminar für indonesische und Südseesprachen.

The groundwork

Dempwolff's first comparative study of AN languages, *Die Lautentsprechungen der indonesischen Lippenlaute in einigen anderen austronesischen Südseesprachen* (LIL) appeared in 1920, when he was nearly 50 years old. It was the earlier of two major contributions preparing the ground for his magnum opus. According to Dahl (1976:6), who studied briefly in Hamburg during the 1930s, 'Dempwolff considered his work as a continuation of Brandstetter's, and Brandstetter had certainly the same view.' A natural starting point for this continuation was the extension of 'Proto Indonesian' to the Oceanic languages that Brandstetter had arbitrarily excluded from his research program. Dempwolff lacked Brandstetter's literary flair, but shared his methodical temperament, aiming from the very beginning at a systematic explanation of sound correspondences in the most diverse AN languages.

LIL is concerned with demonstrating the regularity of labial correspondences between Oceanic and non-Oceanic AN languages. There were at least two reasons why Dempwolff must have considered this demonstration important. First, although von der Gabelentz (1861-1873) appears to have left the question largely open, such influential writers as Codrington (1885:201-202), and Friederici (1912:20) had maintained that the comparative method developed in Indo-European linguistics could not be applied to the languages of Melanesia, as these fail to exhibit regular phonological correspondences. Second, Brandstetter's work represented the epitome of systematisation in comparative AN linguistics in 1920, yet dealt with only half of what was generally recognised as a single language family. By demonstrating that the 'Melanesian' languages are amenable to systematic comparative treatment through an in-depth analysis of a small set of contrasts Dempwolff undoubtedly hoped simultaneously to put to rest the long-held belief that these

languages stand outside the pale of scientific investigation, and to reaffirm Kern's pioneering demonstration that the overwhelming majority of languages in Island Melanesia are indeed Austronesian. At the same time he explicitly observed (1920:5) that the 'Indonesian' languages have preserved the original sound system more faithfully than the more easterly languages.

Dempwolff's point of departure was a statement in Kern (1886) that 'Malayo Polynesian' *p and *w (= *b) appear as Samoan *f*, Fijian *v*, but that *w sometimes is reflected instead as Samoan *p*, Fijian *b* ([mb]). Kern posed this problem almost as a riddle, and Dempwolff set out to solve it. Citing material from twelve 'Indonesian' languages and seven 'South seas' or eastern languages in 350 clearly laid out cognate sets, he was able to demonstrate statistically that the development of the 'Proto Indonesian' labials exhibits compelling regularities in all nineteen languages (1920:89-90). The anomalies noted by Kern were convincingly traced to distinctions of simple and prenasalised consonants in the western languages (hence in 'PIN'): *p/b > Samoan *f*, Fijian *v*, but *mp/mb > Samoan *p*, Fijian *b*. Dempwolff distinguished 'empirical' cases in which prenasalisation was actually attested in cognate IN forms, from 'hypothetical' cases in which the cognate forms in western AN languages exhibit a simple consonant in disagreement with the reflex in eastern languages (the latter presumably reflecting a prenasalised variant). Both the strength of the statistical association in the first category and the often sporadic character of medial homorganic prenasalisation in IN languages appeared to justify the extrapolation of this explanation to the second category of comparisons. Similar anomalies, however, appear in initial position, where few IN languages permit a prenasalised consonant. Dempwolff suggested that these cases could be explained by an appeal to morphology. Homorganic nasal accretion plays a part in the morphology of many IN languages, and the aberrant correspondences in initial position, he reasoned, could be residues of a once active morphological process.

Apart from his treatment of labial correspondences Dempwolff also provided an outline of all other major phonological changes in the seven eastern languages, along with detailed information on phonological conditions as determinants of split. The monograph concludes with three unanswered questions and an important, if somewhat tentative, subgrouping hypothesis. In the first question Dempwolff asks what are the prenasalised correspondents of the IN dentals, palatals and velars in Oceanic languages. In the second he asks why Mota occasionally shows labiovelar m^w and p^w instead of the more frequent plain *m* and *v* corresponding to *m* and *p/b* in IN languages. Since this split cannot be attributed to conditioned change Dempwolff suggests that it may ultimately be due to contact with Papuan languages, in which labiovelars are not uncommon. Dempwolff's third question leads to one of his major conclusions: do the rules of change (i.e. merger of *p/b and *mp/mb) which are stated for the seven eastern languages in his study also apply to other eastern languages? Based on a token sample of suggestive evidence he answers this question tentatively in the affirmative, noting that if true this conclusion is of interest not only to linguistics, but also to ethnology. A phonological merger not typical of other AN languages could hardly have occurred over a vast and continuous area of the Pacific as the result of innumerable independent changes. Rather, it must indicate a period of common development for these eastern languages after their separation from the speech community that gave rise to the entire AN language family.

By any reckoning LIL must be considered a fundamental breakthrough in AN comparative linguistics. In it Dempwolff succeeded not only in demonstrating that the comparative method can be applied to the languages of Melanesia, but also in providing

the first tentative indications of the now well-established Oceanic subgroup. His treatment of phonological correspondences equaled or exceeded Brandstetter's best work, and his breadth of coverage was approached only by Kern among his predecessors.

The second contribution that prepared the ground for Dempwolff's *magnum opus* was *Die l-, r- und d-Laute in austronesischen Sprachen* (1924-1925). In this study Dempwolff sought to bring order to the diverse phonological correspondences that include *l*, *r*, or *d* as constituent segments through an analysis of comparative material from 50 languages representing all major geographical regions of the AN world. To account for these correspondences he proposed seven distinctions, as well as certain conventions for the representation of ambiguous proto segments. These distinctions and their phonetic interpretations are given as 1) *l (alveolar fricative), 2) *ḷ cerebral fricative), 3) *ɣ (voiced velar fricative), 4) *d (voiced alveolar stop), 5) *d′ (voiced coronal-palatal stop), 6) *ḍ voiced cerebral stop, 7) *g′ (voiced dorso-palatal stop). The segments in question are placed in a fully reconstructed sound system consisting of 21 consonants, 11 'nasal clusters' and four vowels (1924-25:37). In effect, then, this study is a re-examination of the two van der Tuuk sound laws which attempts to place them within a wider network of phonological correspondences. Like Conant (1911) and Lafeber (1922) before him, Dempwolff concluded that 'van der Tuuk's Second Law' actually covers more than one sound correspondence.

As in his earlier monograph Dempwolff gives a powerful demonstration of the overwhelming regularity of the correspondences examined. Exceptions to the rules of change are nonetheless treated at great length, and every effort is made to explain them, whether through phonological conditioning, borrowing or reconstructed doublets. It is perhaps in its conscientious acknowledgement of irregularities that Dempwolff's work most clearly transcends that of Brandstetter, whose toleration of mutually corroboratory irregularities in a number of languages prevented him from clearly distinguishing the correspondences that Dempwolff assigned to *ḷ (*ḷibu 'thousand') and *g′ (*ig′uŋ 'nose'). Finally, it is observed (1924-1925:318) that the great majority of languages east of Numfor have merged *d′ and *g′ in a single segment, and *nd′/ŋg′ in another, thus adding further support to the argument for a large eastern subgroup of AN languages.

In addition to these major contributions several shorter studies are noteworthy as steps toward Dempwolff's final reconstruction of 'Uraustronesisch' phonology and his conclusions regarding the phonological evolution of the modern languages. The most important of these probably is his 'Das austronesische Sprachgut in den melanesischen Sprachen' (1927), in which all phonological mergers that Dempwolff ultimately recognised as characteristic of the eastern languages are stated as innovations in a common post-UAN proto language called 'Urmelanesisch' (= Proto Oceanic).

During the same period the German ornithologist Erwin Stresemann published an important comparative study of the languages of the central Moluccas, including Ambon, Buru, Seram and their smaller or less important satellites, which drew heavily on Dempwolff's reconstruction of Uraustronesisch phonology (Stresemann 1927).

The VLAW

Dempwolff's magnum opus, *Vergleichende Lautlehre des austronesischen Wortschatzes* (Comparative phonology of Austronesian vocabularies, hereafter VLAW) is a three-volume study that in every respect represents the culmination of his earlier efforts to bring order into the comparative investigation of the AN languages (remarkably, he also continued to publish in African linguistics during much of this period). Volume I (1934)

proposes an 'inductive' reconstruction of the 'Original IN' sound system based on a comparison of three languages: Tagalog, Toba Batak, and Javanese. In volume II (1937) the adequacy of this reconstruction is tested through a 'deductive' application to material from three other 'Indonesian' languages, Malay, Ngaju Dayak, and the Merina dialect of Malagasy (called 'Hova'), two 'Melanesian' languages (Fijian and Sa'a), and three Polynesian languages (Tongan, Futunan and Samoan). As it is found adequate to explain all phonological developments in these languages, 'Proto Indonesian' is rechristened 'Proto Austronesian' (Uraustronesisch). Volume III (1938) is a comparative dictionary that contains some 2,215 reconstructions with supporting evidence, based almost entirely on the eleven languages examined in volumes one and two.

Dempwolff begins VLAW I with a definition of the term 'Austronesian' which follows Schmidt (1906). He then adds immediately that because the AN languages 'do not have a common grammar,' his investigation will be restricted to a comparison of vocabulary, with special attention to the sound correspondences. Following a generally thorough discussion of his basic assumptions and methodological principles he proceeds to a description of the phonology and verb morphology of Tagalog, Toba Batak and Javanese. On the basis of cognate sets represented in at least two of these languages Dempwolff commences his reconstruction of 'Proto Indonesian' words conforming to the canonical shape CVCVC. He continues with the reconstruction of words of the shape CVCCVC, and concludes with the reconstruction of words of 'other than two syllables.'

Dempwolff's reconstruction is of only historical interest today, and there is no need to document it in detail. Rather, a few general remarks on the conventions he employed and his philosophy of inference should suffice to give an adequate picture of how he worked and what he accomplished. The following cognate sets and the reconstructed forms from which they are derived provide a glimpse into his methods (UAN = 'Uraustronesisch', TAG = Tagalog, TB = Toba Batak, JAV = Javanese): UAN 'apuj > TAG *'apoi*', TB *'api*', JAV *'api*' 'fire', UAN bu'uk > TAG *buhok*, TB *'o|buk*, JAV *wo'* 'head hair', UAN d'aɣum > TAG *ka|rayom*, TB *d'arum*, JAV *dom* 'needle', UAN hatul̹ > TAG *hatol*, TB *'atur*, JAV *'atur* 'regulation', UAN hud̹i' > TAG *hulí*', TB *p-udi*', JAV *b-uri*' 'hind part', UAN nijuɣ > TAG *niyóg*, TB *niur* 'coconut tree', UAN tuva' > TB *ma|tua*', JAV *tuwa*' 'old'.

Dempwolff's orthography

The first thing the modern reader of this material probably will find strange is its orthography. As already noted, Dempwolff took an active interest in phonetics. He worked in the phonetics laboratory of the Kolonial Institut from 1911 to 1919, and in his first comparative study he deplored the fact that he was forced to work with 'letters' (*Buchstaben*) because his sources often gave too little phonetic information to allow a comparison of 'the sounds themselves' (1920:6ff). Given this interest it is surprising that Dempwolff made no use of the International Phonetic Alphabet, which was first published in 1888 and had been considerably refined by the 1930s. Instead, adhering to a convention he had followed consistently since his first publication in 1905, he regularised the orthography of his sources in accordance with the 'Standard Alphabet' of the German Egyptologist Richard Lepsius. The phonetic value of the Lepsius symbols is explained as follows (1934:14-16):

1. *v, r, j, ɣ, l* and *s* are classed as 'fricatives' (*Reibelauten*), of which *v* is labiodental, *r, l* and *s* are alveolar, *j* is palatal, and *ɣ* is velar.

2. Javanese *a̲* is a rounded vowel 'akustisch zwischen *a* und *o*.'

3.   a 'bow' under consecutive vowel symbols indicates that the vowels are not pronounced separately, but as a diphthong (e.g. Tagalog *'apoi'* 'fire'). This is omitted from the reproductions of Dempwolff's transcriptions given here.

4.   a dot under the symbol for an alveolar consonant indicates the corresponding retroflex consonant: *ṇ, ḍ, ṭ, ḷ.*

5.   a slash over *n* and *ŋ* (*ń, ŋ́*), or to the right of *d, t, g,* or *k* (*d', t', g', k'*) indicates a palatal, of which *ń, d', t'* are articulated with the forepart, and *ŋ́, g', k'* with the hindpart of the tongue.

6.   ' (the 'spiritus lenis') indicates tight glottal closure, while ' (the 'spiritus asper') indicates loose glottal closure. Thus, the word for 'fire' shows glottal onset in Tagalog and Toba Batak, but smooth vocalic onset in Javanese.

In a few cases Dempwolff uses one symbol in writing attested forms and another in writing reconstructed forms for the same sound (e.g. Javanese *ě*, but PIN *ə). An acute accent indicates contrastive stress, which Dempwolff marked in Tagalog, but not consistently. What are assumed to be fossilised affixes are separated from the stem by short vertical strokes. All other symbols have their expected values.

Dempwolff's phonological interpretation

Dahl (1976:9) has observed that Dempwolff claimed to be a Neogrammarian, and that in conversation he rejected the concept of the phoneme. This is certainly suggested in his writing of subphonemic detail, as with Tagalog [u] and [o] or Javanese [a] and [ɔ] (written a̲) which, at least in native vocabulary, are best regarded as allophones of single phonemes /u/ and /a/ respectively. But Dempwolff explicitly recognised this complementarity, and further noted that Tagalog [o] alternates with [u], and that Javanese [ɔ] alternates with [a] before a suffix. Other phones which he recognised as being in complementary distribution—despite his orthography—are Tagalog [ɾ], found only intervocalically in native forms, and [d] (found elsewhere), Toba Batak [h] (before a vowel) and [k] (elsewhere), and Javanese [ʔ] (word-finally after any vowel other than schwa), and [k] (elsewhere). In each of these cases as well as with Javanese *b* and *w*, which vary freely in some forms, Dempwolff regarded the segments in question as 'equivalent' for comparative purposes. The sound system that he reconstructed for 'Proto Indonesian' (and ultimately Proto Austronesian), then, is unquestionably a system of phonemic contrasts, and we can feel free to use the term 'phoneme' in referring to its members.

Dempwolff's chart of the PIN sound system, set out to conform more closely to the format customarily used by linguists today, appears in Table 8.6 (diphthongs, which he includes in the descriptions of all attested languages, but not in his reconstruction, have been added; *ń* and *ŋ́* which occur only pre-consonantally, are omitted):

Table 8.6 The 'Proto Indonesian' sound system according to Dempwolff (1937)

	Consonants					
	1	2	3	4	5	6
a)	p	t	ṭ	t′	k′	k
b)	b	d	ḍ	d′	g′	g
c)	m	n		ń		ŋ
d)	v	l	ḷ	j		ɣ

Vowels			**Laryngeals**		**Diphthongs**	
i		u	ʻ	h	uj	iv
	ə					
					aj	av
	a					

1 = labial, 2 = alveolar, 3 = retroflex, 4 = front palatal, 5 = back palatal, 6 = velar; a) = voiceless stop, b) = voiced stop, c) = nasal, d) = fricative. In addition, all stops may be prenasalised medially. Although prenasalised stops are best viewed as clusters in PIN they produced a distinct consonant series in many Oceanic languages, and so are listed separately by Dempwolff. Among important phonotactic constraints may be mentioned:

1.   * g′ does not occur morpheme-initially.

2.   *ṭ, *k′, *d′ and *ń do not occur syllable-finally.

3.   consonant clusters may consist only of a homorganically prenasalised stop or of the abutting consonants in a reduplicated monosyllable.

4.   initial *j is reconstructed in two words: *javak 'monitor lizard', *juju' 'coconut crab'. Both were in fact trisyllables (in modern transcription *bayawak, and *qayuyu).

5.   *ə is not reconstructed in prepenultimate position, nor before final spiritus asper.

The major correspondences that support these contrasts are listed in Table 8.7 (SI = syllable-initial, SF = syllable-final, WI = word-initial, WF = word-final). Where conditions are not stated the correspondence is understood to hold in all positions.[84]

---

[84]   +C and –C symbolise contraction and non-contraction of vowel sequences. In Toba Batak d) 4 (C) stands for the coalescence of *a plus *j before a second *a (*b|uh|aja' > *buea'* 'crocodile', *daja' > *dea'* 'trickery, deceit'), but not before other vowels (*kaju' > *hau'* 'wood'). In Javanese the loss of *ɣ or *ʻ between unlike vowels generally led to mutual assimilation and contraction of the resulting sequence (*ʻaɣut' > *'os|os* 'current', *ʻuɣat > *'ot|ot* 'vein, tendon', *buʻah > *woh* 'fruit', *liʻaŋ > *leŋ* 'cave'). Contraction of these sequences did not occur following loss of *h. Between like vowels loss of a consonant or laryngeal led to simple contraction of the resulting sequence in Toba Batak and Javanese: *d′ahat > TB *d'at* 'evil', *baɣa' > JAV *wa̲|wa̲'* 'ember'.

Table 8.7 Correspondences supporting the 'Proto Indonesian' sound system
(Dempwolff 1937)

		TAG	TB	JAV	PIN	Conditions
Consonants						
a)	1	p	p	p	*p	
	2	t	t	t	*t	
	3	t	t	ṭ	*ṭ	
	4	s	s	s	*t′	
	5	s	s	t′	*k′	
	6	k	h	k	*k	SI
	6	k	k	k	*k	SF after *ə
	6	k	k	ʼ	*k	SF elsewhere
b)	1	b	b	b	*b	SI
	1	b	p	b	*b	SF
	2	r	d	d	*d	/V_V
	2	d	t	d	*d	SF
	2	d	d	d	*d	WI
	3	l	d	ḍ	*ḍ	SI
	3	d	r	d	*ḍ	SF
	4	d	d′	d′	*d′	WI
	4	r	d′	d′	*d′	V_V
	5	l	g	r	*g′	SI
	5	d	k	r	*g′	SF
	6	g	g	g	*g	SI
	6	g	k	g	*g	SF
c)	1	m	m	m	*m	
	2	n	n	n	*n	
	3	n	n	ń	*ń	
	4	ŋ	ŋ	ŋ	*ŋ	
d)	1	w	Ø	w	*v	
	2	l	l	l	*l	
	3	l	r	r	*ḷ	
	4	y	(C)	y	*j	/a_a
	4	y	Ø	y	*j	between other vowels
	5	g	r	Ø(+C)	*ɣ	

	TAG	TB	JAV	PIN	Conditions
Laryngeals					
1	ʼ	ʼ	ʻ	*ʻ	WI
1	ʻ	ʻ	ʻ	*ʻ	WF
1	h	Ø(+C)	Ø(+C)	*ʻ	between like vowels
1	h	Ø	Ø(+C)	*ʻ	between unlike vowels
2	h	ʼ	ʻ	*h	WI
2	ʼ	ʻ	h	*h	WF
2	Ø	Ø(+C)	h	*h	between like vowels
2	Ø	Ø	Ø(-C)	*h	between unlike vowels

	TAG	TB	JAV	PIN	Conditions
Vowels					
1	i	i	i	*i	
2	u/o	u	u	*u	
3	u/o	o	ĕ	*ə	next to *u
3	a	o	ĕ	*ə	next to *a
3	i	o	ĕ	*ə	elsewhere
4	a	a	a̱	*a	/__(Ca)#
4	a	a	a̱	*a	elsewhere

	TAG	TB	JAV	PIN	Conditions
Diphthongs					
1	oiʻ	iʻ	iʻ	*-uj	
2	yoʻ	iʻ	uʻ	*-iv	
3	aiʻ	eʻ	eʻ	*-aj	
4	aoʻ	oʻ	oʻ	*-av	

	TAG	TB	JAV	PIN	Conditions
Prenasalised stops					
1	mp	pp	mp	*mp	
2	nt	tt	nt	*nt	
3	nt	tt	ṇṭ	*ṇṭ	
4	ns	ts	ŋs	*ńtʼ	
5	ns	ts	ńt-	*ńkʼ	
6	ŋk	kk	ŋk	*ŋk	
7	mb	mb	mb	*mb	
8	nd	nd	nd	*nd	
9	nd	nd	ṇḍ	* ṇḍ	
10	nd	ńdʼ	ńdʼ	*ńdʼ	
11	nd	ŋg	nd	*ńgʼ	
12	ŋg	ŋg	ŋg	*ŋg	

Where some languages reflect a nasal cluster but others the corresponding simple stop Dempwolff reconstructed a 'facultative' nasal in parentheses, as with *'a(ŋ́)g'i' 'younger sibling'. While the synchronic interpretation of this convention is clear (some languages reflect the nasal, others do not), its diachronic significance was never made explicit: does it mean 1) that UAN may or may not have had the nasal, or 2) that prenasalised and simple variants were both found in the proto language?

Many consonant clusters were reconstructed in CVCCVC reduplications, as with *bəg'bəg', *dakdak, *gəmgəm, *lunlun or *ŋatŋat. These changed as they would in the corresponding (= SI, SF) non-clustering environment, with the following exceptions:

1.   In Toba Batak the first of two abutting consonants may drop, as in *bəg'bəg' > *bobok* 'wind around', buḍbuḍ > *bubur* 'chop up', or *ḷadḷad > *rarat* 'spread out'.

2.   Toba Batak clusters of nasal plus voiceless stop generally yielded a geminate consonant corresponding in position to the stop (but *ńt', *ŋ́k' > *ts*). However, clusters of velar nasal plus dental or palatal stop show no assimilation (*taŋtaŋ > *taŋtaŋ* 'grasp', *ṭuŋṭuŋ > *tuŋtuŋ* 'dull ringing sound', *t'uŋt'uŋ > *suŋsuŋ* 'go against'.

3.   Javanese nasals normally assimilated in place to a following stop, whereas the initial consonant of other clusters was dropped: *banban > *bamban* 'bast fiber', *d'əŋd'əŋ > *d'ĕnd'ĕŋ* 'stand', *taŋtaŋ > *tantaŋ* 'grasp', but *bakbak > *baba'* 'to skin, flay, peel', *ḍapḍap > *ḍapḍap* 'k.o. shade tree', *pitpit > *pipit* 'pinch', etc.

Dempwolff's treatment of exceptions

As already remarked, one of the hallmarks of Dempwolff's work was his scrupulous attention to exceptions. The most important classes of exceptions are as follows:

1.   Non-final *b is often reflected as *w* in Javanese.

2.   Parallel to the foregoing, *d and *ḍ both frequently appear as Javanese *r*.

3.   In a number of cases Tagalog and Javanese disagree in indicating *d or *ḍ. Where this occurs Dempwolff favored the Tagalog evidence, since Javanese substitutes a retroflex stop for the corresponding alveolar *d* in some loanwords, as with *roḍa* (< Portuguese *roda*) 'wheel'. Parallel to this Tagalog occasionally reflects *d' with *l* for expected *r*.

4.   In Tagalog non-initial *l sometimes appears as *h* or Ø (with the development of automatic glottal stop between like vowels, or homorganic glides between unlike vowels the first of which is high).

5.   The sequence *lVr regularly assimilated to *rVr* in Toba Batak, and the sequence *rVr (regardless of the source of the *r) regularly dissimilated to *lVr* in Javanese.

6.   PIN *ɣ sometimes appears as Tagalog *y*, and as Javanese *r*.

7.   *-' sometimes appears as Tagalog –' (i.e. Tagalog sometimes shows an unexplained final glottal stop).

8.   *-h sometimes appears as Tagalog –'. In both 7) and 8) Dempwolff favored the Javanese evidence.

9.   *i and *u occasionally yield *e* and *o* respectively in both Toba Batak and Javanese.

10.   *-aj is reflected as -i' in some Javanese forms.

Although Javanese *ḍ* < **d* and *r* < **γ* were attributed to borrowing, and the subregularity in 5) to phonological conditioning, the remainder of the above changes were labeled 'unexplained exceptions.' Where irregularities in two or more witnesses were mutually corroboratory, but affected only a few forms, Dempwolff circumvented the need to recognise further exceptions or to reconstruct new phonological distinctions by positing proto doublets instead, as with **hud′an/'uḍan* 'rain'. As will be seen, this expedient was particularly crucial in coping with the laryngeal correspondences.

VLAW:II presents a 'deductive application' of Dempwolff's reconstructed 'Proto Indonesian' to three other 'Indonesian' languages, two 'Melanesian' languages, and three Polynesian languages. A number of ambiguities in the reconstructions are removed and some new lexical reconstructions achieved, but the system of phonological contrasts is found adequate to account for all reflexes in these eight languages. To understand why Dempwolff regarded his reconstruction as justified we must consider his thinking about two kinds of questions: 1) subgrouping, and 2) diagnostic witnesses.

Dempwolff's theory of Austronesian subgrouping

Although he reached a major subgrouping conclusion in VLAW:II, namely that the languages of Polynesia, Melanesia east of Numfor, and Micronesia exclusive of Palauan and Chamorro descend from an immediate common ancestor that he called 'Urmelanesisch' (today called 'Proto Oceanic'), Dempwolff approached the reconstruction of PAN phonology without an explicit subgrouping theory. For convenience he classified the languages as Indonesian, Melanesian (including the Oceanic languages of Micronesia), and Polynesian. No lexical reconstruction was assigned to PAN unless it was reflected in at least two languages, of which at least one was 'Indonesian.' While this approach was consistent with his view that the eastern languages belong to a subgroup apart from those in the west, it implied a much greater degree of genetic independence among the western languages than would usually be admitted today. Whatever his actual but unstated views on the subgrouping relations of the languages of western Indonesia, then, from the standpoint of lexical reconstruction Dempwolff *treated* them as representing primary branches of the AN family. As a result, many reconstructions in VLAW:III are supported only by reflexes in languages of western Indonesia which have either been in a borrowing relationship for centuries or are rather closely related, or both. As many as one-third of these now appear to be better treated as 1) products of previously undetected borrowing from Malay, 2) native forms, but products of relatively late innovation in western Indonesia, or 3) false etymologies (this latter is a very small class).

Test language and criterion language

Dempwolff classified each of the languages he compared as a 'Test Language' (TL) or a 'Criterion Language' (CL). Test languages retain a reconstructed sound unaffected by merger. Thus Tagalog, Toba Batak and Javanese are TLs for **m*, since in all three this sound has a single source. On the other hand, only Javanese is a TL for **ń*, since it alone retains the **n/ń* distinction, which has been lost through merger (as *n*) in the other two languages. A CL is any language that can be used in conjunction with another to distinguish reconstructed sounds through non-identical patterns of merger in the two witnesses. Tagalog non-final *l*, for example, has four possible sources, **ḍ*, **g′*, **l*, and **ḷ*, while Toba Batak non-final *r* has two possible sources, **ḷ* and **γ*. However, the correspondence Tagalog *l* : Toba Batak *r* points unambiguously to **ḷ* (the only area of

overlap), hence Tagalog and Toba Batak taken together are CLs for *l̥. The notion 'Test Language' is thus absolute in the sense that it can be determined once and for all for any given language whether it is a TL for a particular distinction. By contrast, the notion 'Criterion Language' is relative: language A may be a CL in relation to language B, but in relation to language C it may have no diagnostic value in disambiguating phonetically similar proto phonemes, or language A may be a CL with language C for a different distinction.

The representation of ambiguity

Where a cognate set is represented only by languages that are neither TLs nor CLs for the distinction in question Dempwolff enclosed the statistically more frequent proto phoneme within square brackets to indicate an ambiguous choice for two or more reconstructed sounds. Thus *[t]avaʻ 'laugh' lacks a reflex in Javanese, the sole TL for the *t/ṭ distinction, and is therefore ambiguous for *tavaʻ (the statistically preferred choice), or *ṭavaʻ. Because the voiced alveolar and retroflex stops *d and *ḍ were found to have roughly equal frequencies initially and intervocalically in unambiguous reconstructions, an ambiguity involving these segments was written out in full: *[dḍ]aɣah 'blood', *pə[dḍ]ih 'to sting, smart'.

The independent evidence requirement and the principle of symmetry

It is noteworthy that Dempwolff never explicitly discussed what is arguably the central principle in his approach to the comparative method, a principle that can be called the 'independent evidence requirement.' Basically, the independent evidence requirement prevents the reconstruction of a phonological distinction that is not supported by at least two witnesses. To illustrate, Dempwolff recognised that TAG –' generally corresponds to TB –ʻ, JAV –h, but sometimes corresponds instead to –ʻ in both languages. Since only Tagalog distinguishes these two correspondences Dempwolff preferred to treat the appearance of final glottal stop in Tagalog corresponding to Ø in Javanese as an 'unexplained exception' rather than propose a new proto segment (1934:76). Similarly, although Toba Batak and Javanese sometimes agree in showing *d* corresponding to Malay, Ngaju Dayak *d'* (as first noted by van der Tuuk), Dempwolff assigned these sets to doublets rather than posit a new, seemingly extra-systematic proto phoneme in a handful of lexical items (e.g. TB, JAV *dalan* < *dalan, but MAL, NgD *d'alan* < *d'alan 'path, road'). He deviated from this principle in a single instance: the reconstruction of *ṭ, based solely on a distinction in Javanese. The reason for this deviation was Dempwolff's implicit belief in the importance of symmetry in phonological systems. Because he regarded Tagalog and Javanese as jointly supporting the reconstruction of a *d/ḍ distinction, and because Javanese *ṭ* is the voiceless counterpart of *ḍ* he permitted himself in this one case to reconstruct a distinction that is not independently supported.

Canonical shape and the principle of symmetry

Although Dempwolff sometimes invoked morphology to explain phonological irregularitites, he reconstructed only unaffixed word-bases. The great majority of these (over 90%) are disyllabic, while a smaller group is trisyllabic. Four quadrisyllables appear in VLAW:III, all of which appear to be compounds. Finally, a score of monosyllables is posited for PAN. Without exception these are grammatical particles or onomatopoetic roots (*ṭukṭuk = 'knock, pound, beat', but *ṭuk = 'the sound 'thump!''). In many attested

languages these monosyllables may not be affixed, although the corresponding reduplications may.

Dempwolff regarded the unaffixed word-base as intrinsically nominal in meaning, and his glosses often reflect this view. Many disyllabic and trisyllabic word bases may be verbalised by means of affixes, in particular 'nasal substitution' (*nasaler Ersatz*) and 'nasal accretion' (*nasaler Zuwachs*), which were described in Chapter 6. Dempwolff's observations regarding nasal substitution have a direct bearing on his reconstruction of the PAN sound system. For the comparison *s : s : s,* for example, he posited an original palatal because the nasal substitute for *s* in languages that distinguish a palatal from an alveolar series is *ń,* as in Javanese *surat* 'writing' : *ńurat* 'to write'. Although Dempwolff may have distrusted twentieth-century trends in linguistics, then, his analyses of phonological relations took account of all factors generally regarded as relevant to phonemic analyses, including complementary distribution, free variation and alternation.

Perhaps the most curious feature of Dempwolff's system of transcription is his use of the 'spiritus asper' in writing both attested and reconstructed words. In attested words this symbol is said to represent smooth onset, transition and terminus. Initially it is used in some languages to write *h*, but in other positions or in initial position in other languages it signifies nothing at all, as with Malay '*ati*' ([hati]), but Javanese '*ati*' ([ati]) 'liver'. In reconstructed words the spiritus asper gave rise to Malay *h* between like vowels and to Taglaog *h* between any two vowels, but in other positions it served only to regulate canonical shape. This is an important facet of Dempwolff's reconstructional procedure, and it deserves some discussion.

As already seen, Dempwolff's inductive reconstruction of 'Proto Indonesian' is presented in stages, each concerned with base forms conforming to a definite canonical shape. In many languages, however, the assignment of a base form to a recognised canonical pattern was possible only through the addition of the spiritus asper before an initial vowel, between an internal vowel sequence or after a final vowel, or through the reconstruction of a homorganic 'fricative' between a high vowel and a following unlike vowel. Thus the reconstructions *'ambi', *batu', *bu'ah and *'ija' are said to exemplify the canonical shapes CVCCVC, CVCVC, CVCVC and CVCVC, even though they could more plausibly be assigned to the patterns VCCV, CVCV, CVVC and VV respectively. The use of these devices to impose an artificial regularity on the data is only one manifestation of a more general tendency for Dempwolff to approach his material in terms of highly developed a priori ideas of symmetry. Another symptom of this tendency is his classification of *r, l* and *j* ([j]) in attested languages, and of *v, *l, *ḷ, and *j in PAN as 'fricatives,' contrary to their phonetic realisations. Similarly, *a priori* notions of symmetry led Dempwolff to regard *d' and *g' as the voiced counterparts of *t', and *k' respectively, even though both phonetic and phonological evidence suggests that *d' (reflected as a voiced palatal affricate in Javanese and Malay) was the voiced counterpart of *k' (reflected as a voiceless palatal affricate in the same languages), and that *t' and *g' both lacked a counterpart that differed in voicing.

Dempwolff's Austronesian comparative dictionary

VLAW:III contains some 2,215 lexical reconstructions. These are alphabetised in accordance with the following order of symbols: *a, *b, *d, *[dḍ], *ḍ, *d', *ə, *g, *g', *', *h, *i, *j, *k, *k', *l, *ḷ, *m, *n, *ń, *ŋ, *p, *t, *ṭ, *t', *u and *v. Bases beginning with [dḍ] are listed separately from those beginning with *d or *ḍ, whereas ambiguous segments that are represented by a single symbol within square brackets are alphabetised

without regard to the brackets. The spiritus asper and the homorganic nasal of medial nasal clusters are ignored for purposes of alphabetisation, but the nasal of a reduplicated monosyllable is not (thus *kambiŋ 'goat' appears between *ka(m)baŋ and *kabut, but *kuŋkuŋ 'hold firmly' appears between *kunuʻ and *kupatʼ).

One of the most striking and potentially controversial features of Dempwolff's PAN lexicon is its abundance of doublets (*Nebenformen*)—forms that are so similar in shape and meaning that they appear to be variants of the same morpheme, as with *dalan and *dʼalan 'path, road', *tiḍuɣ and *tuḍuɣ 'to sleep', or *tuhaʻ and *tuvaʻ 'old'. Preliminary sampling suggests that around 20% of all Dempwolff reconstructions have doublet forms. In most cases both variants are independently supported, but occasionally a word is assigned to two cognate sets, as where Fijian *kumiʻ* is given as support both for *gumiʻ and for *kumitʼ 'beard', a point to which we will return later.

Dempwolff (1937:119ff) briefly notes submorphemic monosyllabic 'roots' in a number of languages, and refers to Brandstetter's work in this area for a more thorough treatment. Various other submorphemic sound-meaning correlations in his lexicon are also of interest, but are given little or no attention. Most word bases in VLAW:III that begin with a velar nasal, for example, codify meanings that are related in some way to the mouth, most bases that refer to grinding, rubbing, scraping, scratching and the like begin with a velar stop, bases for 'dull, blunt' very often end with *l despite the general rarity of this phoneme in final position, and *ṭ is reconstructed in only a small number of forms, but a large proportion of these are onomatopoetic.

Dempwolff's use of the 'philosophy of the As If'

One last point merits attention, as it has sometimes been misunderstood (as by Uhlenbeck 1955/1956:318). Dempwolff based his reconstruction of 'Proto Indonesian' phonology on the comparison of just three witnesses, yet he himself pointed out that published materials (in many cases very limited, and of poor quality) already existed in the 1930s for some 300 AN languages. Why did he limit himself to only 1% of the available material, and how could a reconstruction based upon such a narrow range of the relevant data possibly be sound?

To answer these questions it is important to recognise that Dempwolff's inductive reconstruction of PIN (volume 1) takes up 124 pages, and that his test of the adequacy of this reconstruction in explaining the phonological development of other languages (volume II) occupies another 194 pages. Together with the comparative dictionary the three-volume work comes to 482 closely written pages replete with data and carefully justified analyses. To have attempted a work of similar thoroughness with even twice the number of languages would have taxed both writer and reader perhaps beyond their limits. Dempwolff had, of course, examined the comparative phonology of many other languages in earlier publications, particularly in Dempwolff (1920) and (1924-25). The real challenge in his final attack on the data was to find the smallest number of languages that would permit him to reconstruct all distinctions necessary to account for the attested correspondences in AN languages generally. It is obvious that three languages chosen at random would have been inadequate for this purpose. The soundness of Dempwolff's reconstruction was thus crucially dependent upon his success in selecting the *right* three witnesses, and this in turn was largely dependent on the wider comparative experience that underlies the specific analyses in VLAW.

As noted by Dahl (1976), in order to simplify his presentation, Dempwolff made important use of Vaihinger's (1911) philosophy of the 'As If.' The core of Vaihinger's

book is an argument for the utility of a distinction between 'hypothesis' and 'fiction' in conducting science. Vaihinger uses the first term in roughly its conventional sense, and applies 'fiction' to any heuristically valuable idealisation. Dempwolff considered his reconstruction of PIN a fiction: his three witnesses represent a far larger number of languages that had previously been examined and which, if included in the investigation, would have had no further effect on its outcome. We now know that Dempwolff's reconstruction is inadequate in certain respects. However, his is not because Tagalog, Toba Batak and Javanese are inadequate as representatives of the languages with which he had previously worked, but rather because he excluded the Formosan languages.

Despite its usefulness in the reconstruction of phonology, Dempwolff's appeal to Vaihinger's concept of fictions had restrictive empirical consequences in lexical reconstruction. Many widely distributed cognate sets could not be used simply because they are not found in at least two of the 'Indonesian' languages consulted in VLAW. Dempwolff sometimes felt constrained by this limitation and on rare occasions introduced evidence from other languages, as with *ha(ŋ)g′av/'a(n)dav) 'day, sun', based on reflexes in Buginese of south Sulawesi, Ibanag of northern Luzon, and the Oceanic languages Sa'a, Tongan, Futunan and Samoan. Some writers have treated VLAW:III as a total or near-total inventory of the PAN lexicon, but this view is far from correct.

Summary of Dempwolff's contributions [85]

Dempwolff's publications from 1920-1938 marked a major advance on all previous comparative work in AN linguistics in several respects. In particular, Dempwolff

1. amplified and corrected Brandstetter's reconstruction of 'Original Indonesian' phonology (Table 8.8)

2. demonstrated that 'Proto Indonesian' was ancestral to all AN languages (outside Taiwan), since it explains the major phonological developments not only in the 'Indonesian' languages that Dempwolff considered, but also in all eastern languages

3. established the existence of a large 'Melanesisch' (Oceanic) subgroup

4. compiled a comparative dictionary with some 2,215 etyma and supporting data

5. provided explanations for many exceptions to the regular correspondences.

Various of the shortcomings of VLAW will become clear as more recent research is considered, but all-in-all there seems to be no reason to question the generally accepted view that with its appearance comparative AN linguistics had at last come of age.

---

[85] Unfortunately, no good English translation of Dempwolff's major work exists to date. Cecilio Lopez's 'Studies on Dempwolff's *Vergleichende Lautlehre des austronesischen Wortschatzes*' (1939) is (despite its somewhat misleading title) a highly condensed translation of VLAW. Sometime in the 1950s this was reissued in undated mimeographed form by the Summer Institute of Linguistics, Philippines Branch, adopting 'Dyen's modifications of Dempwolff's symbols' and adding a translation of volume III which did not appear in Lopez's original work. A full (uncredited) translation of VLAW was issued in mimeographed form by the Ateneo de Manila University in 1971, but this is so riddled with errors that the beginning student is best advised to avoid it.

Table 8.8 Correspondence of Brandstetter's (1916) 'Original Indonesian'
with Dempwolff's (1934-1938) 'Proto Austronesian'

Brandstetter	a	i	u	e	o	ĕ	y	w	r₁	r₁
Dempwolff	a	i	u	--	--	ə	j	v	g′	ḷ

Brandstetter	r₁	r₂	l	q	k	g	n[86]	c	j	ñ
Dempwolff	d/ḍ	ɣ	l	(?)[87]	k	g	ŋ	k′	d′	ń

Brandstetter	t	t	d	d	n	p	b	m	s	h
Dempwolff	t	ṭ	d	ḍ	n	p	b	m	t′	h

Brandstetter	Ø	Ø
Dempwolff	h[88]	'

8.1.6 Revisions to Dempwolff: Dyen

   Dempwolff died less than a year before the outbreak of World War II in Europe, and apart from Arthur Capell's *The linguistic position of South-Eastern Papua* (1943), which used his reconstructions to shed light on the phonological history of various languages of southeastern New Guinea and adjacent islands, no further work in comparative AN phonology was done during the war years. When an interest in systematic comparison and reconstruction emerged again at the end of the war its primary locus was no longer in Europe, but rather in North America.

   Unlike Dempwolff the American linguist Isidore Dyen (1913-2008) began his career as a comparativist, having written a dissertation on Sanskrit grammar at the University of Pennsylvania in 1939 (Blust 2009d). As a result of his training in the America of the 1930s Dyen was thoroughly steeped in the phonemic principle, particularly as espoused by Leonard Bloomfield. In accordance with his Bloomfieldian orientation the philosophical underpinnings of his work can perhaps best be described as a strong form of empiricism. It will be convenient to divide Dyen's AN work into three phases, the first spanning the period 1946-1953, the second centering on his monograph, *The Proto Malayo-Polynesian laryngeals* (1953b), and the third covering the period after 1953.

The early papers

   Dyen began his studies of AN comparative phonology by focusing on exceptions to Dempwolff's rules of change. The analyses that he proposed sometimes affected only the statement of diachronic rules for attested languages, but more often than not they also

---

[86]   Brandstetter symbolised the velar nasal by *n* with a dot on top.

[87]   Brandstetter's discussion of *q is vague. He asserts (1916:249) that only one case of *q 'can with some probability be ascribed to Original IN,' and refers the reader to a later section for documentation. This section, however, states only that 'Original IN' words customarily written with an initial vowel probably began with *q: *atay, or more precisely *qatay 'liver', *atĕp, or more precisely *qatĕp 'roof', etc.

[88]   Brandstetter reconstructed *h only in final position, positing an initial vowel or medial vowel sequence in many cases where Dempwolff proposed a laryngeal consonant. This sometimes forced him to treat the appearance of *h* in attested forms as unexplained (e.g. OIN *añud, but Malay *hañut* 'drift on a current').

required alterations in the shape of reconstructed forms. In the first of these (Dyen 1947a), he suggested that Dempwolff's *ḍuva' be revised to *Dewha 'two'. At the outset Dyen introduced a typographically more convenient orthography for AN reconstructions that included the following symbol changes, all other Dempwolff symbols being retained:

Table 8.9 Modifications to Dempwolff's orthography in Dyen (1947a)

Dempwolff	k′	ḍ	ə	-'-	'-/-'	-h(-)	ń	ḷ	ɣ	t′
Dyen	c	D	e	H	Ø	ʔ	ñ	r	R	s

Dempwolff	ṭ	v	j	D′
Dyen	T	w	y	Z

He added (1947a:50) that 'No difference of phonological interpretation is implied except with regard to the substitutions for Dempwolff's h and '.' This orthography was further modified in subsequent publications up to 1951 as follows:

Table 8.10 Other modifications of Dempwolff's orthography by Dyen up to 1951

Dempwolff	**Dyen**
-'	h (1947b)
g′	j (1949)
-'	Ø (1949)
'-	ʔ (1949)
'	h (1951)
h	q (1951)

In addition, Dyen altered the orthography that Dempwolff used for his sources by dropping the spiritus asper, and by substituting *ò* ([ɔ])for *a* in the transcription of Javanese, and *q* for the spiritus lenis in all languages.

Dyen's revision of *ḍuva' to *Dewha did not conflict with Dempwolff's interpretation of PAN phonology, but it did conflict with his interpretation of PAN canonical shape in two respects. First, *Dewha contains a unique sequence *-ew-, since Dempwolff (1938) contains no instances of *əv in any position. Second, a heterorganic consonant cluster was proposed for the first time in a non-reduplicated base.

Dyen (1947b) and Dyen (1949) test the adequacy of Dempwolff's reconstructed sound system and its exemplification in particular morphemes. The first paper proposes a different set of criteria for *d and *D than those used by Dempwolff, while the latter demonstrates that the nine vowel system of Chuukese can be explained as a product of regular changes from the four vowel system reconstructed by Dempwolff. Although the proposals in Dyen (1947b) bring about a reassignment of correspondence classes so that some words which Dempwolff reconstructed with *d are reconstructed instead with *D and others with *(dD), neither this publication nor Dyen (1949) is directly concerned with structural changes in Dempwolff's phonological system. Nonetheless, in a footnote to the latter paper Dyen (1949:fn. 5) introduced a proposal to split Dempwolff's *-aj into two diphthongs *-ay and *-ey. This proposal will be discussed at greater length below.

The first publication that proposed structural changes in Dempwolff's phonological reconstruction is Dyen (1951). This paper investigates a correspondence noted by van der Tuuk (1865), and called by Blagden (1902) 'van der Tuuk's Third Law.' Briefly,

Dempwolff reconstructed *d′ for the correspondence Tagalog *d-*, *-r-/-nd-* : Toba Batak *d′-*, *-(n)d′-* : Javanese *d′-*, *-(n)d′-*, as in Tagalog *dait*, Toba Batak, Javanese *d′ait* (PAN *d′ahit) 'sew', Tagalog *puri'*, Toba Batak, Javanese *pud′i'* (PAN *pud′i') 'praise', or Tagalog *'indak*, Toba Batak *'ind′ak*, Javanese *'ida'* (PAN *'i(ń)d′ak) 'step, dance'. As observed by van der Tuuk, however, in a few etymologies Toba Batak and Javanese have *d*, and Malagasy has *r* (instead of the expected *d′* : *d′*: *z*) corresponding with Malay, Balinese *d′*. In such cases the first three languages point to *d, and the latter two to *d′. To resolve this apparent contradiction Dempwolff posited five *d/d′ doublets, assigning the Toba Batak, Javanese, and Malagasy cognates (together with some ambiguous reflexes) to the former, and the other cognates to the latter variant, as with *dalan 'road, path': TAG *daan*, TB, JAV *dalan*, MLG *lalană* (assimilation) 'road, path', *andalană* 'arrangement in rows', but *d′alan 'path, road': MAL, Ngaju Dayak *d′alan*, Fijian *sala'*, Sa'a *tala'*, Tongan *hala'*, Futunan, Samoan *'ala'* 'road, path', *'alan-a'i'* 'go with someone', or *'uḍan 'rain': TAG *'ulan*, TB *'udan*, JAV *'udan*, MLG *'urană*, Futunan *'u'a'* 'rain', but *hud′an 'rain': MAL *'ud′an*, Ngaju Dayak *'ud′an*, Fijian *'uza '*, Sa'a *'ute'*, TON *'uha'*, SAM *'ua'* 'rain', FIJ *m-usa'* 'watery'.

Dyen (1951) proposed that each of the five *d/d′ doublets be grouped under a single etymon which contains a previously unrecognised phoneme *Z (distinct from *z). Three other etyma for which Dempwolff did not reconstruct doublets were treated in the same way, as was a doublet pair that differs in two consonants (Dempwolff: *tu(n)duh 'point to' : *tund′uk 'show, point out'). In all, *Z was posited in nine forms: 1) *peZem 'close the eyes', 2) *eZem 'squeeze out', 3) *quZan 'rain', 4) *Zalan 'road, path', 5) *Za(hØ)uq 'far' (the symbols in parentheses signify an indeterminate choice of *h or zero), 6) *Zilat 'lick', 7) *ZeRami(hØ) 'rice stubble', 8) *ZuRuq 'sap, juice, gravy', 9) *tuZuq 'point out, indicate, show'. All other instances of *d′ were replaced by *z.

In his next paper Dyen (1953a) again examined a group of etyma in which Dempwolff had reconstructed a single phoneme to cover several partly overlapping correspondence classes. Dempwolff's *R (= *ɣ) regularly appears as Toba Batak, Malay *r*, Tagalog *g*, Javanese zero, Malagasy *z*, and—with a complication to be discussed below—Ngaju Dayak *h*. This invariant set of reflexes is essentially van der Tuuk's First Law. Apart from the non-problematic reflexes in Malay and Tagalog, however, Dyen drew attention to two reflexes of *R in Javanese and Ngaju Dayak, and three reflexes of *R in Malagasy, with the following distribution in cognate forms (Ø = zero):

Table 8.11 Irregular developments of Dempwolff's *R

	JAV	NgD	MLG
1)	Ø	h	Ø
2)	Ø	h	Z
3)	r	h	Z
4)	r	r	R

Dempwolff handled these irregularities in several ways. The correspondence Javanese zero, Ngaju Dayak *r*, Malagasy *z* was taken to be regular, while Javanese *r* corresponding to Ngaju Dayak *h* or *r* was attributed to borrowing, chiefly from Malay. Ngaju Dayak *h*, together with certain other irregularities, was attributed to an 'old speech stratum' which Dempwolff, following his colleague Walther Aichele, had attempted to derive from a hypothetical ancient Bornean literary language, and Malagasy zero and *r* were left as 'unexplained exceptions' to the rules of change.

To deal with this problem Dyen (1953a:360) suggested that correspondences 1-4) be labeled by 'tentative or 'problematic' reconstructions,' and he accordingly assigned them to $*R_1$, $*R_2$, $*R_3$ and $*R_4$. The number of instances of each of these tentative reconstructions by position within the morpheme is as follows:

Table 8.12 Frequency of Dyen's $*R_1$, $*R_2$, $*R_3$ and $*R_4$ by position

$*R_1$-	:	1	$*R_3$-	:	1
$*$-R_1-	:	8	$*$-R_3-	:	4
$*$-R_1	:	0	$*$-R_3	:	1
$*R_2$-	:	2	$*R_4$-	:	0
$*$-R_2-	:	4	$*$-R_4-	:	3
$*$-R_1	:	0	$*$-R_4	:	?

Many other instances of $*R$ remained ambiguous. Two protoforms with $*R_2$ became Ngaju Dayak *r* rather than *h*, and Dyen therefore raised the prospect of a fifth distinction $*R_5$. As noted already, Dempwolff took Ngaju Dayak *r* to be the directly inherited reflex of $*R$, attributing instances of $*R > h$ to borrowing. Dyen (1953a) reversed this interpretation and avoided the reconstruction of $*R_5$ by assigning Ngaju Dayak morphemes with $r < *R$ to a superstratum derived from Banjarese Malay. At the same time he rejected a suggestion by Dahl (1951) that Malagasy $r < *R$ might be due to early borrowing prior to the Malagasy departure from Indonesia.

Two other reconstructional changes introduced in this paper are noteworthy, although they are tendered as proposals incidental to the main argument. In the first of these Dyen departs from Dempwolff's view of PAN canonical shape by admitting heterorganic consonant clusters with $*q$ in reconstructed forms, as in $*beR_2qaN$ 'molar tooth', $*beR_2qat$ 'heavy', and $*peR_3qes$ 'squeeze out' (362): 'When the Tagalic languages (Tagalog, Bisayan, Bikol) suggest the presence of $*q$ next to a consonant, I follow the practice of inserting it in the position indicated by Tagalog.' In the second (363, fn. 18) he posits $*$-ew in a single form, $*buR_3ew$ 'chase, hunt' thereby filling out the system of diphthongs implied by his earlier (1949) proposal of $*$-ey. Like its precursor, this important structural alternation is presented in a footnote.

The Proto Malayo-Polynesian laryngeals

Easily the most important and widely accepted of Dyen's modifications of Dempwolff concerns the so-called 'laryngeals.' Dempwolff had used this term to distinguish his spiritus asper and $*h$ from both consonants and vowels. From his first publication on comparative AN phonology, Dyen suggested changes in Dempwolff's analysis of the laryngeal correspondences that involved not simply a difference of orthography, but a genuine difference of phonological interpretation. Until 1953, however, no attempt was made to justify these views. Moreover, a comparison of the brief statements in Dyen (1947a, 1947b, 1949 and 1951), all of which appear in footnotes, shows clearly that Dyen's interpretation of the laryngeal correspondences passed through several stages before reaching its final form.[89]

---

[89] Cp. examples such as *abu , *abuh, *abu, *habuh, *abuh 'ashes', *baʔeRu, *baʔeRuh, *baʔeRu, *baqeRuh, *baqeRu(h) 'new', *bayi, *baʔi,, *beyi, *beyih, *bei 'female; woman', or *biyak, *biyak, *biyak, *biyak, *biqak 'split' in Dyen (1947a, 1947b, 1949, 1951 and 1953b) respectively.

It is noteworthy in these publications that Dyen's interpretation of the laryngeals (including Dempwolff's transitional *v and *j) remains constant in certain positions, but shows considerable vacillation in others. Thus, from 1947 to 1951 Dempwolff's *-'-, *v and *j were almost always transcribed as *h, *w and *y, and non-initial *h was replaced by *ʔ until 1951, when by an orthographic convention (not a structural reinterpretation) it was changed to *q. The interpretation of Dempwolff's initial and final spiritus asper and *h-, on the other hand, shows considerable variation. This is especially true of *-', which is dropped in Dyen (1947a), reinstated in Dyen (1947b), dropped again in Dyen (1949), and reinstated a second time as part of a more general change in Dyen (1951). Apart from the splittling of Dempwolff's initial spiritus asper into zero and ʔ-, which will be discussed below, all other variations in Dyen's treatment of the laryngeals between 1947 and 1951 are due to this latter, more general change.

In earlier publications Dyen's analysis of the laryngeal correspondences shows clear differences of phonological interpretation as compared with that of Dempwolff. Thus, in Dyen (1947a) Dempwolff's spiritus asper was rewritten as *h medially but as zero elsewhere, while his *h was maintained initially but written as *ʔ in other positions. In striking contrast to these structural reinterpretations Dyen (1951:535) returns to a simple retranscription of Dempwolff's spiritus asper as *h, and to his *h as *q: 'in a reconstruction quoted from Dempwolff, *h* represents his ' and *q* his *h* in any position. A reconstruction not specifically attributed to Dempwolff and involving *q or *h rests on theories stated in a forthcoming treatment of the Proto Malayo-Polynesian laryngeals.' For convenience of reference these relations are sketched below (changes from an earlier position are marked by parentheses):

Table 8.13 Symbols replacing Dempwolff's 'laryngeals' in Dyen (1947-1951)

	Dempwolff 1934-1938	Dyen 1947a	Dyen 1947b	Dyen 1949	Dyen 1951
1.	'-	Ø	Ø	Ø, (ʔ)	h
2.	-'-	H	h	h	h
3.	-'	Ø	(h)	Ø	(h)
4.	h-	H	h	h	(q)
5.	-h-	ʔ	ʔ	ʔ	q
6.	-h	ʔ	ʔ	ʔ	q
7.	-v-	W	w	w	w
8.	-j-	Y	y	y	y

Although Dyen's earlier statements on the laryngeals are entirely superseded in Dyen (1953b) this brief survey of the development of the laryngeal theory serves two purposes. First, it reveals transcriptional inconsistencies and interpretive vacillations in Dyen's earlier work that may be a source of confusion to the attentive reader. Second, it pinpoints the source of certain misunderstandings of Dyen's laryngeal theory in the wider literature by writers who adopt his *orthography*, but maintain Dempwolff's interpretation of the laryngeals. This problem is especially conspicuous in comparative treatments of the Chamic languages by members of the Summer Institute of Linguistics, and can be traced to a single source: a translation of Dempwolff (1938) by Christine Laurens and Trudy Pauwels which appears in Cecilio Lopez's *Studies on Dempwolff's Vergleichende Lautlehre des austronesischen Wortschatzes*, as reissued in undated mimeographed form

by the Summer Institute of Linguistics, Philippines Branch.[90] Pittman (1959:60) clearly indicates the source he uses ('a translation of Dempwolff's word list of Proto Malayo-Polynesian done by Miss Christine Laurens of Indonesia and Trudy Pauwels of Canada'), noting that 'In the Dempwolff material we have used Dyen's orthography.' In explanation he refers to Dyen (1953b). Other remarks in Pittman's paper show that he did not mean to imply that Dyen's treatment of the laryngeals differs from Dempwolff's only in orthography, despite his use of the Laurens and Pauwels material. Similar, but more perfunctory qualifications are found in Blood (1962) and Lee (1966), where the distinction between orthographic difference and phonological reinterpretation is often blurred. The system of representing the laryngeals in Thomas (1963) is a mixture of Dyen's laryngeal theory (*apuy 'fire', *mata 'eye') and that of Dempwolff written in Dyen's orthography (*hasuh 'dog', *kayuh 'wood; tree').

Before considering Dyen's reconstruction of the laryngeals it will be useful to briefly review Dempwolff's analysis. If for the moment we ignore his 'homorganic fricatives,' which Dyen included in the discussion of the 'laryngeals,' it will be seen that Dempwolff recognised six laryngeal correspondences (Table 8.14). Dempwolff regarded Tagalog, Toba Batak and Javanese as sufficient to reconstruct all PAN phonological distinctions. However, because certain of the laryngeal reconstructions receive additional support from Malay and Tongan these languages are added to the comparison (C = contraction of vowel sequence; NC = no contraction; Tongan lacks final consonants):

Table 8.14 Laryngeal correspondences as recognised by Dempwolff (1934-1938)

	ʻ-	h-	-ʻ-	-h-	-ʻ	-h
TAG	Ø	h	h	ʔ/Ø	Ø(ʔ)	ʔ
TB	Ø	Ø	Ø	Ø	Ø	Ø
JAV	Ø	Ø	C	NC	Ø	h
MAL	Ø	h	Ø	h/Ø	Ø	h
TON	Ø	ʔ	Ø	ʔ	--	--

In initial position Dempwolff recognised only two laryngeal contrasts (ʻ-, h-), whereas Dyen extended the possibilities to three (*q, *h, Ø). He did this through assigning some proposed reflexes of the spiritus asper, together with some proposed reflexes of *h to a new phoneme *q. Specifically, Dempwolff's *ʻ- was generally reinterpreted as zero in reconstructed forms that were not cross-referenced to a doublet with *h-, but was reinterpreted as *q where such a cross-reference is found: * ʻakuʻ > *aku(h) 'I', *ʻanak > *anak 'child', *ʻuɣat > *uRat 'vein, tendon', but *ʻatəp/hatəp > *qatep 'roof; thatch', *ʻataj/hataj > *qatey 'liver', *ʻuluʻ/huluʻ > *qulu(h) 'head', etc. Reconstructed forms with initial spiritus asper that were cross-referenced to a doublet thus represented in effect a third correspondence which Dempwolff had concealed through the artifice of positing

---

[90]   The provisional nature of this source is clearly indicated: "The material must not, in any sense, be regarded as 'published' or available for purchase or review in any periodical." This qualification would raise questions about the propriety of citing this work at all except that references to it appear in several later publications by members of the Summer Institute of Linguistics, Philippines Branch. In effect, the citation of 'not to be cited' material by earlier writers who have thereby introduced this work into the public domain has removed any taint of inappropriateness in citing it now, since there is a need to clarify why the shapes of reconstructions that appear in these publications differ from those in other sources.

doublets where they are not truly justified. The reflexes of Dyen's *q, *h and zero in initial position in the same languages are shown in Table 8.15:

Table 8.15 Initial laryngeal correspondences as recognised by Dyen (1953b)

PMP	q	h	Ø
TAG	Ø	h	Ø
TB	Ø	Ø	Ø
JAV	Ø	Ø	Ø
MAL	h~Ø	h~Ø	Ø
TON	ʔ	Ø	Ø

It can now be seen that Dempwolff's requirements for the reconstruction of *h are contradictory, since Tagalog *h* points unambiguously to *h, while Tongan ʔ points unambiguously to *q. Malay *h~Ø* can reflect either source, agreeing in some morphemes with Tongan ʔ (correspondence 1), and in others with Tagalog *h* (correspondence 2). Its chief comparative value is in distinguishing *q from zero word-initially and in certain intervocalic environments in comparisons that include a Tagalog, but not a Tongan cognate. In effect, then, Dempwolff conflated correspondences (1) and (2) in Table 8.13. Wherever both Tagalog and Tongan cognates were available this led to conflict, since Tongan ʔ- was assigned to *h-, and this implied a Tagalog *h-* that does not correspond with Tongan ʔ-. To avoid this contradiction Dempwolff reconstructed *ʻ-:h- doublets, assigning the Tagalog reflex to the former, and the Tongan reflex to the latter variant.

Dempwolff's source for Malay (Klinkert 1918) usually does not cross-reference variants in *h~Ø*, but records one or the other pronunciation. As a result Dempwolff assigned the Malay reflex now to a variant with *ʻ-, now to a variant with *h-: 1) *ʻatəp : Tagalog *ʼatip*, Javanese *ʻatĕp*, Malay *ʼatap*, Sa'a *s-aoʻ* 'roof; thatch', 2) *hatəp : Ngaju Dayak *hatap* 'palm fronds used as construction material', Tongan *ʼatoʻ*, Futunan, Samoan *ʼatoʻ* 'cover a roof', 3) *ʻataj : Tagalog *ʼatai ʻ*, Toba Batak *ʻate-ateʻ*, Javanese *ʻatiʻ*, Ngaju Dayak *ʻateiʻ*, Malagasy *ʻatiʻ*, Fijian *yate-*, Sa'a *s-aeʻ* 'liver', 4) *hataj : Malay *ʻatiʻ*, Tongan *ʼateʻ*, Futunan, Samoan *ʼateʻ* 'liver'. In all, Dempwolff reconstructed eight instances of *ʻ-:h- doublets which are otherwise identical, as well as several that differ in two or more segments, one of which is a laryngeal. Given his general need to posit doublets in any case, it should not be surprising that Dempwolff fell back on the reconstruction of doublets to reconcile certain problematic sound correspondences. As with the variants that Dyen united under *Z, however, these doublets differ in an important respect from those that are seen today as justified: they exhibit a recurrent pattern of variation: more than 8.5% of Dempwolff's reconstructions with *h- have a doublet with *ʻ-. Dyen correctly saw the recurrent character of Dempwolff's *ʻ-:h- doublets as grounds for suspicion, and he proposed that all such divided cognate sets be united under single etyma with *q (hence *qatep 'thatch; roof', *qatey 'liver', etc.).

Dyen's (1953b) proposal to unite Dempwolff's *ʻ-:h- doublets under a single phoneme was not entirely new. As early as 1949 he had suggested that Dempwolff's spiritus asper be written as *ʔ if a Tongan cognate was available and showed glottal stop, but should otherwise be written as zero. While this was an important first step toward distinguishing two correspondences that Dempwolff had confused, it failed to use all of the evidence available in the languages compared. Until 1953 Dyen relied exclusively on Tongan to distinguish correspondences (1) and (3) in Table 8.15, reconstructing e.g. *ulej 'maggot' in 1949 because no Tongan cognate was available, while ignoring the evidence of Malay

hulat ~ ulat. In effect, to call on Dempwolff's still useful terminology, Dyen was willing to trust only a 'test language' in reconstructing *q-. The chief advance of PMPL with respect to the correspondences for which Dempwolff reconstructed the spiritus asper in initial position, then, was its incorporation of Malay as a second witness (a 'criterion language' together with Tagalog) for the *q:Ø distinction. By thus widening his view of admissible evidence Dyen (1953b) was able to distinguish initial vowels from *q- even in the absence of a Tongan cognate. But he continued to give Tongan priority, as in reconstructing *abuh 'ash'—one of the rare outright errors in PMPL—from Tongan *efu*, Tagalog *abó*, Javanese *awu*, etc., despite Malay *(h)abu* and corroboratory evidence from many other languages which support *qabu (Tongan *efu* apparently reflects PMP *dapuR). The reflexes of Dyen's *q, *h and Ø in non-initial position are shown in Table 8.16 (C = contraction of a vowel sequence, NC = non-contraction):

Table 8.16 Non-initial laryngeal correspondences as recognised by Dyen (1953b)

PMP	-q-	-h-	-Ø-	-q	-h	-Ø
TAG	-ʔ,Ø-	-h-	-Ø-	-ʔ-	-Ø,h-	-Ø
TB	-Ø-	-Ø-	-Ø-	-Ø-	-Ø-	-Ø
JAV	h,Ø+NC	Ø+C	Ø+C	-h-	-Ø-	-Ø
MAL	-h,Ø-	-h,Ø-	-Ø-	-h-	-Ø-	-Ø
TON	-ʔ-	-Ø-	-Ø-	-Ø-	-Ø-	-Ø

In standard Tagalog intervocalic *q became ʔ between identical vowels or unlike vowels the first of which was low; otherwise it disappeared. In Javanese intervocalic *q became *h* between like vowels under certain conditions, but otherwise disappeared. The resulting vowel sequence, however, did not contract. In Malay intervocalic *q became *h* or Ø under conditions closely paralleling those for the reflexes ʔ and Ø in Tagalog.

In Javanese *h disappeared and the abutting vowels contracted; like vowels yielded the corresponding single vowel, sequences of *a + *i or of *i + *a yielded ε, and sequences of *a + *u or of *u + *a yielded ɔ. As can be seen, Toba Batak provides no information on the PMP laryngeals, and is included here only because it is one of the three languages used by Dempwolff in his 'inductive' reconstruction of PAN phonology.

In two situations Dyen found it difficult to distinguish a laryngeal consonant from zero on the basis of the evidence considered by Dempwolff. These situations involved 1) the reconstruction of *q between unlike vowels the first of which is high, and 2) final *h. Among Dempwolff's languages only Tongan consistently distinguishes *q from Ø between unlike vowels the first of which is high, but Tongan cognates are often lacking for Dempwolff reconstructions with *-uva-, *-uvə-, *-uvi-, *-ija-, *-ijə- or *-iju-. To correct this deficiency Dyen added material from Hiligaynon Bisayan, Bikol, and the Pagsanghan dialect of Tagalog, all of which retain the *q : Ø distinction before a last-syllable vowel, as in *RuqaN (*N = velar nasal) > Tagalog (Standard) *guwáŋ*, Tagalog (Pagsanghan), Hiligaynon *guʔáŋ*, Bikol *gúʔaŋ*, Malay *ruaŋ* 'hollow space', next to *buaq > Tagalog (Pagsanghan) *buwáʔ* 'additional growth in fruit', Hiligaynon *búaʔ* 'soft pulp of sprouting coconut', Malay *buah*, Javanese *wɔh*, Tongan *fua* 'fruit'.

The reconstruction of *-h posed special problems, as this segment is lost in all of the languages compared by Dyen. The appearance of 'thematic' consonants before a suffix, however, suggested an original contrast between *-h and Ø, as in Tagalog *tubó* 'sugarcane' : *tubuh-án* 'sugarcane plantation' (< *tebuh) vs. *má:ta* 'eye' : *matá:ʔ -an* 'be looked for' (< *mata). In a few cases an assumed metathesis was taken as supplementary evidence for

*-h, as with (earlier) Tagalog *tahóʔ* < *taquh (Dempwolff's *tahuʻ) 'know'. In many other cases phonological evidence was lacking or questionable, and it was necessary to reconstruct a base as ambiguous for final *h or Ø: *aku(h) 'I', *tuqa(h) 'mature; old', *buqaya(h) 'crocodile', etc. It is now known that the thematic consonants of Tagalog do not provide a reliable basis to infer *-h, since *–h* between unlike vowels is historically secondary (e.g. *abó* 'ash' : *abu-h-án* 'ash pit' < *qabu). Dyen's confidence in positing *-h by internal reconstruction may have depended in part on a preliminary consideration of Formosan evidence, as he notes (1953b:fn. 79b) that the final consonant of Tokuvul Paiwan *tivos* 'sugarcane' supports the final consonant of *tebuh. The fundamental implications of the Formosan evidence for the reconstruction of the PAN laryngeals, however, is only hinted at in PMPL, and did not receive serious consideration until over a decade after its publication.

Dyen's reanalysis of the laryngeal correspondences had consequences on three levels: 1) canonical shape, 2) phoneme inventory, and 3) the phonemic composition of particular morphemes. On the most general level PAN canonical shape was granted a latitude that accords better with the facts in the modern languages than was the case with Dempwolff's reconstruction. Words were no longer forced into a procrustean canonical template, but were allowed to begin or end with a vowel, or to contain a medial vowel sequence if the available evidence ruled out the possibility of reconstructing a laryngeal consonant. In the most extreme cases the formal differences between the two approaches are striking, as with Dempwolff's *ʻijaʻ (CVCVC) vs. Dyen's *ia (VV) 'he/she'. This liberalisation of canonical shape had, of course, begun already in Dyen (1947a), where vowels were reconstructed adjacent to word boundary. As seen above, however, Dyen's views on the laryngeals fluctuated considerably during the period 1947-1951. Moreover, despite his varying flexibility with regard to canonical shape, Dyen steadfastly maintained Dempwolff's 'homorganic fricatives' (as glides *-y-, *-w-) until PMPL, reconstructing e.g. *iya 'he/she' in Dyen (1949). Finally, the reconstruction of postconsonantal *q, proposed in Dyen (1953a) apparently is not even mentioned in the longer work.

In initial and final position Dyen introduced three contrasts (*q, *h and Ø) where Dempwolff had allowed only two (*ʻ and *h). Intervocalically Dempwolff's *ʻ and *h were divided between the same three phonemes, and the 'homorganic fricatives' *j and *v were each split between *q and Ø. A few sequences of like vowels, or of unlike vowels of which one is schwa, were introduced by increasing the number of syllables in a Dempwolff reconstruction: *bə(n)tit- > *betiis 'calf of the leg', *təluk > *te + luuk 'bay', *ʻaliməs > *alimees 'invisible', *ɣabiʻ > *Rabii(h) 'evening'. On the most concrete level—that of the phonemic composition of morphemes—Dyen's revision of Dempwolff was accomplished in two ways: 1) by splitting a Dempwolff phoneme into two, and 2) by reassigning a correspondence from one proto phoneme to another.

In conclusion, Dyen's laryngeal theory was a major achievement of lasting value. Where Dempwolff had confused distinct correspondences and avoided the consequences of his error through the wholesale reconstruction of phonologically recurrent proto doublets ('Nebenformen'), Dyen brought clarity and order. Ironically, however, none of the correspondences that Dempwolff and Dyen considered in this connection derive from phonemes that would now be considered 'laryngeals.' This became clear once attention was belatedly turned to the aboriginal languages of Taiwan.

Formosan evidence

In 1935 the Japanese linguists Naoyoshi Ogawa and Erin Asai published a major survey of the Formosan aborigines which includes comparative vocabularies and some preliminary remarks on phonological correspondences. Ogawa and Asai noted that most Formosan languages have two distinct sets of correspondences for sounds that appear outside Taiwan to reflect *t and *n. They accordingly proposed distinctions $*t_1$ vs. $*t_2$, and $*n_1$ vs. $*n_2$.[91] Because their work was published in Japanese, and only appeared after volume 1 of Dempwolff's VLAW had already gone to press, these proposals had no effect on Dempwolff's conclusions. Dempwolff died in 1938, and the reconstruction of PAN phonology that he left to posterity became the starting point for all further work. Although Brandstetter had considered Formosan evidence, the material available to him was limited, and he made little use of it in the reconstruction of 'Original Indonesian.' As a result, the importance of the Formosan languages to the reconstruction of PAN would go unappreciated for another quarter century.

Footnoted commentary in PMPL shows that by 1953 Dyen was aware of the potential value of Formosan evidence in understanding the history of the 'laryngeals,' but his first use of Formosan language material for purposes of reconstruction did not appear for another twelve years. Dyen (1965c) is remembered for two very different kinds of reasons. On the one hand this paper showed forcibly that it is impossible to arrive at an adequate reconstruction of PAN phonology without taking Formosan evidence into account. Not only did it make Ogawa and Asai's $*t_1$: $*t_2$ and $*n_1$: $*n_2$ distinctions known to the English-speaking world (as *t : *C and *n : *N), but it also showed that *q (= glottal stop?) and *h as reconstructed in 1953 should be reconstructed instead as *q (probably a uvular stop) and *S (a sibilant distinct from *s). This can be considered the positive side of the ledger. Indeed, if Formosan evidence had been available to Dempwolff most of the confusion surrounding the reconstruction of the 'laryngeals' could have been avoided altogether (along with the term).

On the other hand, this paper exacerbated an already existing approach to reconstruction that abandoned phonetic realism in favor of labeling irregularities through reconstructed distinctions. As its title suggests, 'Formosan evidence for some new Proto Austronesian phonemes' was primarily concerned with demonstrating that the segment inventory reconstructed by Dempwolff was insufficient to account for Formosan data. With regard to *C, *N and *S this was not problematic, but the proposals in this paper went much further. Dyen had already shown in positing $*R_1$-$*R_4$ that he was willing to abandon phonetic realism in order to 'account' for minimally distinct sound correspondences. This tendency was carried further in Dyen (1962), and reached its culmination in Dyen (1965c), which completely abandoned phonetic realism (Table 8.17):

---

[91]   Ogawa and Asai also reconstructed $*d_1$ and $*d_2$, but these appear to correspond to the distinction of *d′ and *d/ḍ already recognised by Dempwolff.

Table 8.17 New phonological distinctions introduced in Dyen (1965c)

Previous reconstruction	New reconstructions
*t	*t, *C
*n	*n, *N
*q	*q, $*Q_1$, $*Q_2$
*h	$*S_1$ - $*S_6$
*w	*w, $*w_1$, $*w_2$, *W
*Ø	$*x_1$, $*x_2$, *X, *H, *ʔ

As can be seen, five earlier proto phonemes (plus zero) were transformed into no fewer than 22 phonemic distinctions in Dyen (1965c). Some of these are clearly justified, as with the *t/C and *n/N distinctions first proposed by Ogawa and Asai. Most others, however, are unrealistic, and in many cases are based on irregularities in only one or two words in a single language. Unlike the proposals for $*R_1$ - $*R_4$, which were qualified as 'problematic' reconstructions 'leaving a final solution to the future' (Dyen 1953a:366), the new phonemic distinctions based on Formosan evidence were presented as a *fait accompli*: 'a number of distinctions have appeared that are assignable to proto Austronesian and are not known to appear outside of Formosa' (Dyen 1965c:303). In this paper more than any other, the philosophical differences between Dempwolff and Dyen were laid bare. Whereas Dempwolff was a phonetic realist (sometimes going to extremes with his notions of imposed symmetry), Dyen is a strict constructionist whose approach implies that every observed irregularity can be eliminated by reconstructing more phonemic distinctions. In other words, Dempwolff tried to constrain his theory of reconstruction through phonetic naturalness, while Dyen was concerned with making a theory of reconstruction powerful enough to 'explain' everything with no controls on naturalness, hence draining the notion 'explanation' of any real meaning.

Whatever the shortcomings of his approach to reconstruction, Dyen's orthography has proved its practicality over the years, and it will be used in discussing all PAN reconstructions from this point on except where material is quoted directly from an author using some other system.

8.2   PAN phonology: a critical assessment

Not surprisingly, many younger scholars saw Dyen's approach to reconstruction as a regression. This is not to imply that they regarded Dempwolff's work as flawless, but rather that Dempwolff had at least adhered to notions of explanation that are generally acceptable to philosophers of science: theories must be constrained, inferences must be justified by independent lines of evidence, the present is our best guide to understanding the past, etc. Dyen's work was thoroughly and critically reviewed in Dahl (1976), who was able to draw on Ferrell (1969) as a newer source on Formosan languages. Dahl rejected all of Dyen's new phonemes except *C (which, following Ogawa and Asai, he wrote $*t_2$), *N (which he wrote *ɬ), and *H, which he retained with a question mark. However, he also proposed that Dempwolff's *d/ḍ distinction be replaced by a three-way distinction: $*d_1$, $*d_2$ and $*d_3$, based almost entirely on unexplained irregularities in Paiwan.

The PAN phoneme inventory in Table 8.18 contains 25 consonants and four vowels. In addition, although they were phoneme sequences rather than units, it is useful to list four diphthongs, since these monophthongised in many daughter languages.

Table 8.18 The phoneme inventory of Proto Austronesian

p	t,C	c			k	q
b	d	z	D	j	g	
m	n	ñ			ŋ	
	S	s				h
	l	N				
	r,R					
w		y				

Vowels: i, u, e (schwa), a
Diphthongs (not phonemes, but diachronic sources of single Vowels): *-ay, *-aw, *-uy, *iw

Where they are not obvious, assumed phonetic values are: *C: voiceless alveolar affricate, *c: voiceless palatal affricate, *q: uvular stop, *z: voiced palatal affricate, *D: voiced retroflex stop (only in final position), *j: palatalised voiced velar stop, *S: voiceless alveolar fricative, *s: voiceless palatal fricative, *N: palatalised alveolar lateral, *r: alveolar flap, *R: alveolar or uvular trill. Canonical shape was CVCVC, or CVCCVC, where all consonants were optional, and consonant clusters were allowed only in reduplicated monosyllables such as *buCbuC 'pluck, pull out'. Medial prenasalisation is widespread outside Taiwan, but cannot be reconstructed for PAN.

The one point of general agreement about PAN phonology is that certain proto phonemes are completely non-controversial, while others are problematic in either of two ways: 1) their existence as distinct segments is in doubt, or 2) they are widely accepted, but their phonetic interpretation is unclear. In the following discussion it will be convenient to group the reconstructed segments by manner, departing from this format when it is more economical to treat all members of the same place-defined class together at once. Before treating the segment inventory, however, it will be useful to discuss the question of whether a contrastive accent can be reconstructed for Proto Austronesian.

8.2.1 Did PAN have a phonemic accent?

Dempwolff's reconstruction of 'Uraustronesisch' phonology was methodologically careful, systematic, and generally quite successful. The two areas where he was unable to deal satisfactorily with the data were: 1) his treatment of the 'laryngeals', and 2) his inability to account for phonemic stress in Tagalog. As seen already, Dyen (1953b) addressed the first of these problems. However, the second was left to fester for another two decades.[92]

Zorc (1972) noted that many languages in the central and northern Philippines have phonemic stress. Since penultimate stress is the unmarked pattern elsewhere, he reasoned that the causes of ultimate (oxytone) stress need to be identified. Zorc found a number of environments in which oxytone stress could be predicted in the Tagalic languages: 1) following medial consonant clusters, 2) following penultimate schwa, 3) following consonant loss, and 4) in some form classes that appear to be distinguished by stress

---

[92]   As Zorc (1978:67) pointed out, this shortcoming was a major concern of Laves' (1935) in many ways unduly harsh review of VLAW, vol. 1.

pattern (e.g. topic pronouns, numerals except 'ten', vocative forms, interrogative particles, negatives and deictics are all final-stressed in most Tagalic languages). Finally, stress contrasts carry morphological information in examples such as Tagalog *tápus* 'to finish' vs. *tapús* 'finished'. He concluded (Zorc 1972:53) with a question: "Did Proto-Tagalic inherit or innovate its ultimate stresses?"

Zorc (1978) opens with a similar question, but one directed at Proto Philipppines (PPH). This publication differs from his earlier discussion of Proto Tagalic stress in three respects: 1) it examines contrastive stress in the entire Philippine group of languages, 2) it concludes that "PPH had contrastive word accent" (1978:89), and 3) it argues that PPH contrastive word accent is inherited from 'Proto Hesperonesian.'[93] The last of these arguments seeks to find evidence for contrastive stress outside the Philippine group in languages that no longer have it, and so appeals to an AN version of Verner's Law.

Zorc states that stressed vowels in Philippine languages are long, and unstressed vowels short. If length is assumed to be primary and stress derivative, the search for evidence of contrastive accent can be extended to non-Philippine languages. Dempwolff had observed as early as 1924 that consonants often geminate following schwa, which is inherently short, but Zorc suggests that gemination in Philippine languages can occur after any short vowel. If so, gemination provides a clue to earlier final stress even in languages that do not have (or no longer have) a contrastive accent. As a test case Zorc considers data from Madurese, spoken on the island of Madura, just north of Java. He cites ten Madurese words with medial geminates for which Philippine cognates have final stress (hence short penultimate vowels), and although he acknowledges multiple sources of consonant gemination in Madurese, and recognises a few exceptions to the rules of correspondence, he concludes that there is sufficient evidence to posit contrastive stress in 'Proto Hesperonesian'. A close examination of the Madurese evidence for pre-PPH stress, however, fails to support this conclusion. Table 8.19 lists Zorc's evidence for short vowel-geminate correspondences (Part A), together with counterevidence (Part B). For the sake of consistency PPH forms are written with stress rather than length:

Table 8.19 Philippine accent and Madurese gemination

	PPH	Madurese	English
A)			
1.	laŋúy	laŋŋoy	swim
2.	bukáq	bukkaʔ	open
3.	qasín	assin	salty
4.	basáq	bassa	wet
5.	pitú	pittu	seven
6.	walú	ballu	eight
7.	halíq	alle	move
8.	taqás	attas	above
9.	tudúq	tojju	to point
10.	labúq	labbhu	downwards

---

[93]   The meaning of this term is unclear, although it presumably refers to Dyen's (1965a) 'Hesperonesian Linkage', a lexicostatistically-defined subgroup that includes most languages of western Indonesia and the Philippines, but excludes Tontemboan, Palauan, and Chamorro.

B)

1.	qatép	ataʔ (not –tt-)	roof; thatch
2.	balík	baliʔ (not –ll-)	return
3.	limá	lima (not –mm-)	five
4.	hapúy	apoy (not –pp-)	fire
5.	uRát	uraʔ (not –rr-)	vein
6.	matá	mata (not –tt-)	eye
7.	láŋit	laŋŋeʔ (not –ŋ-)	sky
8.	túkuŋ	tokkoŋ (not –k-)	tailless
9.	túba	tobba (not –b-)	derris root

As seen in 4.1.2, oxytone words outnumber paroxytone words in most Philippine languages. In Madurese all consonants are geminated following schwa, but after other vowels simple consonants outnumber geminate consonants by at least ten to one. By random association, then, we would expect most matches to be 1) oxytone-simple, and this appears to be the case, as literally hundreds of additional examples of the type under B1-6 exist. The remaining associations generated by chance can be ordered hierarchically as 2) paroxytone-simple (many examples), 3) oxytone-geminate, and 4) paroxytone-geminate. Zorc lists only type 3, without reference to the relevant statistics. Yet based on random association we would expect about twice as many instances of Type 3 as of Type 4, and it is necessary to conclude that the counterevidence to this claim is about as strong as the supporting evidence.[94]

Zorc (1983) returns to this topic for a third time, proposing another AN version of Verner's Law. In a number of languages in northern Sarawak the PAN voiced obstruents have split into two series. In some languages one series contains what might be described as phonetically unmarked consonants and the other phonetically marked consonants. In other languages both series contain unmarked consonants, but the crosslinguistic sound correspondences are phonetically unusual, suggesting a derivation from phonetically complex prototypes. Blust (1969, 1974b) attributed these complex consonants to Proto North Sarawak clusters of voiced obstruent plus *S, which were derived by vowel syncope when *S was still a sibilant (Kenyah = Highland Kenyah dialect of Long Anap):

Table 8.20 Hypothesised Proto North Sarawak *S clusters and their reflexes

PNS	Kelabit	Kiput	Kenyah	Miri	Bintulu
*b	b	b	b	b	b/v
*bS	bʰ	s	p	f	ɓ
*d	d	d	d	d	r
*dS	dʰ	s	t	s	ɗ
*j	d	j/c/s	j	j	j
*jS	dʰ	s	c	s	j
*g	g	g/k	g	g	g
*gS	gʰ	k	k	k	g

The Kelabit reflexes of PNS *bS, *dS, *jS and *gS are 'true voiced aspirates' as defined by Ladefoged (1971): they begin voiced but end voiceless, with optional voiceless

---

[94] It must be added that the evidence for consonant gemination after short vowels other than schwa is extremely limited, and such a process is completely absent from the great majority of AN languages.

onset to the following vowel. There were several reasons for positing *S clusters in Proto North Sarawak. Without this hypothesis it appeared particularly difficult to explain the correspondence of Kelabit b^h to Kiput *s*. In addition, it was noted (Blust 1974b) that in Kelabit *i*, *u*, and *a* can precede plain voiced obstruents, and that *i*, *u* and ə can precede the voiced aspirates. The contrast of PAN *a and *ə is thus neutralised in a different way in each of these environments. Since antepenultimate *a* is neutralised with schwa in North Sarawak languages this observation seemed to support the reconstruction of an added syllable with *S: PMP *taliŋa > PNS *təliŋa 'ear', hence PMP *dakdak revised to *da-daSak > PNS *dədSak 'tamp down earth'. In many cases the reconstructions with an additional syllable that contain *S seemed to explain the data well (and presented no problems in most other languages, since *S became h, and often disappeared between like vowels in trisyllables even when it would otherwise have been retained). In other languages, however, it was contradicted, particularly where Formosan cognates failed to support the new instances of *S. The hypothesis was accordingly rejected by Charles (1974), Dahl (1976) and Zorc (1983).

Zorc (1983) claimed that the aberrant obstruents of North Sarawak languages always follow a vowel that is short in cognate Philippine forms. This is generally true, but there are also many examples of normal voiced obstruents in this position. In other words, the following categories appear to correspond: 1) oxytone-normal, 2) oxytone-aberrant, 3) paroxytone-normal. By contrast, paroxytone-aberrant correspondences are rare or absent. The problem with this hypothesis, as with the first version of an AN 'Verner's Law' linking Philippine accent and Madurese gemination, is that the proposed correlation is not bilaterally predictive: aberrant reflexes of the voiced obstruents in North Sarawak languages often predict Philippine oxytonality, but Philippine oxytonality does not predict aberrant reflexes of the voiced obstruents in North Sarawak languages. Moreover, there is a natural explanation why aberrant reflexes of the voiced obstruents in North Sarawak languages so often successfully predict Philippine oxytonality. Adapting suggestions first made by Charles (1974) it appears that the reconstruction of PNS *S clusters should be replaced with geminate consonants which derived primarily from two sources: 1) automatic gemination after *ə, and 2) complete assimilation of the first consonant of a medial cluster in reduplicated monosyllables:

Table 8.21 Revised Proto North Sarawak sources for aberrant voiced obstruents

PMP	PNS	Kelabit	Kiput	Bintulu	English
*qabu	*abu	abuh	abəw	avəw	ashes
*tebuh	*təb^hu	təb^huh	təsəw	təɓəw	sugarcane
*buhek	*əb^huk	əb^huk	suə?	ɓuk	head hair
*bahaq	*əb^ha?	əb^ha?	səi?	ɓa?	water
*bukbuk	*bub^huk	—	busuə?	—	wood weevil
*bahaR	*əb^haR	əb^har	—	—	loincloth
*pajay	*paday	pade	padəy	—	riceplant
*qapeju	*pəd^hu	pəd^huh	pəsəu?	lə-pədɓəw	gall
*dakdak	*dəd^hak	dəd^hak	—	—	tamp earth
*tujuq	*tujuq	tudu?	tucəu?	tuju?	seven
*haRezan	*əj^han	əd^han	asin	k-əjan	ladder

Under this revised interpretation most aberrant reflexes of PMP voiced obtruents in North Sarawak languages reflect geminates (automatic after *ə) that underwent terminal

devoicing before the breakup of PNS. This in itself is not unexpected, since it is difficult to maintain obstruent voicing through the duration of a geminate consonant, hence the greater frequency of gemination with voiceless stops than with voiced stops in languages generally. Since both penultimate schwa and medial consonant clusters predict final stress in Proto Philippines, it is hardly surprising that the aberrant reflexes of the voiced obstruents in North Sarawak languages can be used to predict Philippine oxytonality.

Dahl (1981a:108ff) expressed uncertainty regarding the need to reconstruct contrastive stress for PAN, since there is no clear pattern of agreement between the stress contrasts of Formosan and Philippine languages, but Ross (1992:50) argued that Budai Rukai shows agreement of stress placement in cognate forms: "The oxytones of Budai are apparently the last remnants of PAN contrastive stress. They generally correspond with PPH oxytones, and therefore presumably reflect PAN oxytones." However, the relevant evidence fails to support Ross's claim, since the Budai-PPH correlations he proposes are contradicted nearly as often as they are supported (Blust 1997d).

Following a suggestion of Tsuchida (1976:210ff) that contrastive accent must be posited for Proto Tsouic, Pejros (1994:125) has argued that this accent derives from a similar feature in PAN, and that once it is recognised the *t/C distinction "can be interpreted as a secondary Formosan derivative based on an old accent system that is now reflected only in Tsou." Pejros does not attempt to find correlations between Tsouic vowel reduction and Proto Philippines accent contrasts, and his conclusion that it is possible to reconstruct an "earlier Formosan accent system" is puzzling both in assuming without evidence that there is a Formosan subgroup, and in claiming that is possible to reconstruct contrastive accent in PAN based solely on data from a single language or close-knit subgroup. Because his proposals largely overlap with those of Wolff, which will be examined below in the discussion of the *t/C and *ñ/N contrasts, there appears to be no need to discuss them further here.

8.2.2 Voiceless stops

The uncontroversial voiceless stops are *p, *t, *k and *q. Abundant evidence for the first three is given in Dempwolff (1938), and for *q in Dyen (1953b). There is universal agreement that *p, *t and *k were voiceless and unaspirated. Where adequate information is available the evidence from widely separated languages suggests that *t was postdental, while *d and *n were alveolar. Although Dyen hedged his bets on the phonetic interpretation of *q (1953b:1, 50, fn. 2), he was led to speculate 'that PMP *q* was either (a) a glottal stop, or (b) a glottal spirant.' He later came to reject this view, and it is universally rejected today in favor of the view that *q was a uvular stop. Blust (1999b:43) found *q reflected as a uvular stop in four Formosan witnesses (Proto Atayalic, Paiwan, Thao, Bunun), as glottal stop in four others (Amis, Saisiyat, Kanakanabu, Saaroa), as *h* in Taokas, as a voiceless pharyngeal fricative (written *H*) in the Tamalakaw dialect of Puyuma (Tsuchida 1982:190), and as zero in eleven others (Kulon, Basay, Trobiawan, Kavalan, Siraya, Hoanya, Babuza/Favorlang, Papora, Pazeh, Tsou, Proto Rukai). It has since become clear that in at least Central Amis the reflex of *q is an epiglotto-pharyngeal stop (Edmondson, Esling, Harris and Huang 2005). Taokas *h* was recorded by linguistically untrained observers early in the Japanese colonial period, and its precise phonetic character is unknown.

Although some of these reflexes undoubtedly involve single changes in the immediate common ancestor of lower-level subgroups, as with Basay, Trobiawan and Kavalan, or

Hoanya, Babuza/Favorlang and Papora, failure to take this possibility into account does not affect the balance of evidence in deciding upon a phonetic value for *q. The change paths that can be inferred from the Formosan evidence can be summarised as lenitive: a uvular stop has become backed to glottal stop (sometimes with superimposed epiglottal or pharyngeal features), and this has then disappeared in most witnesses, but become a pharyngeal or glottal fricative in two cases.

In Malayo-Polynesian languages *q is rarely, if ever, reflected as a uvular stop. It is generally lost, and when it is retained it is most often a glottal stop. This is true of most languages in the Philippines and Borneo, of Rejang in southern Sumatra, of Palauan and Chamorro in Micronesia, and a scattering of Oceanic languages including Bonga, Tonga, Makura and Mataso of Vanuatu, and Tongan, Rennell-Bellona, and Rapanui in Polynesia. In much of the Philippines and western Indonesia where *q is reflected as glottal stop in medial and final position, it has merged with zero word-initially. In some of the languages of Sulawesi, as members of the Bungku-Tolaki group (Mead 1998) the glottal stop is preserved only medially, but has left a trace of its former presence word-finally by lowering preceding high vowels. In these languages the glottal stop has only one source, but in some Oceanic languages, as Molima, Bunama, Suau and Wagawaga near the tail of New Guinea *q is reflected as glottal stop, but has merged with the lenis reflex of *k, and so may have passed through a stage in which it was velar.

Probably the second most common reflex of *q outside Taiwan is *k*, usually following the change of *k to glottal stop or zero. This is found in several Philippine languages, including Tagbanwa Kalamian, Agutaynen, Tboli and Bilaan, in Moken-Moklen of the Mergui Archipelago, Watubela of the central Moluccas, several languages of Manus and its satellites in the Admiralty Islands, Kis, Sera, Kairiru, Bam, Medebur and some other poorly described languages of the north coast of New Guinea, Mapos Buang, Mumeng and some other languages of the upper Markham Valley in New Guinea, and in Dawawa, Tubetube and other languages of the Massim region of New Guinea. In a number of the languages of the Markham Valley which show partial merger of *q and *k the reflex of *q is *g* rather than *k* (Ross 1988:139).

Fricative reflexes are somewhat less common, but *q is reflected as *h* in the Chamic languages of mainland Southeast Asia, and positionally as *h* or Ø in Malay and some of its closest relatives in island Southeast Asia, in Sundanese, Javanese, Lampung, Nias and the Northern Batak languages. It is less common for *q to become *h* in Oceanic languages, although this development appears in Lakalai of New Britain, in Longgu of the central Solomons, and occasionally elsewhere; a far more common development in Oceanic languages is for *q to merge with the lenis reflex of *k as ɣ (Ross 1988). A similar change is found in Muna of southeast Sulawesi, where *q is reflected as a voiced velar fricative written *gh* (although here it has not merged with *k). Finally, although Palauan *ch-* (< *q) now represents a glottal stop, the German-based orthography reflects the fact that this sound was recorded as a voiceless fricative somewhat over a century ago. The general assumption is that this was a velar fricative, although the possibility that it was pharyngeal cannot be ruled out.

8.2.2.1. PAN *C

As noted already, the contrast of $*t_1/t_2$ and $*n_1/n_2$ was first proposed by Ogawa and Asai (1935). These distinctions were endorsed and relabeled *t/C and *n/N by Dyen (1965c), and have been accepted by nearly all Austronesian specialists, including Ferrell (1969), Blust (1969, 1970a), Li (1972, 1977b), Dahl (1976, 1981a), Tsuchida (1976), Ho (1978),

Zorc (1982), and Ross (1992). Wolff (1991), however, has argued that *t and *C were allophones of a single phoneme conditioned by stress: the *C allophone occurred in disyllables with final stress, and in trisyllables with penultimate stress, while the *t allophone was found elsewhere. His theory thus assumes that phonemic stress can be reconstructed for PAN.

The evidence given for stress contrasts in PAN is as follows. First, the Formosan languages Kanakanavu and Rukai have contrastive stress (or vowel length). However, in the first of several qualifications Wolff (1991:537) notes that "Rukai stress furnishes no evidence for stress in pre-Formosan." On the other hand, Tsou, Maga Rukai, and Atayal show reduced penultimate vowels in the reflexes of some reconstructed disyllables but not in others, and this is taken as evidence of an earlier oxytone stress pattern that is no longer directly attested in these languages. Wolff's second qualification (1991:537) is that "Atayal is not, however, a good witness for the stress pattern of the root in the proto language." His third qualification holds that "Tsou has lost contrastive stress except on the morphophonemic level" and this language therefore cannot be used as a witness for PAN stress contrasts in verbs. In the next paragraph he gives his fourth qualification: "The stress patterns in all of the Austronesian languages are very heavily subject to analogical changes … Verb forms in Tsou and Kanakanavu rarely provide evidence for root stress because stress shift is part of the verbal morphology … In short, there often is no agreement in stress patterns over languages." Despite these manifold difficulties Wolff concludes (1991:537) that a theory of PAN contrastive stress can be entertained, based on "detailed work with Cebuano and Tagalog and careful scrutiny of Tsou and Atayal data." This statement is surprising in view of his second qualification, but he nonetheless proposes the following guidelines for the reconstruction of PAN stress: "(a) The stress patterns of nouns and some other forms which, other than stative adjectives, occur unaffixed tend to remain unchanged, (b) In the Philippine languages in verbal roots the stress pattern of the actor focus verbs tends to reflect the inherited stress pattern. Verb forms in the Formosan languages rarely provide evidence." These claims are summarised in Table 8.22:

Table 8.22 Sources for inferring PAN contrastive stress according to Wolff (1991)

Language	Verbs	Unaffixed forms (mostly nouns)
Kanakanavu	No	Yes
Rukai	No	No
Tsou	No	Yes
Maga Rukai	No	Yes?
Atayal	No	No?
Tagalog	Yes (AF only)	Yes
Cebuano	Yes (AF only)	Yes

In short, Wolff gives most weight to Central Philippine languages. In fact, his reconstruction strategy essentially requires that Proto Philippines stress be equated with PAN stress: "A common situation is one in which there is disagreement among our Formosan languages, but the Philippine languages show agreement with some of them and confirm the hypothesis which we are seeking to demonstrate here (1991:538).

There are serious problems with this argument. First, since around 90% of all comparata are disyllables there are essentially two possible stress patterns: oxytone and paroxytone. This combination of factors (two possible stress patterns in disyllables, internally inconsistent Formosan evidence but internally consistent Philippine evidence) produces

essentially three conditions on comparisons: (1) the Formosan evidence points unambiguously to one of these patterns and agrees with Philippine evidence, (2) the Formosan evidence is internally inconsistent, but *either of the patterns* can be compared with Philippine evidence, (3) the Formosan evidence is internally consistent but disagrees with Philippine evidence. Wolff takes the first two conditions as evidence for PAN stress. The only counterevidence that his prodecure acknowledges is thus one in which Formosan evidence is internally consistent and disagrees with Philippine evidence. Since Wolff states that the Formosan evidence which he allows (stress contrasts, vowel reductions) tends to be internally contradictory this makes almost any cognate set automatically supportive of his hypothesis that the stress pattern of Philippine languages continues a similar pattern that was present in PAN.

Second, Wolff counts only contrastive stress and vowel reductions in Formosan languages as evidence for PAN stress. This is a strange and unexplained restriction, since his assumptions imply that all languages which distinguish *C from *t provide evidence of stress contrasts in PAN. By removing this arbitrary restriction many other Formosan languages become available to test the hypothesis, including Puyuma, Paiwan, Saisiyat, Pazeh, Thao, all dialects of Rukai, and most of the now extinct languages of the western plains. But these languages are almost always internally *consistent* in pointing to *t or *C. This makes it all the more puzzling why Wolff has excluded inferences about PAN stress based on the *t/C contrast, since these meet condition (3), and hence constitute the only corpus of comparisons that permits a fair test of his claims.

Third, when relevant evidence is assembled the correlation of the *t/C contrast with phonemic stress does not support Wolff's claim that *C occurred in disyllables with final stress and trisyllables with penultimate stress, while *t was found elsewhere. Quite apart from the fact that this is a decidedly odd set of conditions for allophony, differing radically from the straightforward suprasegmental conditioning seen in e.g. Verner's Law, it predicts incorrectly that Thao *t-*, Paiwan *tj-* will correlate with paroxytone stress, and Thao *c-*, Paiwan *ts-* with oxytone stress in Philippine languages. Table 8.23 cites known relevant comparisons for what Wolff calls PAN *t- (minor semantic distinctions in the glosses are ignored, and reduplicated monosyllables are omitted, since these are automatically oxytone in Tagalog and Cebuano):

Table 8.23 Correlation between Wolff's PAN *t and stress in Philippine languages

A. Supportive cases

Thao	Paiwan	Tagalog	Cebuano	
tukus	tjukəz	túkod	túkud	'prop, support'
tusuq	tjuzuq	túloʔ	túluʔ	'drip, leak'
tusuq		túroʔ	túruʔ	'to point'
shaqish	tsaqis	tahíʔ	táhiʔ	'sew'
	tsəvud	tubúd		'spring, source'
	tsiŋas	tiŋá	tiŋá	'food in teeth'
	tsiqaw		tiʔáw	'goatfish'

B. Contradictory cases

Thao	Paiwan	Tagalog	Cebuano	
talhaq		tagáʔ	tágaʔ	'cut down'
	tjalʸək		tanúk	'cook, boil'
tapish	tjapəs	tahíp	tahúp	'winnow'
turu	tjəlu	ta-tló	tulú	'three'
	tjəvas	tibáʔ	tubáʔ	'cut vegetation'
tufúsh	tjəvus	tubó	tubú	'sugarcane'
	tjəza	tirá	tulá	'leftover food'
tiaz	tjialʸ	tiyán	tiyán, tíyan	'belly'
tuza	tjulʸa		tuná	'eel'
	tjuqəz	tuʔód	tuʔúd	'tree stump'
cakaw	tsakaw	tákaw		'steal'
canit	tsaŋitj	táŋis	taŋís	'weep, cry'
	tsalis	táliʔ	táliʔ	'rope'
	tsapa	tápa	tápa	'jerked meat'
	tsapəl	tápal	tápal	'patch'
caqi	tsaqi		táʔi	'feces'
caw	tsau	táʔo	táʔu	'person'
cawa		táwa	tawá	'laugh'
	tsətək	túsok	túsuk	'pierce'
	tsusu	túhog	túhug	'to string'

As seen in Cebuano *tiyán*, *tíyan* 'belly', stress variation occurs even within the same Philippine language. In addition, despite a strong tendency for stress patterns to agree within the Philippines there is cross-linguistic variation, as with Thao *talhaq*, Tagalog *tagáʔ*, Cebuano *tágaʔ* 'cut down', Thao *canit*, Paiwan *tsaŋitj* (for expected **tsaŋit), Tagalog *táŋis*, Cebuano *taŋís* 'weep, cry', Thao *cawa*, Tagalog *táwa*, Cebuano *tawá* 'laugh', or Thao *shaqish* (< caqish), Paiwan *tsaqis*, Tagalog *tahíʔ*, Cebuano *táhiʔ* 'sew'. The indeterminacy introduced by such variation usually can be eliminated by appeal to more distantly related Philippine languages. Thus Ilokano *tián* 'belly' confirms PPH *tián 'belly', Bikol *tagáʔ* 'cut, gash', Ilokano *tagá* 'trim, shape (by cutting)' confirm PPH *taRáq 'cut, hack', Ilokano, Isneg *sáŋit* 'weep, cry' confirm PPH *táŋis 'weep, cry', Ilokano *katáwa* 'laugh' confirms PPH *táwa 'laugh', and Bikol, Hanunóo *tahíʔ* confirm PPH *tahíq 'sew'. We are thus left with seven confirmatory cases and 20 contradictory ones. Chance alone would lead us to expect 6.75 cases for each of the four cells generated by the intersection of stress placement and the *t/C distinction, but the attested values are *t:P = 3, *t:O = 10, *C:P = 10, and *C:O = 4 (where P = paroxytone, and O = oxytone). This suggests a statistically significant *negative* correlation with Wolff's claim which itself requires explanation.

In Wolff's own presentation the correlation of Philippine stress with the *t/C contrast in Formosan languages is more nearly random. For correspondences that other scholars assign to *C-, for example, Wolff finds seven cases that support his hypothesis, but thirteen that show a problem. In medial position there are eight supportive cases and five problematic ones, but only six of the supportive cases involve both Philippine and Formosan cognates. Word-finally there are three supportive cases and four problematic

ones, and much the same can be said for the other examples that he gives. Wolff dismisses this massive counterevidence on the grounds that stress movement is common under affixation, and so secondary developments could have obscured an original pattern of agreement. Without clear evidence for such secondary disturbances, however, the reader is essentially left to accept the argument on faith, and Wolff's attempts to eliminate *C have therefore won few converts.

8.2.2.2 PAN *c

The reconstruction of *c (originally proposed by Dempwolff) presents a very different set of problems. Wolff has suggested that this segment also be dropped, and Ross (1992) agrees with this view, although Dahl (1976:84) does not. As Ross correctly points out, the problem with justifying *c is that it is found in only one primary branch of the AN family. In fact *c is distinguished from *s in only 20-25 languages of western Indonesia and mainland Southeast Asia, including all of the Chamic languages (Thurgood 1999:81), Iban, Malay, Karo and Dairi-Pakpak Batak, Rejang, Lampung, Sundanese, Javanese, Madurese, Balinese, Sasak, Buginese, Makasarese and some other South Sulawesi languages. Few proposed distinctions have so divided AN scholars. Mills (1975) assigned *c to Proto South Sulawesi, Nothofer (1975) assigned *c to his 'Proto Malayo-Javanic' (a pseudo-taxon), and Adelaar (1981) assigned *c to Proto Batak. While these reconstructed languages share cognate forms that contain *c, the problem has always been to justify the assignment of *c to a more remote proto language ancestral not only to them, but also to the languages of the Philippines, eastern Indonesia, Oceania and Taiwan. The arguments for and against proposing a longer ancestry for *c in the AN family are summarised in Figure 8.2:

	Against	For
1.	found only in western Indonesia	*c or massive unconditioned splitting
2.	spread by borrowing from Malay	found in monosyllabic roots
3.	not in basic vocabulary	patterns with other palatals

Figure 8.2 Arguments for and against assigning *c to PMP or PAN

The first argument against *c has already been noted: distinctive reflexes of *c are limited to some two dozen languages in western Indonesia, although these do not form a subgroup, and are geographically discontinuous, including languages in southwest Borneo, northern and southern Sumatra, Java-Bali-Lombok and southern Sulawesi. It is clear that the rarity of evidence for *c cannot be damning in itself. Very few Oceanic languages retain a distinct reflex of the palatal nasal *ñ, but it is universally agreed that this segment formed part of the phoneme inventory of Proto Oceanic (Blust 1978a). As already noted, what is critical in justifying a reconstructed distinction is not cross-linguistic frequency, but rather the distribution of reflexes: if these are confined to a single primary branch there must be cogent extenuating circumstances to justify the inference that an attested correspondence reflects a distinct proto phoneme in the ancestor of the entire group. POC *ñ is certain, since cognate forms containing a phonemic palatal nasal are found not only in several primary branches of Oceanic, but also in non-Oceanic AN languages. To take a more extreme case, there is strong evidence for a Philippine subgroup, but Kapampangan

is the sole member of this group which preserves a reflex of *ñ unmerged with *n. Here an original phonemic distinction is limited to one primary branch, and an appeal to external evidence is needed to strengthen the argument for PPH *ñ.

Apart from the observation that they are generally classified under the rubric 'Western Malayo-Polynesian' the subgrouping relationships of languages that appear to preserve a reflex of *c remain unclear. What would strengthen the argument for PMP or PAN *c, then, would be the discovery of even a single witness from outside western Indonesia that would allow *c to be moved further up the AN family tree. At least two scholars have made such proposals. Dyen (1949:424-425) suggested that Chuukese reflects *c as *s*, but *s, *z/Z and *j as *t*. Somewhat later this suggestion was taken up again by Goodenough (1961:124), who claimed that *c is distinguished from the other palatal obstruents not only by Chuukese, but also by Rotuman and Gilbertese. However, this proposal has not been generally accepted, and plays no part in current reconstructions of Proto Oceanic phonology (Blust 1978a:116-117, Ross 1988). We are thus forced to conclude either that *c did not exist in PMP or PAN, or that it was highly unstable, and merged with *s in nearly all surviving languages.

The second argument against *c is that many of the forms in which it was reconstructed by Dempwolff (1938) are loanwords from Malay. This includes such well-known reconstructions as *candu 'opium', *caŋkul 'hoe', *caremin 'mirror', *campur 'to mix', *cium 'kiss' (from a Mon-Khmer source), *cuba 'to try', *baca 'to read' (from Sanskrit), *kaca 'glass' (from Sanskrit), *ka(n)caŋ 'bean', *kunci 'key', *racun 'poison', and *ucap 'to speak'. Malay loanwords of this kind are found in coastal languages of Borneo, in Sumatra, and the Philippines, and in Sundanese, Javanese and Madurese.

The third argument against *c is that it does not occur in basic vocabulary as represented in the Swadesh 200 word list or suitably adapted variants of this list. This is a serious objection to the reconstruction of any proto phoneme, as it suggests that the evidence for it is a product of borrowing. However, it is not necessarily a fatal objection. Well-supported proto phonemes differ markedly in their frequency in basic vocabulary. In a variant of the Swadesh 200-word list for PMP the voiced velar stop *g is found only in *gemgem 'hold in the fist' and *gurgur 'thunder' (Blust 1993b), yet as will be shown below, its assignment to both PMP and PAN is well-supported.

While these three considerations may seem to rule out PAN *c, at least three others support a PAN voiceless palatal affricate. First, without *c the correspondence that is assigned to it must be assigned to *s, since *c and *s are not distinguished in most languages. But this would result in unconditioned phonemic splitting in scores of forms (Dempwolff posited *c- in 48 etyma, and at least 47 other examples of *c- have been reconstructed in PWMP or PMP forms since). Wolff (1982:3) proposed that *c be eliminated because the comparisons offered for it are flawed in one or more of four ways:

1. they are onomatopoetic, or likely to have been affected by analogy;

2. they are limited to languages in close contact with Malay, and likely to be loans;

3. some comparisons with *c are semantically unrelated, or other phonemes in the form show diachronic irregularities;

4. the protoform could be reconstructed with *s.

The basis for some of these criticisms is clear, and in some cases the skepticism about *c unquestionably is justified. In others, however, the attempts to dismiss evidence for *c appear far-fetched or even motivated by an *a priori* decision. It does not follow, for

example, that because a word is onomatopoetic it is an invalid comparison. Many onomatopoetic words show regular sound correspondences, are as much a part of the lexicon as any other form, and are rarely borrowed. There is thus no compelling reason to dismiss comparisons such as *ciak > Maranao *siak* 'noise made by chicks', Malay *ciak* 'twitter of small bird', *ciap > Tagalog *siyáp* 'chirp of chicks or birdlings', Bikol *siyáp* 'chirp', Hanunóo, Cebuano *siyáp* 'peep, chirp', Lara (Land Dayak), Bekatan *siap*, Kayan *hiap* 'chicken', Malay *ciap* 'twitter of small bird', Bolaang Mongondow *siap* 'peeping of chicks', Hanunóo *suksúk*, Aklanon *súksuk*, Malay *cəcak*, Makasarese *caʔcaʔ*, Buginese *cicaʔ* 'house lizard, gecko', or Iban *cərik-cərik* 'talk in a shrill voice', Toba Batak *sorik* 'let out a shrill scream as a battle-cry', Manggarai *cerik* 'scream shrilly'. One of the obvious results of such a methodological stricture would be to posit proto languages with no onomatopoetic vocabulary, although this clearly would be unnatural, given the rich systems of onomatopoeia found in many of the daughter languages.

Wolff's other objections can be answered by a consideration of specific examples that are 1) not likely to be borrowed from Malay, 2) not semantically questionable, and 3) not assignable to an etymon with *s. Many of these can be found in Blust (1970a, 1973a, 1980b, 1983/84b, 1986, 1989b) and Blust and Trussel (ongoing). The following sample should suffice to make the point. Languages that distinguish *c from *s are Iban, Malay, Karo Batak, Modern and Old Javanese, Balinese, Rhade, Sasak, Makasarese and Buginese:

1. *cakaq : Maranao *saŋkaʔ* 'refuse, decide, oppose', Western Bukidnon Manobo *saŋkaʔ* 'fight each other', Iban *cakah* 'contradict, dispute with, oppose'

2. *ceŋis : Malay *cəŋis* 'odor that spoils the appetite', Bolaang Mongondow *toŋit* 'stench (as of sweaty armpits, burned rice, etc.)'

3. *cikep : Maranao *sikəp* 'catch with the bare hand; grey bird of prey', Malay *cikap* 'chopsticks', Karo Batak *cikəp* 'hold something in the hand', Bolaang Mongondow *sikop* 'catch fish with hand', Tongan *hikof-i* 'pick up with tongs', Rennellese *siko* 'catch, as a ball or a wave', Maori *hiko* 'snatch'

4. *culcul : Maranao *sosol* 'light a fire', Western Bukidnon Manobo *sulsul* 'set fire to s.t. with a torch', Iban *tucul* 'apply fire to'

5. *lucak : Ilokano *lúsak* 'smash, crush', *ma-lúsak* 'fall to the ground and bruise (fruits); be smashed, crushed', Tagalog *lúsak* 'mud, slush', Maranao *losak* 'trample under foot (as a rice paddy prior to planting)', Iban *lucak* 'muddy, soft ground', Karo Batak *pe-lucak* 'have carabaos trample a paddy field to prepare it for planting'

6. *pacek : Amis *pacək* 'a nail', *mi-pacək* 'to nail, hammer a nail in', Kankanaey *pasək* 'wedge', Maranao *pasək* 'post, pillar', Western Bukidnon Manobo *pasək* 'put a post or stick in the ground', Bintulu *pasək* 'enter', Malay *pacak* 'splitting, transfixing, of roasts before the fire or on the spit; fixing pointed stakes in the ground; driving in a boundary post', Nias *fasa* 'stake, nail; to pound in a stake or nail', Old Javanese *pacək* 'spike, pin', Balinese *pacək* 'peg, nail; stick into, make a hole; to plant (seedlings, etc.)', Manggarai *pacək* 'pile, stake; drive in a pile or stake'

7. *peceq : Paiwan *pətəq* 'a break, split (in glass, pottery)', *ma-pətəq* 'become broken', Ilokano *pessá* 'hatch from an egg', Kapampangan *apsáʔ* 'to hatch, of an egg', Tagalog *pisáʔ* 'crushed, pressed, compressed; hatched, of eggs', Cebuano *pusáʔ* 'crush or squash s.t. soft; break s.t. fragile; hatch an egg', Western Bukidnon Manobo *pəsaʔ* 'smash something, as an egg, a caterpillar, or somebody's body or

head', Rhade *mcah* 'broken', Malay *pəcah* 'breakage into bits', Karo Batak *pəcah* 'broken, in pieces; burst out, erupt, of a volcano', Sangir *pəsaʔ* 'crush flat; smash to bits; hatch, of eggs', Bare'e *poso* 'broken, of hard, brittle things; hatched, of eggs', Nggela *posa* 'break, of a boil; burst, of a football', Sa'a *ma-pota* 'broken to pieces', Arosi *bota* 'break by knocking on something else, as an egg on bamboo'

8. *picek : Bontok *písək* 'person who has a white patch in one or both eyes', Maranao *pisək* 'blind', Western Bukidnon Manobo *pisək* 'to blind someone or something in one eye', Iban *picak* 'blind in one eye', Old Javanese *picək* 'blind'

9. *pecel : Tagalog *pisíl* 'squeezing with the hand', Maranao *pəsəl* 'press, as button between fingers', Western Bukidnon Manobo *pəsəl* 'squeeze something between finger and thumb', Ibal *pəcal* 'pinch, squeeze, grasp', Malay *pəcal* 'crush between the fingers, squeeze in the hand'

10. *qapucuk : Amis *ʔapocok* 'the very top of a mountain', Limbang Bisaya *pusuk* 'peak, summit', Iban *pucok* 'top of a tree, the part above the unbranched trunk', Malay *pucok* 'shoot, top branchlet, leaf bud', Old Javanese *pucuk* 'top, highest or foremost point, beginning or end', Balinese *pucuk* 'tip, point, top', Sasak *pucuk/pusuk* 'top, peak', Makasarese *pucuʔ* 'top of a tree, tip of the tongue', Buginese *pucuʔ* 'vegetation; tip of a sprouting plant', Muna *pusu* 'top, tip'

Some of these forms are unknown in Malay (as 1, 4, 5 and 8), and therefore are not likely to be Malay loans. Others are found in Malay, but show none of the telltale traits of Malay loans in languages of western Indonesia or the Philippines (*e > *a* in the final syllable, etc.). There thus appear to be no grounds for dismissing these or similar comparisons as products of chance or borrowing, and if their etyma are reconstructed with *s the palatal affricate that appears in languages such as Iban, Rhade, Malay, Karo Batak, Old Javanese, Balinese, Sasak, Makasarese or Buginese must be regarded not just as an unexplained development, but as a *convergent* irregularity. This observation alone is sufficient to raise serious doubts about the propriety of dismissing *c.

Related to the issue of convergence is the observation that submorphemic roots with *c consistently reflect this segment as a voiceless palatal affricate whenever diagnostic evidence is available. Thus *lucak, which can be glossed 'make muddy through trampling' contains the root *-cak 'muddy; sound of walking in mud', which is also found in: 1) Western Bukidnon Manobo *basak* 'mud', 2) Malay *becak* 'slushy; mud-patch; puddle', 3) Kambera *kapihaku* 'mud, muddy', 4) Javanese *kracak* 'sounds of water pouring or gushing', 5) Malay *ləcak-ləcok* 'sound of a man walking through sticky mud', 6) Banjarese *licak* 'muddy' (+ Standard Malay *lecak* 'moist and slippery, as ground after rain'), 7) Karo Batak *oncak* 'shake, slosh about, as water in a can that someone is carrying', 8) Mansaka *pasak* 'mud', 9) Cebuano *pisák* 'muddy ground, mire', 10) Mandar *ressaʔ* 'mud', 11) Sasak *ricak* 'marshy, swampy, muddy', 12) Aklanon *tamsák* 'splash, splatter', and 13) Bikol *tapsák* 'a splash' (Blust 1988a:90). What is striking is that this and other data sets in Blust (1988a) consistently show *c* in languages that distinguish *c and *s, even though the forms in which a common root is found are not cognate. This distribution clearly rules out borrowing, and reconstruction with *s or dismissal on semantic grounds seem equally inappropriate.

Finally, *c patterns with a larger set of consonants that includes at least *z (a voiced palatal affricate), and *ñ. Taking everything into account, then, the most plausible hypothesis is that PAN had a palatal series that included voiced and voiceless affricates and a nasal. Of these three segments *c was the least stable, and is reflected distinctly from

*s in only 20-25 of the extant AN languages, all of which happen to be in western Indonesia (an area of exceptional phonetic conservatism in other respects as well).

8.2.2.3 *Dempwolff's voiceless retroflex stop*

The reconstruction of *T was problematic from the start. Dempwolff (1934:62) noted that it is distinguished from *t only by Old and modern Javanese, Madurese and Balinese. But Old and modern Javanese are different stages of the same language separated by no more than a millennium, and both Madurese and Balinese have been subjected to heavy Javanese contact influence. *T was the only segment that violated Dempwolff's implicit 'independent evidence requirement', and he allowed this violation only because his method of reconstruction tacitly gave greater weight to symmetry than to anything else. Dahl (1976:66), following an earlier suggestion by André Haudricourt, suggested that the retroflex stops in Javanese may have been acquired under Sanskrit contact influence. Given the occurrence of *ṭ* in onomatopoetic and clearly native forms such as *ṭuṭuk* 'object used for hitting; to hit again and again' (< *CukCuk) this hypothesis must be questioned. Nonetheless, the evidence for *T clearly falls far short of that for *c, since 1) the putative *t/T distinction is restricted to a single language together with two others that have been heavily influenced by it, and 2) the *t/T distinction cross-cuts the *t/C distinction that is far more strongly supported by Formosan evidence. All-in-all, then, the recognition of an unexplained change *t > *ṭ* in a small number of Javanese words seems preferable to reconstructing a phoneme *T on the basis of evidence from essentially a single language.

8.2.2.4 *Did PAN have a phonemic glottal stop?*

Once it became clear that PAN *q and *S were not laryngeals the question naturally arose whether PAN might have had true laryngeal consonants such as *ʔ or *h. The change paths open to glottal stops and glottal fricatives are severely restricted: in general glottal consonants either disappear, or change their manner of articulation while maintaining their place. Of these possibilities loss probably is more common. For this reason glottal stops and fricatives are normally the most difficult consonants to reconstruct, since in most language families they will leave only a faint diachronic signature.

Table 8.24 Correspondences supporting final *ʔ, *q and Ø (after Zorc 1996)

Language	*ʔ	*q	Ø
Tagalog	-ʔ	-ʔ	-Ø/ʔ
Bikol	-ʔ	-ʔ	-Ø/ʔ
Aklanon	-ʔ	-ʔ	-Ø/ʔ
Cebuano	-ʔ	-ʔ	-Ø/ʔ
Western Bukidnon Manobo	-ʔ	-ʔ	-Ø/ʔ
Kalamian Tagbanwa (Northern)	-ʔ	-k	-ʔ
Tboli	-ʔ	-k/ʔ	-h
Iban	-ʔ	-h	-Ø/ʔ
Malay	-Ø/k	-h	-Ø
Old Javanese	-Ø/k?	-h	-Ø
Paiwan	-Ø	-q	-Ø
Bunun (Takituduh)	-Ø/(ʔ)	-q	-Ø/ʔ

Zorc (1982) broached this topic in a seminal paper which raised the prospect that PAN had a phonemic glottal stop, a proposal that was accepted provisionally in Blust (1980b) and subsequent collections of new lexical reconstructions.[95] Most examples of *ʔ are word-final. In Zorc (1996:72) the correspondences supporting the reconstruction of PAN *ʔ, *q, *H, *S and Ø, and PMP *h are listed in tabular form. It is clear from this table that PAN *H and PMP *h are equivalent, and that both can be represented by *h. Languages taken to support the *ʔ : q : Ø distinction are shown in Table 8.24.

As seen in Table 8.24, *q and Ø are distinguished in all of these languages (this will be qualified below for Iban). The major challenge in establishing the reality of *-ʔ is thus to demonstrate that the *-ʔ : -Ø contrast is not due to secondary change. Kalamian Tagbanwa provides no help, since it has regularly added final glottal stop (Reid 1971). Malay and Javanese are also neutral, since the only evidence that Zorc cites for a non-zero reflex of *ʔ in these languages is the appearance of –k ([ʔ]) in several kinship terms, but this is better explained as a retention of *-q 'vocative' as a glottal stop which has resisted the later change *ʔ > h for pragmatic reasons (Blust 1979). Tboli has final glottal stop in transparent loanwords from Malay (*hintuʔ* 'door' < *pintu*) or other languages (*basuʔ* 'drinking glass' < Spanish *vaso*). Moreover, although *-q normally yields Tboli -*k*, it also produces glottal stop in some forms, as in *biRaq > bilaʔ* 'elephant-ear taro' (cf. Malay *birah* 'elephant-ear taro', where –*h* can only reflect *q), *pusuq > hosoʔ* 'heart' (cf. Kalamian Tagbanwa *pusuk*, or Old Javanese *pusuh* 'heart', where the finals can only reflect *q), *salaq > salaʔ* 'mistake, error' (cf. Malay, Old and modern Javanese *salah*, where –*h* can only reflect *q), or *teŋaq > təŋaʔ* 'one-half measurement' (cf. Malay, Old and modern Javanese *təŋah* 'middle'). The use of Tboli as a witness for *ʔ is thus compromised. Finally, neither Paiwan nor Takituduh Bunun provide evidence for a *ʔ : Ø contrast in final position. This leaves the languages of the central and southern Philippines and Iban of southwest Borneo as the principal witnesses for *-ʔ.

Most examples of final glottal stop in the first five languages reflect *q, but Iban presents a different picture. In this language final glottal stop cannot reflect *q or Ø, and is contrastive, as seen in *titi* 'low bridge' : *titiʔ* 'strip off' : *titih* 'follow, go along with, join' : *titik* 'drop; drip'. While final vowels reflect original final vowels, -*h* reflects *q, and -*k* reflects *k, there is no generally accepted source for -ʔ as a product of regular sound change. Final glottal stop in Iban reflects *-R in some forms, as with PMP *wahiR > aiʔ* 'fresh water' (Malay *air*), *ikuR > ikoʔ* (Malay *ekor*) 'tail', *qiliR 'to flow', > iliʔ* 'downstream' (Malay *hilir*), or *qateluR > təluʔ* 'egg' (Malay *təlur*), but this change is irregular (Adelaar 1992:91). In addition, although *q is generally reflected as Iban -*h*, it sometimes yields -ʔ, as in PAN *Cubuq > Iban tubuʔ* 'young edible shoot of bamboo' (cf. Paiwan *tsuvuq* 'bamboo sprout', where the uvular stop rules out the possibility of Zorc's *ʔ), PAN *liseqeS, PMP *lisehaq > Iban linsaʔ* 'nit, egg of a louse' (cf. Paiwan *lʲi-səqəs*, again with a uvular stop, and semi-regular metathesis of PAN *CVS to PMP *hVC), *luaq > luaʔ* 'spit out' (Malay *luah*), *ñatuq > ñatuʔ ~ ñatoh* 'trees that yield gutta percha, esp. *Palaquium* sp.' (Malay *ñatoh*), *sepaq > səmpaʔ* 'residue of food after chewing' (Malay *səpah* 'betel quid'), *sawaʔ ~ sawah* (Malay *sawah*) 'irrigated land', *Sapejiq > pədiʔ* 'sharp pain; sting, smart' (cp. Paiwan *sapədiq* 'be tender-footed, feel hurt', Malay *pədeh* 'smart, ache', which can only reflect *-q), or *ma-etaq> mataʔ ~ mantaʔ* 'raw' (cp. Paiwan *matjaq*, Malay *məntah* 'raw' < *-q). It is possible that these and similar cases are borrowed from a Bornean language in which *q > -ʔ, but if so the source is unknown. Third, some

---

[95] Blust (1980b) did not appear until 1983 or 1984, and a prepublication version of Zorc (1982) was available as a conference paper in January 1981.

Iban forms with final glottal stop appear to be Malay loanwords with -*k* ([ʔ]), as with Iban *badiʔ*, Malay *badek* 'dagger', Iban *biraʔ*, Malay *berak* 'defecate', or Iban *pokoʔ ~ pukuʔ* 'chief, principal', Malay *pokok* 'stem or trunk; 'trunk' as principal part'. This still leaves many examples of Iban -*ʔ* unexplained.

Zorc is careful to note that Philippine languages often show secondary glottal stop word-finally in known loanwords, and these cases are distinguished in his table from reflexes in native vocabulary so as to avoid unnecessary confusion. However, some of the critical witnesses in this list also show a historically secondary glottal stop in native vocabulary. This is particularly clear in Iban, where a glottal stop could only have been added after monophthongisation of some diphthongs, as in PMP *antay > *antiʔ* 'lie in wait for', *baliw > *baliʔ* 'change, alter, fade', *baRiw > *bariʔ* 'musty, 'gone off'', *beRay > *bəriʔ* 'dowry; give', *kahiw > *kayuʔ* 'tree', *um-anduy > *mandiʔ* 'bath, bathe', *gaway > *gawaʔ* 'work, business; anything important or serious that has to be done, sacred rite, essential and solemn part of a festival' ~ *gaway* 'religious rites accompanied by festivity, feast, festival', *punay > *punaʔ ~ punay* 'pigeon', or *reñay 'aftermath of a storm' > *rəñiʔ* 'steady drizzle'.[96]   Adelaar (1992:63ff) has drawn attention to weaknesses in Zorc's use of the Iban evidence for *-ʔ. Zorc (1996) acknowledges these problems, yet maintains that cross-linguistic agreements still support his hypothesis.

Since Zorc acknowledges irregularities in the evidence for *ʔ, what is needed to assess the significance of this material is a better understanding of the strength of association between Iban -*ʔ* and final glottal stop in languages of the central Philippines. To establish such an association it is not enough to selectively draw attention to supporting cases, as Zorc has done. Rather, we need a quantitative evaluation of the expected and attested degrees of correlation between final vowels and final glottal stops in disparate witnesses. Limitations of space and time preclude a complete examination of the evidence that Zorc has assembled. However, a careful consideration of relevant data should show whether or not the correspondence of Iban final glottal stop to final glottal stop in languages of the central Philippines significantly exceeds chance. As already noted, final glottal stop in Iban has several known sources, as well as one or more unknown sources. While the known sources provide an explanation for part of the data no similar control exists for final glottal stop from unknown sources. In these cases the only available control is statistical: if a sound correspondence involving the association of zero with glottal stop or of glottal stop with glottal stop deviates from the null hypothesis it can be taken as evidence of a historical connection and hence of the need to reconstruct a new phoneme. If the observed distribution does not deviate from the null hypothesis, on the other hand, no number of attractive comparisons which appear to support a new proto phoneme can be taken as valid evidence for such a reconstruction. A comparison of the Iban material in Richards (1981) and the Tagalog material in Panganiban (1966) shows the following numbers of cognate sets which exemplify each of the four possible correspondences of final vowels or glottal stop between these languages:

1.   Iban –V : Tagalog –V. There are at least 35 instances of this correspondence in words that appears to be directly inherited in both languages, as in Iban *abu* : Tagalog *abó* 'ash', Iban *dara* : Tagalog *da-lága* 'young girl, virgin', or Iban *laki* : Tagalog *la-láke* 'male'. The same correspondence occurs in only three known loanwords: Iban *baca* :

---

[96]   Iban also has several doublets differing in -*ʔ* vs. –*ay* or –*aw*, where no etymology is known, as with *kəlidoʔ ~ kəlidaw* 'cooking tool', *lambiʔ ~ lambay* 'wave to, beckon', *mərundaʔ ~ mərunday* 'hanging down, slack', *puduʔ ~ pudaw* 'fruit tree: *Artocarpus* sp.', or *rəguʔ ~ rəgaw* 'disturbed'.

Tagalog *bása* 'read', Iban *jala* : Tagalog *dála* 'casting net', and Iban *kaya* 'rich' : Tagalog *kaya* 'able, can do'.

2.   Iban -V : Tagalog -Vʔ. This correspondence has a strikingly different profile from 1) in that almost all examples are found in transparent loanwords from Malay: Iban *baju* 'coat, jacket' : Tagalog *bároʔ* 'dress for upper part of body' (Malay *baju* 'shirt', from Persian), Iban *bərita* : Tagalog *balítaʔ* 'news' (Malay *bərita*, from Sanskrit), Iban *bisa* 'strong, powerful, effective' : Tagalog *bísaʔ* 'efficacy, potency, force' (Malay *bisa* 'venomous; able, can'), Iban *budi* 'kindness, generosity, gratitude' : Tagalog *budhiʔ* 'conscience' (Malay *budi* 'kindly acts and ways, character', from Sanskrit), Iban *daya* 'means, way, dodge' : Tagalog *dáyaʔ* 'deceit, fraud' (Malay *daya* 'artifice, dodge'), Iban *guci* 'small rare jar' : Tagalog *gúsiʔ* 'large China vase' (Malay *guci* 'water vessel'), Iban *kuali* : Tagalog *kawáliʔ* 'frying pan' (Malay *kuali* 'wide-mouthed cooking pot'), Iban *kuta* : Tagalog *kútaʔ* 'fort, fortification' (Malay *kota* 'fortified place', from Sanskrit), Iban *laku* 'widely distributed, in demand' : Tagalog *lákoʔ* 'merchandise being peddled around' (Malay *laku* 'having value, selling well'), Iban *pintu* : Tagalog *pintóʔ* 'door' (Malay *pintu* 'door'), Iban *taji* : Tagalog *táriʔ* 'metal cockspur' (Malay *taji* 'metal cockspur'), Iban *tanda* : Tagalog *tandáʔ* 'sign, mark' (Malay *tanda* 'sign'), etc.

Among the few examples of Iban -V : Tagalog -Vʔ that do not appear in transparent Malay loanwords in Tagalog are: 1) Iban *bunsu* : Tagalog *bunsóʔ* 'youngest child', 2) Iban *əmba* 'scare, threaten' : Tagalog *ambáʔ* 'threatening gesture' (no known Malay cognate), 3) Iban *kaka* 'elder sister' : Tagalog *kakáʔ* 'eldest brother, sister or first cousin', 4) Iban *kəlabu* 'grey, ash-coloured' : Tagalog *kulabóʔ* 'faded in colour, somewhat hazy', 5) Iban *muda* 'young, unripe' : Tagalog *múraʔ* 'unripe; immature' and 6) Iban *tali* : Tagalog *táliʔ* 'rope'. However, the first vowel of Tagalog *ambáʔ* is irregular and so suggests borrowing, and Tagalog *kakáʔ* could reflect *kaka with *-q 'vocative'. Moreover, since Tagalog *táliʔ* reflects PMP *talih this and PAN *tebaS > PMP *tebah > Tagalog *tibáʔ* 'cut down vegetation' may indicate that PMP *-h, which usually became zero, is sometimes reflected instead as glottal stop. The other three items may be Malay loans, but even if they are not the number of examples of this correspondence in clearly native vocabulary is very small.

3.   Iban -Vʔ : Tagalog -V. There are at least 31 examples of this correspondence in forms that appear to be native, and because of their relevance to a longstanding issue in phonological reconstruction they are given here in full: 1) Iban *anuʔ* 'some(thing); interjection' : Tagalog *anó* 'what?', 2) Iban *asuʔ* : Tagalog *áso* 'dog' (but Iban *ŋ-asu* 'to hunt (using dogs)', 3) Iban *ayaʔ* 'uncle, aunt; stepfather' : Tagalog *áya* 'caretaker of children', 4) Iban *baraʔ* 'ember' : Tagalog *bága* 'glowing coal', 5) Iban *bidaʔ* 'kick or stumble against' : Tagalog *birá* 'violent stroke', 6) Iban *bukaʔ* : Tagalog *buká* 'open', 7) Iban *bukuʔ* : Tagalog *bukó* 'node, joint', 8) Iban *dəpaʔ* : Tagalog *dipá* 'fathom', 9) Iban *bə-duaʔ* 'divide' : Tagalog *dalawá* 'two' (but Iban *dua* 'two'), 10) Iban *kənuʔ* : Tagalog *kunó* 'it is said; quotative marker', 11) Iban *kətawaʔ* : Tagalog *táwa* 'laughter', 12) Iban *kitaʔ* '2sg.' : Tagalog *kitá* 'we dual', 13) Iban *kituʔ* 'here, hither' : Tagalog *itó* 'this', *d-itó* 'here', 14) Iban *liaʔ* : Tagalog *lúya* 'ginger', 15) Iban *limaʔ* : Tagalog *limá* 'five', 16) Iban *ñiluʔ* 'set teeth on edge' : Tagalog *ŋiló* 'tooth edge pain', 17) Iban *paŋkuʔ* 'hold in the lap' : Tagalog *paŋkó* 'carried in the arms', 18) Iban *paraʔ* 'rack over the hearth; shelf' : Tagalog *pága* 'storage loft of bamboo', 19) Iban *pəriaʔ* : Tagalog *ampalayá* 'bitter melon: *Momordica charantia*', 20) Iban *pukiʔ* : Tagalog *púki* 'vulva', 21) Iban *ruʔ* : Tagalog *agóʔo* 'shore tree: *Casuarina equisetifolia*', 22) Iban *rusaʔ* : Tagalog *usá* 'Sambhur deer', 23) Iban *saʔ* : Tagalog *isá* 'one', 24) Iban *saŋaʔ* 'fork, branch' : Tagalog *saŋá* 'branch', 25) Iban *sawaʔ* : Tagalog *sawá* 'python', 26) Iban *taiʔ* : Tagalog *táʔe* 'feces', 27) Iban *tasaʔ*

'collect nipa leaves for roofing' : Tagalog *sasá* 'nipa palm', 28) Iban *tədaʔ* 'leavings, remnants, left-overs' : Tagalog *tirá* 'left-over, remainder', 29) Iban *tuliʔ* 'deaf; earwax' : Tagalog *tu-tulí* 'earwax', 30) Iban *tumaʔ* 'louse' : Tagalog *túma* 'clothes louse', 31) Iban *tusuʔ* 'to suck' : Tagalog *s<um>úso* 'to suck' (but Iban *tusu* 'breast'). In giving this number I have omitted comparisons such as Iban *bəriʔ* 'dowry; give' : Tagalog *bigáy* 'give', where Iban has -Vʔ corresponding to a Tagalog diphthong.

The same correspondence is known in four loanwords: 1) Iban *asaʔ* 'deceive' : Tagalog *ása* 'hope' (Malay *asa* 'hope', from Sanskrit), 2) Iban *diriʔ* : Tagalog *saríli* 'self, oneself' (Malay *səndiri* 'self, oneself'), 3) Iban *jəramiʔ* : Tagalog *dayámi* (from Kapampangan) 'rice stubble'), 4) Iban *saguʔ* 'pearl sago' : Tagalog *sagó* 'sago palm' (Malay *sagu* 'mealy pith of the sago palm').

4.   Iban -Vʔ : Tagalog -Vʔ . Zorc (1996:48ff) cites 14 cognate sets that are thought to illustrate this correspondence in forms that appear to be directly inherited and are not subject to vocative marking. However, Iban cognates are not available for two of these, and for others Tagalog cognates are not available or show -V corresponding to -Vʔ in other Central Philippine languages. Several others are very likely Malay loanwords, as with Bikol, Cebuano *dátuʔ* 'chief' (a title that is associated with Muslim societies in the southern Philippines), Tagalog *pákoʔ* 'nail', Tagalog *naŋkaʔ* 'jackfruit' (introduced from tropical America), Tagalog *sípaʔ* 'kick' (from Malay *sepak*, and commonly associated with a game in which a rattan ball is kept aloft by kicking)', and Tagalog *sulambí(ʔ)* 'eaves, gable'. Still others result from conflating apparently distinct etymologies, as with Tagalog *támaʔ* 'hit the mark; correct, right' next to Iban *tamaʔ* 'enter, go in, as in entering a room', or simply fail to cite relevant evidence, as with Iban *kənaʔ* : Malay *kəna*, Kanakanabu *suma-kəna*, Tsou *meʔho*, *əha* 'hit the mark', where Iban is the only language that indicates anything other than a final vowel.

A comparison of forms with final glottal stop in Richards (1981) and Panganiban (1966) suggests the following as valid instances of correspondence 4) in forms that are not transparent loanwords: 1) Iban *bulaʔ* 'lying' : Tagalog *búlaʔ* 'fib, untruth', 2) Iban *dampaʔ* 'temporary longhouse' : Tagalog *dampáʔ* 'hut, hovel', 3) Iban *əmpə-lawaʔ* 'spider' : Tagalog *la-láwaʔ* 'spider', 4) Iban *gapaʔ* 'grope one's way' : Tagalog *kapáʔ* 'groping in the dark', 5) Iban *gawaʔ* 'work, business' : Tagalog *gawáʔ* 'work', 6) Iban *pakuʔ* : Tagalog *pakóʔ* 'fern', 7) Iban *paluʔ* 'hit, beat, strike' : Tagalog *páloʔ* 'stroke with hand or stick', 8) Iban *sudaʔ* 'bamboo spike set in ground' : Tagalog *suláʔ* 'empale', 9) Iban *suduʔ* : Tagalog *súroʔ* 'spoon', 10) Iban *təbaʔ* : Tagalog *tibáʔ* 'cut down, clear vegetation', 11) Iban *tikuʔ* 'bend sharply' : Tagalog *tikóʔ* 'curved, bent', 12. Iban *timbaʔ* 'bail' : Tagalog *timbáʔ* 'water pail, bucket', 13. Iban *tuŋkuʔ* 'trivet for cooking pot' : Tagalog *tuŋkóʔ* 'tripod'. Even in this short list there are several problems (*bulaʔ* appears to be confined to Brunei and Sarawak Malay; a priori reflexes of *gawaʔ co-exist with reflexes of *gaway, Tagalog *tuŋkóʔ* seems to be a borrowing which replaced the inherited *dalikan). Nonetheless, if we take the examples of these four sound correspondences as our data base we have 35 + 3 + 31 + 13 = 82 Iban-Tagalog cognate sets. A count of all lexical bases in the first 100 pages of Richards (1981) shows about 570 Iban forms that end either in a vowel or in glottal stop (minor complications are introduced by the use of cross-referencing). Of these forms 334, or 58.6% end with a vowel, and 236, or 41.4% end with glottal stop. A similar count of all lexical bases in the first 100 pages of Panganiban (1966) shows about 455 Tagalog forms that end either in a vowel or in glottal stop (known Spanish loans were excluded, as this would artificially inflate the number of vowel-final bases; transparent loanwords from Malay with final glottal stop have similarly been

discarded). Of these 455 forms 218, or 47.9% end with a vowel, and 237, or 52.1% end with glottal stop. The cross-linguistic product of these relative frequencies produces a set of expected frequencies of association (EF) between final vowels and glottal stop in the two languages. These appear in Table 8.25 together with the attested frequencies of association (AF):

Table 8.25: Expected and attested frequencies of Iban-Tagalog correspondences with final vowel or final glottal stop

Iban		Tagalog	EF	AF
-V (.586)	x	-V (.479)	23/82 = 28.1%	35/82 = 42.7%
-V (.586)	x	-Vʔ (.521)	25/82 = 30.5 %	3/82 = 3.6%
-Vʔ (.414)	x	-V (.479)	16/82 = 19.8%	31/82 = 37.8%
-Vʔ (.414)	x	-Vʔ (.521)	18/82 = 21.6%	13/82 = 15.9%

Table 8.25 provides several types of information. First, -V : -V correspondences exceed EF by over 50%, and so are clearly non-random, a reflection of the fact that they go back to PMP etyma which contained a final vowel. Second, -V : -Vʔ correspondences fall far below the level expected by chance, indicating that Tagalog glottal stop in directly inherited forms almost always reflects a consonant (*q) that did not usually become Iban glottal stop. Third, -Vʔ : -V correspondences occur at nearly twice EF, suggesting that the glottal stop in these Iban forms probably was added to earlier final vowels. Zorc (1996) attributes this correspondence to PMP *-h, as in PAN/PMP *baRah > Iban *baraʔ* 'ember', PAN *CiŋaS > PMP *tiŋah > Iban *tiŋaʔ* 'food stuck between teeth after eating', PAN *paRiS > PMP *paRih > Iban *pariʔ* 'stingray', and PAN *tumeS > PMP *tumah > Iban *tumaʔ* 'louse', but there appear to be roughly as many other cases in which PMP *-h yielded a final vowel in Iban, as with PAN/PMP *baqeRuh > Iban *baru* 'new', PAN *kuSkuS > PMP *kuhkuh > Iban *kuku* 'claw, finger- or toenail', PAN *ma-tuqaS > PMP *ma-tuqah > Iban *əntua* 'parent-in-law', PAN *CaliS > PMP *talih > Iban *tali* 'rope', or PAN *tebuS/CebuS > PMP *tebuh > Iban *təbu* 'sugarcane'. Given these forms it is not clear that final glottal stop in Iban words such as *baraʔ* or *tumaʔ* reflects PMP *-h rather than Ø. Finally, the -Vʔ : -Vʔ correspondence that Zorc takes as evidence for PAN *ʔ occurs with *lower* than chance frequency. The simplest explanation for all these observations is therefore that the final -Vʔ in Iban corresponding to a final vowel in most other AN languages is historically secondary.

This conclusion is consistent with the inescapable inference that the glottal stop in Iban forms such as *antiʔ* 'lie in wait for' (PMP *antay), *baliʔ* 'change, alter, fade' (PMP *baliw), or *kayuʔ* 'tree' (PMP*kahiw) must also be secondary. It is, moreover, consistent with variation between final glottal stop and zero within Iban morphological paradigms, as with *asuʔ* 'dog' : *ŋ-asu* 'hunt (using dogs)', or *dua* 'two' : *bə-duaʔ* 'divide'. It does not, however, explain why Iban developed a final glottal stop after some vowels but not others. As with other apparently unconditioned phonemic splits in AN languages, the appearance of -Vʔ for expected -V in Iban cannot support *-ʔ without corroboratory evidence from independent witnesses. Zorc has attempted to present such evidence, but a systematic comparison of cognate forms that end with a glottal stop in Tagalog or Iban does not support his conclusions. If there is no statistical support that Iban -ʔ : Tagalog -ʔ reflects *-ʔ, it is reasonable to assume that the same conclusion will hold in comparing Iban with other languages in the Philippines. Since Zorc's evidence for *ʔ is more strongly supported

in final position than word-medially the reconstruction of *ʔ as a phoneme in any position thus appears to rest on insecure foundations.

Iban is not alone among Malayic languages and their close relatives in having a phonemic final glottal stop of obscure origin. Other languages with an apparently intrusive glottal stop are Sarawak Malay and Sasak. Table 8.26 gives an overview of comparisons that contain a final glottal stop of obscure origin in one or another of these languages (the material available for Sarawak Malay is very limited):

Table 8.26 Examples of unexplained -ʔ in Iban, Sarawak Malay and/or Sasak

	Iban	**Sarawak Malay**	**Sasak**	
1.	adiʔ	adiʔ	adiʔ	younger sibling
2.	ŋ-akuʔ		ŋ-akuʔ	claim
3.	asuʔ	asuʔ	asu/asuʔ	dog
4.	auʔ		aoʔ	yes
5.	bapaʔ	bapaʔ	bapa	father
6.	bəriʔ	bəriʔ		give
7.	bəkaʔ		bukaʔ	open
8.	bukuʔ		buku	joint
9.	dua/	duaʔ	dua/	two
	bə-duaʔ 'divide'		pə-duaʔ	divide
10.	isiʔ	isi	isi	contents
11.	dituʔ	situʔ	ito	here; there
12.		juaʔ	juaʔ	too, also
13.	kakaʔ	kakaʔ	kakaʔ	older sibling
14.	kali	kaliʔ	kali	dig
15.	kami	kameʔ	kami	1pl incl.
16.	kitaʔ	kitaʔ	kita	2pl; 1pl incl.
17.	labuʔ	labuʔ		gourd
18.	agiʔ/lagiʔ		lagi	more
19.	layuʔ	layuʔ	layu	wither
20.	limaʔ	limaʔ	lima	five
21.		kereʔ	kiri	left side
22.	lupa	lupaʔ	lupaʔ	forget
23.	mandiʔ	mandiʔ	mandiʔ	bathe
24.	mudaʔ	mudaʔ		young
25.	palaʔ	palaʔ	kəpala	head (L)
26.	paluʔ	paluʔ	palu	hit; hammer
27.	sidaʔ	sidaʔ	sida	3pl; 2pl.
28.	taliʔ	tali	tali/	rope
			taliʔ	to bind
29.	tanda	tandaʔ	tanda	sign
30.	tauʔ	tauʔ	tao	know

The list in Table 8.26 could be extended considerably, but there would be little point in doing so. Iban and Sarawak Malay are closely related, and the high degree of agreement that they show in the appearance of unexpected final glottal stops is best explained as a product of inheritance from an immediate common ancestor. However, even between these closely related languages there are disagreements in the appearance of final glottal stop, as

seen in items 9, 10, 14, 15, 22, 28 and 29, and this is still more marked in comparing either Iban or Sarawak Malay with Sasak. Moreover, in some cases where all three languages agree the form in question did not originally end in a vowel, as with PMP *wahiR> Iban, SM, Sasak *aiʔ* 'water', or *um-anduy > Iban, SM, Sasak *mandiʔ* 'bathe'. Like Iban, Sasak shows a number of -V/Vʔ doublets which suggest that glottal stop may have been introduced as part of some morphological process that has survived in only fragmentary form: Sasak *baru* 'new', but *baruʔ* 'awhile ago', *batu* 'stone', but *batuʔ* 'whetstone', *dua* 'two' : *pə-duaʔ* 'divide', *tali* 'rope' : *taliʔ* 'to bind', etc. Taken as a whole, then, the evidence does not support the reconstruction of PAN *ʔ in final position, or by implication, in any position.

8.2.3 Voiced stops

Dempwolff reconstructed six voiced stops: *b, *d, *D, *z, *j and *g. All of these except *j have been questioned in one way or another. The major issues that have arisen in the comparative literature are summarised below.

8.2.3.1 *Were there two *b phonemes?*

Given his implicit appeal to an 'independent evidence' requirement Dempwolff allowed unconditioned phonemic splits in individual languages. Prominent among these was the split of *b into *b*/*w* and of *d and *ḍ into *d*/*ḍ*/*r* in Javanese. So long as no other language showed a similar split which was correlated with that of Javanese this irregularity did not threaten his reconstruction of 'Uraustronesisch'. But the great increase in descriptive work on the AN languages following Dempwolff's death led in time to new data that challenged this interpretation.

Drawing on new data from the Dusunic and Murutic languages of Sabah (called 'Idahan'), Prentice (1974) argued that Dempwolff's *b represents two phonemes $*b_1$ and $*b_2$, the former reflected as *b* and the latter as *w* in Idahan and Javanese. Both $*b_1$ and $*b_2$ presumably were bilabial stops, as they are distinct from *w, and structural and phonetic arguments weigh against interpreting $*b_2$ as a fricative. Although Prentice is silent on the assumed phonetics of his contrast, this would almost force the interpretation that one of these stops was implosive or otherwise 'complex'. This interpretation is supported by evidence that the more lenis reflex often corresponds to Proto North Sarawak *b, and the more fortis reflex to $*b^h$. The correspondences within Borneo can thus generally be explained by assuming that the immediate common ancestor of the Sabahan and North Sarawak languages geminated consonants following schwa, and that geminate voiced obstruents underwent terminal devoicing, giving rise to a $*b : *b^h$ contrast that is well-preserved in the North Sarawak languages, and is reflected in a generalised form in Ida'an Begak of eastern Sabah (Blust 1974b, 2006a, Goudswaard 2005). In the Dusunic and Murutic languages $*b^h$ and *b have lenited to *b* and *v*/*w* respectively.

The real challenge that Prentice poses is how to explain the apparently high degree of correlation between Idahan *b*/*w* and Javanese *b*/*w* without projecting a distinction between two types of voiced bilabial stops onto a fairly early proto language. He presents 17 comparisons that support Idahan -*b*- : Javanese -*b*-. However, 14 of these contain penultimate schwa, and at least two of the remaining three (*bibi 'duck', *labu 'gourd') are likely loanwords. In addition he presents 14 examples of Idahan -*w*- : Javanese -*w*-, none of which follow schwa. Eleven other examples of labial stop/glide correspondences in medial position are contradictory, one witness indicating $*b_1$ and the other $*b_2$, or the

Idahan languages showing an invariant reflex and Javanese showing variation between *b* and *w*. The evidence for a $*b_1$: $*b_2$ distinction in medial position is thus most simply explained by vocalic conditioning: as in many other AN languages, gemination of consonants after schwa produced segments that were more resistant to lenition than their simplex counterparts.

In initial position conditioning cannot provide an alternative explanation. Here Prentice found 15 examples of Idahan *b-* : Javanese *b-*, and 20 examples of Idahan *w-* : Javanese *w*. However, he also found 22 examples in which these correspondences are violated in one way or another. He dismissed nine forms with initial labials (1974:51), four of which support his case and five of which do not. This leaves 31 supporting cases and 17 conflicting ones for initial $*b_1$ or $*b_2$. Given the absence of known evidence for a similar distinction elsewhere in the AN language family, and the possibility that several examples of Idahan *b* : Javanese *b* are products of borrowing from Malay, the argument for a $*b_1$: $*b_2$ distinction is unconvincing in its present form.

8.2.3.2 Was there a *d/D distinction?

The weakest part of Dempwolff's UAN sound system was the set of retroflex stops. His *T was supported only by Javanese and languages such as Madurese and Balinese which have borrowed heavily from Javanese. The situation with *D, however, is more complex. In final position several languages in western Indonesia reflect *d as *-t*, but *D as *-r*, and this appears to be rather consistent. In non-final position, the evidence is much more problematic. Dempwolff believed he had found statistical evidence that *d became Javanese *d-*, *-d-* : Tagalog *d-*, *-r-*, but that *D became Javanese *ḍ-*, *-ḍ-* : Tagalog *l-*, *-l-*. As Dahl (1976:55) has shown, however, there are many exceptions to these correspondences, since Javanese not uncommonly has variants differing in *d ~ ḍ*, or in one of these stops and *r*. Given this unsatisfactory situation Dahl (1976:66ff), following an earlier suggestion by Haudricourt, proposed that the Javanese evidence for a dental/retroflex distinction be attributed to contact influence from Sanskrit. As already noted in connection with *T, there are problems with this explanation. However, there can be no denying that the Javanese evidence for a *d/D distinction is contradictory, and Dempwolff consequently relied more heavily on Tagalog as a critical witness. Zorc (1987) has effectively demolished the argument that Tagalog (or any other Central Philippine language) provides evidence for a *d/D distinction, and with the elimination of this part of the comparisons the entire argument collapses.[97]

While there is no clear evidence for a *d/D distinction in non-final position, the evidence for *-D involves a separate set of correspondences. Javanese does not permit final retroflex stops, and Tagalog, like other Philippine languages, reflects word-final *d and *D identically. Dempwolff's *-d : *-D distinction was therefore based on the correspondence Tagalog *-d* : Malay, Toba Batak *-t*, as against Tagalog *-d* : Malay, Toba Batak *-r*. Although some of these are doublets, at least 26 comparisons are now known to support *-D: 1) *bayaD 'price; pay', 2) *be(ŋ)kaD 'unfold, as a blossom', 3) *bitaD 'spread out', 4) *bu(n)tuD 'bloated, as with gas', 5) *da(m)paD 'flat, level', 6) *gaDgaD 'fall apart', 7) *hantaD 'visible, exposed', 8) *hateD 'escort, accompany', 9) *i(ŋ)suD 'inch away, budge', 10) *kiDkiD 'scrape', 11) *kiliD 'sharpen an edge', 12) *laqaD 'dry streambed',

---

[97] Mahdi (1996) has suggested that agreements between Old Javanese, Balinese, and Madurese on the one hand, with Paiwan and Puyuma on the other, provides support for the *d/D distinction that is at least as strong as support for the *t/C distinction, but the data he presents for this argument can be interpreted in other ways.

13) *le(m)paD 'soar, fly (as a missile)', 14) li(ŋ)keD 'coil', 15) *nataD 'cleared area around house', 16) *puseD 'whorl, eddy', 17) *qa(m)buD 'strew, scatter, as grains', 18) *sabeD 'obstacle', 19) *sabuD 'strew, scatter, as grains', 20) *saluD 'funnel or channel water', 21) *sapaD 'flattened', 22) *siDsiD 'sail along a coastline', 23) *sikaD 'energetic, active', 24) *tabuD 'strew, scatter, as grains', 25) *tawaD 'haggle, bargain', and 26) *tuaD 'kind of fish trap'. In reflexes of these forms Iban, Malay, the Batak languages as a group, Balinese, and possibly a few other languages in western Indonesia consistently reflect *D as -r, but *-d as -t.

Some of Dempwolff's reconstructions with *-D have proved to be erroneous, as with his *buDbuD 'chop up; porridge', which must be corrected to *buRbuR 'porridge', *cemeD 'impure', which must be dismissed as ill-conceived, or *luluD 'shin', which must be corrected to *luluj. Despite these reductions in supporting evidence the 26 etyma given above are generally well-supported, and there can be no doubt about the reality of the sound correspondence in question. The principal problem with positing *-D is that it is distinguished from *-d (and *-j) by only a handful of languages in western Indonesia, some of which (Iban, Malay, Balinese) may belong to a subgroup with a time-depth not much in excess of 2,500 years. There are, however, strong counter-arguments to this objection. First, the Batak languages of northern Sumatra, which are not closely related to Malay, also distinguish *-d, from *-D (and *-j). Although these languages have been exposed to Malay for centuries, contact apparently was relatively light in Toba Batak until the twentieth century, and in most cases where Toba Bataj shows *-D > *r* the word appears to be native. Second, if the 26 forms cited above are reconstructed with *-d we must conclude that *-d usually became -t, but sometimes -r in the Batak and Malayic languages, and in Balinese. Not only would this be an unconditioned phonemic split, but it is one in which the two outcomes are phonetically very distinct (final voiced stops normally devoiced in Iban, Malay and Toba Batak, although not in northern Batak, where they merged with the homorganic nasals). Third, where a monosyllabic root can be identified it is consistent in reflecting *-D as -r, as with Iban *papar* 'cut smooth (with adze, etc.)', Malay *papar* 'flat, smooth, blunt', next to Tagalog *sapád* 'flat, of irregular flatness (said of things that are expected to be more rounded)', Iban *sapar* 'flat side, of a post, etc.; square (not round)' < *sapaD), or Malay *aŋsur* 'a little at a time, in action or motion', next to Tagalog *ísod* 'act of moving something up or away in position', Iban *insur* 'adjust, move something' < *i(ŋ)suD. The problems with trying to eliminate *-D are thus similar to those for *c. The principal difference is that *c fits a structural gap in the palatal series, while *-D does not appear to pair with any other proto phoneme, and is found only in final position.

It is impossible not to feel uncomfortable with the reconstruction of a single retroflex consonant that occurs only in final position. One alternative is that PAN *-D was *-d, and apparent instances of a contrastive *-d are actually *-j. This often works where the languages compared do not distinguish *-d and *-j, and even sometimes where they do, as with Dempwolff's *bukid 'forested mountain areas', but Ibanag *vukig* 'mountain', Itawis *hukíg* 'thickly forested mountain', Pangasinan *bokíg* 'east; eastern section of town or province', which indicate *bukij, and force a reconsideration of Toba Batak *buhit* 'high, splendid, majestic', *buŋkit* 'high' (expected **buhik or **buŋkik) either as a loan or an unrelated form. In other cases, however, this hypothesis breaks down, as in *qañud > Thao *qazus*, Ilokano *ánud*, Tboli *konul*, Malay *hañut* 'drift with a current, be carried off by water', where Thao -s, Ilokano -d cannot be reconciled with *j (> Thao z, Ilokano g), and Malay -t cannot be reconciled with *-D (> Malay –r), or in *kukud > Ilokano *kúkod* 'shank, shin of animals', Kelabit *kukud* 'foot, leg', Achenese *kukuët* 'lower leg of an

animal, from the hoof to the ankle', where again Ilokano *-d* cannot regularly reflect *-j, and Achenese *-t* cannot be reconciled with *-D (> Ø).

8.2.3.3   *How many types of *d?*

Following an earlier proposal by Ogawa and Asai, Dahl (1976:58ff) reconstructed $*d_1$, $*d_2$, and $*d_3$. This distinction is based almost exclusively on data from Paiwan and Puyuma of southeast Taiwan, two neighbouring languages that have been in a borrowing relationship for many centuries. Although Dahl suggested that Paiwan and Puyuma are mutually supportive in distinguishing these three types of *d, the evidence is in fact contradictory (Blust 1999b:49ff). Moreover, the $*d_1$-d_3 distinction crosscuts Dempwolff's *d/D, so that if both sets of distinctions are accepted the number of reconstructed *d phonemes must be greater than three. All-in-all, then, the puzzling apical stop correspondences in Paiwan and Puyuma are probably best explained as products of a long and complex history of borrowing.

8.2.3.4   *Was there a *z/Z distinction?*

Following a proposal first made by van der Tuuk, Dyen (1951) split Dempwolff's *z into *z and *Z. The evidence for *Z was a mutually corroboratory irregularity in Toba Batak and Javanese, both of which normally have a voiced palatal affricate corresponding to Malay *j*, but sometimes show *d* instead. Dahl (1976:82) questioned many of Dyen's comparisons, but accepted his argument.

Dyen (1951) reconstructed *Z in eight forms: 1) *Zalan 'path, road', 2) *Za(hØ)uq 'far', 3) *ZeRami(hØ) 'rice straw', 4) *Zilat 'lick', 5) *ZuRuq 'sap, gravy', 6) *peZem 'close the eyes', 7) *eZen 'press, squeeze out', 8) *quZan 'rain'. For all of these except *peZem (Toba Batak *podom*, Malay *pəjam*, Javanese *mərəm*) one or more Batak languages shows *d*, and Javanese shows *d* or *r* for expected *j*. However, there are also examples in which Toba Batak and Javanese disagree: 1) Karo Batak, Toba Batak *jarum*, Old and modern Javanese *dom* 'needle', 2) Toba Batak *injak* 'set foot on, step on', Javanese *idak* 'get stepped on', 3) Karo Batak, Toba Batak *jadi*, Old and modern Javanese *dadi* 'be, become', 4) Toba Batak *ijur* 'spittle, saliva', Old Javanese *hidu/idu* 'spittle, saliva', modern Javanese *idu* 'saliva; salivate'. In these cases Batak evidence points to *z, and Javanese to *Z. This correspondence may thus have arisen through an incipient depalatalisation process that has gone further in Javanese than in the Batak languages. Depalatalisation may have begun with the most commonly used lexical items, leading to an apparently significant cross-linguistic agreement in the morphemes affected which in reality is driven by shared patterns of word frequency.

The tendency for *z to irregularly merge with *d is in fact more widespread than van der Tuuk or Dempwolff apparently realised. Although Malay *padam* 'extinguished, put out' unambiguously supports Dempwolff's *padem 'to extinguish', for example, Mukah Melanau has *pajəm* 'close the eyes; extinguish a fire', suggesting that even Malay sometimes reflects *z as *d*.[98] Similar problems arise in the North Sarawak languages, as in Kiput, which usually reflects Proto North Sarawak *j as a palatal affricate (8 known cases), but sometimes reflects it instead as *d-* (4 known cases), or more rarely as *s-* (Blust 2002c:418). Moreover, both Javanese and the Batak languages have *d/j* doublets, as in Javanese *pidak/pinjak* 'pedal; step on' (Malay *pijak* 'set foot upon'), or Karo Batak

---

[98]   Probably in the same category is Dempwolff's *dilaq 'tongue', supported by Malay *lidah* (met.) 'tongue', while many of the languages of central and western Borneo instead reflect *zelaq.

dalan/dalin/jalan 'path, road', *pidəm/pijəm* 'hollow tree trunk used as a beehive'. Again, an initially attractive hypothesis that some of Dempwolff's symbols are cover terms for multiple proto phonemes turns out on closer examination to be problematic, and it seems best not to distinguish *z and *Z. Since *Z appears in the two most stable comparisons that contain a voiced palatal affricate (Dyen's *Zalan 'path, road', and *quZan 'rain') some scholars who have abandoned the *z/Z distinction use *Z to represent both. Once the distinction is abandoned, however, there is no reason not to write *z.

8.2.3.5 *The phonetic value of *j*

The phonetic value of some proto phonemes is hardly controversial, as with *m, *n, or *ŋ, which in nonfinal position show changes in relatively few languages. In other cases the phonetic value of a proto phoneme can be determined with little controversy even though there has been extensive change, as with *p, *t or *k which often undergo modification, but normally along fairly well-defined paths of lenition. In a few cases, however, a well-established proto phoneme continues to resist phonetic interpretation. This is the case with PAN *j and its reflexes in many lower-order proto languages.

While virtually everyone accepts *j as a distinct phoneme, the phonetic value of this segment is problematic due to great variation in the content of its reflexes. These include at least the following developments:

Table 8.27 Reflexes of PAN *j

c ([ʧ])	=	Seediq (-c only)
d	=	Paiwan, Puyuma, Pazeh (-d only), Tagalog (-d only), Kelabit, Proto Sangiric, Proto Minahasan (-d only)
g	=	Atayal (sporadic), Rukai, Ilokano, Toba Batak, Rejang (-g- in all dialects)
j	=	Bare'e/Pamona (probably through earlier *y*)
k	=	Palauan (/_C only), Toba Batak (word-final only)
l	=	Kanakanabu, Tagalog (-l- only), Botolan Sambal
ɬ	=	Saaroa (-ɬ- only)
n	=	Amis, Kavalan, Siraya
ŋ	=	Karo Batak (word-final only)
ʔ	=	Chamorro
r	=	Squliq Atayal, Proto Minahasan (-r- only), Tetun, Proto Admiralty
s	=	Atayal (sporadic), Maloh, Buginese, Palauan (-s only), Proto Oceanic (minus the Admiralties)
t	=	Malay, Rejang, Rawas dialect (-t only)
x	=	Nias
y	=	Seediq (-y- only)
z ([ð])	=	Saisiyat, Thao
z ([dz])	=	Pazeh (-z- only)
Ø	=	Bunun, Tsou, Pamona, Proto Bungku-Tolaki (probably through earlier *y*), Muna (probably through earlier *y*)

Unless stated otherwise languages with general final devoicing reflect an intervocalic voiced obstruent as the corresponding voiceless obstruent word-finally (Malay *j > -*t*, Toba Batak *j > -*k*, etc.). The challenge of determining a phonetic value for *j should be

obvious from this sample of reflexes, which covers the entire range of manner features, and the entire range of place features apart from labial and uvular.

The majority of reflexes suggest an obstruent articulation, and probably a stop, since in known changes the lenition of stops to fricatives is much more common than the fortition of fricatives to stops. Moreover, when not subject to general devoicing in word-final position reflexes of *j are almost invariably voiced. We can conclude, then, that *j probably was a voiced stop. But what was its place of articulation? There is no question that *j was distinct from both *d and *g, yet it has recurrently merged with each of these stops. If *j was alveolar it is difficult to imagine any secondary feature of articulation that could have caused it to be backed to velar position again and again (Atayal, Rukai, Ilocano and other languages of northern Luzon, the Batak languages of northern Sumatra, all dialects of Rejang in southern Sumatra except Rawas, and Palauan). If *j was velar, on the other hand, it could have been fronted by a secondary palatalisation, hence [g^y]. This hypothesis has the additional advantage of providing a simple explanation of palatal reflexes such as -*y*- or -*j*- in some languages. Our best guess, then, would appear to be that *j was a palatalised voiced velar stop which formed an 'island' in the phonological system in that there was no matching voiceless stop or nasal.

8.2.3.6   *Dempwolff's *g*

Wolff (1982, 1997, 2003) has questioned the evidence for Dempwoff's *g, although the details of his position vary from one publication to the next. Wolff (1982) eliminates all but 15 Dempwolff reconstructions with *g on the grounds that the forms on which they are based are non-cognate, or are products of borrowing from Malay into other languages of western Indonesia and the Philippines. With regard to the 15 remaining forms he then concludes (1982:13) that 'since none of these forms show cognates in Oceania or Formosa, I believe that they are likely to be related by borrowing, and I do not consider them sufficient evidence to reconstruct *g.' Wolff (1997:581), on the other hand, states that 'I reconstruct with *-g- and *-g the forms which are traditionally reconstructed with *-j- and *-j (Dempwolff's medial and final *g'). I reconstruct initial *g- in a few of the forms which have traditionally been reconstructed in *g- (although most of them are not inherited, as documented in Wolff 1982).' In this publication, then, he evidently retreats from his earlier decision to eliminate *g- entirely. Wolff (2003:2) substitutes *g for the traditional *j, and eliminates the traditional *g once more.

Given the range of reflexes for *j, the hypothesis that this proto phoneme was a voiced velar stop [g] must be considered highly unlikely. It is doubtful whether there are attested instances in any language family in which a well-established *g is regularly reflected as *d*, *ð*, *ł*, or *s* to name only a few of the more problematic outcomes of this hypothesised development. This, however, is not the only problem with Wolff's proposed revisions of the PAN consonant system. His claim that plausible Dempwolff reconstructions with *g are not reflected in Oceanic languages is contrary to well-known evidence, although these are indistinguishable from forms with *k (Blust 1977b). But a refusal to recognise PAN *g because Oceanic languages do not distinguish *g and *k is pointless: the merger of PAN *b/p and *g/k is one of the defining features of the Oceanic subgroup, and one could as easily deny the evidence for *b on similar grounds. This still leaves the claim that distinct reflexes of *g are unknown in Formosan languages, but this is false, as both Paiwan and Puyuma reflect *g differently from all other segments (*g > Paiwan *g*, Puyuma *h*, *k > Paiwan *k*, Puyuma *k*):

1. PAN *gaCel 'itch' : Paiwan *gatsəl*, Puyuma (Tamalakaw) *hatər*, Ilokano *gatél*, Bontok *gatəl*, Kapampangan *gatál*, Hanunóo *gatúl*, Kelabit *gatəl*, Kiput, Long Anap Kenyah *gatən*, Iban, Malay *gatal*, Karo Batak *gatəl*, Old and modern Javanese, Balinese *gatəl* 'itch, itchy'.

2. PAN : *gayaŋ 'hunting spear' : Paiwan *gayaŋ* 'hunting spear with detachable harpoon-like barbed iron point', Ilokano *gayáŋ* 'pointed weapon; spear, lance, arrow', Casiguran Dumagat *gayáŋ* 'spear, lance; throw a spear (referring esp. to an Ilongot spear; the Dumagats themselves do not use spears)', Bontok *gayáŋ* 'kind of spear, having upward curving projections from the blade', Maranao *gayaŋ* 'weapon with a blade; trowel-like flat-bladed tool', Tausug *gayaŋ* 'a bladed weapon (similar to a bolo, having a long blade)', Kadazan *gazaŋ* 'a long parang, about two and one half feet in length', Iban *gayaŋ* 'pierce, stab, as by thrusting a spear or knife into the neck of a pig as a ritual act', Bolaang Mongondow *gayaŋ* 'kind of sword', Tae' *gaaŋ* 'golden kris'; *gayaŋ* 'stab with a kris', Makasarese *gayaŋ* 'kind of weapon, long, pointed and sharp on both sides', Buginese *gajaŋ* 'kris'; *pa-gajaŋ* 'stab'.

3. PAN *gemgem : Paiwan *gəmgəm*, Ilokano *gemgém* 'fist', Bontok *gəmgəm* 'hold in one's closed hand; a fistful', Iban *gəŋgam* 'handful, fist; grasp', Malay *gəŋgam* 'grasp, grip; holding in the closed hand or claw', Karo Batak *gəmgəm* 'have something under one's control; guide, direct', Sundanese *gəŋgəm* 'grip; handful', Balinese *gəmgəm* 'squeeze between two things', Makasarese *gaŋgaŋ* 'hold tightly in the fist'.

4. PAN *gerger 'shake, shiver, tremble': Paiwan *i-gərgər* 'tremble', Ilokano *pi-gergér* 'tremble, shake, shiver, shudder', Malay *gəgar* 'vibration, quivering, shuddering', Manggarai *gəgər* 'shiver with chills, tremble', Buruese *gege* 'tremble'.

5. PAN *geriC 'sound of ripping, screaming, etc.' : Paiwan *pa-gərits* 'scream in fright', *g<alʸ>ərits* 'sound of ripping cloth or paper', Iban *gərit* 'gnaw', *gərit-gərit* 'noise of gnawing', Malay *gərit* 'scraping sound; gnawing by a mouse or rat', Sundanese *gərit* 'squeaking of the wheels of a *padati* (cart for transporting goods), or the creaking of a door or gate on its hinges', Balinese *gərit* 'scour, scrub', Manggarai *gərit* 'scratch, claw; scream'.

6. PAN *gisgis/gisagis 'rub, scrape against' : Paiwan *gisagis* 'something rubbed against; tree used by wild pigs for rubbing themselves',[99] Puyuma *gisgis* 'shave; graze, brush against', Casiguran Dumagat *gisagis* 'to scratch (as for a pig or carabao to rub his side back and forth against a tree)', Bikol *gisgís*, *giságis* 'rub (as against a wall, tree) to relieve an itch', Malay (Brunei) *gigis* 'scratch, make a mark with a nail or marking gauge', Balinese *gisgis* 'scratch with the nails'.

7. PAN *guSam 'a throat infection: thrush' : Puyuma *huwam* 'suffering from thrush', Tagalog *guham* 'cutaneous eruption', Malay *guam* 'thrush, a disease of the sprue type attacking children', Javanese *gom* 'scurvy-like disease of the mouth resulting from a vitamin C deficiency', Balinese *guwam* 'a disease of the mouth in children, quinsy'.

8. PAN *guCguC 'pull out, uproot' : Paiwan *gutsguts* 'weed a paddy field', Iban *gugut* 'pull out violently'.

---

[99]   Paiwan normally shows *s > *t*, but not in this form.

9.　PAN *gutgut 'nibble with front teeth' : Puyuma (Tamalakaw) *huthut* 'nibble with the front teeth', Kelabit *gugut* 'bottom incisor teeth', Karo Batak *gugut* 'nibble or bite off with the front teeth', Sundanese *gugut* 'bite', Old Javanese *gugut* 'bite, nip', Balinese *gutgut* 'bite (men and animals); snap; nibble; gnaw'.

10.　PAN *tageRaŋ 'ribs' : Puyuma (Tamalakaw) *tahRaŋ* 'chest, breast', Ifugaw *taglán* 'ribs; side of a human chest (what is enclosed by the ribs)', Pangasinan *taglán* 'rib, frame', Botolan Sambal, Kapampangan *tagyán*, Miri *tagreŋ* , Long Anap Kenyah *təgaaŋ* 'ribs'.

In addition, although Oceanic languages have merged *g and *k, some Central Malayo-Polynesian languages preserve the distinction, and these languages are part of a Central-Eastern Malayo-Polynesian subgroup which includes the Oceanic languages but not the languages of Taiwan, the Philippines or western Indonesia. A few examples have already been given, as with Manggarai *gəgər, gərit,* Buruese *gege,* and these can be extended with comparisons such as the following:

11.　PMP *gagar 'bold' : Minangkabau *gagar* 'brave, plucky', Manggarai *gagar* 'to like, have an appetite for (fighting, talk, sex), brave, bold, lusty, desirous'.

12.　PMP *gemi 'hold on by biting' : Malay *ikan gəmi,* Makasarese *gammi* 'sucker fish: *Echineis naucrates'*, Sika *gəmi* 'pinch, shut, close (as the mouth)'.

13.　PMP *pa(ŋ)gal 'neck shackle on domestic animal' : Ilokano *paŋgál* 'stick or bamboo transversely tied to the head or neck of an animal … to keep the wearer from passing through bamboo fences', Manggarai *pagal* 'hobble; heavy block hung from buffalo's neck to impede his movements'.

14.　PMP *pager 'palisade' : Ilokano *pagér* 'palisade', Miri *fager* 'fence', Iban *pagar* 'fence, railing', *pagar ruyon* 'stockaded fort', Jarai *pəga* 'fence, palisade', Old Javanese *pagər* 'hedge, fence, enclosure', Balinese (Low register) *pagəh* 'be fenced'; *pagəh-an* 'hedge, fence, wall, fortification', Buginese *pagəʔ* 'fence, wall (as a stone wall around a planted field)', Buruese *pager* 'fence'.

Wolff has questioned the validity of reconstructions with *g on the grounds that some reflexes are irregular. Voicing crossover in the velar stops (*g > *k* and *k > *g*) is found in a number of AN languages (Blust 1996d). Since the potential duration of voicing is shorter for velar stops than prevelar stops the actual duration of voicing is likely to be shorter, hence giving weakened cues to the listener to distinguish *g* from *k*. In AN languages sporadic voicing crossover is about equally common in both directions, but for any well-attested lexical item the preponderance of evidence almost always favors one or the other velar stop. In the cases cited here the evidence overwhelmingly supports the reconstruction of *g, and as a result there appears to be no responsible alternative to the traditional analysis of *g as a voiced velar stop and *j as a different voiced obstruent.

8.2.4 Nasals

In general the nasals are non-controversial. The basic triad *m, *n, *ŋ is universally accepted. However, there has been some controversy with regard to *ñ and *N. The core evidence for *ñ was presented by Dempwolff (1934-1938). Relatively few languages retain the *n/ñ distinction, but these are widely scattered and the evidence from them is quite clear. Witnesses for *ñ include Kapampangan in central Luzon, many of the languages of Borneo south of Sabah (Miri, Kiput, Berawan, Bintulu, the Melanau dialect complex, Ukit,

Bekatan, Bidayuh, Maloh, Kayan, Kenyah, Iban, Ngaju Dayak, Ma'anyan), Rejang, Lampung and all Malayic languages of Sumatra, Malay, Moken, all Chamic languages of mainland Southeast Asia, Sundanese, Javanese, Madurese, Balinese, Sasak, most of the Tomini-Tolitoli languages of northwest Sulawesi, several of the Kaili-Pamona languages of central Sulawesi, Makasarese, Buginese, Mandar, Duri and other South Sulawesi languages, Hawu in the Lesser Sundas, Chamorro in western Micronesia, and a scattering of Oceanic languages including most languages of the Admiralty Islands, Wogeo and Kairiru on the north coast of New Guinea, Bugotu and other languages of the western and central Solomons, some Nuclear Micronesian languages and Western Fijian. All of these languages distinguish *ñ from *n, generally as palatal and alveolar nasals respectively, but sometimes in other ways (as *y* vs. *n*). Examples include Kapampangan *yamuk*, Malay *ñamuk*, Chamorro *ñamu*, Bugotu *ñamu* < PMP *ñamuk 'mosquito' (cp. Kapampangan *tanám*, Malay *tanam*, Chamorro *tanom* < PMP *tanem 'plant, bury', or Malay *tanah*, Chamorro *tanoʔ*, Bugotu *tano* < PMP *taneq 'earth, soil'), and Malay *pəñu*, Bugotu *voñu* < PMP *peñu 'green turtle' (cp. Malay *pənuh*, Bugotu *vonu* < PMP *penuq 'full'). In addition, a few languages such as Ivatan and Itabayaten, in the Batanes Islands north of Luzon, have developed a palatal nasal from *n adjacent to a high front vowel, but do not distinguish *n from *ñ.

This distinction, which is clearly reflected outside Taiwan, is obscured in Formosan languages by a cross-cutting distinction, the contrast of *n and *N. While *n usually is reflected as *n*, *N and *ñ are generally distinguished from it, and have merged in all Formosan languages for which data is available with the possible exception of Tsou and Kanakanabu.[100] Most reflexes of *N/ñ are alveolar or palatal, but they are otherwise very diverse (where not indicated the reflex of *ñ is unknown): 1) *N/ñ > *ð* (written *z*): Thao, 2) *N > *t*: Taokas, 3) *N > *n* (final position only): Pazeh, 4) *N > *n* (reflex of *ñ unknown): Basay, Trobiawan 5) *N/ñ > *n*: Bunun, Kavalan, 6) *N > *n* but *ñ > *ŋ*: Kanakanabu?, 7) *N > *s*: Hoanya, Babuza, 8) *N > *l*: Pazeh (medial position only), 9) *N > *l* (reflex of *ñ unknown): Papora, Rukai, 10) *N/ñ > *l*: Atayal, Seediq, Saisiyat, Siraya, Puyuma, 11) *N/ñ > *ɬ*: Saaroa, Amis, 12) *N > *lʸ* : Paiwan, 13) *N > *r* : Kulon, 14) *N > *k/f,s_*, *h* elsewhere, but * ñ > *n*: Tsou. Outside Taiwan medial and final *n and *N have merged in all languages. In initial position *N may have merged with *l rather than *n, as rejection of this hypothesis entails the acceptance of a number of doublets in non-Formosan languages that differ in *l- vs. *n- (*laŋuy/naŋuy 'swim', *lesuŋ/nesuŋ 'rice mortar', *luka/nuka 'wound', etc.). There is, in addition, one doublet pair that differs in *ñ vs. *l: *ñamuk/lamuk 'mosquito'.

Partly for this reason Wolff (1993) has argued that *N and *ñ were one phoneme (written *ñ), which shows phonemic splitting conditioned by contrastive stress. As with his attempt to eliminate *C, this proposal appeals to an AN equivalent of Verner's Law: whereas the placement of the accent conditioned voicing of intervocalic fricatives in the Germanic languages, Wolff proposes that the placement of the accent conditioned depalatalisation in AN languages. The basis for this argument is understandable, but the proposal is surprising for two reasons: 1) it depends crucially on the problematic

---

[100]   Cf. Blust (1999b:43), where the reflexes of *N and *ñ are tabulated for 21 Formosan languages, two of them proto languages. Identical reflexes are found in Proto Atayalic (*l), Kavalan (n), Amis (ɬ), Puyuma (l), Paiwan (ʎ), Siraya (l), Saisiyat (l), Thao (z), Bunun (n), and Saaroa (ɬ). These proto phonemes appear to be distinguished by Tsou, where *N > *k* after a sibilant, but *h* elsewhere, while *ñ > *n-, -h-, -h*, and by Kanakanabu, where *N > *n*, but * ñ > *ŋ* (Tsuchida 1976:138-143, 307).

assumption that contrastive stress can be assigned to PAN, and 2) counterevidence to the hypothesis is nearly as great as supporting evidence.

Wolff begins (1993:49) by noting that since *N is reconstructed in final position, but *ñ is not, 'nothing is changed by stating that final *N is in fact final *ñ.' This may be true, but it sets the typology of the proto language at odds with the typology of its descendants, since attested AN languages that reflect *c, *z or *ñ as palatals never allow them word-finally. So strong is the avoidance of final palatals in attested languages that even where *s became a palatal, as in Manggarai, this happened only syllable-initially: *salaq > *cala* 'wrong', *sebu > *cəbu* 'spray', *siku > *ciku* 'elbow', *susu > *cucu* 'breast', *baseq > *baca* 'wet', *asu > *acu* 'dog', but *beties > *wətis* 'calf of the leg', *qaRus 'current' > *arus* 'come quickly, come flowing', *Ratus > *ratus* 'hundred', etc. Similarly, where a medial palatal nasal came to be final in languages of the Admiralty Islands, there was a recurrent tendency for it to change, even though it was stable syllable-initially: POC *poñu > Bipi *puy*, Lindrou, Sori *boy* 'green turtle' (cp. *ñatuq > Bipi *ñak*, Lindrou *ñek*, Sori *ña* 'k.o. hardwood tree'). Although the typology of proto languages should respect the typology of their descendants, these observations do not entirely rule out Wolff's proposal to reconstruct *ñ in final position, since one might argue that *-ñ was so unstable that it merged with *n syllable-finally, but not syllable-initially in PMP. On the other hand, the general restriction against final palatal obstruents and nasals in modern languages makes it likely that none of these segments could occur word-finally in PAN.

Wolff (1993:47ff) notes that of the nasals *N, *n and *ñ only *N and *n are distinguished in Formosan languages, and only *n and *ñ in non-Formosan languages. *A priori* this supports a three-way contrast, but he argues that PAN had only *n and *ñ, and that the latter segment split under the influence of a preceding or following stressed vowel such that *ñ in oxytones = *N in Formosan languages and *ñ in non-Formosan languages both initially and medially, whereas *ñ in paroxytones = *N in Formosan languages but *l-, *-n- in non-Formosan languages (MP-1 = Non-Formosan ñ-retaining languages, MP-2 = Non-Formosan ñ-losing languages):

	Oxytones		Paroxytones	
	Initial	**Medial**	**Initial**	**Medial**
F	ɬ	ɬ	ɬ	ɬ
MP-1	ñ	ñ	l	n
MP-2	n	n	l	n

Figure 8.3 Schematised reflexes of PAN *ñ in Formosan and
Malayo-Polynesian languages (adapted from Wolff 1993)

Wolff divides his discussion into 1) *ñ in medial position, and 2) *ñ in initial position, and claims to find a high degree of correlation between accent placement in PAN and the reflex of *ñ. For reasons of space we will consider only *ñ in medial position, where it is most richly attested. There are two ways to test the adequacy of this claim. First, since the conditioning factor is stress, the relative frequency of penultimate and final stresses can be calculated, yielding a measure of the percentage of cases in which *ñ should become *n* or

ñ. Second, since Wolff cites cognate sets in support of his argument the primary sources for the data can be consulted to confirm or disconfirm his claims, and additional examples that he has not cited can be added to the list.

As seen earlier, attempts to project the phonemic stress of Proto Philippines onto PAN have not met with wide acceptance, and this raises questions about the explanation Wolff proposes. However, the clearest counterevidence to Wolff's claims comes from his own predictions, which produce results that do not deviate significantly from chance. Relevant comparisons must meet each of the following conditions: 1) there must be a Formosan cognate that distinguishes *N from *n, 2) at least one language must provide evidence for placement of the accent, 3) there must be a cognate in at least one MP language that distinguishes *ñ from *n. Wolff presents 18 comparisons in support of the above correlations, and recognises only two exceptions. An unbiased examination of the relevant evidence, however, does not support his interpretation. Table 8.28 lists all known forms that meet the above three conditions. I exclude Wolff's * keñáŋ 'recognise, remember', and *ʃuñud 'move forward, follow' as unconvincing (+ = prediction confirmed, - = prediction falsified; *qaNuNaŋ is counted twice).

Table 8.28 Correlation of proposed PAN stress with *N > n or *N > ñ
in Malayo-Polynesian languages (adapted from Wolff 1993)

No.	PAN	PAN-Wolff	PMP	Predicted		
1.	aNak	añák	anak	n	+	child
2.	aNay/SayaN	áñay	anay	ñ	-	termite
3.	baNaR	bañáɣ	banaR	n	+	*Smilax* sp.
4.	baNaS	báñas	banah	ñ	-	male (animal)
5.	buNi	búñi	buni	ñ	-	hide, conceal
6.	CaNem	tañém	tanem	n	+	plant, bury
7.	CuNuh	túñu	tunu	ñ	-	roast
8.	daNum	dañúm	danum	n	+	water
9.	keNa	keñá	kena	n	+	be struck
10.	paNaw	páñaw	panaw	ñ	-	walk, go
11.	paNij	páñig	panij	ñ	-	wing
12.	qaNiC	qáñit	qanit	ñ	-	hide, leather
13.	qaNiCu	qañítu	qanitu	n	+	ghost
14.	qaNuaŋ	qañuáŋ	qanuaŋ	n	+	carabao
15.	qaNuNaŋ	qañúñaŋ	qanunaŋ	n	+	tree sp.
16.	qaNuNaŋ	qañúñaŋ	qanunaŋ	ñ	-	tree sp
17.	qaNup	qañúp	qanup	n	+	to hunt
18.	qeNeb	qeñéb	qeneb	n	+	to close
19.	siNaR	ʃĭñaɣ	sinaR	ñ	-	to shine
20.	SiNuq	siñúq	hinuq	n	+	beads
21.	SuNus	súñuʃ	hunus	ñ	-	withdraw
22.	taNek	tañék	tanek	n	+	to boil
23.	tuNa	tuñá	tuna	n	+	eel
24.	zaNi	jáñi	zani	ñ	-	near
25.	bañaw	báñaw	bañaw	ñ	+	to wash
26.	qañud	qáñud	qañuj	ñ	+	drift
27.	Siñaw	siñáw	hiñaw	n	-	to wash

In this set of reconstructions Wolff's predictions are supported in 15 of 27 cases, hence 55.5% of the time. This is nowhere near the 90% success rate (18 of 20 cases) that he claims for correlating reconstructed stress with depalatalisation, nor does it depart significantly from chance. Together with the absence of a generally accepted basis for reconstructing contrastive stress in PAN we are therefore left with little choice but to accept the reconstruction of *n, *ñ and *N as distinct Proto Austronesian phonemes.

There is one last point that must be made in connection with *N. Among the Formosan languages the change *N > *n* is attested in just three historically independent cases: 1) in a language immediately ancestral to Basay, Trobiawan and Kavalan, 2) in Bunun, and 3) in Kanakanabu. In eight of the nine primary branches of the AN family represented in Taiwan some or all languages reflect *N as a lateral. In Paiwan this is palatalised, in Saaroa and Amis it is voiceless or partially devoiced, while in the other languages it is reportedly [l]. This strongly suggests that *N was not a nasal, but rather some type of lateral distinct from *l. Given its recurrent tendency to merge with the palatal nasal *ñ in Formosan languages there are strong phonetic grounds for assuming that *N was a palatal lateral. If so, it fits into a well-established series *c, *z, *ñ, *N ([ʎ]) and *j ([gʸ]) which—apart from *c—is reasonably well-preserved in MP languages, but generally lost through mergers in Formosan languages. This interpretation may also help to understand unusual mergers such as that of *ñ, *N and *j in Thao (all three may have shared a palatal component). It is, moreover, further strengthened by the absence of clear evidence for *N-, a gap that parallels the gap for *j-. However, since the symbol *N is thoroughly embedded in the comparative literature there is no point in changing it.

8.2.5 Fricatives

PAN contrasted two sibilants, *S (*h in Dyen 1953b) and *s. The first of these is reflected as a sibilant in all Formosan languages except Kulon (extinct), where it became *s* or *h*, Taokas (extinct), where it became *s* or Ø, Siraya (extinct), where it became *g* (apparently [x]) and Puyuma, where it disappeared. Outside Taiwan *S became *h* in one or more environments in a number of Philippine languages (Bashiic, Central Philippines), in Kayan of central Borneo, in Malay, and in Soboyo of the central Moluccas. Elsewhere it became glottal stop, or most commonly zero. This is a common diachronic profile for voiceless alveolar fricatives in many language families, and strongly supports the view of Ross (1992:38ff) that *S was [s]. This means that *s must have had some other value. Dempwolff (1934-1938) interpreted *s as a voiceless palatal stop. He considered it palatal because in languages that preserve *ñ as a palatal nasal and that have active systems of nasal substitution *s* is replaced by *ñ* in certain word-formation processes: Kapampangan *saklub* 'a cover' : *ma-ñaklub* 'to cover', *sulúʔ* 'torch' : *ma-ñulúʔ* 'light a torch or candle', Kenyah *salut* 'captive, prisoner' : *ñalut* 'capture, take prisoner', *səŋit* 'urine' : *ñəŋit* 'urinate', Javanese *sorot* 'ray, beam' : *ñorot* 'shine light on', *sikəp-an* 'an embrace' : *ñikəp* 'to embrace', Chamorro *saga* 'stay, rest' : *ma-ñaga* 'to stay', *sotsot* 'contrite, repentant' : *ma-ñotsot* 'to repent'. He considered it a stop, because *s is often reflected as a stop or affricate. Despite these observations the reflex of *s in the great majority of AN languages is a voiceless alveolar or laminal fricative.

Ross (1992) has attempted to reconcile these seemingly contradictory features of *s: if it was a palatal, and takes *ñ* in nasal substitution, why is it almost never reflected as a palatal, especially in MP languages? The solution he proposes has considerable explanatory value. PAN *s was a voiceless palatal affricate (*c) that remained palatal until the process of

homorganic nasal substitution was innovated in PMP. During this period *S was becoming PMP *h, and *s shifted to an alveolar sibilant to fill the gap created by this change, but not until a pattern of alternation with *ñ under prefixation was firmly established. This does, however, create a problem, since we have already found it necessary to reconstruct PAN *c on the basis of a different sound correspondence. If we treat *s as a palatal fricative this problem can be resolved, and the reconstruction is still consistent with the scenario Ross has proposed.

The one remaining problem concerns the apparent fortition of *s in many daughter languages. If *s was a sibilant it is surprising that it is so often reflected as a stop or affricate. Examples include *ts* (Basay, Trobiawan, Amis), *c* (Manggarai), and *t* (Proto Western Plains, Paiwan, Agta, Atta, Gaddang, Isneg, Tboli and Bilaan, Palauan, Wuvulu-Aua, Sa'a, Arosi and other languages of the southeast Solomons, Chuukese and other languages of Micronesia). This seems to support the view that *s was not a fricative, but reflexes of this proto phoneme have just as often become *h* (Proto Atayalic, Kulon, Saisiyat, Ifugaw, Kallahan, Balangaw, Botolan Sambal, Samihim, Ma'anyan, Proto Bungku-Tolaki, Hawu, Mwotlap, Raga, Sakao, Tongan) or Ø (Miri, Kelabit, Hawaiian). All-in-all, then, it seems best to assume that *s was a palatal fricative in PAN that became alveolar after the change *S > *h* and the innovation of homorganic nasal substitution. Why it was prone to repeated fortition is unclear, although it is possible that fortitions of [s] are more common in the world's languages than is often assumed.

The third fricative that must be reconstructed for PAN generally appears in the literature as *H, but since this is now the only glottal fricative it can be written as *h. This proto phoneme is very weakly attested, which is what we might expect after 5000-6000 years of linguistic evolution. Four languages in Taiwan (Proto Atayalic, Saisiyat, Pazeh, Amis) provide the primary evidence for *h, which is found word-finally in seven items, and medially in two more (Table 8.29):

Table 8.29 Evidence for PAN *h

PAN	PATY	Saisiyat	Pazeh	Amis	
*baqeRuh	—	—	—	faʔloh	'new'
*baRah	*bagah	baLah	bahah	—	'ember'
*Capuh	—	—	sapuh	—	'sweep'
*CuNuh	—	s-om-olœh	—	—	'roast'
*nunuh	*nunuh	—	nunuh	—	'breast'
*qasiRah	—	—	—	cilah	'salt'
*qumah	*qumah	—	u-ma-mah	omah	'swidden'
*bahi	—	—	—	fafahi	'woman'
*buhet	*buhut	ka-bhœt	buhut	fohət	'squirrel'

The strength of this inference rests on the agreements between these languages, which appear to represent four primary subgroups of the AN family. Two sets are also supported by Philippine evidence: *baRah > Itbayaten *vayah* 'redness, red colour', *vayah-ən* 'to heat red-hot, as iron', *bahi > Binukid, Western Bukidnon Manobo *bahi* 'female; woman', although others are contradicted, as with Itbayaten *vaʔyo* 'new, fresh', with final vowel. Within Taiwan the evidence is sometimes contradictory as well, as with PATY *kari 'dig', Pazeh *sa-kari* 'digging stick', but Saisiyat *kaLih* 'dig up', or PATY *mama 'uncle; FB/MB', Amis *mama* 'father', but Pazeh *mamah* 'eB, wife's EB'.

8.2.6. Liquids

Setting aside *N, which probably was a palatal lateral, PAN had three liquids. The least problematic of these is *l, which (despite some irregularities in Central Philippine languages) is supported as a distinct phoneme by hundreds of languages, and needs no further discussion.

Wolff (1974) claimed that Dempwolff's *r was based entirely on correspondences produced by borrowing (primarily from Malay), and on erroneous cognate identifications. As with his similar claims about *c and *g, however, a closer inspection of the available evidence fails to support this position. Dempwolff's *r and *R have similar frequencies of occurrence. Chrétien (1965) lists 138 examples of the former and 160 of the latter.

While many of Dempwolff's reconstructions with *r clearly are products of borrowing from Malay, as Wolff has argued, others cannot easily be dismissed in this way. Moreover, a number of newer comparisons also support *r:

1.   *rauC 'peel rattan or bamboo' : Paiwan *r<m>auts* 'split wood', Isneg *raut* 'strips of rattan used in making baskets', Bontok *láot* 'a strip of *bíkal* bamboo used for tying', *laót-an* 'reeds used as ties to press down thatching', Kankanaey *láot* 'scraped rattan', Iban *raut* 'pare, peel, scrape', Malay *raut* 'whittle off', Javanese *rot-an* 'rattan'.

2.   *rawan 'strong emotion' : Hanunóo *rawán* 'grudge', Iban *rawan* 'experience emotion, be agitated', Malay *rawan* 'emotion, tender feeling; be stirred', Manggarai *rawaŋ* 'sad; anxious'

3.   *ruit 'barb of a hook' : Iban *ruit* 'a barb', Malay *ruit* 'bent, hook-shaped; (Brunei) hook or barb', Javanese *ruit* 'barb (of a hook)', Bolaang Mongondow *ruit* 'sharp, pointed'.

4.   *kure(n)dut 'wrinkled, as the skin' : Bikol *kurundót* 'wrinkled (used only of the skin)', Mukah Melanau *kərədut* 'wrinkled', Malay *kərdut* 'creased, furrowed, deeply lined', *k. muka* 'with a frowning face'.

5.   *periŋ 'bamboo sp.' : Rungus Dusun, Dohoi *poriŋ*, Malay, Manggarai *pəriŋ*, Banggai *peliŋ*, Lio *pəri* 'type of slender bamboo', Pohnpeian *pe:ri* '*Bambusa vulgaris*'

6.   *samir 'leaf awning' : Ilokano *sámir* 'pinnae of coconut palm leaves woven into basketwork and used for a temporary shade', Maranao *samir* 'awning', Malay *samir* 'nipah leaf dried but not worked up into roofing thatch (atap). Plain dried leaves are used for temporary screens or shelters', Toba Batak *samir* 'anything used to provide shelter—as leaves', Javanese *samir* 'round banana leaf fixed in a shallow inverted cone shape and used for covering foods'

None of the liquid correspondences in these comparisons, or many others like them, can easily be explained as reflexes of other proto phonemes. Paiwan *r* sometimes reflects *R, but this is not the usual development, and Isneg *r* cannot regularly derive from either *d or *R. Similar remarks apply to Hanunóo, where *R > *g*, and *d > *d*, while for *ruit Iban and Malay could reflect *Ruit (but not *duit), while Javanese and Bolaang Mongondow could reflect *duit (but not *Ruit), etc. Borrowing may account for an occasional comparison of this type, but cannot be invoked for all or even a majority of them. We therefore have no alternative but to posit *r, which presumably was an alveolar tap (as these examples suggest, reflexes in most languages are *r* or *l*).

While *r has a very narrow range of reflexes, the historical development of *R probably shows greater phonetic variety than that of any other PAN phoneme with the possible

exception of *j. Somewhat surprisingly, in view of its inferred phonetics, only a minority of these reflexes are liquids:

Table 8.30 Reflexes of PAN *R

d	=	Ibaloy (through *R > l), Inati
ð	=	Bugotu (before non-high vowels)
g	=	Atayal, Northern Cordilleran languages, Greater Central Philippines languages, Sabahan languages, Berawan, Chamorro, Tigak
h	=	Samal, Kayan, Ngaju Dayak, Kove
j	=	Dangal (before *u* only)
k	=	Chamorro, Tigak (word-final only)
l	=	Taokas, Bunun, Basay, Amis, Bontok, Pangasinan, Helong
lh	=	Thao (voiceless lateral)
L	=	Saisiyat (retroflex flap)
l^y	=	Sissano (before *i*)
n	=	Mekeo, Solos
ŋ	=	Taiof (word-final only)
r	=	Atayal (before *i*), Seediq, Tiruray, Kelabit, Malay, Toba Batak, Palauan (only before dentals), Motu
R	=	Kavalan (contrastive uvular rhotic)
s	=	Maloh, Buginese, Palauan
w	=	Bisaya (word-final only)
x	=	Pazeh
y	=	Bashiic, Botolan Sambal, Kapampangan, northern Mangyan, Melanau languages (word-final only), Ma'anyan, Lampung, Sundanese, most languages of Manus, Mwotlap, Olrat
z	=	Malagasy (from intermediate *y)
Ø	=	Paiwan, Kenyah, Javanese, Bimanese, Hawu, Roti, Proto Central Pacific

Despite this wide range of reflexes there has been greater agreement regarding the phonetic properties of *R than of *j. Most writers from Demwolff on have assumed that *R was a uvular fricative or trill, and this is one reasonable possibility. However, coronal reflexes are much more common than uvular or velar reflexes. Given this observation it is reasonable to suggest that *r was an alveolar flap and *R an alveolar trill, a hypothesis that is supported by known sound changes, since the shift of an alveolar trill to a uvular trill is well-attested in the history of French and other European languages, while a shift in the opposite direction is unknown.

PAN *R has been the subject of extensive discussion. Originally recognised by van der Tuuk and then named van der Tuuk's 'first law' by Brandes, it was later split into four subtypes by Dyen (1953a), as noted in Table 8.12. Dahl (1976:86ff) attributed Javanese, Ngaju Dayak, Malagasy *r* < *R to borrowing from a language in which *R became *r* (most likely Malay) and Malagasy *z* < *R to a language in which *R became *y* (most likely Ma'anyan or some similar Southeast Barito language), and so concluded that *R$_1$ represents the only directly inherited reflex of *R in Javanese, Ngaju Dayak and possibly Malagasy. Borrowing undoubtedly has played an important role in determining the *R

reflexes in many AN languages. Dempwolff and Dyen both recognised that Ngaju Dayak has two speech strata distinguished by several sets of double reflexes, one of these being *R > h/r. Reid (1973b) has called attention to a similar situation in Kankanaey of northern Luzon, where *R is reflected as *l*, *g*, or Ø. Following Conant (1911), he notes that the apparently unconditioned split of *R into *g* or a liquid is widespread in the northern and southern Philippines. In some languages, as Tiruray, this is clearly due to borrowing (Blust 1992). In other languages, however, the split of *R cannot easily be explained either through phonological conditioning or borrowing.

*R has been reconstructed in approximately 200 morphemes. Dyen was able to provide a subscript for some 24 of these. Dahl (1976:91), however, reduces this to 'only 16-18 instances which correspond more or less with Dyen's hypothesis.' Although systematic investigation of the *R reflexes has not advanced to the point where we can be certain that a full account of the correspondences would lead to the reconstruction of a distinct *R in each reconstructed morpheme (and hence well over a hundred subscripted varieties of *R), it has become clear that as more languages are integrated into the comparison Dyen's reconstructional procedure leads inevitably in this direction. Thus, Nothofer (1975:160ff) points out that in preconsonantal and final positions $*R_3$ and $*R_4$ merged with *D and *r as 'Proto Malayo-Javanic' *r, but that $*R_1$ and $*R_2$ must be further subdivided ($*R_1$ into two, $*R_2$ into three proto phonemes) to account for Sundanese evidence. This means that at some higher node in the AN family tree acceptance of his procedure would require us to posit at least seven subscripted types of *R. Similarly, in Limbang Bisaya of northern Sarawak *R appears as initial Ø or *r-* (uvular), medial Ø, *-g-* or *-r-*, and final *-w*, *-g*, or *-r*. All instances of *r* apparently can be attributed to borrowing from Brunei or Sarawak Malay, and there clearly is some phonological conditioning (after *e the only reflex of *R is *g*), but this still leaves an unexplained split of *R in a number of forms. Where the etyma contain a subscripted *R this shows no correlation with Dyen's $*R_1$ - $*R_4$: $*baqeR_1u$ > *agu* 'new', but $*daR_1aq$ > *raaʔ* 'blood', $*teR_2as$ > *tagas* 'hardwood tree', but $*uR_2at$ > *uat* 'vein', etc. A similar pattern is found in Oceanic languages (Geraghty 1990, Lynch 2009a, François 2011). This leaves little choice but to conclude that *R shows unconditioned phonemic splits in a variety of widely scattered languages from southern Taiwan (Paiwan, with Ø and *r*) through the Philippines and western Borneo, to Vanuatu and Micronesia. Why this particular proto phoneme would be so subject to irregular change is unexplained, but the alternative of reconstructing a potentially unlimited number of subscripted *R phonemes justified by progressively smaller sets of cognates clearly is unacceptable.

Finally, several languages in the Philippines and Indonesia share apparent cognates that contain a previously unrecognised liquid : sibilant correspondence which currently resists analysis. In most of these cognate sets Philippine languages have an *l* or *r* that corresponds with *s* in a language of western Indonesia (Blust 2006b). These comparisons are puzzling, as they do not conform to any known comparative formula. The novel correspondence that they appear to contain could be labeled by *L, but the phonetic value of such a symbol would be obscure.

8.2.7 Glides

There is little to say about glides. Both *y and *w must be reconstructed, and there are very few problems with their reflexes. Dahl (1976:14) proposed that glides be replaced by the corresponding vowels *i and *u, but this proposal obscures important features of canonical shape in AN languages, makes it difficult to state the rules of stress in many

languages, and leads to peculiar statements of diachronic change where glide fortition has taken place. Dyen (1962) suggested subscripted varieties of *w, but this was largely a product of failing to understand fortitive processes in a number of widely separated languages.

8.2.8 Vowels and diphthongs

The PAN vowels require little discussion. There is universal agreement regarding *i, *u, *e (schwa) and *a. In AN linguistics the term 'diphthong' is applied to the -VC sequences *-ay, *-aw, *-uy and *-iw, since these monophthongised in many daughter languages, and so must be treated differently from other -VC sequences. In addition, the sequences *-ey and *-ew have been proposed, the first only in final position, and the second both medially (in *dewha 'two') and finally.

Dyen (1949:fn. 5) proposed *-ey casually in a footnote: 'I reconstruct *-ey to explain final correspondences such as those of Tg. *paláy*, Ml. *padi* as against the reconstruction of *-ay to explain the correspondences of Tg. *á:nay*, Ml. *anay-anay* 'termite', To. *ane* 'moth'. Dempwolff reconstructs *-ay in both cases.' In a second footnote (1953a:fn. 18) he elaborated this interpretation, holding that *-ay yields Tagalog, Malay *-ay*, Toba Batak, Javanese *-e*, while *-ey yields Tagalog *-ay*, Toba Batak *-e*, Javanese, Malay *-i*. In addition, Hendon (1964) reconstructed *-ew on the basis of a somewhat different set of criteria (Tagalog *-oʔ* : Javanese, Malay *-u*). Dahl (1976:40ff) believed that the evidence for *-ey is best explained by an appeal to borrowing, while that for *-ew probably is a product of faulty analysis of the data. In addition, Blust (1982a) pointed out that forms such as Malay *bini* 'wife' almost certainly reflect PAN *b<in>ahi, and so exemplify an irregular development of *-ay to *-i* (*b<in>ahi > binay > *bini*).

The most thorough attempt to justify these reconstructions is that of Nothofer (1984), who drew attention to evidence for an *-ay/ey and *-aw/ew distinction in four languages of western Indonesia:

Table 8.31 Evidence for *-ey and *-ew (after Nothofer 1984)

	Tag.	TBt.	Jav.	Mal.	Snd.	Mad.
*-ay	-ay	-e	-e	-ay	-ay	-ay
*-ey	-ay	-e	-i	-i	-e	-i(h), e(h)
*-i	-i	-i	-i	-i	-i	-i(h), e(h)
*-aw	-aw	-o	-o	-aw	-o	-aw
*-ew	-aw	-o	-u	-u	-o	-u(h), o(h)
*-u	-o	-u	-u	-u	-u	-u(h), o(h)

Table 8.31 shows that *-ay and *-ey are distinguished by Javanese, Malay, Sundanese and Madurese, and that *-aw and *-ew are distinguished by the same set of languages less Sundanese. These correspondences are puzzling. On the one hand, the languages that appear to distinguish *-ay from *-ey and *-aw from *-ew show a high degree of agreement, supporting Nothofer's claim that this distinction was found in their putative immediate common ancestor (which he calls 'Proto Malayo-Javanic'). On the other hand, these languages are contiguous (Javanese and Sundanese, Javanese and Madurese), or have long been in a borrowing relationship (Malay with all of the others). It is therefore tempting to follow Dahl (1976:40ff) in dismissing these correspondences as due to diffusion. However, some words with *-ey are basic and rarely borrowed (liver, to die,

riceplant). Moreover, it would be unlikely that the same words would be borrowed independently into several different languages.

This leaves us in methodological limbo. It would be helpful if evidence of these distinctions could be found in more distant languages. Javanese does not appear to be closely related to Malay, Sundanese, or Madurese, so the distribution of this comparative feature is geographically compact but genetically disparate. To raise further suspicions, the distinction is not found in the closest relatives of Malay that lie outside the 'diffusion zone' of Malay influence in western Indonesia, namely the Chamic languages (Thurgood 1999:124ff). It might appear that the reconstruction of *-ey and *-ew presents problems similar to those for e.g. *c, but there are important differences. Whereas *c fills a structural gap in a system that must be reconstructed with a corresponding voiced palatal stop and nasal, *-ey and *-ew fill no gaps, but simply elaborate a system of diphthongs that is symmetrical and intrinsically complete without them. Similarly, although *c is consistently distinguished from *s even in submorphemic roots, *-ey and *-ew do not Figure in any known root (Blust 1988a). Until a better basis can be discovered for assigning *-ey or *-ew to PAN or any other early proto language, then, it seems best to treat the correspondences in question as simply unexplained.

For a very different view of the PAN sound system and vocabulary than that offered here the reader is referred to Wolff (2010), and several reviews of that work (Blust 2011b, Adelaar 2012, Mahdi 2012).

8.3   Phonological reconstruction below the level of PAN

Space does not permit more than a few passing remarks about phonological reconstruction below the level of PAN. Apart from the reconstruction of *C, *N and *S, most of the problems discussed in connection with the reconstruction of PAN also arise with Proto Malayo-Polynesian. The principal difference is that PMP permitted medial prenasalisation of obstruents, while PAN apparently did not.

Medial prenasalisation figured prominently in Dempwolff's methodology of reconstruction. Where there was agreement in pointing to medial prenasalisation he reconstructed a prenasalised obstruent, as with *punti 'banana'. Where the languages disagreed in indicating a simple or prenasalised obstruent Dempwolff reconstructed a 'facultative' nasal, as in *tu(m)buq: Tagalog *tubóʔ*, Toba Batak *tubu*, Javanese *tuwuh*, but Malay *tumbuh*, Fijian *tubu* ([tumbu]), Tongan *tupu* (*p* < *mb*) 'to grow'. Dempwolff's convention for representing the facultative nasal (use of parentheses) differed from his convention for representing ambiguity (use of square brackets). The proper inference appears to be that he did not regard forms like *tu(m)buq as ambiguous. However, it was never made clear what claim the facultative nasal embodies: 1) a claim that both forms were found in PAN, 2) a claim that only the CVCVC form was found, with historically secondary prenasalisation in some daughter languages, or 3) a claim that only the CVNCVC form was found, with historically secondary denasalisation in some daughter languages. In particular cases all three of these interpretations can be defended. The evidence for Dempwolff's *tu(m)buq, for example, presents roughly equal support for *tubuq or *tumbuq, and both variants may well have co-existed in Proto Malayo-Polynesian. On the other hand, the addition of other languages to some of Dempwolff's cognate sets shows that (adopting his conventions) a 'facultative nasal' must be reconstructed in many cases where he posited a simple medial obstruent: *pusej, but Balinese *puŋsəd* 'navel' (hence *pu(ŋ)sej), *betuŋ, but Western Bukidnon Manobo *bəntuŋ*

'large bamboo sp.' (hence *be(n)tuŋ), *pija, but Maloh *insa* 'how much/how many?' (hence *pi(n)ja), etc. In still other cases the expansion of a Dempwolff cognate set undermines the certainty with which he regarded certain preconsonantal medial nasals: *buŋkus, but Western Bukidnon Manobo *bukus* 'roll or wrap s.t. around s.t. else, as in bandaging a wound', Sasak *bukus* 'to wrap, bundle' (hence *bu(ŋ)kus). It now appears likely that although some PMP forms contained prenasalised medial obstruents, sporadic prenasalisation and denasalisation both operated throughout the history of the non-Formosan AN languages (Blust 1996d).

The reconstruction of Proto Oceanic phonology presents some of the same problems as these, and some new ones. Oceanic languages may reflect obstruents in either an oral grade (*p, *t, *c, *s, *k) or a nasal grade (*b, *d, *dr, *j, *g), and this is true both medially and morpheme-initially. Table 8.32 provides examples of reconstructions in PMP and their continuations in POC to illustrate the relationship between prenasalisation and consonant 'grade' (OG = oral grade, NG = nasal grade):

Table 8.32 Relationship between PMP prenasalisation and POC consonant grade

PMP	POC		
pitu	pitu	(OG)	seven
hapuy	api	(OG)	fire
bulan	pulan	(OG)	moon
qabu	qapu	(OG)	ashes
punay	bune	(NG)	pigeon
t-umpu	tubu	(NG)	grandparent/ancestor
beRek	boRok	(NG)	pig
tumbuq	tubuq	(NG)	grow
taqun	taqun	(OG)	year
qutin	qutin	(OG)	penis
duha	rua	(OG)	two
kuden	kuron	(OG)	cooking pot
-nta	-da	(NG)	ours (incl.)
punti	pudi	(NG)	banana
danum	danum	(NG)	fresh water
pandan	padran	(NG)	pandanus
susu	susu	(OG)	breast
quzan	qusan	(OG)	rain
zalan	jalan	(NG)	path, road
tazim	tajim	(NG)	sharp

Given the general correlation between medial prenasalisation in non-Oceanic languages and nasal grade reflex in Oceanic languages (*tumpu : *tubu, *tumbuq : *tubuq, *-nta : *-da, *punti : *pudi, *pandan : *padran) Dempwolff was able to show that although PMP *b and *p (and *g and *k) had merged in POC, simple and prenasalised stops yielded different reflexes. Starting from this point he reasoned that nasal grade reflexes in initial position probably have the same origin as those in medial position. Since he did not reconstruct prenasalised obstruents word-initially he speculated that the appearance of nasal grade initials in OC languages corresponding to simple initials in the western languages might be a residue of the process of homorganic nasal accretion in forms once prefixed with *maŋ- or *paŋ-. This explanation is superficially plausible, but is difficult to

apply to words like POC *bune 'pigeon', or boRok 'pig'. Moreover, although Dempwolff failed to notice it, many Oceanic languages have three consonant grade reflexes, and scholars in the 1970s began to sense that Dempwolff's theory of consonant grades is incomplete (Blust 1976a). In an attempt to mend the problem Ross (1988) split the oral grade reflexes into fortis and lenis types, with no clear basis for conditioning. This results in massive unconditioned phonemic splitting, but to date no better solution has been found for this longstanding problem. Despite earlier reports of 'crossover' in consonant grade for the same lexical item in different languages, Ross (1989) has also argued that consonant grade is largely regular, and can be reconstructed with considerable confidence for POC. A close look at the comparative picture for Oceanic languages supports this view in many cases, but not in others, and the matter is still in need of further investigation.

A problem in reconstruction that is peculiar to the Oceanic languages within the AN family is the need to posit a separate set of labiovelar consonants. Dempwolff thought that Oceanic languages had acquired labiovelars through contact with Papuan languages, but was unable to propose specific plausible scenarios of borrowing. Goodenough (1962:406ff) suggested that labiovelars in Oceanic languages continue a labiovelar series that was found in PAN, but this proposal has not been accepted. Blust (1981a) showed that although labiovelars must be reconstructed in some POC forms, labiovelar reflexes are historically secondary in many others, as in *Rumaq > Hiw *em^w*, Mota, Raga, Valpei *im^wa*, Ngwatua *iŋ^wa*, Soa *im^w*, Chuukese *iim^w*, Gilbertese *uum^wa* 'house', Chuukese *im^wa-n* , Gilbertese *um^wa-n* 'his/her house', Chuukese *mapo 'heal' : Wuvulu-Aua *mafo*, Seimat *ma-mahu-a*, Gedaged *mao*, Lau *mafo*, Arosi *maho*, Mota *mawo*, Pohnpei, Mokilese *mo*, Fijian *mavo*, Samoan *mafu*, but Baluan, Lou *m^wap* 'to heal', or *ma-saŋa 'bifurcated, forked; twins': Lau, Sa'a *mataŋa* 'forked, branched', Arosi *mataŋa* 'doubled, forked', Rotuman *majaŋa* 'forking, bifurcation', Tongan *māhaŋa*, Samoan *māsaŋa* 'twins', but Aua *wataa* (< *mwasaŋa) 'twins'. While phonological conditioning is clear in some forms, as *Rumaq > *im^wa*, etc., with transfer of rounding from the vowel to the labial nasal, the appearance of labiovelars in other forms is unexplained. The whole question of labiovelars in Oceanic languages has recently been subjected to a thorough review by Lynch (2002), and most recently Ross (2011) has proposed the reconstruction of POC *k^w in addition to the previously accepted *m^w, *p^w, and *b^w.

One other feature of POC that distinguishes it from PAN is its five vowel system: *i, *u, *e, *o, *a (where *e represents a mid-front vowel, not a schwa as in PAN or PMP). Most instances of POC *e derive from *-ay. Most instances of POC *o derive from PMP *e (schwa), with others reflecting *-aw. In addition to these major sources of mid-front and mid-back vowels, a small number of forms contain *e or *o either in apparent lexical innovations or as a result of sporadic changes in earlier *i, *u, or *a. This change is shared with the Central Malayo-Polynesian and South Halmahera-West New Guinea languages, and so must predate POC (Blust 1983/84a).

8.4 Lexical reconstruction

Since some aspects of morphosyntactic reconstruction are touched on, at least briefly, in Chapter 6, the last topic that will be discussed here is lexical reconstruction. As noted already, Dempwolff (1938) contains some 2,215 lexical bases. When this publication appeared comparative dictionaries were available for very few language families, and Dempwolff's collection of reconstructed bases advanced Austronesian comparative linguistics to a position that probably was second only to Indo-European. For over thirty

years after Dempwolff's comparative dictionary appeared almost no new AN lexical reconstructions were proposed apart from some work on Proto Oceanic by the German linguist Wilhelm Milke (1958, 1961, 1968). In 1970 this situation began to change, and in the period 1970-1995 about 3,000 new reconstructions on the PAN, PMP and PWMP levels were published, together with supporting evidence from nearly 200 languages. Much of this material has been brought together and combined with a critical reassessment of Dempwolff reconstructions in the *Austronesian Comparative Dictionary* (Blust and Trussel ongoing), an online resource that currently (April, 2013) is only about 33% complete, but which contains around 4,767 base forms and more than 13,000 proto words (bases plus affixed forms).

The major features of the *Austronesian Comparative Dictionary* (ACD) are as follows. The first line of each entry contains three types of information: a code number in parentheses that indicates the proto language to which an etymon is assigned, the reconstructed form, and its gloss. Code numbers employed are (1) = Proto Austronesian, (2) = Proto Malayo-Polynesian, (3) = Proto Western Malayo-Polynesian, (3a) = Proto Philippines, (4) Proto Central-Eastern Malayo-Polynesian, (5) = Proto Central Malayo-Polynesian, (6) = Proto Eastern Malayo-Polynesian, (7) Proto South Halmahera-West New Guinea, and (8) = Proto Oceanic. Reconstructions with the same shape are distinguished by hyphenated numeral. Below this line is the supporting evidence, organised by major subgroup or collection of subgroups, where F = Formosan (nine primary branches of AN), WMP = Western Malayo-Polynesian (possibly more than one primary branch of MP), CMP = Central Malayo-Polynesian, SHWNG = South Halmahera-West New Guinea, OC = Oceanic. Language names are indicated with three-letter abbreviations unless the name contains four characters, in which case it is written in full. The last line is an optional note, which is reserved for apparently related forms that show phonological irregularities, or for any other matter requiring discussion. Two successive entries from the 'A' section of the dictionary will serve to illustrate these conventions in more concrete terms:

Table 8.33 Sample entries from the *Austronesian Comparative Dictionary*

(3) **abat-1**			give a supporting hand
WMP:	HAN	ábat	leading by the hand
	CEB	ábat	hold onto s.t. fixed to support o.s.
		abat-án	railing in the shape of a ladder for baby to cling to when he starts to walk
	IBAN	ambat	lend a hand (to hold steady)
(3a) **abat-2**			spirit that causes sickness
WMP:	ISG	ábat	a spirit who brings sickness to people
	BON	ábat	perform a ceremony for s.o. who has had a spirit encounter; such a ceremony
	CEB	ábat	any supernatural being or human with supernatural powers which shows itself in an unexpected and startling way

Note: Also DIB *abot* 'sickness in which the victim vomits red vomitus, caused by an evil spirit'

In the first comparison (3) marks the reconstruction as Proto Western Malayo-Polynesian, and the hyphenated numeral shows that it is the first of two or more protoforms with the same shape. WMP indicates that the languages cited in this section are classified as Western Malayo-Polynesian, while HAN and CEB are abbreviations for Hanunóo and Cebuano (Iban is written out in full). In the second comparison (3a) marks the reconstruction as Proto Philippines, and the abbreviations ISG, BON and DIB represent Isneg, Bontok, and Dibabawon. Where there are disagreements in simple vs. prenasalised medial consonants it is generally assumed that prenasalisation is historically secondary unless the great majority of language support it. Contrary to Dempwolff's practice, then, the ACD does not automatically include 'facultative' nasals in reconstructions that represent such correspondences.

It should be emphasised that these two entries are very simple. Many entries are much longer, and in addition to the base contain all affixed, reduplicated and compound forms of a base that can be reconstructed. In addition, some are heavily annotated. The entry (1) *ama 'father', for example, is roughly six and one half pages long, and contains sixteen affixed forms of this base that are assigned to various proto languages: *da-ama 'father', *maR-ama 'father and child', *ta-ama 'father (ref.)', assigned to (1), *pa(ka)-ama 'treat like a father', assigned to (2), *si-ama 'father (ref.)' , *ka-ama-en 'father's brother; fatherhood', assigned to (3), *paRi-tama 'relationship of father and children', assigned to (8), etc. The entry (1) *aNak 'child, offspring; son, daughter; interest on a loan' is fifteen pages long, contains 46 affixed, reduplicated or compounded forms of the base, and nearly a full page of additional annotation. Even in its present unfinished state, the printed version of the ACD is thus over 2,700 single-spaced pages.

In addition, since reconstruction is carried out for nine different proto languages at once, some entries contain reconstructions at multiple nodes in the AN family tree. In the interest of simplicity reconstructions are explicitly indicated at subordinate nodes only where they differ either in form or meaning (or both) from the primary entry. Thus (1) *aNak 'child, offspring; son, daughter; interest on a loan' appears after the Formosan reflexes as (2) *anak 'child, offspring; son, daughter; BC (m.s.), ZC (w.s.); young animal or plant; young; small (for its kind); dependent or component part of something larger; native, resident, inhabitant; interest on a loan' because of the change of phonological form and larger number of meanings that can be associated with it. Following the WMP reflexes, however, the CMP reflexes are given without a further reconstruction, since (4) *anak is essentially identical to the PMP form. Glosses are assigned on the basis of distribution in primary branches of a given node, even where these may be controversial (as with 'interest on a loan', which currently is interpreted in monetary terms, but which may well have had a similar meaning at an earlier time in relation to barter).

As in Dempwolff's material, approximately 90% of all reconstructed content morphemes are disyllabic. In PAN this means that the great majority of bases are CVCVC (where all consonants are optional), while in PMP they are either CVCVC or CVNCVC, with homorganic medial prenasalisation. Both PAN and PMP also have a smaller class of CVCCVC reduplicated monosyllables such as PAN *buCbuC 'pluck, pull out', which are unitary morphemes in both languages. Nearly one-third of Dempwolff's reconstructed bases are now considered to be later innovations, many of them spread through western Indonesia and the Philippines by borrowing from Malay. To deal with the re-evaluation of this material the ACD contains an appendix of rejected comparisons, as well as tempting new comparisons that appear to be false leads. Since the Formosan languages have been in close contact for a very long time and early borrowing may be difficult to detect, it was felt

that Formosan-only cognate sets should be treated with caution, even though in many cases they may well point to PAN forms that have disappeared outside Taiwan. Consequently, Formosan-only cognate sets that are found either throughout Taiwan, or with a geographically discontinuous distribution on the island, are compiled separately. Finally, the dictionary also contains an appendix of submorphemic roots, incorporating and expanding upon the material in Blust (1988a).

Reconstruction has also been done on many lower-level proto languages, most notably Proto Polynesian (Walsh and Biggs 1966 and ongoing online additions), Proto Oceanic (Grace 1969, Ross, Pawley and Osmond 1998, 2003, 2008, 2011), Proto Philippines (Zorc 1971), Proto South Sulawesi (Mills 1975), Proto Tsouic (Tsuchida 1976), Proto Minahasan (Sneddon 1978), Proto Batak (Adelaar 1981), Proto Sangiric (Sneddon 1984), Proto Malayic (Adelaar 1992), Proto Bungku-Tolaki (Mead 1998), Proto Chamic (Thurgood 1999), and most recently Proto Micronesian (Bender et al. 2003a, 2003b). Three of these are described at somewhat greater length below.

Two features of the lexicons of PAN and PMP can be singled out for brief discussion: 1) the problem of doubleting, 2) the problem of submorphemic roots. The term 'doublet' normally refers to two (or more) historically related forms that are found in the same language as a result of different historical trajectories. In English, for example, *shirt* and *skirt* have a common origin, but *shirt* is native and *skirt* an early Scandinavian loanword. Somewhat different are the loanwords *wine* and *vine*, which were borrowed from Latin *vīnum* at different historical periods, leading to different phonemic adaptations. Doublets of this standard type are as common in AN languages as they are elsewhere in the world. Tiruray in the southern Philippines, for example, reflects *Ratas 'milk' both as *ratah* 'human breast milk' (native), and as *gatas* 'store-bought milk' (borrowed from a neighbouring Danaw language). However, comparative work in AN must also deal with many reconstructed doublets that are not amenable to the same type of explanation, as shown in Table 8.34:

Table 8.34 Sample patterns of doubleting in Austronesian reconstructions

1)	d : n	(3) *adaduq : (2) *anaduq 'long, of objects'
2)	b : l	(1) *baŋaw : (1) *laŋaw 'blowfly'
3)	R : w	(1) *baNaR : (1) *banaw 'a plant: *Smilax* spp.'
4)	ŋ : q	(3) *bintaŋuR : (2) *bitaquR 'a tree: *Calophyllum inophyllum*'
5)	q : t	(1) *beriq : (3) *berit 'tear open'
6)	ŋ : Ø	(3) *bakukuŋ : (3) *bakuku 'fish sp., sea bream'
7)	c : t	(3) *cekcek : (2) *tektek 'gecko'
8)	mb : nd	(3) *kambiŋ : (3) *kandiŋ 'goat'
9)	s : Ø	(2) *ŋisŋis : (3) *ŋiŋi 'grin, show the teeth'
10)	s : ñ	(1) *sepsep : (2) ñepñep 'sip, slurp'
11)	aw : u	(2) *qali-maŋaw : (2) *qali-maŋu 'mangrove crab'
12)	i : u	(2) *ma-tiduR : (2) *ma-tuduR 'sleep'
13)	e : i	(1) *esa : (1) *isa 'one'
14)	a : e	(3) *puŋal : (3) *puŋel 'prune, cut off the tip'
15)	i/e : u/a	(3) *bileR : (1) *bulaR 'ocular cataract; dim vision'

These 15 patterns of doubleting are a small sample from what is essentially an open-ended set (Blust 2011a). Through chance a pattern is occasionally repeated in more than one base, but this shows no obvious correlation with meaning. As can be seen, doubleting

affects content morphemes of almost every conceivable type, including nouns, verbs, numerals, and descriptive terms. It includes low-frequency terms for flora and fauna (*Smilax* spp., sea bream), and the most high-frequency types of basic vocabulary (one, sleep). To give some idea of how common lexical doubleting is in reconstructed forms, at least 20% of the 443 bases proposed in Blust (1980b) are cross-referenced to doublets. Although the term 'doublet' naturally suggests that lexical variation tends to be expressed through *pairs* of formally and semantically similar word bases, doubleting can be more exuberant, as with (2) *ipen, (2) *lipen, (1) *nipen, (3) ŋipen 'tooth', none of which contains an identifiable fossilised affix.

Blust (1970a) distinguished 'doublets,' in which there is independent evidence for the reconstruction of variant forms of what might reasonably be called the 'same' base, from 'disjuncts,' in which doublets can be posited only by allowing the overlap of cognate sets. Dempwolff (1938), for example, posited *gumi and *kumis 'beard', but assigned Fijian *kumi* to both etyma. In this case the reconstruction of PMP doublets depends crucially on allowing the Fijian form to have a dual etymology. The two cognate sets thus 'overlap', and doubleting is possible only by arbitrarily assigning Fijian *kumi* to both protoforms. Based just on the evidence considered by Dempwolff, then, *gumi and *kumis are PMP disjuncts, not doublets.

What is different about the pattern of doubleting seen here, and ordinary doublets like English *shirt* : *skirt*, or Tiruray *ratah* : *gatas*, is that the phonological variation seen in reconstructed doublets cannot be explained as a product of borrowing. First, many of these phonological differences, as *d* : *n*, *b* : *l*, *R* : *w*, or *ŋ* : *q*, are rare or unknown as sound correspondences between AN languages. Second, in contrast with ordinary doubleting, there is little recurrence in the patterns of variation. Third, even if these cautions were ignored the most basic constraint on hypotheses of borrowing, namely, that a plausible source language be identified, can rarely be satisfied in any of these cases. Fourth, many attested AN languages, especially those in the Philippines and western Indonesia, show cognate doublets that do not appear to involve loanwords, as with Cebuano *biŋáʔ*/*biŋág* next to Malay *biŋah*/*biŋar* 'k.o. volute shell: *Voluta diadema*'. The reason for such extreme variation and its relative absence in Oceanic languages is unclear, but in general doubleting has no effect on the recognition of recurrent sound correspondences.

The number of doublets recognised in PAN, PMP or PWMP reconstructions would be even larger if bases that share a common submorphemic root were treated in the same way, but this analysis would miss an important generalisation. As noted in 6.2, one of the most striking features of both attested and reconstructed AN languages is the presence of recurrent sound-meaning associations in the last –CVC of longer word bases. This creates a great deal of variation in reconstructions based on a shared submorphemic root, as with (2) *ekeb 'cover up, hide', (3) *kebkeb 'to cover', (2) *Ruŋkeb 'cover over (as a dish containing food)', (3) *se(ŋ)keb 'cover', (3) *si(ŋ)keb 'cover', (3) *ta(ŋ)keb 'cover, overlapping part'. These reconstructions of similar shape and meaning clearly differ from variants of the type seen in Table 8.34, since they exemplify a pattern of recurrent partials, one that is further seen in attested forms for which no etymon is known (Balinese *aŋkeb* 'covering, lid; bedspread, tablecloth', Javanese *dəkəb* 'catch by covering with the hand, as a bird', Paiwan *tsukəv* 'cover', etc.).

Because of the apparent irregularities introduced by doubleting and by the sharing of monosyllabic roots in non-cognate morphemes, lexical comparison in AN is not always a straightforward matter. When we add to this mix widespread loanwords between AN languages that might be mistaken for native forms, it may occasionally happen that proto-

forms are posited that had no historical reality. This was done quite consciously by Dempwolff (1938), who marked known loanwords from non-AN sources with an 'x', but included them for their value in illustrating recurrent sound correspondences, as with ^xkak′a' 'glass' (< Sanskrit). Mahdi (1994a,b) calls such pseudo-reconstructions 'maverick protoforms', and he explores the history of a number of these. However, like Wolff, who disallows any etymon with *c, *g or *r, because he does not believe these phonemes ever existed, Mahdi often goes too far and throws the baby out with the bath, as with the rejection of PAN *paŋudaN 'pandanus', or *bulaw-an 'gold', even though both are supported by widely distributed forms that appear to be native (Blust and Trussel ongoing).

8.4.1 The Proto Oceanic lexicon

Undoubtedly the most spectacular recent achievement in AN lexical reconstruction has been the appearance of the first five volumes of a projected seven volume work, *The lexicon of Proto Oceanic* (LPOC; Ross, Pawley and Osmond 1998, 2003, 2008, 2011, 2013). Unlike the ACD, which is organised alphabetically by highest-level reconstruction, LPOC is organised by semantic field, an arrangement that greatly facilitates teasing out semantic distinctions in the glosses of protoforms. Each volume consists of an interlocking set of semantic fields which together form a larger 'super-field'. Volume 1, 'Material culture', for example, contains the following chapters: 1) 'Introduction', reviewing earlier work, and discussing reconstructional methodology and conventions, 2) 'Proto Oceanic phonology and morphology', 3) 'Architectural forms and settlement patterns', 4) 'Household artefacts', 5) 'Horticultural practices', 6) 'Food preparation', 7) 'Canoes and seafaring', 8) 'Fishing and hunting implements', and 9) 'Acts of impact, force and change of state.' In addition, there are two appendices: 1) data sources and collation, and 2) languages. Individual chapters strive to paint a picture of the life and environment of Proto Oceanic speakers through collections of etymologies that permit reconstructed forms relevant to the semantic field treated (34 cognate sets used in the chapter on 'Architectural forms and settlement patterns', 46 in the chapter on 'Canoes and seafaring', etc.). The results are arguably on a par with the best Indo-European work in this vein, as Benveniste (1973), or Mallory and Adams (1997).

8.4.2 The Proto Polynesian lexicon

Walsh and Biggs (1966) initiated a Proto Polynesian dictionary that initially contained about 1,200 cognate sets and associated reconstructions on various levels, including Proto Polynesian, Proto Nuclear Polynesian and Proto Eastern Polynesian. The material was presented as an alphabetical listing of protoforms together with supporting evidence, similar to that used in the ACD. Work on the comparative Polynesian lexicon, known in its online form as POLLEX (http://pollex.org.nz) has continued for over 40 years, with the result that the project, which is currently administered by Ross Clark of the University of Auckland, "has grown to include over 55,000 reflexes of over 4,700 reconstructed forms in 68 languages" (Greenhill and Clark 2011). An outstanding example of how this data can be used to augment the material available from other disciplines in shedding light on Polynesian culture history is seen in Kirch and Green (2001), an extremely insightful book written by two leading Pacific archaeologists which is guided throughout, and in some cases almost exclusively, by reliance on the reconstructed vocabulary.

8.4.3 The Proto Micronesian lexicon

Bender et al (2003a) have proposed about 900 lexical reconstructions for Proto Micronesian, Proto Central Micronesian, and Proto Western Micronesian. In a subsequent publication they have added a smaller number of forms for more restricted subgroups within Micronesian (Proto Pohnpeic-Chuukic, Proto Pohnpeic, Proto Chuukic), together with an appendix of 91 apparent loanwords that have a fairly wide distribution across Micronesian languages. As with POLLEX, reconstructions are listed alphabetically and given together with supporting evidence. Neither POLLEX nor the Proto Micronesian reconstructions proposed to date are annotated, and unlike the ACD, they do not reconstruct affixed forms.

9 *Sound change*

9.0 Introduction

A survey of sound change in AN languages that is at all adequate must review data in a very large number of distinct communities. Not only are there at least 1,000 languages in this family, but many of these have numerous dialects that may differ in interesting details of phonological evolution. James T. Collins believes that there are up to 65 dialects of Malay scattered from Sri Lanka to Irian Jaya. These exhibit a wide range of typological variation, and reflect many of the sound changes that are found in related languages. While Malay may be exceptional in its dialect diversity, even small languages such as Kelabit, with about 5,000 speakers on both sides of the Sarawak-Kalimantan border, can be dialectally quite complex. Apart from the standard dialect of Bario, at least six other dialects are known from the Sarawak side of the border alone. This does not include Sa'ban, which would be considered a dialect of the same language on lexicostatistical grounds, but which has undergone such rapid and drastic changes as to require a separate status (Blust 2001e). Lun Dayeh, so closely related to Kelabit that it might be regarded as a divergent dialect of the same language, also has a number of dialects. Under these circumstances we face a difficult choice: should we try to look at vast numbers of details which are sometimes fascinating in their own right, or should we concentrate on broad patterns of change, and refer to details only in illustration of these more general developments? For practical reasons the latter choice is made here, but even so the reader will see how long and complex a discussion of sound change is required. The presentation will cover five topics: 1. normal sound change, 2. bizarre sound change, 3. quantitative aspects of sound change, 4. the Regularity Hypothesis, and 5. drift.

Table 9.1 provides an overview of changes from PAN, attributed to the Neolithic founding population on Taiwan around 5,500-6,000 BP, to two key daughter languages: Proto Malayo-Polynesian, spoken in the northern Philippines between 4,000 and 4,500 BP, and Proto Oceanic, associated with the Lapita pottery complex, which appears suddenly in the archaeolgical record of the Bismarck Archipelago around 3,350 BP or slightly earlier. Due to limited data the POC reflex of PAN/PMP *r is unclear.

Table 9.1 Changes from Proto Austronesian to Proto Malayo-Polynesian
to Proto Oceanic

PAN	PMP	POC
*p	*p	*p/p^w
	*mp	*b/b^w
*t	*t	*t
*C	*t	*t
	*nt	*d
*c	*c	*s

PAN	PMP	POC
	*nc	*j
*k	*k	*k
	*ŋk	*g
*q	*q	*q
*b	*b	*p/pʷ
	*mb	*b/bʷ
*d	*d	*r
	*nd	*dr
*z	*z	*s
	*nz	*j
*D	*D	*r
*j	*j	*c
	*nj	?
*g	*g	*k
	*ŋg	*g
*m	*m	*m/mʷ
*n	*n	*n
*N	*n	*n
*ñ	*ñ	*ñ
*ŋ	*ŋ	*ŋ
*S	*h	Ø
*s	*s	*s
	*ns	*j
*h	*h/Ø	Ø
*l	*l	*l
*r	*r	*r(?)
*R	*R	*R
*w	*w	*w
*y	*y	*y
*a	*a	*a
*e	*e	*o
*i	*i	*i
*u	*u	*u
*-ay	*-ay	*e
*-aw	*-aw	*o
*-uy	*-uy	*i
*-iw	*-iw	*i

To recapitulate from Chapter 8, *C probably was a voiceless alveolar affricate [ts], *c a voiceless palatal affricate [ʧ], *q a uvular stop, *z a voiced palatal affricate, *D a voiced retroflex stop (found only word-finally), PAN/PMP *j a voiced palatalised velar stop [gʸ] (but POC *j a prenasalised voiced palatal affricate), *N has nasal reflexes in most languages, but may have been a palatal lateral, the phonetic values of *S and *s are unclear, apart from stating that both were sibilants, *r probably was an alveolar tap and *R an alveolar trill (that became uvular in many languages). POC voiced stops were automatically prenasalised. It should also be noted that PMP permitted both simple and

prenasalised obstruents medially, but this distinction cannot be assigned to PAN. In citing etymologies reconstructions are generally labeled 'PAN', 'PMP', 'POC', etc., but may sometimes shift between levels of temporal reference without explicit indication.

Before looking at particulars it should be noted that the term 'sound change' is often used ambiguously for: 1. a change from one segment, including zero, to another (e.g. *p > *f*), or 2. the cumulative result of changes from one segment to another (e.g. *p > *h*, through intermediate *f). More properly both *f* and *h* should be called 'reflexes' of *p, but only *p > *f* should be called a 'sound change' (arguably, even this may be considered the cumulative product of more fine-grained changes). Since historical linguists work with the results of change and generally must infer individual changes from their end-products the distinction between 'sound change' and 'reflex' can easily become blurred, and often is. In the following discussion every effort will be made to keep this distinction in mind in looking both at broad patterns of change and at individual changes that seem to challenge important aspects of general theory.

9.1   Normal sound change

The cross-linguistic study of sound change as an aspect of linguistic typology has lagged behind other kinds of language universals. Nonetheless, through experience historical linguists have developed widely-shared notions about 'normal' sound change. Many of the sound changes familiar to Indo-Europeanists are found in AN languages, but some widespread sound changes in AN are uncommon or unknown in Indo-European or other well-studied language families, and some widespread sound changes in other language families are uncommon or unknown in AN. We will begin with changes that should come as no surprise to most historical linguists. Needless to say, the distinction between 'normal' and 'aberrant' sound changes is not always easy to maintain, and this section will include certain normal processes of change that involve unusual details of content, especially in relation to dissimilation.

9.1.1 Lenition and fortition

The terms 'lenition' and 'fortition' are not used identically by all linguists. As used here, 'lenition' means a decrease of articulatoruy constriction, or a movement toward zero, and fortition an increase of articulatory constriction, or a movement away from zero.[101] It will be applied mostly to unconditioned sound changes, since some assimilations can be viewed as movements toward zero, and assimilation will be treated separately below. Although lenition can apply to any segment, core lenitive processes apply prototypically to stops. Among the voiceless stops of AN languages two prominent lenitive processes, or 'erosion sequences' can be identified: 1) *p > *f* > *h* > Ø, and 2) *k > *h* > Ø. The first of these can be illustrated with reflexes of PAN *pitu 'seven' and *Sapuy 'fire':

---

[101]   Needless to say, this does not imply that *p or *k will inevitably disappear over time, only that *if* change occurs (and millennia may pass with no change) lenitive changes move a segment in the direction of increasingly weaker constriction until zero is reached. Thus, the sequence *p > f > h > zero may stabilize at any point, but if change resumes again it is most likely to continue rightward in this erosion sequence.

PAN	pitu 'seven'	Sapuy 'fire'
Paiwan	pitju	sapuy
Samoan	fitu	afi
Hawaiian	hiku	ahi
Helong	itu	ai

Figure 9.1 The erosion sequence *p > f > h > Ø

These examples from four languages represent stages that each language presumably has passed through in its evolution from PAN. In other words, it is assumed that Helong did not lose *p directly, but passed through stages in which *p lenited to *f and then to *h before disappearing. Similarly, it is assumed that in Hawaiian *p lenited to *f before becoming *h*. Although we usually lack documentary evidence of such intermediate stages, in some cases evidence for them can be derived from the nesting of subgroups. PAN *p became Proto Polynesian *f, for example, which further lenited in Hawaiian, and Helong, spoken in western Timor, has lost *p, but its close relative Tetun preserves a less lenited reflex in *hitu* 'seven', and *hai* (with metathesis) 'fire', showing that loss of *p probably passed through the stages *p > *f* > *h* > Ø.

The sequence of changes presented here is, of course, a generalised schema. Although it accurately describes the trajectory through which many languages have passed in leniting *p, it does not exclude other change paths. The Polynesian language Rarotongan, for example, reflects PAN *p as glottal stop: ʔitu 'seven', aʔi 'fire'. Here the glottal stop has no intrinsic relationship to the *p from which it originated. Rather, it has developed from a historically anterior *h of multiple origins, as seen from the fact that Rarotongan also reflects PAN *s as glottal stop. Many other Oceanic languages reflect PAN *p as *v* rather than *f*, and some languages (as Palauan) reflect it as *w*. While all of these changes are lenitions of *p they represent forks in the road of phonological development, since *v or *w are less likely than *f to produce *h* as a proximate successor.

A parallel series of changes is seen in the lenition of *k to *h* and Ø. This can be illustrated with reflexes of PAN *kuCu 'hair louse' and *aku 'I':

PAN	kuCu 'hair louse'	aku 'I'
Thao	kucu	y-aku
Toba Batak	hutu	ahu
Roti	utu	au

Figure 9.2 The erosion sequence *k > h > Ø

As with the erosion sequence for *p, some languages have followed lenitive change paths for *k that depart from this schema. In Samoan and Hawaiian, for example *k became glottal stop following the loss of Proto Polynesian *ʔ < PAN *q). Erosion sequences of this kind may remind some readers of that part of Grimm's Law which deals with the voiceless stops. A major difference (apart from the fact that the erosion sequences continue to zero in some AN languages) is that *k > *x* is rare in AN languages. Rather,

unlike PAN *p, which probably has always passed through an /f/ stage on its way to /h/, PAN *k appears to have often changed directly to a glottal spirant. Where a phoneme *x* does exist in AN languages it has usually developed from other sources, as PAN *q (Southern Bunun), PAN *R (Pazeh, Seraway, Tobi), PAN *j (Nias), POC *r (Bipi), or PAN *s (Seimat). The major exception appears to be in Vanuatu, where some of the languages of Malakula and southern Santo reflect *k as *k* or *x* (Tryon 1976). A second difference between the lenition of voiceless stops in AN and that part of Grimm's Law which converted PIE *p, *t, *k to Proto Germanic *f, *θ, *x will be noted below.

In keeping with the distinct behavior of coronal consonants in many languages, PAN *t shows a very different evolutionary trajectory than *p or *k. This stop generally did not change, and when it did change it rarely lenited. Among the few examples of what is arguably lenition rather than assimilation is *t > ʔ in Nusalaut of the central Moluccas, Moor of Irian Jaya, Wuvulu-Aua of the Admiralty Islands, and at least some dialects of Mekeo in southeast New Guinea (PMP *mata > Nusalaut *maʔa* 'eye', Wuvulu *maʔa-ia* 'see', *m-atay > Moor *maʔa*, Wuvulu *maʔe* 'die', *qatep > Nusalaut *aʔo-l* 'roof; thatch', *kutu > Moor *kuʔa* 'louse'), *t > ð (written *d*) in Palauan, where it is unconditioned, *t > s in Micronesian languages such as Chuukese (*mata > *maas* 'eye, face', *pitu > *fisu* 'seven'), and *t > Ø throughout the southeast Solomons (*taliŋa > To'ambaita, Lau, Sa'a *aliŋa-na* 'ear', *m-atay > Toqambaqita, Lau, Sa'a *mae* 'die'). Since the latter languages have also lost *q, which is reflected as a glottal stop in some Oceanic languages, it is possible that loss of *t passed through an intermediate stage *t > ʔ. The change of *t to *h* in Roro of southeast New Guinea (PMP *taliŋa > *haia* 'ear', *taŋis > *hai* 'cry, weep', *tina > *hina* 'mother', *qate > *ahe-na* 'liver', *qateluR > *ahoi* 'egg') appears to be unique; Roro *h* has no other known source, and any intermediate steps that might have occurred in this lenition are obscure. By far the most common lenitive change of *t is the assimilative change *t > s/__ i, which will be treated in a later section.

The last of the PAN voiceless stops is *q. While *q has remained a uvular stop in several Formosan languages, outside Taiwan it almost always became glottal stop, zero, or more rarely *k*, or *h*. The great majority of languages have lost it entirely, perhaps through an intermediate change to glottal stop. Many of the languages of island Southeast Asia which reflect *q as a glottal stop in medial and final position reflect it as phonemic zero initially, either because it has been lost or because initial vowels have acquired an automatic glottal onset. This is true over much of the Philippines, Borneo, Sulawesi, and eastern Indonesia. It is, however, not true of Oceanic languages, where *q always appears as a phonemic glottal stop in initial position if it is reflected as a glottal stop in medial position. The change *q > h occurs in a number of the languages of western Indonesia and mainland Southeast Asia (Chamic, Malay, Acehnese, Northern Batak, Nias, Lampung, Sundanese, Javanese, Balinese), and in Lakalai of New Britain. Somewhat different types of lenition are found in Southern (Isbukun) Bunun, where *q > *x*, and in most dialects of Puyuma, where *q became a voiced glottal fricative.

As noted by Ross (1991), and Blust (1991c), AN languages show a puzzling correlation between degree of lenition and migration distance. To date this correlation has been explored carefully only for *p. If reflexes of *p are given numerical values corresponding to degrees of lenition such that $p = \emptyset$ (no change), $f = 1$, $h = 2$ and $\emptyset = 3$, an 'erosion value' can be calculated for sets of languages that represent major geographical areas. This has been done for 930 languages, and the results are clearly non-random. Table 9.2 provides a simplified picture of the results, using binary rather than scalar values. For

statistical reasons areas such as Madagascar and mainland Southeast Asia, with very small numbers of AN languages, have been omitted:

Table 9.2 Erosion values for PAN *p in languages representing major geographical areas

No.	Area	No. lgs.	% changed
1.	Taiwan	22	10
2.	Philippines	129	17
3.	Borneo	96	3
4.	Sumatra-Sumbawa	21	11
5.	Sulawesi	67	11
6.	Lesser Sundas	46	50
7.	Moluccas	53	95
8.	New Guinea	161	99
9.	Bismarck Archipelago	74	84
10.	Solomons-Santa Cruz	70	100
11.	Micronesia	24	100
12.	Vanuatu	105	100
13.	New Caledonia-Loyalties	28	100
14.	Rotuma, Fiji, Polynesia	23	100

Only about 10% of the 22 Formosan languages sampled have changed *p. This figure rises to 17% in the 129 languages of the Philippines that were sampled, and so on as one moves further south and east. Where all languages show some change to *p, as in the last five regions, lenition probably had begun in their immediate common ancestor, and the statistics are thus compromised by what is commonly known as 'Galton's problem.' But this explanation can be maintained only where a subgroup can be justified on independent grounds. Surprisingly, although most Oceanic languages show some lenition of PAN/PMP *p/b, these phonemes apparently were retained as *p in Proto Oceanic: PMP *hapuy > POC *api > Sori *jap*, Lehalurup *ep*, Mwotlap *n-ep* 'fire', PMP *tabuRi > POC *tapuRi > Sori *dap* 'conch shell', PMP *qapuR > POC *qapuR > Lou *kɔp* 'lime', PMP *habaRat > POC *apaRat > Sori *japay* 'northwest monsoon'. What we find then, is a tendency for PMP *p to show heightened erosion values in eastern Indonesia and the Pacific. Since the historical movement of AN speakers evidently was from Taiwan into the Philippines, with a subsequent split into eastern and western streams, the pattern of erosion seen in Table 9.2 exhibits a high correlation with migration distance. However, there is no obvious reason why this correlation would exist if PMP *p was still a stop in Proto Oceanic. In fact the correlation is imperfect, since the languages of western Indonesia have lower erosion values for *p than those of eastern Indonesia, even though the distance from Taiwan is in some cases (Sumatra, Java, Bali, Lombok, Sumbawa) roughly the same. Perhaps degree of contact with non-AN languages will help to explain these observations, since this could have been a factor in the Pacific and some parts of eastern Indonesia, but is not likely in western Indonesia. Casual inspection of the available data suggests that similar results will be obtained for *k, but the investigation of other segments has not yet been carried out systematically.

Another erosion sequence in AN languages that is common in other language families is *s > *h* > Ø. This can be illustrated with reflexes of PAN *susu 'breast' and *asu 'dog':

PAN	susu 'breast'	asu 'dog'
Bunun	susu	asu
Kambera	huhu	ahu
Kédang	—	au
Hawaiian	ū	—

Figure 9.3 The erosion sequence *s > h > Ø

Although *s > *h* is widespread, it does not appear as common as the lenitions of *p and *k. In some languages *s lenites only word-finally, as in some dialects of Spanish, and here it often produces *-ih* rather than *-h*. This is true of a number of the Melanau languages of coastal Sarawak, and of some of the Chamic languages of Vietnam, as Jarai: *Ratus > Mukah Melanau *ratuih*, Jarai *rətuih* 'hundred', *panas 'hot' > Mukah *panaih* 'flash of anger', *beRas > Jarai *braih* 'husked rice'. In Minangkabau of southwest Sumatra *-s became *-ih* after *u, but not after *a: Proto Malayic *haus > Minangkabau *auih* 'thirsty', but PM *pəras > *parah* 'squeeze, press' (Adelaar 1992).

Most lenitive changes of the voiced stops are commonplace. *b often becomes *v* or *w*, and *d lenites to a liquid, either intervocalically (as in Tagalog and other Central Philippine languages) or unconditionally (as in Oceanic languages). A less common, but surprisingly recurrent change is *b > *f*, found in Tsou, Thao, and Amis of Taiwan, Simalur of the Barrier Islands west of Sumatra, Taje of northwest Sulawesi, Roti, Atoni, Tetun and some Lamaholot dialects of the Lesser Sundas, and in Elat, Buruese and Soboyo of the central Moluccas. Attention to dialect data reveals other examples, as with the Ulibulibuk dialect of Puyuma, which Ting (1978) reports as showing *b >*f*, although other Puyuma dialects show *b > *b*. In Bontok it is part of a larger allophonic pattern in which *b, d, g* lenite syllable-initially (Reid 1976:ix). Although this change may be a shared retention in Roti and Atoni, it appears to be independent in the other languages, and is distinct from reflexes of *p, and usually of *w in all of these languages. The less common change *b > *h* in languages that distinguish reflexes of *p and *b, as in Erai of the Lesser Sundas, probably passed through stages *b > *f* > *h*. Among the rare lenitive changes of *d that are noteworthy is *d > *h* in Chamorro, which may have passed through stages *d > *r* > *h*.

Among the liquids *l shows two recurrent lenitive patterns: to a palatal glide *y*, or to a velar or glottal fricative. Neither of these innovations is common, but they are of interest due to their conditioning environment. The first (*l > *y*) is found in Bare'e (Pamona) and some other languages of central Sulawesi, Palauan of western Micronesia, Nali and Lele of eastern Manus, and Hiw of the Banks Islands in northern Vanuatu, where it is unconditioned. In some Bisayan dialects of the central Philippines *l remained unchanged when adjacent to a coronal vowel or consonant (*i, y, d, t, n* or *s*), but otherwise lenited to *y* or *ɣ* (Zorc 1977:209ff, Lobel 2013:249). The second innovation is also found in Itbayaten and Ivatan of the northern Philippines, where *l remained unchanged adjacent to *i, but otherwise became a voiced velar fricative in Itbayaten and *h* in Ivatan. Adjacent coronal segments, especially *i* thus appear to increase resistance to lenition among laterals. As seen in 4.3.1.6, this is an environment in which the fortition of *l also takes place in several widely separated languages.

A few other languages show *l > *h*, as Kadazan of western Sabah, where *l regularly lenited to *h*, which then disappeared word-finally (*lima > *himo* 'five', *balik > *bahik*

'reverse', *putul > *putu* 'cut'). Tagalog shows a similar change that is sporadic, as in *balay > *báhay* 'house' or *luslus > *hushós* 'slipping downward', next to *l > *l* in most forms. In some Philippine languages, as Northern Kankanaey, Botolan Sambal, Tagalog, Dibabawon Manobo, Samal, Tausug and Sangil intervocalic *l disappeared regularly or sporadically, presumably by first leniting to *h*: *zalan > Tagalog *daán* ([daʔán]) 'road, path', *taliŋa > *taiŋa* ([taʔíŋa]) 'ear' (glottal stop between like vowels or unlike vowels the first of which is low, is automatic). In some of the Wemale dialects of western Seram in the central Moluccas *l disappeared before *i, *u and *o, but remained a lateral before *a or *e (Stresemann 1927:30). Since the change *l > *w* is unknown in AN languages but *l > *y* is recurrent we might conclude that PAN *l was a 'bright' lateral and that the same phonetic property was retained in most daughter languages. This interpretation may help to explain a peculiar sound change in Motu and some other languages of southeast New Guinea, where *l disappeared before high vowels and occasionally in other forms, but *-lu > -*i*: POC *lima > *ima* 'hand, arm', *pulan > *hua* 'moon', *poli > *hoi* 'buy, sell', but *pulu > *hui* 'hair', *sa-ŋa-puluq > *ahui* 'ten', *qatoluR > *gatoi* 'egg', *tolu *toi* 'three'. The less common changes of *l to ɣ or *h* in other languages, however, could be taken as evidence for prior 'dark' *l*.

Lenition of semivowels is relatively uninteresting: prevocalic *w disappeared in some languages of western Indonesia such as Toba Batak, and in many languages of the western Solomons. A similar change occured in Malay except in the sequence *-awa-. *y disappeared unconditionally in Proto Polynesian, and in a number of the languages of the Solomon Islands.

Vowel lenition may be expressed in terms of sonority, in which case raising is a lenition process, since high vowels are more likely than mid or low vowels to disappear from vulnerable positions. However, raising processes in AN languages are generally difficult to associate with increased probability of loss. This leaves centralisation, devoicing and deletion as the primary evidence for vowel lenition.

The weakening and centralisation of unstressed vowels is common to much of western Indonesia and southern Mindanao. Throughout this area, in which stress is predominantly penultimate, antepenultimate vowels are weakened to schwa, a vowel that is shorter than other unstressed vowels in similar positions. In some languages, as Western Bukidnon Manobo of Mindanao, or Minangkabau of southern Sumatra, the lenition of antepenultimate vowels affects only *a, leaving high vowels intact. In other languages all antepenultimate vowels merge as schwa. Because many languages disallow initial schwa in antepenultimate position the first syllable of PMP trisyllables with *a-, *qa-, or *ha- has been lost, as in *qasawa 'spouse' > Kadazan *savo*, Kelabit *awa-n*, Ngaju Dayak *sawe* 'spouse', Kiput *safəh*, Tunjung *saga-n* 'wife' (cp. Tagalog *asáwa*, Chamorro *asagʷa*, Western Bukidnon Manobo *əsawa* 'spouse'). This lenition process (*a/qa/ha- > *a- > ə- > Ø), is found in every known language of Borneo south of Sabah, but not in Malagasy, suggesting that the merger of antepenultimate vowels as schwa took place recurrently over a large geographical area after the departure of the Malagasy from Southeast Borneo. Unlike pretonic vowels, posttonic vowels do not reduce in any language of western Indonesia (where final consonant clusters are unknown). In Palauan, on the other hand, all unstressed vowels reduced to schwa or deleted, and in the Chamic languages, which have become oxytone in adaptating to the dominant Mon-Khmer languages of the region, unstressed vowels following an initial consonant have reduced to schwa or dropped. By contrast, initial high vowels generally became *a-*, and *a- generally remained unchanged (Thurgood 1999:280ff, where it is written *ʔa-).

Vowel devoicing is most common word-finally, and this process tends to affect vowels of lower sonority (*i, u*) first. Nearly all known AN examples are found in Oceanic languages. Rehg (1991) provides a comprehensive survey of final vowel lenition in Micronesian languages. He divides this into two categories: 1. full apocope, and 2. partial apocope and devoicing. Typically, full apocope is realised as the loss of a final short vowel or the shortening of a final long vowel, two processes that can be combined as the loss of a vowel after a vowel followed by zero or more consonants. Partial apocope and devoicing is found in Woleaian—a member of the Chuukic dialect continuum—, and in Gilbertese. In Woleaian phrase-final short vowels are devoiced, and phrase-final long vowels are shortened. In Gilbertese the conditions for apocope are quite complex. First, following a nasal phrase-final short high vowels delete and long vowels shorten. The first of these processes is said to be irregular and the second optional. In addition, Gilbertese devoices phrase-final short high vowels after *t*, which has an allophone [s] in this environment, phrase-final short vowels optionally devoice after other consonants, and phrase-final short non-high vowels optionally devoice after geminate nasals, with mid-vowels being more likely to do so than low vowels. In general, then, there appears to be a clear correlation between devoicing and sonority: the less sonorous vowels are the most likely to devoice and ultimately delete. A similar pattern is seen in Mota of northern Vanuatu, where high vowels that became final have historically deleted, but mid and low vowels have not: POC *kamali > ɣamal* 'men's house', *mata-gu > mata-k* 'my eye', *ñamuk* (> *ñamu*) > *nam* 'mosquito', *pusuR > us* 'bow (weapon)', *qumun > um* 'earth oven', but *papine > vavine* 'woman', *qone > one* 'sand', *pose > wose* 'canoe paddle', *mate > mate* 'dead', *lipon > liwo* 'tooth', *Rumaq > im^w a* 'house', *kuRita > wirita* 'squid, octopus', *pulan > vula* 'moon', *ma-maja > mamasa* 'dry', *rua > ni-rua* 'two'. Other languages of the Banks and Torres Islands that subgroup most closely with Mota, including Hiw, Toga, Lehali, Lehalurup and Mwotlap appear to have lost all word-final vowels, suggesting that apocope in these languages proceeded by stages, starting with the least sonorous segments. Although Polynesian languages are known to permit no final consonants, similar vowel deletion processes are observable in colloquial speech. In colloquial Samoan, for example, final high vowels are commonly devoiced, particularly after certain voiceless obstruents, as in *tasi* 'one'. Finally, Sonsorol, spoken near the western extremity of the Chuukic dialect continuum, has what have been called 'furtive vowels'. Capell (1969:13), who represents them with raised vowel symbols, says that these segments are normally not whispered, but are 'only slightly heard and sometimes not heard at all.' They may occur: 1. as finals, after a consonant (*ɣametak^i* 'sick', *lil^i* 'to marry', *rabut^o* 'snake', *ŋaidir^e* 'edge of canoe'), 2. after a full, generally long vowel, and before a consonant, when they are acoustically similar to falling diphthongs (*ma^il* 'forehead decoration', *ita^il* 'their names'), or 3. after non-final consonants, causing palatalisation or velarisation (with the vowel essentially disappearing). Where etymologies are available these show that the 'furtive' vowel reflects a POC last-syllable vowel that has dropped or devoiced in most other Micronesian languages, as in *masakit > Woleaian *ɣametak^i* 'sick'.

One last question that will reward consideration is 'Do implicational relations exist among lenited consonants?' The easiest way to answer this is by comparing reflexes of *p, *t, *k on the one hand, with *b, *d, *g on the other. Since *t behaves differently from *p and *k in relation to lenition, only the latter two segments need be examined, and since the lenition patterns for voiced and voiceless stops are strikingly different in AN languages, it will be best to treat them separately. Table 9.3 summarises the frequency of patterns in which *p or *k, or both lenite. Patterns above the line show lenition of both *p and *k;

those below the line show lenition of one segment only. Where lenition is found in some lexical items but not others I have counted it as positive:

Table 9.3 Lenition patterns for PAN *p and *k in Austronesian languages

*p	*k	
p, f	h	: Bimanese
p, f	h, Ø	: Miri
f	k, Ø	: Masiwang
f	ʔ	: Bonfia, Kwaio
f	ʔ, Ø	: Nias, Lau
f	h	: Malagasy, Chamorro
f	Ø	: Koiwai, Minyaifuin, Buli, Wuvulu, Taiof
v	ɣ	: Hoava, Roviana, Bugotu, Ghari
v	Ø	: Magori
h	k, ʔ	: Arosi
h	k, Ø	: Roti
h	ʔ	: Soboyo, Hawaiian
h	Ø	: Tetun, Kayeli, Motu, Seimat, Mono-Alu
Ø	ɣ	: Banoni
Ø	Ø	: Simalur, Mussau
p	k, Ø	: Kambera
p	ʔ	: Ifugaw, Gorontalo, Pendau, Manam
p	h	: Toba Batak
p	Ø	: Talaud, Taje, Kemak, Mekeo
f	g	: Tiruray
f	k	: Bilaan, Watubela, Proto Polynesian
v	k	: Gapapaiwa
w	k	: Palauan
y	k	: Marshallese
h	k	: Tboli
Ø	k	: Gilbertese, Kosraean

Table 9.3 is meant to be representative, not complete. For conditioned changes a lenited reflex is cited only where it is found in at least intervocalic position (this turns out to be more relevant for voiced than voiceless stops). Kelabit, for example, reflects *k as ʔ between unlike vowels provided that the first is not schwa, but as *k* elsewhere. For this reason it is excluded from the table. Toba Batak, on the other hand, reflects *k as *h* before a vowel, but as k elsewhere. It is therefore cited as a *p > *p*, *k > *h* language, even though languages such as Chamorro show *k > *h* in all environments (with subsequent loss of *-h*). In Oceanic languages the situation is complicated by the presence of oral grade and nasal grade reflexes, and the further distinction between fortis and lenis reflexes of the oral grade (Ross 1988). Table 9.3 cites only lenis oral grade reflexes.

The results of this comparison are of some interest. As noted above, erosion sequences such as *p > *f* > *h* > Ø or *k > *h* > Ø resemble the lenition of *p, *t, *k in Grimm's Law, but there are important differences between them. The transformation of PIE *p/t/k to Proto Germanic *f/θ/x, for example, is usually treated as a single change that lenited

voiceless stops as a class, but this is not true of AN languages. Not only was *t almost always excluded from this process, but the lenition of *p and *k is usually due to independent changes. The only parallelism between lenited reflexes of *p and *k that might be attributed to a single change is *p > *v*, and *k > γ in Hoava, Roviana, Bugotu, Ghari and other languages of the western and central Solomons. The available evidence thus suggests that the lenition of voiceless stops in AN languages almost always targets individual segments *p or *k rather than the class *p, *t, *k or even the smaller class *p, *k. There is also no clear preference for leniting labials or velars: 16 languages in Table 9.3 retain *p in at least some forms while leniting *k, and 14 languages retain *k in at least some forms while leniting *p. Although 27 languages reflect *k as zero in some or all etymologies, and only 5 reflect *p as zero, this difference may be due to the longer erosion sequence that must be assumed for *p, with two intermediate steps (*f* and *h*) that normally precede loss, as opposed to one intermediate step (*h*) that normally precedes loss of *k. Languages with a voiced reflex of *p (generally *v* or *w*) at Step 1 appear to be much less likely to develop a zero reflex through subsequent change.

Table 9.4 presents a similar data sample for patterns of lenition in PAN *b and *d. Slashes separate initial vs. intervocalic position; commas separate unconditioned reflexes:

Table 9.4 Lenition patterns for PAN *b and *d in Austronesian languages

*b	*d	
b/v	d/z	: Western Bukidnon Manobo
b, w	d, r	: Maranao, Javanese
b/w	d/r	: Sangir
b/v	d/r	: Kayan
b/v	r	: Bintulu
b/β	r	: Nias
β	r	: Ratahan
f	s	: Thao
f	r	: Amis, Simalur, Taje, Tetun, Ujir, Elat, Buruese
f	c, r	: Tsou
f	ts	: Bontok
f	n	: Atoni
f	l	: Bonfia
f	h	: Manusela, Soboyo
f, h	Ø	: Nuaulu
v	d, r	: Paiwan
v	r	: Yami, Kejaman, Malagasy, Bimanese, Fordata
v	l	: Amahei
v	Ø	: Kadazan
w	r	: Tiruray, Tombonuwo, Tondano, Pamona, Proto Bungku-Tolaki, Ende, Sika, Lamaholot, Wetan
ɓ, w	d, r	: Muna
w	s	: Manggarai
w	s, d	: Kédang
w	z, r	: Ngadha
w	d, r	: Kambera, Kodi, Hawu, Kisar
h	d, r	: Dhao
h	n	: Hatue

*b	*d	
h	r	: Erai
h	l	: Nusalaut, Paulohi, Asilulu
b	r	: Saisiyat, Tagalog, Palauan, Numfor
b	l	: Amblau
p	l	: Buli
p	h	: Chamorro
v	d	: Puyuma

These tables show at least four differences in the lenition patterns for voiced and voiceless stops. First, the lenition of voiced stops is more likely to be conditioned, often affecting *b and *d only intervocalically. Second, the voiceless stop that is least likely to show a lenitive change is *t, but the voiced stop that is least likely to show a lenitive change is *g[102]. Third, while zero reflexes of *p and *k are common (especially for *k), zero reflexes of *b and *d are rare (even though *b > *f* is fairly common, and *b > *f* > *h* > Ø might therefore be expected in some languages). Finally, as already noted, in most languages the lenition of *p and *k appears to reflect historically independent changes, while the lenition of *b and *d appears to reflect a single change. Although this may not always be possible to determine, a gross estimate of association is possible by measuring the amount of disagreement between reflexes of *p and *k or reflexes of *b and *d in lenition behavior. Some 24 of 55 languages in Table 9.3, or about 44% (those below the line) show a 'disharmonic' lenition regime. In Table 9.4, by contrast, only 8 of 62, or about 13% show disagreements in lenition.

Like 'assimilation' and 'dissimilation', the terms 'lenition' and 'fortition' form a conceptual dyad which implies that one process is in some sense the 'opposite' or 'reverse' of the other. This view of the relationships between these processes, however, can be misleading, since if fortition were the mirror-image of lenition we might expect erosion sequences such as *p > *f > *h > Ø or *s > *h* > Ø to be reversible when fortition applies, but this is generally not the case.

There are, however, two well-attested cases of *p > *f > *p*. The first of these is seen in Pohnpeian and Mokilese, where POC *p became Proto Micronesian *f and then returned to a stop, as in POC *pali > PMC *fali 'taboo' > PON *pɛl* 'be in a taboo relationship', or POC *piliq > PMC *fili, *filifili > PON *pilipil* 'choose, select' (Bender et al. 2003a). The second is found in the Polynesian Outlier of Anuta, where POC *p became Proto Polynesian *f and then returned to a stop, as in POC *paRi > PPN *fai > ANU *pai* 'stingray', POC *puaq > PPN *fua > ANU *pua* 'fruit', or POC *api > PPN *afi > ANU *api* 'fire'. Although these forms contained *p in PAN, PMP and even POC, it is uneconomical to assume that the Ponapeic languages or Anuta preserve this stop unchanged, since all other Micronesian and Polynesian languages, as well as Fijian and Rotuman—generally accepted as the closest relatives of the Polynesian languages—reflect it as a fricative, glottal stop or zero. This argument could be extended to the relatively few other Oceanic

---

[102]   This is generally true, although *b*, *d* and *g* all have intervocalic fricative variants in some Philippine languages, as Kalamian Tagbanwa, Western Bukidnon Manobo and Sangir. Note, however, that Guinaang Bontok has prevocalic fricative allophones of *b* [f], and *d* [ts], but not of *g*, which in prevocalic position is pronounced as a 'fronted, slightly aspirated kind of *k* similar to the first sound of English 'keep'' (Reid 1976:viii). In the dialect of Mainit and some other villages *b* and *d* have voiceless fricative allophones, but *g* remains a voiced velar stop in all environments.

languages that appear to preserve a reflex of PMP *p as a stop, as with PMP *paRi > Sori *bay* 'stingray', *peñu > *boy* 'green sea turtle', *tapuRi > *dap* 'conch shell', *hapuy > *jap* 'fire', *sa-ŋa-puluq > *saŋop* 'ten', or PMP *paRi > Lou *pe* 'stingray', *peñu > *puon* 'green sea turtle', *panakaw > *panak* 'steal', *puqun > *pu-n* 'base, foundation', *qapuR > *kɔp* 'lime, limestone'. If the Ponapeic languages and Anuta show *p > f > p why not assume the same for the few dozen other Oceanic languages that have a stop reflex of PMP *p, since most of the 450 or more Oceanic languages reflect PMP *p as *f, v, h*, or Ø? Although such a viewpoint might be advanced, it is not currently dominant. In part this is because languages that lenite *p in other environments often show no lenition word-initially in a noun, as with PMP *paŋan > Bipi *hak* 'to feed', *epat > *ha-h* 'four', *tapuRi > *drah* 'conch shell', but *paRi > *pay* 'stingray', *peñu > *poy* 'green sea turtle'. As noted by Ross (1988), many of the languages of the Admiralty Islands show 'secondary nasal grade' reflexes of obstruents as a result of syncope of the vowel and fusion of the nasal of the common noun article *na with the consonant onset of a following free morpheme. If PMP *p had become POC *f, it would be necessary to assume that in languages such as Bipi *f lenited to *h* when it was not prenasalised, but otherwise strengthened to p. Before a stressed vowel, however, a development *na+f > *mf > p is less likely than *na+p > *mp > p or *b*, and it therefore appears that *p, like *k was preserved as a stop in POC, and only later underwent widespread lenition.

In AN languages fortition is most prominent in the development of *s and the glides *w and *y. PAN *s is reflected as *t* in at least two extant Formosan languages (Thao and Paiwan), as well as in several extinct languages of western and northeast Taiwan, in several Philippine languages, including Agta, Ilongot, and Isneg, in Wuvulu and Aua of the Admiralty Islands (with allophonic [ʧ] before high vowels), in the Malaita-Cristobal languages of the southeastern Solomons (where it remained *s* before high vowels but became *t* before non-high vowels), and in many Micronesian languages. In Manggarai of western Flores *s became a voiceless palatal affricate, as in *susu > *cucu* 'breast'. The frequency of such stop reflexes of *s has caused some scholars to speculate that *s was not a fricative, but as already noted, this interpretation creates as many problems as it solves.

Arguably the most interesting fortitive changes in the AN family are those that affected the semivowels *w and *y in prevocalic position and their non-phonemic counterparts between *u or *i and a following unlike vowel. Glide fortition generally yielded voiced labiovelar and palatal obstruents g^w and *j*. However, subsequent change has obscured some of these innovations. Bintulu, of northern Sarawak, has strengthened *w and *y to *b* and *z* respectively, as in *qasawa 'spouse' > *saba* 'wife', and *buqaya (> *baya) > *baza* 'crocodile'. Data in Ray (1913), however, show that the word for 'wife' in Bintulu was sag^wa around 1900. It appears, then, that *w first strengthened by increasing the velar gesture of the labiovelar glide, and only two or three generations later the primary and secondary articulations mutually assimilated to produce a labial stop. Similar changes have occurred in other languages of northern Sarawak, as Miri (*w > *b*, *y > *j*), and Kiput (*w > *f* [f^w], *y > *č*). In Tunjung of southeast Borneo glide fortition and reduction to a consonant without secondary articulation has produced the changes *w > *g* (presumably through earlier *g^w), and *y > *j*. Beyond Borneo similar types of glide fortition are known from some of the languages of the Aru Islands, as Warloy (*w > k^w), and Ngaibor (*w > *g*), Chamorro of western Micronesia (*w > g^w, *y > *dz*), and some languages of western Manus, including Likum, Lindrou (*w > g^w, *y > *j*) and Sori (*w > *g*, *y > *j*). In other Oceanic languages, as Kwaio and Lau in the southeast Solomons, and some languages of Aoba Island in north-central Vanuatu *w strengthened to k^w, but *y did not change. Alune,

spoken in western Seram in the central Moluccas, has strengthened *w to k^w before a vowel. In nouns that acquired a suffix *-e*, even the coda of diphthongs was affected: *kasaw > *ʔasak^w e* 'rafter', *kalaw > *ʔalak^w e* 'hornbill', *labaw > *ma-laβak^w e* 'rat'.

Some of these changes will be familiar from the historical study of other language families. The fortition of *w to g^w is reminiscent of the development of labiovelar glides in the history of Gothic and Icelandic, and the fortition of *w to g^w (or [ɣ^w]) and of *y to a voiced palatal affricate is very similar to the treatment of contemporary English *w* and *y* by many speakers of Spanish in words such as 'Washington' ([ɣ^w ásiŋton]), or 'mayonnaise' ([mǽdʒonez]). The further reduction of g^w to *b* in Bintulu and some other languages of northern Sarawak recalls comparable trajectories of change in 'P-Celtic' languages, and the reduction of g^w to *g* in Tunjung, Ngaibor or Sori parallels the treatment of Germanic loanwords in the history of French.

Perhaps the most surprising aspect of glide fortition in AN languages is its application to non-phonemic glides, a development with no known counterpart in Indo-European, or other well-studied language families. Chamorro illustrates the simplest case. In Chamorro non-phonemic [w] and [j] strengthened to g^w and *dz* just like *w and *y (before rounded vowels *w became *g*): *buaq > *pug^w aʔ* 'betel nut', *zauq > *chagoʔ* 'far', *niuR > *nidzok* 'coconut tree', *ia > *g^w idza* '3sg.'. As the last form shows, labiovelar stops were also added to words that originally had a vocalic onset, a consequence of *w*-accretion followed by glide fortition, hence *ia > *wia > *g^w idza* (Blust 2000c). Before initial glide accretion was possible some consonants had to be lost, as with PAN *Sapuy (> PMP *hapuy > *api > *wapi) > *g^w afi* 'fire.' The same holds for intervocalic glides, which in some cases could not have developed before the loss of a medial consonant, as with PAN *duSa (> PMP *duha > *dua [duwa]) > *hug^w a* 'two', or PMP *dahun (> *daun > *dawən) > *hagon* 'leaf'. Similar fortitions of historically secondary glides are found in some other languages, as Sori of western Manus, where after the loss of certain initial consonants *w* developed before *o-* or *u-*, and then was strengthened to *g*: POC *onom (> *wono) > *gono-p* 'six', *Rumaq (> *um > *wum) > *gum* 'house'.

Certain cross-linguistic differences in glide fortition are puzzling. Unlike Chamorro, which strengthened phonemic and nonphonemic glides in the same way, Lau strengthened *w to k^w, but did nothing with the similar nonphonemic glide: PMP *wahiR > *k^w ai* 'fresh water', *siwa > *sik^w a* 'nine', but *buaq > *fua* 'fruit', *duha (> *dua) > *rua* 'two'. Even more surprising, in Narum of northern Sarawak nonphonemic glides underwent fortition, but phonemic glides did not: *laqia (> *lia) > *ləjeəh* 'ginger', *tian > *tijiən* 'belly', *quay (> *uay) > *bi* 'rattan', *buat > *biət* 'long', (note that the first syllable has been lost in the last two forms; b reflects [w]), *duha (> *dua) > *dəbeh* 'two', but *ayam > *ayam* 'grandchild', *buqaya (> *baya) > *bayeəh* 'crocodile', *daya > *dayeəh* 'landwards, inland', *kaSiw (> *kahiw > *kayu) > *hayeəw* 'wood, tree', *jaway > *jaweəy* 'face', *pawat > *pawat* 'fruit bat, flying fox', *qasawa 'spouse' (> *sawa) > *awəh* 'wife'.

The etymology *duha > Narum *dəbeh*, Long Terawan Berawan *ləbih*, Bintulu *ba* (*dua > *dəba > *ba*) raises another issue. Most languages of northern Sarawak that have strengthened non-phonemic glides have centralised the high vowel from which they are derived. Together with the recurrent but unpredictable loss of initial syllables these changes produce such bizarre etymologies as Long Terawan *kəbiŋ* 'Malayan sun bear' (a cognate of e.g. Malay *bəruaŋ*, or Taboyan *biaŋ* < *biRuaŋ), and *kəjin* 'durian' (a cognate of e.g. Malay *durian*, or Taboyan *duyan* < *duRian), where *k* reflects *R, the phonetic glide has strengthened, the penultimate high vowel has been centralised, and the low vowel has been fronted under conditions that will be discussed below (hence *biRuaŋ > biguaŋ >

bikuaŋ > kuaŋ [kuwaŋ] > kəbaŋ > *kəbiŋ*; *duRian > dugian > dukian > kian [kijan] > kəjan > *kəjin*, with medial voiced palatal affricate). Bizarre as these etymologies are, they are only marginally the product of bizarre sound changes that fronted low vowels after voiced obstruents and devoiced intervocalic obstuents prior to the fortition of *w and *y. Even without the latter changes these developments in Long Terawan would be nearly as strange (**gəbaŋ, *gəjan).

The centralisation of *i and *u in these and other forms is puzzling since high vowels do not otherwise centralise. Glide fortition and high vowel centralisation might thus be considered aspects of a single complex sound change. The absence of centralisation in etymologies such as *tian > Narum *tijiə*n 'belly' or *duha (> *dua) > Kiput *dufih*, Lahanan *lug^wa*, Bekatan *dug^wo*, 'two', however, suggests that fortition of nonphonemic glides in Narum and Berawan occurred first, and that this change somehow 'marked' the preceding vowel for later centralisation.

One other segment shows interesting fortitive developments. The change *l > *d* before *i is found in Long Terawan Berawan (*lima > *dimməh* 'five', *lalej > *dilən* 'housefly'), and Malagasy (*lima > *dimy* 'five', *kali > *mi-hady* 'to dig'). A superficially distinct but clearly related phenomenon is the lenition of *l before any vowel except *i in a number of the Bisayan languages of the central Philippines (Lobel 2013). In Tonsea of northern Sulawesi *l usually became *d* regardless of the following vowel, although this is not entirely consistent. Stresemann (1927:24ff) also noted a curious set of patterns involving changes in liquids that he used to classify central Moluccan languages as falling into five sound change groups, called the 'lilolo', 'rilolo', 'rirolo', 'diroro' and 'riroro' types. In a language that he called 'Sub-Ambon' earlier *l, *r, and *d merged as *l. This lateral then developed various reflexes conditioned by the vocalic environment. In the 'lilolo type' there is no change to *l. In the 'rilolo type, covering Nusalaut, Sepa and Paulohi, *l became *r* before a high vowel, but otherwise remained unchanged. In the 'riroro type', which includes Amahei, Haruku, Hatusua, Kaibobo and Tihulale, *l became *r* before or after a high vowel, but otherwise remained unchanged. In the 'diroro type', which includes the West Sapalewa languages, the changes were more complex: initial *l did not change before *a, *e and *o; medial *l did not change before *e, but became *r* before *a and *o; both initial and medial *l became *d* before high vowels. Finally, in the 'riroro type', which includes Saparua, Hila and Hitulama, the pattern of 'lilolo type' languages is preserved, but *l often becomes *r* with no statable evidence of vocalic conditioning. While only Stresemann's 'diroro type' is overtly fortitive, it is possible that the change *l > *r* adjacent to a high vowel began as *l > *d* with subsequent lenition.

Other fortitions of *l are rare. Selaru of the southern Moluccas shows *l > *s* adjacent to a high vowel (PMP *lima > *sim-* 'five', *luheq > *su* 'tears', *talih > *tasi-* 'rope', *qulu > *usu* 'head/headwaters'), but *l > *l* elsewhere (*dalem > *rala* 'inside', *salaq > *sal* 'wrong, in error'), a change that is otherwise unknown. Finally, the change of *l to *t* syllable-finally in Chamorro may have passed through a stage in which it first became a voiced alveolar stop. The possibility that this innovation involved a direct transition from *l to *t* cannot be ruled out, however, as it also affected Spanish loans, where both *l* and *r* were replaced by *t*: *átgidon, átgodon* 'cotton' (Span. *algodon*), *atmas* 'weapon, fire-arms' (Span. *armas*); *debet* 'haggard, debilitated' (Span. *débil*) 'weak'), *rumót* 'rumor' (Span. *rumor*).

9.1.2 Assimilation and dissimilation

Although there are areas of overlap, assimilation is distinct from lenition. When a low vowel is raised and fronted through the influence of a high vowel in an adjacent syllable the change clearly is assimilatory, but it cannot be called lenitive in the sense adopted here, since mid-vowels are no more likely than low vowels to be lost through phonological erosion. On the other hand, the change *p > *f* clearly is lenitive, but it cannot be called assimilatory since it is almost invariably unconditioned.

Many 'normal' sound changes in the AN languages are assimilatory in nature. One of the most widespread of these is *t > *s* before *i. A diachronic change of this form is found in several languages of the northern Philippines, including Agta, Atta, Isneg, and Gaddang. Some of these languages may have inherited this change from a common ancestor, but it appears unlikely that this is true of all of them. In Borneo a similar change is found in Kelabit, where it continues to operate as a synchronic rule, as in *tanəm* 'grave' : *nanəm* 'to bury' : *s<in>anəm* 'was buried by someone'. In northern Sulawesi the same change appears in Bolaang Mongondow (*tandoy* 'bamboo sp.' : *mo-nandoy* 'cook in a bamboo tube' : *s<in>andoy* 'was cooked in a bamboo tube'). Further east, all of the 30-odd South Halmahera-West New Guinea languages reflect *t > s/_i, which presumably had already taken place in their immediate common ancestor. A change of nearly the same form is well-attested in Motu and other closely related languages of southeast New Guinea, such as Kuni, Lala, Gabadi and Doura, where *t spirantised before POC *i and *e (*tina > Motu *sina* 'mother', *tiRom > *siro* 'oyster', *qutin > *usi* 'penis', *qate > *ase* 'liver'). Ross (1988) reports variations on this change in Ubir, Anuki, Dobu and other languages of southeast New Guinea (with *t > *h* before *i in Tawala and a few other languages resulting from the later change *s > *h*), the Mengen languages in New Britain, several languages of the western Solomons, including Uruava and Varisi, in Bulu and Bola of northern New Britain, and in Lihir, Sursurunga and some other languages of New Ireland. Finally, in western Polynesia Tongan and its close relative Niue show *t > s/_i; in Tongan this is invariant, but in Niue it appears to be optional.

Several features of this change are noteworthy. First, *t > *s*/__i usually took place only after *s had changed to some other sound (*t* in many northern Philippine languages, Bolaang Mongondow, and Southeast Solomonic languages, *d* in Motu, *h* in Tongan, generally zero in Kelabit). However, in Proto South Halmahera-West New Guinea and Dobuan *t and *s merged before *i. Second, *t > *s*/__i has the same form as typical palatalisations in many language families, yet the resulting segment is alveolar, not palatal. It is thus a process of assibilation rather than palatalisation, similar to that found in Finnic languages. This deviation from a typical rule of palatalisation possibly follows from the rarity of voiceless palatal affricates, and more particularly, fricatives in AN languages.[103] Third, *t > s/_i is much more common in Oceanic than non-Oceanic languages. There are perhaps four or five historically independent examples of this change in the whole of insular Southeast Asia, but fully twice that number in the Pacific. Since the number of Oceanic and non-Oceanic AN languages is nearly equal, and since sound change is generally much more extensive in Oceanic, the rate at which *t spirantises before a high front vowel appears to correlate with the overall rate of phonological change.

---

[103]   Among the rare examples of palatal fricatives is the segment written *j* in the Polynesian Outlier of West Futuna. Palatal affricates are not uncommon in western Indonesia, the Admiralty Islands, and New Caledonia, but are rare elsewhere. Ross (1988) also gives examples of the palatalisation of *t before a high front vowel in Petats, Halia, Selau (a divergent dialect of Halia), and some other languages of the western Solomons. My own fieldnotes show *t* > [ts]/_i in Selau.

Unlike Indo-European languages, where velars are more likely than alveolars to palatalise before front vowels, the change *k > č/_i is almost unknown in AN languages. There is also little tendency for languages to show a parallel change for *d, whether or not they change *t to *s* before *i. In view of typical palatalisation patterns in other language families, these negative patterns in AN languages are puzzling. Nonetheless, a few languages in western Indonesia that have not undergone the change *t > s/_i, including Malay, Ngaju Dayak and Kayan, have palatalised *n before a high front vowel. In Long Jegan Berawan of northern Sarawak word-final *t and *k have both palatalised *after* *i, but this is rare: *sakit > *cakəic* 'sick, painful', *betik > *bətiəc* 'tattoo'.

In some languages of the eastern Admiralty Islands including Nali, Ere, Lele, Ponam, Pak, Lou, Penchal, Lenkau and Nauna, in Mendak (= Madak) of central New Ireland, and in some of the languages of central and southern Vanuatu, the change *t > r has occurred in intervocalic position in a manner reminiscent of flapping in American English. Ross (1988) reports a similar change in some of the languages of the Markham Valley of northeast New Guinea in prevocalic position. Somewhat more common, but still rare is the change of intervocalic voiced stops to fricatives, which usually remain as allophones of the stops. Examples are seen in Western Bukidnon Manobo of the southern Philippines, where intervocalic spirantisation of *b* to [v] and *d to [z] remains as a synchronic residue (*babaw > *bavəw* 'top surface' : *di-vavəw* 'upon, over, above', *dumpəl* 'to dull the cutting edge' : *mə-zumpəl* 'dull, of a cutting edge'), in Tagalog, where intervocalic weakening of *d to [ɾ] also remains as part of the synchronic grammar (*bukid > *bukid* 'remote mountain areas' : *ka-bukir-an* 'farmland, fields'), and in Bintulu, Kayan, Kejaman and some other languages of central and western Borneo, where no synchronic residue remains. Changes of this kind can be viewed either as lenitions or as assimilations.

Liquids show assimilatory changes in a few languages. Two patterns are noteworthy: 1) sequences of dissimilar liquids assimilate to produce identical liquids, 2) *l becomes *n* if there is a nasal in an adjacent syllable. The first type of liquid assimilation is known only from a few languages in the southern Philippines and western Indonesia, and always involves the change *lVr to rVr. This is seen in Tiruray, of southwest Mindanao, where *l assimilated to a following *r* from more than one historical source: *bulud (> *bulur) > *burur* 'hill', *lujan (> *luran) > *ruran* 'load something', *luqar > *ruwar* 'loose, spacious'. Kelabit of northern Sarawak, which normally reflects *l as *l*, has assimilated earlier *l...r sequences (from PMP *l...R) to *r...r*, as in *aluR > *arur* 'flow, current', *liqeR > *riʔər* 'neck', or *qateluR > *tərur* 'egg'. A similar conditioned change is seen in Maloh of west Kalimantan (PMP *laRiw > *rari* 'run', *libaR > *ribar* 'wide', *luqar > *ruar* 'outside' next to *lima > *lima* 'five', *talih > *tali* 'rope', etc.), and in Toba Batak of northern Sumatra (*raraŋ* 'forbidden', next to Malay *laraŋ*, *rura* 'valley' next to Malay *lurah*, etc.). Most Philippine languages have a single liquid, and the rarity of liquid assimilation in the Philippines can be explained in part by this fact.

The second pattern of liquid assimilation is found almost exclusively in Oceanic languages, where it is fairly widespread and generally sporadic. Mussau, of the St. Mathias Islands north of New Ireland, reflects *l as *n* in POC *lima > *nima* 'hand, arm', *roŋoR (> *loŋoR) > *noŋo-noŋo* 'to hear', but not elsewhere (*laqia > *laia* 'ginger', *laŋo > *laŋo* 'housefly', *pale > *ale* 'house', etc.). A similar change occurs in reflexes of POC *lima 'five' and/or *taliŋa 'ear' in several languages of the Solomon Islands, including Langalanga, Kwaio, Sa'a and Arosi, and in some languages of New Caledonia, including at least Pije, Fwâi, Nemi and Jawe. Tongan shows the change *l > *n*, apparently just in

*lima > *nima* 'hand, arm'. Since Mussau and Tongan reflect POC *lima 'five' with no change of *l, the assimilation in 'hand, arm' may already have occurred in Proto Oceanic.

Sibilant assimilation was described under 4.3.1.2. as a recurrent assimilatory change found in three geographically and genetically separated Formosan languages, but only limited examples of its historical development were given there. Although the phonetic difference between PAN *s and *S is unknown, it is assumed that both were sibilants. Where these segments were found in the same morpheme they assimilated in the history of Saisiyat and Paiwan. In Thao, which has an unusually large number of fricative phonemes, *s*, *sh* (voiceless palatal fricative), *c* (voiceless postdental fricative), *z* (voiced postdental fricative) and *lh* (voiceless lateral) all interfered with one another historically through a process that—by an extension of the conventional natural class of sibilants—can be called 'sibilant assimilation' (Blust 1995b). The unconditioned reflexes of PAN *C, *d, *z, *S, *R and *j are Thao c [θ], s [s], s, sh [ʃ], lh [ɬ], and z [ð] respectively, but where two of these reflexes are expected in the same morpheme one assimilates to the other:

Table 9.5 Sibilant assimilation in Thao

PAN	Expected form	Actual form	Meaning
*CaqiS	caqish	shaqish	to sew
*dakeS	sakish	shakish	camphor laurel
*daqiS	saqish	shaqish	face
*daRa	salha	lhalha	Formosan maple
*diRi	silhi	mu-lhilhi	to stand
*zaRum	salhum	lhalhum	needle
*Sidi	shisi	Sisi	Formosan serow (wild goat)
*baRuj	falhuz	falhuz/falhulh	Formosan green pigeon

Sibilant assimilation in Thao persists synchronically in three different forms. First, it leads to allomorphy in some affixes with a sibilant phoneme when they are added to bases that contain a different sibilant: *masa-shdu > masha-shdu* 'agree with, have the same point of view as someone else' (progressive assimilation), *shi-suhuy > si-suhuy* 'was over there', *shi-sasaz > si-sasaz* 'was old' (regressive assimilation). Second, some bases with dissimilar sibilants show neutralisation on the surface, but the distinction reappears when the sibilants are separated by a syllable nucleus through infixation, as with *s<m>as* 'to deliver' next to *sh<in>as-ik* 'I delivered it' (hence base *shas*), or *m<in>i-susu* 'warmed oneself by a fire' next to *sh<m>in-usu* 'warmed oneself by a fire' (hence base *shusu*). Third, as *falhuz/falhulh* suggests, sibilant assimilation shows interpersonal variation or individual speaker variation at different times, as with *filhaq ~ lhilhaq* 'to spit', *ma-lhacas ~ ma-lhalhas* 'cooked, done', and *lh<m>aushin ~ lh<m>aulhin ~ sh<m>aushin* 'to swing'. It is notable that the class of segments affected by this change can be characterised as 'sibilants', since it includes [s] and [ʃ], but is much broader than classic characterisations of this natural class.

Another sound change of general theoretical interest is coronal place assimilation. Standard Malay contrasts *-in* and *-iŋ*, as well as *-it* and *–ik*, but in Sarawak Malay velar nasals and stops have become dental just in this environment: *dagiŋ : dagin* 'meat', *giliŋ : gilin* 'roll', *guntiŋ : guntin* 'scissors', *kambiŋ : kambin* 'goat'; *balik : balit* 'return', *bilik : bilit* 'room', *sisik : sisit* 'fish scale', *tarik : tarit* 'pull'. In other environments no change occurs, showing that final velar consonants have assimilated to the coronal place of articulation of a preceding high front vowel (Blust 1994b). A related change is found in

Sumbawanese of western Sumbawa, where *u fronted only before final dental consonants, as seen in the following comparisons with Malay (preceding the colon): *aŋkut : aŋkit* 'transport', *(h)arus : aris* 'water current', *kəntut : əntit* 'fart', *gugur : gugir* 'fall out', *kukus : kukis* 'steam food', *lamun : lamin* 'if, provided', *ratus : ratis* 'hundred', *(h)ulun : ulin* 'slave, servant', next to e.g. *tutup : tutup* 'shut, close', *minum : inum* 'drink', *hiduŋ : iduŋ* 'nose'. Underlying the differences between these changes, one affecting the place features of consonants and the other of vowels, is a common process: the mutual assimilation of consonants and vowels with respect to the feature [coronal].

Several languages also show the change *u > i in the final syllable, and here the phonetic basis for conditioning is obscure. This change occurs in many languages of the South Halmahera-West New Guinea group, but apparently was not present in their immediate proto language: PMP *manuk > Gimán *manik*, Buli, Waropen *mani* (but Moor *manu*) 'bird', *susu > Munggui, Waropen *susi* (but Pom *huhu*, Wandamen *susu*) 'breast', *tuduR > Gimán *im-tuli* 'to sleep'. In addition, it appears in Bobot (Bonfia), of the central Moluccas, as in PMP *batu > *fati* 'stone', *qulu > *uli-n* 'head', or *susu > *susi-n* 'breast', and in Wetan of the Lesser Sundas, as in *batu > *wati* 'stone', *susu > *ui* 'breast', *asu > *ai* 'dog', *tuktuk > *tuti* 'knock, bump', or *suluq > *uli* 'torch'. Since these changes are historically independent, and no language is known to show the opposite pattern of *u > i only in the penult, they suggest that last syllable position favors vowel fronting.

A change that is ubiquitous in Oceanic, but unknown elsewhere in the AN family, is the tranfer of rounding from a vowel to an adjacent, or sometimes non-adjacent consonant. In POC *Rumaq > Chuukese *im^w a* 'house' rounding is transferred from the vowel to the labial consonant, introducing a new labiovelar nasal. Similar changes are common in the languages of Vanuatu: *Rumaq > Hiw *eŋ^w*, Mota, Navenevene, Narovorovo, Baetora Valpei *im^w a*, Sowa *im^w*, Ngwatua *iŋ^w a*, Seke *im*, Tasmate, Malmariv, Fortsenal *ima* (the last several with subsequent unrounding of the nasal). In other languages rounding is transferred to an adjacent consonant without being lost from the vowel, as in Mate *yum^w a*, Burumba *yum^w o*, Vowa *ni-um^w a* 'house'. In Loniu of eastern Manus rounding spreads rightward through vowels and consonants, as in *kaman* [kaman] 'men's house', but *lo kaman* [lo kom^w an] 'in the men's house.' Similar tendencies for rounding to spread are rare outside Oceanic, but can be seen in such Bornean languages as Sa'ban or Lahanan, where *Cuqelaŋ (> *tulaŋ) > S *hloəŋ*, L *tulwaŋ* 'bone'.

The tendency for rounding to spread rightward in Oceanic languages is only part of a larger tendency for the labial or palatal features of vowels to be transferred to adjacent consonants. This interpenetration of vowel and consonant features has been carried to perhaps the greatest extremes among the Nuclear Micronesian languages. Marshallese has been analyzed as having twelve surface vowels which arise from just four underlying segments distinguished only by height (Bender 1968). The consonants fall into three sets: velarised, labialised, and palatalised. As a result of conditioning the four underlying vowels surface as twelve. Historically, the vowels conditioned the labiality and palatality of the consonants, but the consonants may now be considered the conditioning segments and the vowels the products of their assimilatory influence. While this may seem like an extreme case, the general problem that it raises—how to distinguish the boundaries between consonant and vowel features—is common to many Oceanic languages.

No discussion of sound change would be complete if it were limited to changes that occur. As already noted briefly in connection with palatalisation, in some ways of equal interest is the *absence* of changes which are common in other language families. In many language families, for example, intervocalic stops tend to assimilate to the voicing and

continuative features of adjacent vowels, as in the transition from Latin to Spanish, where voiceless stops became voiced fricatives. Intervocalic voice assimilation, however, is rare in AN languages. Although Oceanic languages often reflect PAN *p and *k as voiced fricatives provided that they do not follow a nasal (in which case they became voiced stops), it is very difficult to find any AN language in which *p, *t or *k has voiced only intervocalically. Among the few promising cases are seen in Bimanese and Hawu of the Lesser Sunda Islands: PMP *qaRta > B *ada* 'slave', *qatay > B, S *ade* 'liver', *kutu > B *hudu*, S *udu* 'hair louse', *mata > B *mada*, S *na-mada* 'eye', *m-atay > B, S *made* 'die, dead', *pitu > B, S *pidu* 'seven', *batu > B *wadu*, S *wo wadu* 'stone'. However, *ma-takut > B *dahu*, S *mə-daʔu* 'afraid', *qateluR > B *dolu*, S *dəlu* 'egg', *tau > B *dou*, S *doʔu* 'person', or *tebuh > B *doɓu* 'sugargane' show that a similar change has occurred in initial position. It is possible that Bimanese *dahu* and *dolu* underwent intervocalic voicing before loss of the stative prefix *ma-, or of prepenultimate initial syllables beginning with *(q)a-. But Bimanese *dou* and *doɓu* find no ready explanation under this hypothesis. Moreover, comparisons such as PMP *hipaR > B *ipa* 'opposite bank', *ma-nipis > B *nipi* 'thin (materials)', or *epat > B *upa*, S *əpa* 'four' show no intervocalic voicing. What the evidence suggests, then, is that Bimanese and Hawu voiced *t in prevocalic position. Intervocalically this change was exceptionless or nearly so, but in initial position it was not, as witnessed by PMP *telu > B *tolu*, S *təlu* 'three', or *tuqah > B *tua* 'old'.

Unlike assimilation, which may reflect the desire of speakers to compress a message into increasingly more compact articulatory form, the motivation behind dissimilation is often far more obscure. It has been suggested that dissimilations assist the hearer by increasing the distinctness of the speech signal, but most dissimilations in AN languages do not fit comfortably into this account. Probably the first widespread dissimilation recognised in AN languages is a change known somewhat misleadingly as 'Eastern Polynesian labial dissimilation,' an innovation that is found in most Eastern Polynesian languages, but is sporadic in Marquesan and rare in Rapanui. In Eastern Polynesian labial dissimilation the first of two labiodental fricatives in successive syllables became *w* and the second became *h*, as in PPN *fafa > Hawaiian, Maori *waha*, Tuamotuan *kaa-vaha*, *taa-vaha* 'carry on the back', *fafa > Hawaiian, Maori *waha*, Rarotongan *vaʔa* 'mouth', *fafie > Hawaiian, Maori *wahie*, Marquesan *vehie*, Rarotongan *vaʔie* 'firewood', *fafine > Hawaiian, Maori *wahine*, Marquesan *vahine*, *vehine*, Rarotongan *vaʔine* 'woman', or *fafo > Hawaiian, Maori *waho*, Marquesan, Tuamotuan *vaho*, Rarotongan *vaʔo* 'outside'. The exact form of labial dissimilation remains unclear; it is attested only before low vowels, and it did not take place before a rounded vowel: *fofonu 'deep, full' > Hawaiian, Marquesan *hohonu*, Maori *ho(o)honu*, Rarotongan *ʔoʔonu* 'deep', *fufu > Hawaiian *huhu* 'wood weevil', Maori *huhu*, Rarotongan *ʔuʔu* 'larva of the beetle *Prionoplus reticularis*, found in decayed timber'.[104]

A second dissimilation of some interest changed the first of two sibilants in successive syllables to *t* (again, extending the term 'sibilant' beyond its customary use in general phonetics). This change affected several of the languages of southern Borneo, including Iban and Ngaju Dayak. In both of these languages sibilant dissimilation affected s: PMP *sasa > Iban *tasa* ('collect material and plait, as palm thatch', *selsel > *təsal* 'regret', *sisir > *tisir* 'comb, *susu > *tusu* 'breast', PMP *salsal > NgD *tasal* 'hammer of a smithy', *susu > *tuso* 'breast'. In Iban the change extended to palatal affricates, as in *tacat* 'incomplete,

---

deformed' (Malay *cacat*), *ticak* 'gecko' (Malay *cicak*), *tuci* 'pure (Malay *suci*, ultimately from Sanskrit), *dajaʔ* 'to hawk, peddle (Malay *jaja*), *dəjal* 'stopper, cork; stop up' (Malay *jəjal*), or *dijir* 'put in a row' (Malay *jijir*). Ngaju Dayak lacks c in native vocabulary, and where it shows dissimilation for *c, as in *tisin* 'finger ring' (Malay *cincin*), this presumably followed the change *c > s, since voiced palatal affricates did not dissimilate: *jəjəl* 'stop up'. It is possible that a similar change is seen in Proto Chamic *tasi 'comb; hand of bananas' (Malay *sisir*), or PMP *susu > Proto Chamic *tasəw 'breast'.

A more ordinary type of change well-known to Romance linguists is liquid dissimilation. In Javanese if sound change produced a sequence of *rVrV* the first rhotic dissimilated to *l*, as in PMP *daRa (> *ra) > Old Javanese *rara* > modern Javanese *lɔrɔ* 'virgin, preadolescent girl', or *duha (> *dua > *ro) > Old Javanese *roro* > modern Javanese *loro* 'two'. Since liquids also condition one another in assimilations, this suggests that the processing of laterals and at least some types of rhotics (alveolar flaps) in sequence is inherently difficult.

Perhaps the most enigmatic type of dissimilatory change in AN languages is low vowel dissimilation. In a number of geographically separated languages within the Oceanic subgroup *a is raised if and only if there is another *a in the following syllable. Low vowel dissimilation was first reported in Nuclear Micronesian languages, where it appears to reflect several historically independent changes. It can be illustrated with data from Marshallese, where it remains an active process in the synchronic grammar (Bender 1969a): POC *ma-sakit > *metak* 'pain', *ma-ramaR > *meram* 'light, luminosity', *kataman > *kejam* 'door', *tama > *jema-* 'father', *mata > *maj* 'eye' : *mata-ña > *meja-n* 'his/her eye', *acan > *yat* 'name' : *acan-ña > *yeta-n* 'his/her name'. A very similar change has occurred in Ere of eastern Manus (*kanase > *kinas* 'mullet', *katapa > *kirah* 'frigate bird', *mata-ña > *mira-n* 'his/her eye', *tama-ña > *tima-n* 'his/her father'), irregularly in Leipon, spoken on an islet off the north coast of Manus (POC *mata-gu > *minde*-w 'my eye', *tama-gu > *time-w* 'my father', *katapa > *kitah* 'frigate bird', but *tamata > *dramat* 'person', *padran > *padr* 'pandanus'), and in several parts of central and southern Vanuatu (Blust 1996a,b, Lynch 2003a). What is puzzling is that this development is recurrent over a wide geographical area, yet within the AN family it is confined to Oceanic languages. The vowel system of POC (*a, *e, *o, *i, *u) differed marginally from that of PAN (*a, *e = schwa, *i, *u), but the distribution of *a was identical in both languages, and it is therefore not at all clear why a drift of this kind would operate within one subgroup of AN while leaving other members of the same language family untouched.[105]

9.1.3 Final devoicing

Final devoicing can arguably be classified as a type of lenition, since the merger of voiced and voiceless stops word-finally usually precedes the reduction of final stops to glottal stop.[106] Final devoicing can also be classified as a type of assimilation to the

---

[105]   Since the first edition of this book went to press Blevins (2009) pointed out that a similar change is found in Alamblak, a member of the East Sepik Hills family of 'Papuan' languages.

[106]   Among the rare exceptions is Bunun, where *p, *t, *k are retained in all positions in all dialects, but *b and *d (which are preglottalized in the modern languages) are retained as voiced stops only in non-final position. In final position *b and *d are reflected as glottal stop or zero in all Bunun dialects except Isbukun, where they are unchanged (Li 1988). A partially similar situation is found in some Atayal dialects, as Squliq, Skikun and Mnawyan, where *b and *p merge as -*p*, but *d becomes -*ʔ*, and *t remains -*t*.

silence that follows word endings. Although some linguists argue against a phonetic basis for this change since word endings in connected speech are not followed by silence, it clearly is recurrent in the world's languages, and Blevins (2006), who cites only one AN example (Malay), argues that this recurrence is a product of phonetic motivation.

Careful attention to historical change shows that final devoicing has occurred in a number of AN languages, often in a less straightforward manner than is theoretically expected. Because it is transitional between lenition, assimilation, and what will be termed 'erosion from the right, left and center' it is treated separately here. Table 9.6 lists all cases of AN final devoicing that can be established with reasonable confidence. It should be noted that *-g is rare; *-b is nearly as rare, evidence often depending on the reflex of PAN *Suab 'to yawn'; *-d and *-j, which merge as *d in many languages, are more common, and are found in several stable forms (*likud 'back', *qelad 'wing', *qañud 'drift on a current', *qulej 'maggot, caterpillar', *pusej 'navel', *panij 'wing', etc):

Table 9.6 Final devoicing in Austronesian languages

Language	Alternation
Atayal	yes
Seediq	yes
Pazeh	yes
Atta	?
Sama-Bajaw	no
Sa'ban	no
Pa' Dalih Kelabit	?
Tring	?
Kenyah	no
Murik	?
Proto Kiput-Narum	no
Lahanan	?
Bukat	?
Maloh	?
Proto Southeast Barito	?
Proto Northwest Barito	?
Proto Malayo-Chamic	no
Gayō	no
Toba Batak	no
Lampung	no
Javanese	yes/no
Sasak	no
Tonsawang	no
Chamorro	no

The following remarks are needed to clarify some of these 24 cases:

Atayal and Seediq

The Atayalic group consists of two languages, Atayal and Seediq, each with a number of dialects. Atayal is conventionally divided into two dialect clusters: Squliq, which is relatively uniform, and Cʔuliʔ, which is much more heterogeneous (Li 1981). All named

dialects below are in the CʔuliʔD group. According to Li (1982b), Proto Atayalic *-b remained -*b* in Mayrinax Atayal, but became -*k* in Palŋawan Atayal and all Seediq dialects, and -*p* in other Atayal dialects. Final *-d devoiced in all dialects of both Atayal and Seediq, being reflected as -*ʔ* (Squliq, Skikun, Mnawyan, Mayrinax), -*t* (Maspaziʔ, Matabalay), or -*c* (all others). Surprisingly, in the Matabalay dialect of Atayal final *b and *d devoiced to -*p* and -*t* but *g remained unchanged; in most other dialects of both Atayal and Seediq final *g (< *R) is generally lenited to -*w*. In both Atayal and Seediq alternations of *p* ~ *b*, *c* ~ *d* and *w* ~ *g* occur before vowel-initial suffixes. Even if *b > -*k* in Palŋawan is due to contact with Seediq, *-b > -*p* must have taken place in at least Proto Squliq, Proto Seediq, and a language ancestral to all CʔuliʔD dialects of Atayal except Mayrinax. The picture for devoicing of final *d is even more complex, suggesting a change *-d > -*ʔ* (without intermediate *t) for Squliq and some CʔuliʔD dialects (Skikun, Mnawyan, Mayrinax), of *-d > -*t* in other CʔuliʔD dialects (Maspaziʔ, Matabalay), and of

*-d > -*c* in Palŋawan Atayal and Proto Seediq. Needless to say, contact has almost certainly played a major role in the diffusion of these changes, and the application of a strict family tree model to determine how many historically independent changes are reflected in this complex set of data may be misguided. Minimally, however, final devoicing probably took place independently in both Proto Squliq and Proto Seediq. In Proto Seediq this resulted in the merger of *p/b as -*k* and *d/t as -*c*; in Proto Squliq it resulted in the merger of *p/b as -*p*, but *t > -*t* and *-d > -*ʔ*. The history of final devoicing in CʔuliʔD dialects appears to reflect a complex process of contact and inter-influence that is yet to be worked out in detail.

Pazeh

Historically Pazeh shows final devoicing, but synchronically alternations of voiced and voiceless base-final consonants are better treated as instances of intervocalic voicing, since voiced-voiceless alternations affect reflexes of both voiced and voiceless stops, as with *sepsep > *zəzəp* 'suck' : *zəzəb-i* 'suck it!', *qaNeb 'close; door' > *a-aləp* 'door', *ta-aləb-i* 'Let's close it!' (Blust 1999a:326, Li and Tsuchida 2001).

Atta

Relatively little has been published on Atta, or Northern Cagayan Negrito. The material in Reid (1971), however, shows clearly that earlier *g from PMP *g, *j and *R was devoiced word-finally. Evidence for similar devoicing of final *b and *d is lacking: *mana:ddak* 'to stand' (cp. e.g. Guinaang Bontok *takdəg* 'to stand up', with monosyllabic root *-zeg), *ikej > *ikak* 'to cough' (cp. *qalejaw > *a:ggaw* 'day'), *timij > *simik* 'chin', *niuR (> *niug) > *niuk* 'coconut', *bibiR > *bibik* 'lip', *bahaR > *ba:k* 'loincloth', *ma-besuR > *mabattuk* 'satiated', *qiteluR > *illuk* 'egg' (cp. *tageRaŋ > *ta:gga:ŋ* 'rib', *uRat > *uga:ʔ* 'vein'). Atta is noteworthy in permitting voiced geminate stops yet undergoing final devoicing, a developmental combination that is disfavored within some theoretical frameworks (Blevins 2006).

Sama-Bajaw

The historical phonology of the Samalan languages is complicated by the presence of at least three distinct speech strata, one consisting of loanwords from Malay, another of Central Philippine loanwords, mostly from Tausug or Waray, and the third of native terms. So far as is known, all Samalan languages show evidence of final devoicing in the native

vocabulary, but permit voiced stops word-finally in the numerous Central Philippine loans. Examples include PMP *huab > Yakan *ohap* 'yawn', *qulej > Mapun *uwot*, Yakan *olet* 'maggot', *pusej > Mapun, Yakan *ponsot* 'navel', and *tuhud > Mapun *tuut*, Central Sama, Yakan *tuʔut* 'knee'.

Sa'ban

Sa'ban provides extensive evidence of final devoicing: Proto Kelabit-Lun Dayeh *aleb > *aləp* 'knee', *ukab > *wap* 'opened', *alud > *alut* 'boat', *lalid > *alit* 'ear', *eleg > *ləp* 'to stop, as work', *ileg > *eləp* 'to separate, divorce', etc. Like Atta, it is of special interest in allowing both voiced geminate stops and final devoicing (Blust 2001e).

Pa' Dalih Kelabit

This dialect of Kelabit differs from most others in showing an incipient process of final devoicing. In some cases this appears to be correlated with the fronting and raising of a preceding schwa, as with Bario Kelabit *puəd* 'navel' : *puət* 'base, bottom', but Pa' Dalih *puit* 'navel' : *puət* 'base, bottom', Bario Kelabit *atəb*, but Pa' Dalih *atip* 'deadfall trap', Bario Kelabit *kəkəb*, but Pa' Dalih *kəkip* 'lid, cover', Bario Kelabit *tənəb*, but Pa' Dalih *tənip* 'cold', Bario Kelabit *tuʔəd*, but Pa' Dalih *tuʔit* 'tree stump'. However, other examples suggest that the raising of schwa to i and final devoicing are independent changes that have intersected one another in forms where they co-occur: Pa' Dalih *kibit* (Bario *kibət*) 'heal', Pa' Dalih *dadim* (Bario *dadəm*) 'cold and shivering', Pa' Dalih *gatil* (Bario *gatəl*) 'itch'. In several cases, particularly with final labial stops, both voiced and voiceless variants were recorded, as with Pa' Dalih *kərib/kərip* 'can, able' (Bario *kərəb*).

Tring

Proto Kelabit-Lun Dayeh *b, *d and *g all devoiced word-finally in Long Terawan Tring, a Kelabit dialect that has for some generations been spoken in close proximity with Long Terawan Berawan. As noted in Blust (1984b) Tring and Sa'ban share two changes that are not known to be present in other Kelabit dialects (complete final devoicing, and *r > l-, -r-, -l). It is thus possible, although uncertain that final devoicing occurred in an exclusive common ancestor.

Kenyah

All known Kenyah languages/dialects show final devoicing, which must therefore be attributed to Proto Kenyah. This covers both languages that are conventionally called 'Kenyah', and some of those known as 'Penan', including at least the Long Labid, Long Lamai and Long Merigam dialects of Penan spoken in northern Sarawak.

Murik

Murik is a close relative of Kayan, spoken over a large area in central Borneo. Neither language has devoiced *-d: *qulej > *uləd > Uma Juman Kayan *ulər*, Uma Bawang, Long Atip Kayan, Murik *ulən* 'maggot, caterpillar'. However, all known Kayan dialects reflect *-b as -v or -m, while Murik cognates have -p. This implies that after the lenition of *-d to -r in Proto Kayan-Murik (PKM), Murik separately devoiced the remaining word-final voiced stops. Known evidence is available only for *-b: PKM *keleb > Uma Juman *kələv*, Murik *kələp* 'tortoise', PKM *kaheb 'capsize, collapse', *ŋaheb 'demolish' > Uma Juman,

Long Tebangan Kayan *ŋahəm*, Murik *ŋahəp* 'demolish, as an old longhouse', and PMP *huab > Uma Juman *uhav*, Long Tebangan Kayan *uham*, Murik *t-uap* 'yawn'.

Kiput

Both Kiput and Narum show final devoicing, which can be attributed to their immediate common ancestor. Kiput is most fully documented (Blust 2002c), with final devoicing most extensively attested for earlier *-d (Proto North Sarawak *alud > *alot* 'boat', *likud > *cut* 'back', *lulud > *lulot* 'shin', etc.), but also including *b (*Ruab > *lufiəp* 'high tide').

Lahanan and Bukat

Both of these languages are spoken in the upper Rejang River basin of south-central Sarawak. They do not appear to subgroup closely, but both show final devoicing of at least earlier *d (< PMP *d and *j): *likud > L *likut* 'back', *qulej > L, B *ulət* 'maggot', *qelad > L *lat* 'wing', *dalij > B *lalit* 'buttress root'.

Maloh

The evidence for final devoicing in Maloh is slender but convincing, since the relevant etymologies cannot easily be attributed to chance or borrowing: *lahud > *i-laut* 'downriver', *tuhud > *liŋku-tut* 'knee', *sumaŋed > *sumaŋat* 'soul, spirit'.

Proto Southeast Barito

Hudson (1967) proposed a subgroup of AN languages in southeast Borneo called the 'Barito Family', and further subdivided into: 1. Barito-Mahakam (Tunjung), 2. West Barito (with Northwest Barito and Southwest Barito as primary divisions), and 3. East Barito (with Northeast, Central-East, and Southeast Barito as primary divisions). Northwest Barito includes Dohoi, Murung-1, Murung-2, and Siang (collectively known as 'Ot Danum'), Southwest Barito includes Ba'amang, Kapuas, and Katingan (collectively known as 'Ngaju Dayak'), Northeast Barito includes Taboyan and Lawangan, Central-East Barito includes only Dusun Deyah, and Southeast Barito includes Dusun Malang, Samihim, Dusun Witu, Paku, and Ma'anyan. Final devoicing is found in all Southeast Barito languages, as in *qulej > Dusun Malang, Samihim, Dusun Witu, Paku, Ma'anyan *ulet* 'maggot, caterpillar', *qelad > Paku, Ma'anyan *elat* 'wing', or *tukad > Dusun Malang, Dusun Witu, Ma'anyan *tukat* 'ladder'. It is absent in Northeast Barito, where *-d > *r*, and is inconsistently reflected in Dusun Deyah (*likud > *likut* 'back', but *qulej > *ulor* 'caterpillar', *tukad > *tukar* 'ladder'). Given this distribution I assume a single change in Proto Southeast Barito. Malagasy, which also shows final devoicing prior to the addition of supporting vowels, presumably inherited this change from Proto Southeast Barito.

Proto Northwest Barito

In addition to the presence of historical final devoicing in all Southeast Barito languages, a similar change is attested in nearly all Northwest Barito languages: PMP *likud > Dohoi, Murung-2, Siang *likut* 'back', *qulej > Dohoi *ulʸat*, Murung-2 *ulot*, Siang *ulʸot* 'caterpillar'.

Proto Malayo-Chamic

Malay shows abundant evidence of final devoicing (Dempwolff 1937:13-45), and the same is true of the closely related Chamic languages (Thurgood 1999). Voiceless final stops that derive from voiced originals do not alternate under suffixation.

Gayō

The historical phonology of Gayō is similar to that of Toba Batak, but this is clearly a result of parallel changes. Final devoicing is seen in etymologies such as PMP *ma-huab > *map* 'yawn', *qañud > *anut* 'drift on a current', and *pusej > *pusok* 'navel'.

Toba Batak

As in Malay, final devoicing in Toba Batak is well-attested (Dempwolff 1934), and did not result in voicing alternations under suffixation. Final devoicing did not occur in the Northern Batak languages (Adelaar 1981), and since these intervene between Gayō and Southern Batak the two sets of innovations must be considered independent.

Lampung

The evidence for final devoicing in Lampung is not abundant, but appears to be valid: PMP *hated 'carry piecemeal' > *atot* 'carry (rice)', *lahud 'downriver, toward the sea', *laoʔ* 'sea', *surud > *suxut* 'ebb, recede, of water', *taŋkub 'cover, lid' > *taŋkup* 'trap' (Walker 1976, Anderbeck 2007).

Javanese

Although Dempwolff (1934-1938) stated that Javanese reflects *-b, *-d, *-j and *-g without devoicing, Nothofer (1975:107ff) showed that this is true only of the western dialects. By contrast, Central Javanese has voiceless stops before pause but their voiced counterparts before a suffix, and Eastern Javanese has voiceless stops both before pause and before a suffix. He illustrates this only for *-d and *-g, and implies in his discussion of *-b that devoicing did not affect the labial stop in any Javanese dialect (1975:142).

Sasak

Sasak shows final devoicing of at least *b and *j: PMP *huab > *uap* 'yawn', *pusej > *pusət* 'navel', *qulej > *ulət* 'maggot, caterpillar', *luluj > *lulut* 'shin, shinbone'. Although it may prove to subgroup closely with Malayo-Chamic, the evidence suggests that it is more closely related to Balinese, which preserves final voiced stops (Adelaar 2005c). The changes in Malayo-Chamic and Sasak are thus assumed to be independent.

Tonsawang

Tonsawang of northern Sulawesi has [ʔp], [ʔt] and [ʔk] as word-final allophones of *b*, *d* and *g* (Sneddon 1978:54ff). Preglottalisation identifies these voiceless stops as underlyingly voiced, even without alternations with their voiced counterparts.

Chamorro

PAN *b unconditionally became Chamorro *p*, masking any final devoicing that might have occurred, and there are few reflexes of *g. However, *d became *h* prevocalically, and

was lost word-finally: *daRaq > *haga?* 'blood', *dalem > *halom* 'in, into', *qudaŋ > *uhaŋ* 'shrimp', but *lahud 'seaward, downriver' > *lagu* 'north (in Guam); west (in Saipan)'. Since *-t was not lost, this suggests that *-d never devoiced. Moreover, since Chamorro shows clear evidence of final devoicing in reflexes of *R (*Rabut > *gapot* 'pull out', *uRat > *gugat* 'vein, tendon', but *deŋeR > *huŋok* 'to hear', *niuR > *nidzok* 'coconut palm', *peRes > *foks-e* 'squeeze, press out'), it evidently developed final devoicing after *d > *h* > Ø word-finally, and *R > *g*.

In addition to these cases some other languages do not permit word-final voiced obstruents, but etymologies that support an inference of final devoicing are difficult to find. In Manggarai of west Flores, for example, voiceless stops, nasals, liquids and *s did not change in final position, but reflexes of *-b, *-d, *-j or *-g are elusive, and the absence of final voiced obstruents suggests, but does not demonstrate a history of final devoicing. In other languages voiced and voiceless stops merged as glottal stop, as in Proto South Sulawesi (Mills 1975). Although it is possible that final voiced stops first became glottal stop, and were later followed in this change by voiceless stops, it is far more likely that final devoicing preceded the lenition of all final stops to glottal stop.

As with some other types of phonological change examined here, the distribution of final devoicing in AN languages is geographically skewed. Devoicing is rare in Philippine languages, especially when considering that the Sama-Bajaw languages are an adventitious group that almost certainly arrived in the southern Philippines from Borneo within the past 1,200 years (Pallesen 1985, Blust 2007d). By contrast, at least eleven historically independent cases of final devoicing are known from languages in Borneo (more if we include Sama-Bajaw, or the indigenous Malayic languages of southwest Borneo). Apart from Tonsawang no examples have been confirmed from Sulawesi or eastern Indonesia, although the loss of final consonants in many of the languages of Sulawesi presumably included a stage in which voicing distinctions for obstruents first disappeared word-finally. Most Oceanic languages have lost all final consonants, and have unconditionally merged *b/p, *g/k, and *j/c/s/z. Where a reflex of *-d is preserved through the addition of a supporting vowel, however, it does not show devoicing, as with *teka lahud > Roviana *togarauru* 'wind from the north-northwest'. Finally, it is striking that all Formosan cases of final devoicing have resulted in synchronic voicing alternations, whereas this is very rare elsewhere in AN, being reported only for Central Javanese, but not for western or eastern dialects. Since it is a reasonable supposition that final devoicing initially produces voicing alternations, it is likely that the leveling of voicing alternations has happened repeatedly through independent historical changes.

9.1.4 Erosion from the right, left and center

Erosion sequences such as *p > *f* > *h* > Ø are often unconditioned changes that affect phoneme inventories, but have no effect on canonical shape (until the final stage). By contrast, erosion from the right, left and center is conditioned. Initially erosion from the right and left may simply increase allophony, as by adding a synchronic rule of final devoicing, but over time the accumulation of such lenitive changes will have an effect on canonical shape, as by eliminating closed syllables. Unlike erosion from the right and left, erosion from the center tends to have immediate consequences for canonical shape, giving rise to medial consonant clusters that were not previously permitted in non-reduplicated bases. Over time these can be simplified, eliminating the canonical innovations that were introduced by syncope. Erosion from the right and left, then, tend to begin with non-

canonical effects and to move toward canonical change, while erosion from the center tends to begin with canonical effects and to move toward non-canonical change.

9.1.4.1 *Erosion from the right*

As noted already, the wearing off of word endings, commonly called 'erosion from the right', can be seen as at once a type of lenition and a type of assimilation. When stops devoice before word- or syllable-boundaries, the histories of many languages show that these changes are lenitive, as they often precede the merger of final stops to glottal stop, and the subsequent merger of glottal stop with zero. Since these changes are limited to a particular position it can also be argued that they represent assimilations to a non-segmental condition (in this case silence at word boundaries). Erosion from the right can thus be seen as intermediate between lenition and assimilation.

One of the most striking typological differences between AN languages with regard to morpheme structure concerns the range of permitted final consonants. As noted in Chapter 4 (Table 4.28), AN languages can be arranged along a cline from those that permit all consonants in final position to those that allow only open syllables. This is a historical product of the weakening and loss of final consonants. Erosion from the right is a natural, and hence recurrent development that over time may eliminate closed syllables. But it is not enough to observe that some languages have lost final consonants. The geographical distribution of such languages is skewed in ways that resist explanation. None of the languages of Taiwan, the Philippines, Borneo, mainland Sumatra, Java, Bali or Lombok, for example, have lost final consonants, but this is rather common in central and southern Sulawesi (Sneddon 1993), in parts of eastern Indonesia, and in Oceanic languages.[107]

Careful comparative analysis has shown that this change is recurrent in areas where it is common, as in Sulawesi (Mead 1996). This tendency is carried farthest in the Pacific, where perhaps 90% of the more than 450 OC languages have lost all final consonants. In some cases the original last-syllable vowel has further eroded, giving rise to new closed syllables, as in the languages of the eastern Admiralty Islands, the Micronesian languages, or most languages of Vanuatu. Although it was not initially recognised in the literature, POC retained almost all final consonants from PMP, apparently losing only *h. Final consonant loss in Oceanic languages is thus evidently a recurrent change, as it is in Sulawesi. In short, Oceanic languages as a group show a markedly higher degree of erosion from the right than is typical of other AN languages. This difference agrees with the profile for other types of erosion noted earlier, as in the erosion sequence *p > *f* > *h* > Ø, where the highest erosion values are found in eastern Indonesia and Oceanic. These facts have occasionally inspired claims for a subgrouping connection between the languages of Sulawesi and those further east, but clearly cannot be used for such a purpose.

Blevins (2004a) has drawn attention to a puzzle in connection with final consonant loss in Oceanic languages. General theoretical expectations are that final consonant loss proceeds by steps (first final devoicing, then stops merge as glottal stop and nasals as *ŋ*, etc.), yet all known Oceanic languages in which a lost final consonant reappears under suffixation show no lenition of that segment beyond what happened through unconditioned sound change, as with PMP *inum > Samoan *inu* 'drink' : *inu-mia* (not **inu-nia or **inu-ŋia) 'be drunk by someone', or *qutup > Samoan utu 'submerge a container to fill it' : *utu-*

---

[107]   Blevins (2004a:210) states that 'CMP is the only major Austronesian subgroup where final C-loss is unattested.' She nonetheless cites Bimanese, Ngadha, Ende and other languages of Flores, and Sawu as languages that have lost all final consonants, but mistakenly labels them 'Western Malayo-Polynesian'.

fia (not **utu-ʔia or **utu-hia) 'be filled by submerging'. The first of these forms shows no reduction of *-m to *n or *ŋ, which would be the usual path followed where final nasals have been lost. While the second form shows a lenitive change of *p > f, this change happened unconditionally in Proto Polynesian, and it would be expected that true erosion from the right would show further lenition. The implication of this kind of pattern in many Oceanic languages, is that final consonants were dropped abruptly, without intermediate stages of lenition. As Blevins points out, the abrupt loss of final consonants is not a change that can be called phonetically motivated, and the fact that it has happened repeatedly in the history of Oceanic languages constitutes a still unsolved mystery.

Finally, although changes to the PAN sequences *-ay, *-aw, *-uy and *-iw in most AN languages fall under the general rubric of 'monophthongisation', an unusual pattern of erosion from the right in which the syllable coda is lost with no change to the nucleus is found in western Taiwan and eastern Indonesia. Within Taiwan it can be seen in the rather poor materials available for three extinct languages of the western plains: *m-aCay > Hoanya *maθa* 'die', *Cumay > Babuza *choma* 'bear', *pajay > Babuza *adda*, Papora *pada*, Hoanya *padza* 'rice plant', *ma-Cakaw > Papora *matsāha* 'steal', *babuy > Babuza, Papora, Hoanya *babu* 'pig', or Sapuy > Papora *tapū*, Hoanya *dzapu* 'fire' (Tsuchida 1982). In addition, it is an ongoing change in Thao, as seen in *Sapuy > *apuy ~ apu* 'fire', *i-saháy* : [isáj] ~ [isá] 'there', or *ma-cuaw* : [maθuwaw] ~ [maθuwa] 'intensive marker; very' (Blust 2003a:32). This change, which can be called 'diphthong truncation', is best attested in languages of western and central Flores, as Manggarai, Ngadha and Lio, and in languages of the central Moluccas, as Elat, Watubela, most of the languages of Seram and Ambon, Buru (but not the Sula Archipelago) and Sekar and Koiwai of the Bomberai Peninsula of New Guinea. It appears to be completely absent from the languages of Timor, Roti, Hawu and Sumba, and occurs only occasionally as an irregular change in Lamaholot, Erai, Kisar and Leti (Blust 1993b:265). Within eastern Indonesia diphthong truncation is thus attested in geographically discontinuous areas, with total absence (full monophthongisation of diphthongs) in the central zone. Table 9.7 provides examples of diphthong truncation in Manggarai and Ngadha of Flores, and in Masiwang and Buruese of the central Moluccas. PMP *b<in>ahi is assumed to have developed an intermediate form *binay before this change. Cognate forms that do not show diphthong truncation appear in parentheses; non-cognates are omitted:

Table 9.7 Diphthong truncation in four Central-Malayo-Polynesian languages

PMP	Manggarai	Ngadha	Masiwang	Buruese	English
*qatay	(ati)	(ate)	yata-n	—	liver
*matay	mata	mata	mata	mata	die
*ma-Ruqanay	rona	—	mnana	ana mhana	male
*b<in>ahi	wina	(fai)	vina	ana fina	female
*qenay	—	əna	yəna	ena	sand
*lakaw	(lako)	laa	—	—	walk, go
*-labaw	(bəlawo)	—	mi-lava	—	rat
*takaw	(tako)	naka	manaa	ʔnaka	steal
*qalejaw	(ləso)	ləza	—	—	day
*naŋuy	—	naŋu	naku	(naŋo)	swim

A priori it would seem natural to relate diphthong truncation to erosion from the right. In Ngadha this works, since all final consonants were lost, and it might be argued that

diphthongal codas were treated like any final consonant (although a few forms, such as *qatay > *ate* show monophthongisation). While this works in Ngadha and to some extent in Buruese, it is very difficult to make it work in Manggarai, which shows almost no final consonant loss. Similarly, Masiwang retains most final consonants, as do Babuza, Papora, Hoanya and Thao. Reflexes of medial *y and *w (the onset of final syllables) also differ from the pattern of final glide loss in most languages: *layaR > Manggarai *lajar*, Ngadha *ladza* 'sail', *hawak > Manggarai *awak*, Buruese *awa-n* 'waist'. In conclusion, diphthong truncation in both Formosan and CMP languages must be considered a change that targeted sequences of vowel + final glide, rather than a by-product of more general changes such as final consonant loss or glide lenition.

9.1.4.2 *Erosion from the left*

In AN languages erosion from the left generally operates on vowels rather than consonants. A number of languages reaching from Mindanao in the southern Philippines through the entire island of Borneo, and including Malay and a few other extra-Bornean languages of western Indonesia, weaken vowels and merge them as schwa in antepenultimate position. This process appears to have happened in four steps, one of which was repeated: 1. PMP *a merged with schwa, 2. *h- and *q- disappeared, 3. antepenultimate schwa was dropped word-initially, 4. high vowels merged with schwa, 5. step 3 was repeated. Steps 1 and 2 occur in either order. The first change in this sequence is seen in Western Bukidnon Manobo of Mindanao, where it has restructured some forms, but has left a synchronic alternation between *a* and *ə* in others: PMP *qanibuŋ > *ənivuŋ* 'a palm: *Oncosperma* spp.', *qanunaŋ > *ənunaŋ* 'a tree: *Cordia dichotoma*', *qasawa > *əsawa* 'spouse', *balatik > *bəlatik* 'bamboo spear trap', *taliŋa > *təliŋa* 'ear'; *amut* 'to contribute' : *əmut-aʔ* 'a contribution', *apuʔ* 'grandparent, grandchild' : *əpuʔ-an* 'line of descent', *balak* 'of persons, to come together' : *bəlak-an* 'a crossroads'. In Borneo PMP *h- and *q- probably had already been lost when antepenultimate *a merged with schwa. Given this chronological ordering we can thus speak more generally of the loss of antepenultimate initial *a.

In Western Bukidnon Manobo the contrast between antepenultimate *a* and schwa is neutralised, but canonical shape is not affected. This incipient tendency to erosion from the left in the southern Philippines is carried one step further in Borneo, where every known indigenous language (but not recent immigrant languages from the southern Philippines, such as Iranun) has passed through stages 1 and 2. As a result of this change, words that began with antepenultimate *a- (following *h-, *q- > Ø) have lost the first syllable. Syllable loss did not normally occur if a word began with a non-laryngeal consonant, or under any condition in disyllables. A number of morphemes could be used to illustrate this process, but since these have very unequal retention rates and would present a fragmentary picture, a single form which tends to be preserved in most languages will suffice. Reflexes of PMP *qasawa 'spouse' from widely separated parts of the island of Borneo include: (Sabah) Rungus Dusun, Kadazan *savo*, Ida'an *sawa*, (northern Sarawak) Lun Dayeh *awa-n*, Kiput *safəh*, Bintulu *saba*, (central Sarawak) Mukah Melanau *sawa*, Uma Juman Kayan *hawa-n*, Lahanan *sawa*, (southern Sarawak) Kuap *sawɨdn*, (southeast Kalimantan) Kapuas *sawa*, Tunjung *saga-n* 'spouse'. Some languages in Borneo, as Miri of northern Sarawak, have shifted schwa back to *a*, and others, as Rungus Dusun of northern Sabah, have shifted schwa to *o*, but must have passed through stages 1) and 2), since the initial syllable of trisyllables that began with *a-, *qa-, or *ha- has been lost: PMP *taliŋa > Miri *taliŋah*, Rungus Dusun *toliŋow* 'ear', but *qasawa > Miri *abah*, Rungus Dusun *savo* 'spouse'.

Dempwolff (1934-1938) failed to understand this change and consequently reconstructed such forms as *qanibuŋ 'a palm: *Oncosperma* spp.' and *qasawa 'spouse' as disyllables *nibuŋ and *sawa, marking the Tagalog reflexes *aníboŋ* and *asáwa* as containing an unidentified (and spurious) prefix *a-*.

As these examples suggest, and further evidence confirms, with a single exception every reported language of Borneo south of Sabah has lost the first syllable of trisyllables that began with PMP *a, *ha- or *qa-. The one known exception is instructive. It is now almost universally accepted that Malagasy reached Madagascar from southeast Borneo sometime after the seventh century AD. Unlike Bornean languages, Malagasy preserves antepenultimate *a-, *ha- and *qa- as a-: *qapeju > *aféro* 'bile, gall', *qanibuŋ 'a palm: *Oncosperma* spp.' > *anívona* 'palm used in housebuilding', *qateluR > *atódy* 'egg'. This implies that the loss of antepenultimate *a-, *ha- and *qa- is an areal feature. The Malagasy evidently departed at a time when some languages in southeast Borneo still preserved the initial syllables of such forms, and they consequently escaped the areal leveling that affected all other known languages. Although Malagasy phonology is in general innovative, this is a phonological conservatism that stands out. Moreover, as seen in Chapter 7, it agrees with the conservative nature of the Malagasy voice system, another typological feature that has been lost in Borneo south of Sabah through widespread drift since the Malagasy migration.

Minangkabau and Malay, which are closely related languages or divergent dialects of the same language, can be used to illustrate the progress from stages 1 and 2 to stages 3 and 4. Proto Malayic, immediately ancestral to both languages and some others, had already passed through stages 1) and 2), thus losing antepenultimate initial syllables that began with *a-, *qa- or *ha-, as seen, for example, in PMP *qanunaŋ > Proto Malayic *nunaŋ > Minangkabau, Malay *nunaŋ* 'a tree: *Cordia dichotoma*' (Adelaar 1992). Following the breakup of Proto Malayic, Minangkabau generally preserved antepenultimate high vowels, but Malay merged them with *a as schwa: Proto Malayic *kuliliŋ > Minangkabau *kuliliŋ*, Malay *kəliliŋ* 'go or turn around', *sumaŋet > Minangkabau *sumaŋeʔ*, Malay *səmaŋat* 'spirit, soul', *biRuaŋ > Minangkabau *biruaŋ*, Malay *bəruaŋ* 'sun bear', *tiŋadah > Minangkabau *tiŋadah*, Malay *təŋadah* 'look upward.' Minangkabau has thus reached stage 3, but Malay has reached stage 5.

Many languages in eastern Indonesia preserve *a in, e.g. reflexes of PMP *taliŋa 'ear', but have lost the initial syllable of *qasawa or other trisyllables that began with a vowel, *q or *h: PMP *qanunaŋ > Manggarai *nunaŋ*, Tetun *nunan* 'a tree: *Cordia dichotoma*', *qanitu > Roti, Asilulu, Buruese *nitu* 'ghost, spirit of the dead', *qasawa > Roti *sao* 'marry; spouse'. This suggests that antepenultimate *a also weakened to schwa in Proto Central Malayo-Polynesian, but was restored to a low vowel following the loss of initial schwa in trisyllables. The alternative is to assume that initial *a was lost in trisyllables without prior lenition, a change that would be phonetically surprising, but would parallel the abrupt loss of final consonants in Oceanic languages.

As with other phonological innovations reviewed here, it remains unclear why erosion from the left has an areal character, affecting many languages in Indonesia, while leaving Formosan languages, Oceanic Languages, and nearly all Philippine languages untouched. In most languages of the northern and central Philippines stress is phonemic, and the antiquity of this pattern remains a major unsolved puzzle. With some notable exceptions, in the southern Philippines and in most languages of Indonesia and Oceania stress is penultimate. The neutralisation of antepenultimate vowels is thus equivalent to pretonic neutralisation. The languages of the southern Philippines that have undergone this change

fall into three groups: 1. Bilic (Bilaan, Tboli, Tiruray), 2. Manobo (Western Bukidnon, Ilianen, Cotabato, Sarangani), and 3. Subanun (Sindangan and Siocon dialects as reported in Reid 1971). Some, like Western Bukidnon Manobo, are at stage 1/2. Others, like Sindangan Subanun, appear to be at stage 3. All Bornean languages have reached at least stage 3, and most have reached stage 5. Since Bilic, Manobo and Subanun subgroup with other languages in the Philippines, not with the Bornean languages with which they share the feature of vocalic erosion from the left, it follows that this change was independently acquired in languages that developed a fixed penultimate stress, or was spread by contact. Given the first interpretation we have no explanation why a similar erosion process did not affect any Oceanic language, many of which also have penultimate stress. Given the second interpretation there is no plausible contact scenario which would include both the languages of Borneo and those of eastern Indonesia. Perhaps both factors were operative: erosion from the left would have been a natural response to the fixing of penultimate stress, but could have spread by contact once it became established in a given geographical area. In any case this explanation assumes, perhaps incorrectly, that penultimate stress was historically secondary, a result of the loss of earlier phonemic stress.

In a few cases erosion from the left takes other forms, operating on consonants rather than vowels. Perhaps the most notable example of this is seen in Sa'ban of northern Sarawak which, as noted earlier, is lexicostatistically a dialect of Kelabit, as it shares some 82% of its basic vocabulary with the standard dialect of Bario. However, rapid and sometimes surprising sound change has given this language a very different appearance than other Kelabit dialects on every level from phonetics and phonology to morpheme structure, morphology and syntax. Again, a change in the placement of stress appears to be implicated in these structurally far-reaching changes (Blust 2001e).

Unlike other Kelabit dialects Sa'ban has word-final stress. Why this stress pattern developed is unknown, but once it did many other changes followed. Among these changes was a tendency for consonantal erosion from the left. These changes are complex and sometimes quite surprising as illustrated by the etymologies in Table 9.8 (PKLD = Proto Kelabit-Lun Dayeh):

Table 9.8 Erosion from the left in Sa'ban

PKLD	Sa'ban	Gloss
*baka	aka	wild pig
*batuh	ataw	stone
*bəŋar	ŋal	plank
*bərək	rək	domesticated pig
*bibir	ibiəl	lip
*buaq	wəiʔ	fruit[108]
*butuq	tuʔ	penis
*daqun	un	leaf
*dayəh	ayəh	toward the interior
*dəlaw	liəw	freshwater eel
*dilaq	iliʔ	tongue
*duruq	ruʔ	honey
*gaiŋ	ayəŋ	spinning top
*guta	toə	ford a river
*kamih	amay	we, us (excl.)
*kayuh	ayəw	wood, tree
*kəduit	wiət	ladle, scoop
*kilat	ilat ilat	lightning
*kulat	loət	mushroom
*lalid	alit	ear
*lipən	epən	tooth
*matəh	atah	eye
*mulaq	loəʔ	many
*namuk	muək	sandfly
*nubaq	biʔ	cooked rice
*ŋadan	adin	name
*pahaqən	ahan	shoulder pole
*ranih	anay	harvest
*rəraq	raʔ	ant
*riruh	eraw	a laugh
*rumaq	maʔ	house
*sagət	ajɪt	quickly
*taruq	aroʔ	to make
*tədʰak	seək	pumpkin
*tidhuq	səuʔ	hand
*tukəd	kət	prop

The changes in these forms must be distinguished from those in etymologies such as *pənuq > *hnoʔ* 'full', *pudut > *dduət* 'way, manner; shape', or *tulaŋ > *hloəŋ* 'bone', where a penultimate vowel has deleted, giving rise to voiceless sonorants or initial geminates. These forms do not show initial consonant loss any more than do examples where a derived cluster remained unchanged, as in *bulan > *blin* 'moon'.

---

[108]   The labiovelar glides in *wəiʔ* and *wiət*, and the palatal glide in *ayəŋ* are phonemic developments of the non-phonemic glides between *u* and a following unlike vowel, or between *a* and *i*.

Sa'ban also shows vocalic erosion from the left, but not in the typical form described for other languages. Instead, penultimate *a tends to remain unchanged whether it was initial or whether it followed an initial consonant, *ə invariably disappears when initial, and usually disappears when following an initial consonant, *i often lowers to a mid-front vowel, but is not lost, and *u generally disappears whether initial or when following an initial consonant. Both consonantal and vocalic erosion from the left in Sa'ban appear to be unpredictable despite an abundance of well-supported PKLD reconstructions. Thus, while *baquŋ > *uəŋ* 'banana' or *təluh > *law* 'three' show loss of the entire first syllable, *baqaw > *biʔiəw* 'beads', or *tələn > *hlən* 'swallowed' do not. There are, nonetheless, some general patterns in consonantal erosion from the left that are worth noting. Although virtually all initial consonants are lost in at least one form, the rate of loss varies markedly, as seen in Table 9.9:

Table 9.9 Patterns of initial consonant loss in Sa'ban

Consonant	No. lost	No. retained	% lost
b	36	42	46
d	23	12	66
g	5	8	38
k	34	11	76
l	16	52	24
m	6	31	16
n	4	13	24
ŋ	2	7	22
p	6	54	10
r	12	17	41
s	1	10	9
t	17	73	19
w	1	0	100

The loss of initial consonants in Sa'ban is patterned, but in ways that remain explanatorily opaque. Voiced stops show about a 50/50 chance of being lost, with *d the most likely to disappear. Voiceless stops show about a 35% chance of being lost, but with a huge disparity between *k (76%) as opposed to *t (19%) or *p (10%). These figures do not take into account the independent operation of other variables such as canonical shape, or the tendency of most initial consonants to be retained before *a. Initial consonants for example, almost always drop in trisyllables, and never in monosyllables. But since over 90% of PKLD vocabulary is disyllabic and the initial consonant of trisyllables shows no statistical bias these factors do not greatly distort the results.

A surprisingly similar set of phonological innovations is found in Modang, a language of east Kalimantan that subgroups with Kayan and Murik (Revel-Macdonald 1982). Like Sa'ban, Modang has developed a pervasive monosyllabism through loss of the first syllable or vowel, and permits a large number of initial consonant clusters that were disallowed in its immediate common ancestor, and in nearly all other languages of Borneo: PMP *quzan > *si:n* (Sa'ban *din*) 'rain', *tuqelaŋ (> *tulaŋ) > *tluaŋ* (Sa'ban *hloəŋ*) 'bone', *tuzuq 'to point' > *tsuʔ* (Sa'ban *ddəuʔ*) 'seven', *manuk > *mnɔk* (Sa'ban *manok*) 'bird' (Revel-Macdonald 1982). Despite the obvious resemblance of these distinctive innovations in Sa'ban and Modang, it is clear that the languages belong to different subgroups. In terms of details, Modang tends to permit a wider range of initial consonant clusters, while Sa'ban

has converted many of these into initial geminates or voiceless sonorants. In both languages stress shift apparently triggered erosion from the left, but stress shift can be cited only as a necessary, not a sufficient condition for these changes. Many of the languages of coastal Sarawak have shifted stress to the final syllable, at least in citation forms, without setting into motion corresponding trajectories of exuberant innovation. Why this happened in strikingly similar form in two languages that are spoken about 200 km. from one another, and – so far as is known -- not in any intervening language, remains a mystery. Blevins (2004b) has noted a similar recurrent pattern of change in many Australian languages, and has tried to find explanations for it in terms of phonetic characteristics peculiar to them, but it seems unlikely that the same explanation would work for Sa'ban.

9.1.4.3 *Erosion from the center*

Many AN languages scattered over a wide geographical range show a process of canonical reduction that can be called 'erosion from the center.' In this process an unstressed vowel, usually *e, deletes in the environment VC_CV. The result of syncope is a medial consonant cluster that typically consists of heterorganic segments. Many languages in the central Philippines show such changes both as historical restructurings (in trisyllabic bases) and as active processes in the synchronic grammar (in affixed forms of disyllabic bases). In Tagalog, for example, the changes PMP *binehiq > *binhíʔ* 'seed rice' or *qiteluR > *itlóg* 'egg' produced consonant clusters that were impossible in PAN or PMP unreduplicated bases. In *qatep > *atíp* 'roof thatch' *e does not delete, since the environment for deletion is not satisfied, but *qatep-an > *apt-án* 'to thatch a roof' shows both syncope and metathesis. Similar historical changes and synchronic residues of these changes are found in many other languages in the central Philippines. A similar change led to the emergence of medial heterorganic consonant clusters in some Formosan languages, as Bunun and Thao: PAN *baqeRu > Bunun *baqlu*, Thao *faqlhu* 'new', *bineSiq > Bunun *binsiq*, Thao *finshiq* 'seed rice'. No synchronic residue of this change is known in Formosan languages, but in Chamorro the same type of change produced phonemic restructuring and alternations of full and reduced stems parallel to those in Tagalog, although they are historically independent: PMP *baqeRu > *paʔgo* 'new', *qalejaw > *atdaw* 'day', *qatep > *atof* 'roof thatch', *qatep-i > *aft-e* 'to roof a house'.

In Malay and other languages of western Indonesia a similar change took place that was then masked by cluster reduction. Citing Malay, Ngaju Dayak, Javanese *timah*, Toba Batak *tima* 'tin, lead', for example, Dempwolff reconstructed *timah 'tin', a form that appears to be in need of no further comment (tin has been surface mined in insular Southeast Asia for centuries, and the term is both native and ancient). Given this reconstruction Tagalog *tiŋgáʔ* 'lead (metal)' appears unrelated, and was consequently ignored by Dempwolff and others after him. But the variation in Bikol *timgáʔ, tiŋgáʔ* 'lead (metal)' suggests that the Tagalog word reflects *timeRaq, with regular *R > g, schwa syncope and nasal place assimilation. A reconstruction of this shape is, in fact, supported by cognates in other languages such as Puyuma *timRa*, and Kelabit *səmәraʔ* 'tin'.

Given *timeRaq, it follows that Malay (and many other languages of western Indonesia) collapsed original trisyllables by 1) medial schwa syncope, and 2) subsequent cluster reduction. Other examples of medial vowel syncope, cluster reduction or both, include PMP *qali-metaq > Isneg *alimtá*, Singhi *rimotah*, Malay *lintah* 'paddy leech' (Dempwolff's *lintah), *tuqelaŋ > Palawan Batak *tuʔlaŋ*, Chamorro *toʔlaŋ*, Malay *tulaŋ* 'bone' (Dempwolff's *tulaŋ), *bineSiq > Thao *finshiq*, Tagalog *binhíʔ*, Malay *bәneh* 'seed rice' (Dempwolff's *benih/binih), *bakelad > Tagalog *baklád* 'fish corral', Malay *bәlat*

'screen-trap for fish', *saŋelaR > Cebuano *saŋlag* 'roast something in a pan with little or no oil', Minangkabau *saŋlar* 'broiling, cooking at an open fire', Malay *səlar* 'branding' (with schwa showing that the penultimate vowel in Malay was once antepenultimate), or *uteŋaq > Cebuano *utŋáʔ* 'come off, let go of s.t. attaching firmly to s.t. else', Malay *uŋah* 'to be shaking loose, as a tooth, or insecure stake' (Blust 1982c, 2001d). As a result of these etymologies it appears very likely that Malay and many other languages in western Indonesia passed through a stage in which medial vowel syncope gave rise to heterorganic consonant clusters such as those in Philippine languages. The difference is that such historically derived clusters have been retained in Philippine languages, but almost everywhere in western Indonesia they were reduced, generally leaving no consonantal trace of the original third syllable, but sometimes leaving a trace on the original prepenultimate vowel in the form of reduction to schwa.

One other area in which a change of this type has taken place quite independently is in eastern Manus in the Admiralty Islands. Here any unstressed vowel was deleted in the environment VC_CV, producing medial heterorganic consonant clusters. In some languages these further assimilated to produce geminate consonants. Examples include Proto Admiralty *na taliŋa > Nali *drayiŋa-*, Kuruti *dralŋa-*, Ere *dralŋʷa-* 'ear', and *papanako > Nali *pahana*, Kuruti *pahna* (with syncope), Ere *panna* (with syncope and assimilation) 'to steal'.

9.1.5 Epenthesis

Many of the changes considered so far involve segment loss. There are, however, changes in a number of AN languages that involve segment addition. These can be discussed collectively under the general heading 'epenthesis.'

A particularly widespread change is laryngeal epenthesis, in which a glottal stop or *h* is added after final vowels. Many of the Formosan languages, including Atayal, Saisiyat, Pazeh, Rukai, Bunun, Kavalan, Paiwan, Puyuma and Amis show a largely predictable final glottal stop. In some languages, as Atayal, marginal evidence of contrast has led to this segment being written in most transcriptions of the language, although it is predictable in all but a small number of cases.

Most Philippine languages reflect PAN *q as glottal stop and final vowels as final vowels, thus contrasting glottal stop with zero in final position. However, the Bashiic languages Yami, Itbayaten and Ivatan as well as Casiguran Dumagat merge the two as glottal stop, and Kalamian Tagbanwa reflects *q as *k*, freeing the way for the use of a non-phonemic glottal stop after all underlying final vowels. In Borneo many of the Barito languages of southeast Kalimantan have added glottal stop after original final vowels, as has Sundanese of west Java. In Kayan of central Borneo earlier *-V is reflected as -Vʔ, and *-Vʔ as -V. Dialect evidence shows that glottal stop was first added after final vowels, which were distinguished from inherited vowel + glottal stop sequences by vowel length (low vowels) or height (non-low vowels). Given the number of languages that have a non-phonemic final glottal stop and their positions in the AN family tree, it is possible that this feature is inherited from PAN, which had a uvular stop *q, but no glottal stop.

In a number of other languages *-h* has been added after original final vowels. This change is seen in Aklanon and other Bisayan dialects of the central Philippines, in Tausug of the southern Philippines, and in various languages of Sarawak, including Miri, Narum, Kiput, Berawan, Western Penan, Long Wat Kenyah, Sebop, Kelabit, Dalat, Matu and Serikei Melanau, and some of the Land Dayak languages: PMP *mata > Tausug *matah*,

Kelabit, Narum, Matu Melanau *matəh* 'eye', *telu > Tausug *tuuh*, Kelabit *təluh*, Paus, Siburan Land Dayak *taruh* 'three'. Given its discontinuous distribution (southern Philippines and Sarawak, but not Sabah), *h*-accretion must be considered the product of at least two historically unconnected changes. Even in Sarawak, it is clear that *-h* addition must be recurrent. Whereas PMP *-a, *-i and *-u became Kelabit *-əh*, *-ih* and *-uh*, for example, final high vowels diphthongised in many coastal languages of Sarawak, preventing the addition of *-h*, hence *telu > Narum *tələw*, Matu Melanau *tələw* 'three'. In other languages, as Miri, spoken on the northernmost coast of Sarawak, final high vowels were diphthongised, but *-h* was still added: *mata > *matah* 'eye', *telu > *təlauh* ~ *təloh* 'three', *nupi > *nupaih* 'dream'. The similar diphthongisation in *putiq > *futaiʔ* 'white', or *kulit > *ulait* 'skin, bark' suggests that *-h* was added before vowel breaking.

Although virtually all known examples of *h*-epenthesis in AN languages are word-final, Hudson (1967) recorded a vocabulary of about 330 words for Dohoi in southeast Kalimantan which shows the addition of *h* before medial voiceless obstruents: PMP *utaq > *ŋ-uhtaʔ* 'to vomit', *mata > *mahtaʔ* 'eye', *m-atay > *mahtoy* 'to die', *tektek > *nohtok* 'to cut, hack', *ikuR > *ihkuh* 'tail', *kutu > *kuhtuʔ* 'louse', *batu > *bahtuʔ* 'stone', *putiq > *puhtiʔ* 'white', *nipis > *mihpih* 'thin', *kapal > *kahpan* 'thick', *itu > *ihtuʔ* 'this', *ma-zauq > *mahcuʔ* 'far', *aku > *ahkuʔ* 'I', *beken > *buhkon* 'other', *epat > *ohpat* 'four', *pitu > *pihtuʔ* 'seven', *tugal > *tuhkan* 'dibble stick', *seput > *sohput* 'blowgun', *betaw > *bohtow* 'sister'. There are a few exceptions, as *ka-taqu > *kotouʔ* 'right side', *likud > *likut* 'back', *qatay > *atoy* 'liver', *quzan > *ucan* 'rain', *ma-qitem > *mitom* 'black', and *punti > *putiʔ* 'banana' (cp. *puhtiʔ* 'white'). It is unclear whether these are due to transcriptional inconsistency, conditioned change (addition of preconsonantal *h* before reduction of the medial cluster in *punti), or are exceptions. It is noteworthy, however, that *h*-epenthesis does not appear before any medial consonant that is not a voiceless stop.

Several of the Land Dayak languages of southern Sarawak exhibit a remarkable pattern of obstruent epenthesis. Singhi can be taken as representative. In Singhi *a and *u merged as *-ux*, and *i became *-is*: PMP *Raya > *ayux* 'great, large', *depa > *dopux* 'fathom', *lima > *rimux* 'five', *qabu > *abux* 'ash; fireplace', *batu > *batux* 'stone', *tunu > *ninux* 'burn'; *qubi > *bis* 'yam', *kali > *karis* 'dig', *suligi > *sirugis* 'spear'. Obstruent epenthesis was a late change that followed the loss of certain PMP final consonants, including *q and *h: *putiq 'white' > *bi-putis* 'European', *piliq > *piris* 'choose', *talih > *taris* 'rope'. Rensch et al (2006) suggest that this change began with *-h* epenthesis, and that the quality of this fricative was then coloured by the preceding vowel. If this explanation is correct, *h* following *i was palatalised to [hʸ], and then strengthened to a sibilant, a development that is otherwise unknown in AN languages. This interpretation depends on the reconstruction of epenthetic *-h in Proto Land Dayak, an inference that is complicated by the fact that *-h* accretion is an areal phenomenon in Sarawak (Proto Kenyah, for example, apparently lacked *-h but such Kenyah languages as Long Wat, Sebop, and Western Penan added final glottal spirants as a result of areal diffusion). For this reason the possibility that obstruents were added directly to final high vowels in Singhi cannot be excluded, as the changes *-i > *-it* and *-u > *-uk* are generally accepted for Maru, a Tibeto-Burman language spoken in northern Burma (Blust 1994b).

One of the most widespread examples of epenthesis in AN languages is the addition of a palatal glide before initial *a. This change appears in scattered parts of eastern Indonesia, including the CMP languages Fordata, Kei, and Masiwang, and in As, Buli and Numfor among the South Halmahera-West New Guinea languages. It is widely attested in Oceanic languages, including Likum, Levei, Pelipowai, Drehet, Mondropolon and Bipi of western

Manus, Gedaged of the north coast of New Guinea, Keapara, Tubetube, Gapapaiwa, Molima, Dobuan and Kilivila of southeast New Guinea, Mortlockese, Puluwat, Woleaian, Ulithian, Sonsorol and Pulo Annian of Micronesia. In a number of cases palatal glide epenthesis was followed by changes that converted *y* from any source to some other segment, thus somewhat obscuring the primary change. Woleaian, spoken in the Caroline Islands of Micronesia, provides examples of the unobscured change in POC *api > *yaf(i)* 'fire', *aŋin > *yaŋ (i)* 'wind', *qate > *yas(e)* 'liver' and *qatop > *yas(o)* 'roof; sago leaf thatch' (cp. *ikan > *igal* 'fish' and *qusan > *ut* 'rain', with no glide epenthesis, and *onom > *wolo* 'six', with labiovelar glide epenthesis). Fijian shows a similar change which appears to have been intersected mid-course by the change *y > *c* (a voiced interdental fricative): POC *acan > *yaca-* 'name', *asaq > *yaca* 'grind, sharpen', *qalop > *yalo* 'beckon', *qate > *yate-* 'liver', *Rabia 'sago' > *yabia* 'arrowroot', but *aŋin > *caɲi* 'wind', *aRu > *cau* 'a shore tree: *Casuarina equisetifolia*', *apaRat 'west monsoon' > *cavā* 'storm wind'. As several of these examples illustrate, *y*-epenthesis in Fijian followed the loss of certain initial consonants, including POC *q and *R. A similar intersection of competing sound changes evidently must lie behind the appearance of either *y-* or *j-* before the reflex of *a-* in Bobot (Bonfia) of eastern Seram in the central Moluccas: PMP *hapuy > *yāf* 'fire', *ama > *yama* 'father', but *haŋin > *jakin* 'wind', *qatay > *jata-n* 'chin, jaw'.

A more obscure but still clearly retrievable example of the same change is seen in Motu, where at first sight it appears that *l* was added before an initial low vowel: POC *acan > *lada* 'name', *asaŋ > *lada* 'gills', *apaRat > *lahara* 'northwest wind and season', *api > *lahi* 'fire', *aŋin > *lai* 'wind', *aku > *lau* 'I'. Comparative evidence show that this is the same type of change seen in Woleaian and Fijian, since 1) *l* was not added before other initial vowels: *inum > *inu-a* 'to drink', *upe > *uhe* 'seed for planting', and 2) POC *y became *l*: *maya > *mala* 'tongue', *puqaya > *huala* 'crocodile'.

The languages of the southeast Solomon Islands show an even greater range of obfuscating secondary changes, reflecting POC *y as *l* (Kwaio), *r* ('Āre'āre), *s* (Lau, Kwara'ae, Sa'a, Arosi), *θ* (Baelelea, Toqabaqita), or *ð* (Longgu), and showing a similar epenthetic consonant before initial *a: POC *qatop > Kwaio *lao*, 'Āre'āre *rāo*, Lau, Kwara'ae, Sa'a, *sao*, Toqabaqita *θao*, Longgu *ðao* 'sago palm; palm thatch'; POC *puqaya > Kwaio *huala*, 'Āre'āre *huara*, Lau *fuasa*, Sa'a, Arosi *huasa* 'crocodile'. Again, the patterning of reflexes leaves no doubt that this change began as *y*-epenthesis (Lichtenberk 1994). Similar examples of *y*-accretion with secondary fortition to an obstruent are found in Lindrou and Sori of western Manus, where the reflex of POC *a is *ja-*, as in *apaRat > Lindrou *jaha*, Sori *japay* 'west monsoon'. In the Polynesian Outlier of Aniwa in Vanuatu, POC *a- and *qa- generally became *cia-*, as in *api > *ciafi* 'fire', *qapu > *ciafu* 'ashes', *saman (> PPN *hama) > *ciama* 'outrigger float', *qajo (> PPN *ʔaho) > *ciao* 'day'. Since POC *y had already disappeared in Proto Polynesian it is unclear whether Aniwa *ci-* in bases that begin with a low vowel reflects an earlier palatal glide. POC *t is reflected as Aniwa *c* before front vowels, and it might be suspected that the secondary element *ci-* in these nouns is a reflex of the PPN common noun article *te. However, no such segment occurs in nouns that begin with other vowels: POC *ikan > Aniwa *ika* 'fish', *qusan > *ua* 'rain', *qupi > *ufi* 'yam'. The historically epenthetic initial consonants of Aniwa thus remain problematic, since they could only reflect a historically secondary palatal glide which developed after the loss of POC *y.

Given its wide distribution in eastern Indonesia and the Pacific it is puzzling that this change is unattested elsewhere in Austronesian. Again, we see a recurrent innovation that has an areal character, but one that is both discontinuous and broad. This change is so

widespread in Oceanic languages that it is sometimes attributed to Proto Oceanic. Certain observations, however, weigh against this explanation. Unlike Fijian, for example, which shows glide epenthesis even in POC bases that began with *q or *R, Motu contrasted *a- and *qa- at the time of *y*-epenthesis, since no glide was added to words that began with *q, which disappeared: PMP *qasawa > Motu *adava* 'spouse', *qazay > *ade* 'chin', *qaRus > *aru* 'current in river or sea', *qatay > *ase* 'liver'.

The addition of other glides before initial vowels is attested from a variety of languages. Puluwat and some other languages of the Caroline Islands in central Micronesia have added homorganic glides before initial high or mid vowels, as in POC *ikan > *yiik, yika-n* 'fish', *ina > *yiin, yina-n* 'mother', *ican > *yiit, yita-n* 'name', *onom > *woon* 'six', *qumaŋ > *wum^wó-wum* 'hermit crab' or *quraŋ > *wúúr* 'lobster'. As with *y*-epenthesis before *a-, subsequent changes have obscured homorganic glide accretion in some languages. Babatana, Simbo and other languages of the western Solomons, for example, show an epenthetic *v*-, and Sori of northwest Manus, an epenthetic *g*- before historically primary or secondary rounded vowels: POC *onom > Babatana, Simbo *vonomo*, Sori *gono-p* 'six', *kutu (> *utu) > Babatana *vutu* 'louse', *qone > Sori *goŋ* 'sand', *Rumaq > Sori *gum* 'house' (cp. POC *waiR > Sori *gay* 'fresh water' for evidence that *g* in these cases reflects earlier *w*). Wogeo, spoken on a small island of the same name off the north coast of New Guinea, has added either *y* or *w* unpredictably before the reflex of POC *a-, although *w*, which was subsequently strengthened to *v*, is more common: POC *apaRat > *yavara* 'west monsoon', *qasawa > *yawa-* 'spouse'; *ayawan > *vaiawa* 'banyan', *saman (> *ama) > *vama* 'outrigger float', *anunu > *vanunu* 'shadow, reflection', *qatop > *vato* 'thatch; roof', *qawa > *wawa-* (with assimilation of *v*- to -*w*-) 'mouth'. Before other initial vowels no glide was added.

Gorontalo of northern Sulawesi added *w*- before *(q)a-, as in PMP *qabu > *wahu* 'ashes', *añam > *walamo* 'plait', *anak > *walaʔo* 'child', or *anay > *wale* 'termite' (Sneddon and Usup 1986). Following loss of *h, Chamorro added *w*- before all initial vowels which then underwent fortition to *g^w* before non-round vowels, and *g* before round vowels. This change did not happen before glottal stop, which disappeared phonemically: PMP *hapuy > *g^wafi* 'fire', *aku > *g^wahu* 'I', *ini > *g^wini* 'here; this'; *hunus > *gunos* 'wean, withdraw', *enem > *gunum* 'six', but *qazay > *achay* 'chin', *qalep > *alof* 'beckon', *qipil > *ifet* 'a tree: *Intsia bijuga*', *quzan > *uchan* 'rain'.

In Palauan, words that originally began with a vowel or a consonant that dropped now begin with a velar nasal: PMP *hated 'accompany, escort' > *ŋádər* 'gift of food accompanying the bride when she is brought to the prospective husband's family', *anay > *ŋal* 'termite', *wada (> *ada) > *ŋar* 'be located; exist; be alive', *aRuhu > *ŋas* 'ironwood; *Casuarina*', *hapuy > *ŋaw* 'fire', *ia > *ŋíy* '3sg', *hikan > *ŋíkəl* 'fish', *uRat > *ŋurd* 'vein, artery'. While it is tempting to propose that both the Palauan and Chamorro innovations result from captured morphology, the weight of evidence suggests instead that they are sound changes. In both languages these segments appear before morphemes of all word classes, a distribution that is unlikely to result from metanalysis. Palauan presents additional problems for a theory that traces *ŋ*- in words such as *ŋal* or *ŋar* to an earlier grammatical marker, as many loanwords show final velar nasal accretion:

Table 9.10 Loanwords in Palauan with final velar nasal accretion

Form	Source	Original	Gloss
baŋderáŋ	Spanish	bandera	flag, banner
biáŋ	English	[bija]	beer
blauáŋ	English	[flawa] (flour)	bread
butiliáŋ	Spanish	botella	bottle
kámaŋ	Japanese	Kama	sickle
kambaláŋ	Spanish	campana	bell
karróŋ	Spanish	Carro	wagon
kawáŋ	Malay	kawah	cauldron
kuábaŋ	English	Guava	guava
mastáŋ	English	[masta]	master
stoáŋ	English	[stoa]	store
tóktaŋ	English	doctor [dɔkta]	doctor

Although many loanwords with a final vowel have not added the velar nasal, this change is clearly recurrent, and is not phonologically conditioned. Velar nasal accretion in Palauan shows a curious asymmetry. In native forms it is almost always word-initial, presumably because unstressed final vowels generally deleted before velar nasals were added. In loanwords, on the other hand, there are a number of examples after final vowels, but none before initial vowels (*adios* 'good-bye', *iíŋs* 'inch; hinge', *ikelésia* 'church', *osbitár* 'hospital', *uós* 'horse'). However, a small number of native word bases retain a vowel in final position, and here we also find a historically secondary velar nasal. As it happens, all of these occur in low numerals: PMP *esa > *e-taŋ* 'one', *duha > *e-ruŋ* 'two', *epat > *e-waŋ* 'four'. Josephs (1975:470) lists these numerals without the final velar nasal, but McManus and Josephs (1977) give the dictionary entries with the velar nasal, which drops when the numeral is not phrase-final: *taŋ* 'one (unit of time, person)' : *ta el sils* 'one day', *eruŋ* 'two (units of time)' : *eru el sikaŋ* 'two hours', *eru el sils* 'two days'. A vigorous but inconclusive discussion of this puzzling historical development appears in Blust (2009b), where a phonological analysis is proposed, and Reid (2010), where a morphological analysis is proposed on inadequate external grounds. More recently Blevins and Kaufman (2012) have proposed a morphological source for the intrusive velar nasal based on a careful internal analysis of Palauan, and it appears likely that their hypothesis will ultimately win general acceptance.

Many of the Dusun languages of Sabah reflect a number of etyma with an accretive initial *t*, as in PMP *anay > Kadazan *tanay* 'termite', *hasaŋ > *tasaŋ* 'gills', *hawak > *tavak* 'waist, loins', *qabu > *tavu* 'ash', *qayam > *tazam* 'domesticated animal', *ikuR > *tikiw* 'tail', *hipun > *tipun* 'small shrimp', *buhek (> *buk > *əbuk > *obuk) > *toɓuk* 'hair of the head', *qanibuŋ > *toniɓuŋ* 'nibung palm', *qulin > *tuhin* 'rudder', *qulu > *tuhu* 'head', or *uban > *tuvan* 'grey hair'. This change superficially resembles *w*-epenthesis in Chamorro, and *ŋ*-epenthesis in Palauan, but differs in a crucial respect: it occurs only with nouns (cp. *isa > *isoʔ* 'one', *epat > *apat* 'four', *enem > *onom* 'six', *esak > *ansak* 'cook; be ripe', *hasaq > *asaʔ* 'whet, sharpen', *ikej > *ikod* 'to cough', uliq > *uhiʔ* 'return, go home', without t-accretion). Given the syntactic complementation of forms with and without accretive *t-*, a morphological explanation for this segment thus seems preferable to a phonological explanation, and this has been ably argued by Kroeger (1990).

Vowel epenthesis takes essentially two forms: 1) the addition of an initial vowel (usually schwa) to restore a lost disyllabism, 2) the addition of a final supporting vowel or

echo vowel to satisfy an open syllable constraint. The latter change conflicts with the common AN disyllabic canonical target, and this conflict is resolved in various ways. Both types of epenthesis are best treated in the more general discussion of 'drift', but can be mentioned briefly here.

In Anejom of southern Vanuatu nouns beginning with a historically derived nC- acquired an initial *i*, presumably to facilitate the pronunciation of the initial consonant cluster, but this is an isolated change: POC *na patu (> *na hatu > *nhatu) > *inhat* 'stone', *na waiR > *inwai* 'water'.

A number of languages in western Melanesia added echo vowels to POC final consonants, as with Mussau of the St. Matthias Archipelago north of New Ireland, or Roviana, spoken in the New Georgia Archipelago of the western Solomons: POC *panas (> *pa-panas) > Mussau *aanasa* 'hot', *kiRam > *iema* 'axe, knife', *ma-takut > *(ma)matautu* 'afraid', *onom > *(o)nomo* 'six', *pulan > *ulana* 'moon, month'; *asaŋ > Roviana *asaŋa-na* 'it's gills', *qapij > *avisi* 'twins of the same sex', *kaRat > *garata* 'to bite', *ma-takut > *matagutu* 'afraid', *onom > *onomo* 'six'. Several languages of the D'Entrecasteaux Archipelago southeast of New Guinea have instead added a fixed supporting vowel, *a*. The best known of these is Dobuan, in which *p and *k were subsequently lost: POC *pat > *ata* 'four', *qatop > *ʔatoa* 'roof, thatch', *ikan > *iana* 'fish', *inum > *numa* 'drink', *manuk > *manua* 'bird', *sinaR 'shine' > *sinara* 'sun'.

A few languages in island Southeast Asia, including Tsou, Kanakanabu, Saaroa and Rukai of southern Taiwan, Malagasy, Gorontalo of northern Sulawesi, Kambera of eastern Sumba, and Moor of Irian Jaya have also added supporting vowels. These correspond to echo vowels in some cases, but the correspondence is incomplete, and it must be concluded that true echo vowels are rare in AN language outside the Oceanic group. The Tsouic languages and Rukai copy PAN *i, *u and *e after a final consonant, but support *aC with an epenthetic schwa, as in PAN *qayam > Tsou *zomə*, Saaroa *aɬamə*, Maga Rukai *arámə*, 'bird'. The quality of this vowel is not a product of canonical constraints, since *-a is reflected differently, as in *lima > Tsou *eimo*, Saaroa *(k)u-lima*, or Maga Rukai *rima* 'five'. In Malagasy *-a* was added after most final consonants (PMP *epat > *éfatra* 'four', *enem > *énina* 'six', *laŋit > lánitra 'sky'), and a similar change took place optionally in Moor, a SHWNG language spoken in the Bird's Head Peninsula of New Guinea: PMP *niuR > *nera* 'coconut', *taŋis > *ʔanita* 'weep', *danum > *rarum(a)* 'fresh water', *zalan > *rarin(a)* 'road'. In Gorontalo a schwa was added word-finally and then underwent the general change of schwa to *o* (PMP *bulan > *hulalo* 'moon', *enem > *olomo* 'six', *tuqelaN > *tulalo* 'bone'), and in Kambera final consonants were supported by the addition of *-u*: PMP *alas > *alahu* 'forest', *hikan > *iyaŋu* 'fish', *ma-qitem > *mítiŋu* 'black', *epat > *patu* 'four'.

A different form of vowel epenthesis is found in some of the languages of Sulawesi. As noted by Sneddon (1993), in Sangir and Bantik of northern Sulawesi a vowel plus glottal stop was added after word-final voiced stops, *s and liquids. This has the form schwa + glottal stop in both Sangir and the Sangil dialect of southern Mindanao, but echo vowel + glottal stop in Bantik: PMP *lahud 'toward the sea' > Sangir, Sangil *laudəʔ*, Bantik *láodoʔ* 'ocean', *tapis > Sangir *tapisəʔ*, Bantik *tápisiʔ* 'strain, filter', *qateluR > Sangir *təluhəʔ* 'egg'. Sneddon (1984) posits a North Sangiric subgroup that includes Sangir, Sangil and Talaud, and a South Sangiric subgroup that includes Bantik and Ratahan. Since Sangir/Sangil and Bantik are in different primary branches of the Sangiric group and since neither Talaud nor Ratahan show this change the addition of -Vʔ in Sangir/Sangil and Bantik appears to be a product of distinct historical changes (although one of these may

have been contact-induced). Surprisingly, a very similar type of syllable addition occurs in Makasarese of southwest Sulawesi, where affected final *s and liquids: PMP *nipis > Makasarese *nípisiʔ* 'thin, of materials', *tuŋgal > *túŋgalaʔ* 'alone', *huluR > *úloroʔ* 'to lower, as by rope'. General linguistic theory suggests that the addition of supporting vowels provides a means to unmark syllables, but cases such as these are difficult to reconcile with this interpretation. Vowel epenthesis in Talaud, which added -*a* 'after all final consonants, usually with doubling of the preceding consonant' (Sneddon 1984:33) presents special problems that will be addressed in the discussion of gemination below.

9.1.6 Metathesis

Metathesis generally is viewed as sporadic, and sporadic metathesis can be identified in many AN languages. Although sporadic changes are not normally regarded as part of sound change, Table 9.11 lists some examples of sporadic metathesis so as to provide a basis for later generalisations about the phenomenon of metathesis as a whole. Protoforms are of varying time-depths, and to simplify the discussion a search was made only in CVCVC bases, or bases that probably became disyllabic before metathesis occurred (thus excluding cases such as *balian > Western Bukidnon Manobo *bəylan* 'shaman', or the recurrent metatheses of the second and third consonants in reflexes of POC *palisi 'grass'). Consonants are optional, and consonant positions are counted in relation to a CVCVC template so that the first consonant in e.g. *iluR is C_2, not C_1:

Table 9.11 Examples of sporadic metathesis in Austronesian languages

Protoform	Reflex	Segments	Language	Gloss
*dilaq	lidah	C_1/C_2	Malay	tongue
*laŋo	nalo	C_1/C_2	Hawaiian	housefly
*ŋipen	wiŋəl	C_1/C_2	Palauan	tooth
*qañuj	pa-a-laHud	C_1/C_2	Puyuma	adrift; set adrift
*Ratus	dart	C_1/C_2	Palauan	hundred
*qalunan (> lunan)	nulaŋu	C_1/C_2	Kambera	headrest, pillow
*qanilaw (> nilaw)	lino	C_1/C_2	Roti	tree: *Grewia* spp.
*qitik	siʔi	C_1/C_2	Tongan	small, little
*salaq (> lasaq)	hasaʔ	C1/C2	Kadazan	wrong, in error
*tuhud (> hutud)	utut	C1/C2	Ngaju Dayak	knee
*iluR	liur	V1/C2	Malay	spittle, saliva
*ma-iraŋ	ma-hiaŋ	V1/C2	Ngaju Dayak	red
*iRik	giʔík	V_1/C_2	Tagalog	thresh grain
*bañen	bənnaŋ	V1/V2	Sangir	sneeze
*dakep	dəkap	V1/V2	Malay	catch, hold
*qinep (> *qenip)	one	V1/V2	Ngadha	lie down
*bales	wahal	C2/C3	Roma	answer
*baŋun	n-vanuŋ	C2/C3	Fordat	awake; erect
*haŋin	aniŋ	C2/C3	Sika	wind
*liseqah (> *lisa)	lias	C2/V2	Ngaju Dayak	nit, louse egg

These are only a few examples of sporadic metathesis in AN languages, but they have been chosen randomly (mostly from Dempwolff 1934-1938, and Blust and Trussel ongoing), and so should not show sample bias with regard to pattern frequency.

Nonetheless, of the 20 examples cited here half involve the interchange of C1 and C2, about 15% affect the interchange of V1/C2, V1/V2, and C2/C3, while only one example was recorded of C2/V2 metathesis. It is noteworthy, although perhaps accidental given the small sample size, that the dominant pattern is attested throughout the AN language family, the V1/C2 pattern only in WMP languages, and the C2/C3 pattern only in CMP languages. Sporadic metathesis in AN languages thus appears to pattern in ways that have not previously been noticed, and for which no explanation is currently available.

In addition to sporadic metathesis what appears to be recurrent metathesis is also found, although on closer inspection some cases turn out to be products of other types of change that cumulatively mimic metathesis. Before discussing regular or recurrent metathesis, however, one other point should be mentioned. Where it is sporadic we would not expect metathesis to target the same lexical item in more than one language unless these languages have inherited the innovation from an immediate common ancestor. Oddly, however, reflexes of PAN *Caliŋa 'ear' appear with transposition of the liquid and nasal consonants in a number of quite diverse and widely separated languages, including: (Taiwan) Mayrinax Atayal *caŋiyaʔ* Pazeh *sariŋa, saŋira*, Bunun *taiŋa/taŋia*, Maga Rukai *cŋira*, Siraya *taŋira*, (Philippines) Ibaloy *taŋida*, Keleyqiq Kallahan *taŋilaq*, Maranao *taŋila*, (New Guinea) Irarutu *tegra*, Kilivila *tegila*, (Solomon Islands) Alu *taŋna-na*, Banoni *taŋina*, and Kokota *tagla-na*. This distribution suggests that next to *Caliŋa, PAN probably had a less favored variant *Caŋila which has survived in some languages as an alternative pronunciation or as the only pronunciation of this morpheme.

There is one well-established pattern of regular (or at least recurrent) metathesis that can only be seen by comparing Formosan and non-Formosan languages. PAN permitted both the phoneme sequence *CVS and the sequence *SVC, where C = any consonant, V = any vowel, and S = the PAN sibilant *S. This contrast is preserved in Formosan languages, but outside Taiwan *CVS has often metathesised to PMP *hVC: PAN *bukeS > PMP *buhek 'head hair' (cp. PAN *ma-buSek > PMP *ma-buhek 'drunk', where no change took place), PAN *CaqiS > PMP *tahiq 'sew', PAN *liseqeS > PMP liseheq 'nit, egg of a louse', PAN *quSeNap > PMP *huqenap 'fish scale', PAN *tapeS > PMP *tahep 'winnow', PAN *tuduS > PMP *tuhud 'knee'.

Many of the languages of the northern Philippines show a regular metathesis of *t and *s in the sequence *tVs or *tVCVs. Ilokano can be taken as a type case: PMP *Ratus > *gasút* 'hundred', *Retas > *gessát* 'to snap, of a string', *lásat* 'pass, cross over' (with monosyllabic root *tas 'cut through, sever; take a short cut'), *utas 'cut through, sever' > *úsat* 'open a road, clear a path in the jungle', *tastas > *satsát* 'unravel; tear, rip garments', *tebus > *sambut* 'redeem', *tameqis > *samʔít* 'sweet' *taŋis > *sáŋit* 'weep, cry', *tiRis > *sígit* 'to pour, decant', *timus 'salt' > *símut* 'dip in salt or sauce'. All known examples of *tVs metathesis are found in final syllables, although the absence of this pattern with the first two consonants of the base probably is accidental. A similar innovation has occurred in Bontok, Kankanaey, Kalinga, Isneg, Itneg, Isinai, Dupaningan Agta (Robinson 2011) and Pangasinan, and less regularly in Bolinao and Botolan Sambal. Other languages of northern Luzon that may have undergone the same change have merged *s with *t, thus masking the evidence.

Surprisingly, an innovation of essentially the same form is also reported for Sangir, Sangil and Talaud of northern Sulawesi. Sneddon (1984:31ff) notes that these three languages 'reflect metathesis of word-final *s with a preceding *t,' as seen in *Ratus > Sangir *hasuʔ*, Sangil *rasuʔ*, Talaud *žasutta* (but Bantik *hátusuʔ*) 'hundred', or *bitis > Sangir, Sangil *bisiʔ*, Talaud *bisitta* (but Bantik *bítisiʔ*) 'calf of the leg'. Metathesis of *-tVs

also occurred where *t was prenasalised: Proto Sangiric *pəntas > Sangir *pənsaʔ* (but Bantak *pantasaʔ*) 'to harvest'. Sneddon states that metathesis of *-tVs is regular, but where these consonants are separated by two syllable peaks it is sporadic, thus differing from the similar change in northern Luzon. In addition, Sangir and Sangil (but not Talaud) show a complementary change whereby initial *t assimilated completely to a following *s, regardless of the number of intervening syllable peaks: *tasik > Sangir, Sangil *sasiʔ* (but Bantik *tasiʔ*) 'sea', *tasak > Sangir, Sangil *sasaʔ* (but Bantik *tasaʔ*) 'ripe', *talisay > Sangir *saḷise*, Sangil *saḷisay* (but Bantik *talisey*) 'tree sp: *Terminalia catappa*'. Given their geographical separation and the details of the change, then, it seems clear that *tVs metathesis in northern Luzon and northern Sulawesi is a product of convergence.

In addition, laryngeal metathesis is widespread in the central Philippines. Where glottal stop or *h* came to stand next to another consonant as a result of 'erosion from the center' it metathesised so that laryngeals were allowed only preconsonantally or postconsonantally in any given language. Cebuano Bisayan permits only postconsonantal laryngeals in consonant clusters: PMP *baqeRu (> *baʔgu) > *bagʔu* 'new', *tuqelan 'bone' (> *tuʔlan) > *tulʔan* 'knee joint of human beings or large animals', *kihkih (> *kihki) > *kikhi* 'scrape, scratch off', *kuhkuh (> *kuhku) > *kukhu* 'scrape, scratch off'. Palawan Batak, on the other hand, permits only preconsonantal laryngeals: PMP *baqeRu > *baʔgu* 'new', *ma-beReqat > *ma-bəʔgat* 'heavy'.

A restricted form of metathesis that appears to be motivated by language universals is the reversal of derived consonant sequences in which the first consonant is coronal and the second non-coronal. This is seen in Tagalog suffixed forms such as *qatep > *atíp* 'roof thatch' : *qatep-an > *apt-án* 'thatch a roof', or *tanem > *taním* 'plant, bury' : *tanem-an > *tamn-án* 'plant on, have a place planted with'. That these metatheses are motivated by avoidance of specific consonant sequences is supported by at least three observations: 1) derived consonant clusters in other orders do not change, 2) coronal-non-coronal consonant clusters in which the segments differ in nasality are also changed, but by assimilation rather than metathesis: *baníg* 'mat' : *baŋ-án* 'put a mat on', *dateŋ > *datíŋ* 'arrive' : *datn-án* 'find someone upon arrival', 3) coronal-non-coronal consonant clusters in many other languages show a similar lability (Blust 1971). Similar metatheses are found in Cebuano Bisayan and some other Central Philippine languages, in Chamorro (*qatep > *atof* 'roof thatch' : *qatep-i > *aft-e* 'thatch a roof'), and in Leti, Moa, Wetan and other languages of the Luang-Kisar group (PMP *tanem > Leti *tomna*, Moa *tamna*, Wetan *tutni-tamna* 'to plant', where the expected medial cluster is -nm-).

Kapampangan, spoken in north-central Luzon, appears to show regular CV- metathesis in forms such as PMP *bales > *ablás* 'return s.t., answer back', *besuR > *absíʔ* 'satiated', *beRas > *abyás* 'milled rice', *lebuR > *álbug* 'flood', *letiq > *altí* 'thunder', *tebuh > *atbú* 'sugarcane', and *telu > *atlú* 'three'. Since the vowel in all of these examples is PMP *e, it is possible that a prefix *a* was added, placing *e in the environment VC__CV where, as we have seen, syncope has taken place in many languages. This interpretation is strengthened by independent evidence for syncope of *e in *qalejaw > *aldó* 'sun; day', and *qapeju 'gall' > *atdúʔ* 'bile', and by evidence for a nominal prefix *a-* in *puluq > *apúluʔ* '(group of) ten', *tian (> *tyan) > *atcan* 'stomach', and in the Spanish loanword *aplaya* 'beach' (Spanish *playa*). It is weakened, however, by the need to assume vocalic metathesis in *bales > *ablás*, and by the observation that neither *ablás* nor *absíʔ* is a noun.

The historical phonology of Kapampangan raises the issue of pseudo-metathesis, a recurrent problem in the history of AN languages. A second language that appears to have a historical rule of metathesis which is actually the cumulative product of other changes is

Letinese, spoken in the Leti-Moa Archipelago east of Timor. Examples of this change appear in Table 9.12:

Table 9.12 Letinese pseudo-metathesis

PMP	Leti	English
*haŋin	anni	wind
*kempuŋ	apnu	belly
*ijuŋ	irnu	nose
*garut	kartu	scratch
*gatel	katla	itch
*laŋit	lianti	sky
*lumut	lumtu	moss
*ma-qitem	metma	black
*panas	pansa	hot
*zalan	talla	path, road
*tenun	tennu	weave
*tanem	tomna	plant, bury
*kulit	ulti	skin
*quzan	utna	rain
*habaRat	warta	west
*bulan	wulla	moon

A number of loanwords from Malay also show an apparent metathesis, as with Leti *derku* 'citrus fruit' (Malay *jəruk*), *lemnu* 'lemon' (Malay *limun*, from Portuguese), *mapku* 'drunk' (probably from Malay *mabuk*), *riwtu* 'storm' (Malay *ribut*), or *surta* (Malay *surat*) 'write'. These forms suggest either that metathesis was innovated after the introduction of Malay loans, or that structural constraints in the language required loanwords to undergo adaptations to possible word shapes. The fact that a similar pattern of historical metathesis is also reported in the Moa dialect, and in Wetan and Kisar, which are distinct languages, suggests that Malay loans were adapted to existing structural constraints. Many of these words also occur in non-metathesised form. At first glance the distribution of metathesised and non-metathesised variants in Leti appears to involve grammatical conditioning, since the same morpheme may occur in one form phrase-medially and another phrase-finally. The conditions governing these alternations are fairly complex, but appear nonetheless to be statable entirely in phonological terms.

Although the facts in Leti may be treated synchronically as instances of metathesis, Mills and Grima (1980) have shown that the observed permutation of consonants probably arose through two distinct changes: 1) addition of an echo vowel, and 2) syncope of a vowel in the environment VC_CV. The evidence for this interpretation comes from forms that lacked the environment for vowel syncope, as with *kawil (> *kail) > Leti *a:li*, Moa *áili* 'fishhook', *likud > Leti *li:ru*, Moa *líuru* 'back', or Leti *ta:li*, Moa *táili* (Malay *tahil*, ultimately from Southern Min) 'a commercial unit of weight'. In these forms the Moa dialect clearly retains the -VC order, but with the addition of an echo vowel. Given this observation etymologies such as *kulit > *ulti* can be reinterpreted as the result of a series of changes *kulit > *kuliti > uliti > *ulti*. Every Moa form with a medial vowel sequence also has a metathesised variant (*ail*, *liur*, *tail*), suggesting that once vowel copying and syncope had applied historically the resulting variants were reinterepreted as related by metathesis.

Other languages in which systematic metathesis has been reported include Dawan (Atoni) of western Timor, Kiandarat of eastern Seram, Sissano, on the north coast of New Guinea, Ririo, on Choiseul Island in the western Solomons, Kwara'ae, on Malaita in the southeast Solomons, Rowa, in the Banks Islands of northern Vanuatu, Lolsiwoi, on the island of Aoba in central Vanuatu, and Rotuman, spoken north of Fiji in the central Pacific.

In Kiandarat and Sissano the sequences *-Ci and *-Cu merged as -Ci before metathesising: PMP *hapuy > Kiandarat *aif-a* 'fire', *susu > *suis-a* 'breast', *batu > *wait-a* 'stone'; POC *kani > Sissano *ʔain* 'eat', *boŋi > *poin* 'night', *manuk > *main* 'bird', *ranum > *rain* 'water', *kutu > *te-ʔuit* 'louse'. In Ririo -CV sequences metathesised with contraction or coalescence of a resulting vowel sequence, as seen in comparing the following forms in the closely related Babatana : B *madaka* : R *madak* 'blood', B *pade* : Ririo *pɛd* 'house', B *saŋgi* : R *sɛŋg* 'bear young', B *vato* : R *vɔt* 'manner', B *boko* : R *boʔ* 'pig', B *vumi* : R *vuim* 'beard', B *mata-ŋgu* : R *matóŋg* 'my eye', B *vati* : R *vɛc* 'four'. In Kwara'ae final -CV sequences are said to metathesise in rapid speech through a process that begins with anticipatory vowel copying. Certain resulting vowel sequences contract, while others do not (careful speech forms precede the colon, rapid speech forms follow): *faŋa* : *faŋ* 'food', *leka* : *leak* 'go', *kusi* : *kuis* 'negative', *pita* : *piat* 'Peter'. A similar type of metathesis is found in Rowa, of the Banks Islands, which when compared with the closely related Mota shows transposition of the final -CV and invariable change of the vowel to *e*: M *liwo* : R *liew* 'tooth', M *lito* : R *liet* 'firewood', M *siŋa* : R *sieŋ* 'shine', M *siwo* : R *siew* 'down', and a similar innovation in which metathesis and a qualitative change in the metathesising vowel are associated, is found in the more distantly related Lolsiwoi: POC *pulan > *vuol* 'moon', *qusan > *wuos* 'rain', *pilak > *viel* 'lightning', *pose > *voas* 'paddle', *pusuR > *vus* 'bow', *maqetom > *maeat* 'black', *ma-maja > *mamas* 'dry', *mapat > *mav* 'heavy', *tolu > *ke-tol*, *pati > *ke-vet* 'four', *lima > *ke-lim* 'five', *onom > *ke-on* 'six', *boŋi > *mboŋ* 'night'. The relationship between these apparently associated phonological processes in several widely separated languages remains unclear.

The Kiandarat change described by Collins (1982:119) raises another issue. Collins separates the final low vowel with a colon in etymologies such as *hapuy (> *api) > *aif-a* 'fire'. In a synchronic analysis this would indicate a morpheme boundary, but in a diachronic analysis it may mean only that the segment in question is not etymologically accountable. Once again, as with a few other changes that have been discussed previously, there is a striking parallel with a language that is genetically and geographically remote, namely Mambai of central Timor. In Mambai, a close relative of Tetun, *-CV also became -VCa: PMP *hapuy (> *api) > *aifa* 'fire', *asu > *ausa* 'dog', *batu > *hauta* 'stone', *manuk (> *manu) > *mauna* 'bird', *hisi > s-*isa* 'meat', *tali > *taila* 'rope', *tasik > *taisa* 'sea', *kutu > *uta* 'louse'. The available material for both Kiandarat and Mambai makes it clear that the change *-CV > -VCa took place only in nouns, and it is therefore possible that −*a* is a separate morpheme, although one that is yet to be identified or clearly described in either language. In non-nouns metathesis occurred without the addition of -*a*: *matay (> *mate) > *maet* 'die', *inum (> *eno) > *eon* (cf. Kemak *enu*) 'drink'.

Metathesis is sometimes regarded as an innovation that is incompatible with the notion of gradual sound change. But this is true only in theories of sound change that do not recognise the essential role of variation. As seen in Chapter 4 (Table 4.45) in Dawan (Atoni) of western Timor the metathesis of -CV syllables appears to be a change in progress. In general, -CV metathesis in Dawan appears more likely if the last vowel is high. Some published analyses, as Steinhauer (1993, 1996a) state that Dawan metathesis is syntactically conditioned, with underlying forms occurring only in phrase-final position.

During elicitation in 1973, however, Urias Ba'it, a speaker of the Molo dialect, gave both metathesised and non-metathesised variants in serial counting.

Metathesis as a change in progress has interesting implications for general theory. Although it is usually assumed that the drastic consequences of metathesis must intrude on a speaker's attention, speakers of Dawan appear to be as unconscious of the difference between metathesised and non-metathesised forms as speakers of American English are of released vs. unreleased final stops. When confronted with a metathesised form, for example, Urias Ba'it insisted that he was producing the unmetathesised form, and it was difficult to convince him of what he actually said.

Perhaps the most widely discussed case of metathesis in the AN language family is that of Rotuman, a case that has received repeated attention in the general phonological literature. Since the general outlines of this process have been discussed in chapter 4, little need be added here. As noted earlier, Blevins and Garrett (1998) conclude that Rotuman metathesis was a two-step process involving 1) anticipatory vowel copying, and 2) deletion of the final unstressed vowel. If so, Rotuman, like Letinese, has a synchronic rule of metathesis that originated from a sequence of two changes neither of which involved a permutation of segments.

Finally, a number of languages show metathesis across a morpheme boundary with the infixes *-um- and *-in- (*C<um>VCVC > *mu-CVCVC, *C<in>VCVC > *ni-CVCVC). Although this type of change is sometimes called 'infix metathesis' there is some question as to whether the infix itself moves, or whether the first consonant of the base and the consonant of the infix change places with a subsequent adjustment of the morpheme boundary: *C<um>VCVC > *m<uC>VCVC > *mu-CVCVC . As noted in chapter 6, the most likely explanation for this type of innovation is that infixes are marked and children tend to convert them to prefixes so as to create an unbroken base form. Most innovations in child language leave no trace in adult grammars, perhaps because they are so idiosyncratic, but is possible that changes such as the metathesis of infixal consonants are so common that they persist into adult speech.

To summarise, the following generalisations can be justified. First, sporadic metathesis generally transposes C_1 and C_2. Second, no recorded example of regular metathesis adheres to this pattern except metathesis of the infixes *-um- and *-in-. Third, regular consonant metathesis falls into three categories: 1) PAN *CVS to PMP *hvC, where C was a stop, and *CVS could be either the last CVC or the first CVC of the word base, 2) *tVs to sVt or or *tVCVs to sVCVt in many languages of northern Luzon and northern Sulawesi, 3) *-ʔC- to -Cʔ- (or vice versa) in the central Philippines. To these we can add metathesis of derived coronal-noncoronal consonant sequences that have the same value for nasality (two stops, two nasals), as in *qatep > Cebuano *atúp* 'roof' : *apt-án* 'roofed', or *inum > Cebuano *inúm* 'drink' : *imn-án-an* 'drink somewhat habitually'. Most cases of systematic metathesis in AN languages involve the transposition of -CV- sequences, and so may be amenable to a unitary explanation. It seems doubtful, however, that this explanation can be applied to *CVS metathesis in Proto Malayo-Polynesian, *tVs to *sVt, or metathesis of the infixal consonants in *-um- and *-in-. The AN languages are an exceptionally rich source of data on metathesis, and doubtless will continue to provide important evidence for theories concerning this type of change long into the future.

9.1.7 Preglottalisation and implosion

Preglottalised and implosive consonants have developed in a number of AN languages. These tend to have an areal character.

Thao, Bunun and Tsou of central Taiwan have preglottalised labial and alveolar stops which show little or no implosion. In Bunun and Tsou they result from automatic preglottalisation of *b* and *d* in prevocalic position, and in Thao they are due to borrowing from Bunun. Historically, PAN *b and *d remained *b* and *d* in Bunun, but became Tsou *f* and *c*; Tsou *b* and *d* are thus historically secondary segments. The distribution of preglottalised stops in central Taiwan suggests an origin in Bunun or Tsou, with later diffusion to neighbouring languages. Since Kanakanabu and Saaroa lack such segments it is most likely that preglottalisation is a feature that arose first in Bunun, and then spread north to Thao and west to Tsou.

In the southern Philippines implosive allophones of *b*, *d* and *g* are reported for Central Sama of the Sulu Archipelago, and of *b* and *d* (but not *g*) for Sindangan Subanun (Central Subanen) of western Mindanao (Reid 1971). Maranao *b*, *d*, *g* reportedly have implosive allophones only in initial position (Lobel 2013:286). In all three cases implosion apparently arose spontaneously as an allophonic feature of voiced stops.

More significantly, the Proto North Sarawak voiced aspirates $*b^h$, $*d^h$, $*j^h$, $*g^h$, which gave rise to standard Kelabit b^h, d^h, d^h, g^h, Kiput *s*, *s*, *s*, *k*, and Highland Kenyah *p*, *t*, *c*, *k* produced phonemically contrastive labial and alveolar implosives ɓ, ɗ in Bintulu, as well as a non-implosive *j* and *g* (Blust 1973b). As noted in 4.1.3, Lowland Kenyah dialects such as Long Ikang, Long San, and Long Sela'an have generalised implosion to all voiced stops. In some of these dialects it has remained allophonic, while in those that have begun to reduce prenasalised stops to simple stops, implosion is becoming phonemic at all four points of articulation (Blust 1980a). Although both the Bintulu and Lowland Kenyah implosives ultimately reflect PNS voiced aspirates, implosion has developed independently in the two groups of languages.

The Chamic languages of mainland Southeast Asia show many areal adaptations to their Mon-Khmer neighbours (Thurgood 1999). Among these, clusters of voiced stop + PMP *h became phonemically preglottalised stops, presumably through a prior change of *h to glottal stop, as in PMP *buhek (> *buʔək) > Jarai ʔbuk 'head hair'. Because of these very narrowly-defined conditions, there are relatively few examples of such segments.

As Klamer (2002) has pointed out, implosive stops are an areal feature in southern Sulawesi and adjacent parts of the Lesser Sundas. In several South Sulawesi languages voiced geminates are realised phonetically as preglottalised stops. In Buginese, Makasarese, and Mandar these include all voiced stops: /bb/ [ʔb], /dd/ [ʔd], /jj/ [ʔdʒ], /gg/ [ʔg], and occur both within a morpheme and at morpheme boundaries when a consonant-initial base is prefixed with *mar-*. However, neither preglottalisation nor implosion is phonemic in any South Sulawesi language (Mills 1975). Table 9.13 shows the distribution of phonemic implosives in languages of this area:

Table 9.13 Distribution of implosive stops in southern Sulawesi and the Lesser Sundas

Language	Labial	Alveolar	Palatal	Velar
Muna	x			
Kulisusu	x	x		
Wolio	x	x		
Tukang Besi	x	x		
Bimanese	x	x		
Komodo		x	x	
Ngadha	x	x		
Kambera	x	x		
Hawu	x	x	x	x

Table 9.13 is organised by number of implosive consonants. Muna has a single bilabial implosive that apparently developed by simple shift of a phonetic norm, although the change was conditioned, leaving *b unaffected before *u, but shifting its phonetic realisation to an implosive before other vowels (van den Berg 1991b:10). Mead (1998:22) suggests that Kulisusu *ɓ* and *ɗ* are products of simple phonetic shift in inherited forms, and that their contrastive status arose as a result of borrowing non-implosive voiced stops from other languages. In Bimanese implosion may be automatic for *b* (*b became *v*, and there is no other source for [b] within native words), but *d* and *ɗ* clearly contrast, as in *eda* 'see' : *eɗi* 'foot/leg'. The origin of the Bimanese implosives is obscure, but where comparative information sheds light on the question it appears that implosives in most languages of this geographical region reflect plain voiced stops with shift of the phonetic norm.

In addition to the languages listed here, implosive stops probably are found in other languages of southeast Sulawesi and central and western Flores. Riung, of west-central Flores, for example, is said to have implosive allophones of the voiced stops *b*, *d* and *g*, and van den Berg (1991b:10) states that "Implosive consonants are found in all the languages of the Muna-Buton area, and even as far north as Tolaki on the southeast Sulawesi mainland." Mead (1998:19) qualifies this remark: Tolaki voiced stops *b* and *d* are optionally imploded, and "Implosion is more frequent preceding mid and low vowels, and more characteristic of rustic than urban speech."

Finally, as noted in passing in chapter 4, true implosives appear to be absent in Oceanic languages.

9.1.8 Gemination

As noted in chapter 4, consonant gemination is widespread in the AN language family, but as with other historically secondary phonological traits it often has an areal character. Among Formosan languages geminates have been confidently reported only for Kavalan and possibly Basay; however, they are found in most languages of northern Luzon apart from the South Cordilleran group. Elsewhere in the Philippines geminate consonants are found only in the Samalan languages of the far south and in Tausug, which has been in close contact with them for centuries. None of the languages of Sabah are said to have geminates, but gemination is found in several of the North Sarawak languages, including Sa'ban, Berawan, and Kiput. Further south geminates are found in Toba Batak of northern Sumatra, Madurese, Talaud of northern Sulawesi, and in Tae', Buginese, Makasarese and other South Sulawesi languages. In eastern Indonesia they are found in some of the

languages of Sumba and some Lamaholot dialects, but are otherwise rare. In the Pacific geminates appear in various languages of Manus, in many of the Nuclear Micronesian languages, and in some Polynesian Outliers in Micronesia and Melanesia.

Geminate consonants in AN languages most commonly arise through one of four processes: 1) assimilation of a nasal to a following homorganic stop, 2) assimilation of heterorganic consonants in a cluster, 3) compensatory lengthening after schwa, followed by merger of schwa with some other vowel, 4) syncope between identical consonants or consonants that share a common place feature. A fifth process will be discussed in 9.2.2.3.

The first case, which is attested only for nasal + voiceless stop, can be illustrated by Toba Batak of northern Sumatra, which shows the developments *mp > -pp-, *nt > -tt-, and *ŋk > -kk- as well as *ns > -ts-): *ampeRij > *apporik* (written *amporik*) 'rice bird', *tentu > *tottu* (written *tontu*) 'certain', *luŋkas > *lukkas* (written *luŋkas*) 'be open'. Heterorganic sequences of nasal + voiceless stop sometimes show geminating assimilation and sometimes do not: *kamkam > *hakkam* 'seize with the hand, grasp', but *tiŋtiŋ 'ring a bell' > *tiŋtiŋ* 'announce', *tuŋtuŋ > *tuŋtuŋ* 'slitgong'. It is possible that preservation of the nasal in both halves of *tiŋtiŋ* and *tuŋtuŋ* is due to onomatopoetic retention. Sequences of nasal + voiced obstruent remained unchanged.

The situation in Kiput of northern Sarawak is somewhat different. In Kiput geminate stops appear to reflect earlier clusters of nasal + voiceless stop, but only in loanwords from Malay, as in *lappiəw* 'kind of fish trap' (< Malay *səlambaw*, with probable postnasal devoicing before nasal assimilation), *lappuŋ* 'lamp' (< Malay *lampu*), *guttiŋ* 'scissors' (< Malay *guntiŋ*), *lattaay* 'chain' (< Malay *rantay*), *lattiŋ* 'raft' (< Malay *rantiŋ*), *mattay* 'kingfisher' (< Brunei Malay *mantis*), *baccɪʔ*, 'to hate' (< Malay *bənci*), *kacciŋ* ('button' (< Malay *kanciŋ*), *jacciʔ*, 'promise' (< Malay *janji*), and *bassəʔ* 'race, nationality' (< Malay *baŋsa*). There are few examples of geminate consonants in native words, and where these occur, as in *daccih* 'crocodile' (cp. *dacih* 'large'), or *durrəy* 'escape' (cp. *durəy* 'thorn'), they are etymologically obscure.

The second case can be divided into two subtypes: one in which a heterorganic consonant cluster existed previously, and another in which a cluster arose through vowel syncope. The first subtype is seen in Madurese, where some geminates have arisen from the complete assimilation of earlier heterorganic consonants in CVCCVC reduplicated monosyllables, as in PMP *bakbak > *babbaʔ* 'tree bark', or *paqpaq > *pappa* 'frond of coconut or banana tree'. The second subtype is seen in Atta of northern Luzon, Buginese and Makasarese of southern Sulawesi, and a number of the languages of Manus in the Admiralty Islands of western Melanesia. Vowel syncope produced heterorganic consonant clusters that remain unassimilated in many Philippine languages, but in Atta these show complete assimilation of the first consonant to the second: *qalejaw > *a:ggaw* 'day' (cp. Batad Ifugaw *algáw*). Similar processes lie behind the geminate consonants of some of the languages of central and eastern Manus, as with *pa-panako (> pahanak) > Kuruti *pahna*, but Nali *panna* 'to steal'. In addition, assimilation of the coda of the active verb prefix *mag-* in Atta has produced heteromorphemic geminates in many verbs, as with Atta *majjiguq* 'bathe' (< *maR-diRuq), or *mappi:li* 'to choose' (< *maR-piliq). Very similar examples of consonant gemination across a morpheme boundary are found in South Sulawesi languages such as Buginese, Makasarese or Mandar (Mills 1975), except that voiced geminates are realised as preglottalised stops.

The third case can be illustrated by Isneg, Makasarese, or Sri Lankan Malay: consonants geminated allophonically after schwa, which then merged with another vowel. As seen already, consonant gemination after schwa is common in insular Southeast Asia, where it

usually remains part of the synchronic phonology. When sound change eliminates the predictability of gemination, however, allophonic consonant length becomes phonemic. In the first two languages *e merged with *a, producing singleton : geminate contrasts after /a/: *enem > Isneg *annám*, Makasarese *annaŋ* 'six', *teken > Isneg *takkán*, Makasarese *takkaŋ* 'punting pole; staff' (cp. *anak > Isneg *anáʔ*, Makasarese *anaʔ* 'child'). In Sri Lankan Malay (Adelaar 1991) consonants automatically geminated after schwa, which then became *i* if there was a front vowel in the next syllable, but *u* if the next vowel was back, giving rise to phonemic geminates (SM = Standard Malay, SLM = Sri Lankan Malay): SM *kəcil* : SLM *kiccil* 'small', SM *ləbih* : SLM *libbi* 'more', SM *pərgi* 'go' : SLM *peggi, piggi* 'gone, last', SM *təman* : SLM *tumman* 'friend', SM *pənuh* : SLM *punnu* 'full', SM *təbu* : SLM *tubbu* 'sugarcane', SM *təlor* : SLM *tullor* 'egg'.

The last case can be illustrated by a recurrent change in Oceanic languages whereby a vowel between identical consonants drops to produce a geminate consonant. In many languages this happens only with CV- reduplication followed by syncope, as in Kapingamarangi: *piki > *piki* 'sticky' : *pi-piki > *ppiki* 'stuck', *tuki > *tuki tuki* 'pounder, mallet' : *tu-tuki > *ttuki* 'to pound', *laka > *laka laka* 'stride' : *la-laka > *llaka* 'stride'. Very similar developments appear in other Polynesian Outlier languages, in many Nuclear Micronesian languages, and in some of the Oceanic languages of Melanesia, as Mussau of the St. Matthias Archipelago. In Mussau, which added echo vowels and so has many underlying trisyllabic bases, syncope between identical consonants is common within a morpheme for younger speakers, but not for older speakers: *mumuko* : *mmuko* 'sea cucumber', *papasa* : *ppasa* 'outrigger poles', *kabitoto* > *kabitto* 'nit', *mumumu* > *mummu* 'suck'. The pattern recorded in Blust (1984c), whereby syncope occurred only in words of more than two syllables, apparently has been extended to words of two syllables by younger speakers (Brownie and Brownie 2007:16). Finally, in Nuclear Micronesian languages such as Chuukese, syncope has occurred between consonants that share place features, but not necessarily manner features (Blust 2007b).

Table 9.14 Gemination of final consonants in Talaud

PMP	Talaud	
*-p	-ppa	*qatep > atuppa 'roof, thatch'
*-t	-tta	*Ramut > žamutta 'root'
*-k	-ʔa	*anak > anaʔa 'child'
*-b	-bba	*tutub > tutubba 'to close, as a door'
*-d	-dda	*likud > liʔudda 'back'
*-g	?	
*-m	-mma	*inum > inumma 'drink'
*-n	-nna	*kaen > anna 'eat; food'
*-ŋ	-ŋŋa	*daŋdaŋ (> dadaŋ) > daraŋŋa 'warm by a fire'
*-s	-ssa	*teRas > tohassa 'hard, strong'
*-l	-lla	*beŋel > beŋella 'deaf'
*-R	-kka	niuR > niukka 'coconut'

One other case that does not fit neatly into any of these, but is not clearly 'aberrant' is found in Talaud of northern Sulawesi. Sneddon (1984) reports three sources of gemination in Talaud. The first of these is an underdescribed and poorly understood process of consonant doubling under prefixation, as where Cu- + *daḷanna* 'to walk' > *duddaḷanna* 'is walking'. The second is the now familiar one in which consonants were phonetically

geminated after schwa, which then merged with *a*: PMP *enem > *annuma* 'six', *epat > *appata* 'four', *lemes > *lammisa* 'drown', *qateluR (> *teluR) > *talluka* 'egg' (cp. *anak > *anaʔa* 'child', or *mata > *mata* 'eye', without gemination). The third and most interesting is summarised in Table 9.14.

In short, all final consonants except glides (which fused with the preceding vowel), and *-k (which became a glottal stop that could not geminate) are doubled and followed by *-a*. Sneddon notes certain constraints on this process. First, as seen in *enem > *annama* 'six', gemination after schwa preceded the doubling of final consonants, and blocked the process in this environment. Second, gemination is also often blocked following consonant clusters, as in PMP *demdem > Talaud *danduma* 'dark', *puŋgut > *puŋguta* 'tailless', or *sandeR > *sandaka* 'lean against'. Together these conditions suggest a general constraint against more than one long consonant or consonant cluster in a morpheme. Multiple geminates are permitted, however, within the same word, as in *duddaḷanna* 'is walking'. More surprisingly, gemination is also blocked if the onset of the final syllable is *ž* (< *R): PMP *hiRup > *ižupa* 'sip, suck up', *huRas > *užasa* 'to wash', *habaRat 'west monsoon' (> *baRat) > *bažata* 'west', PS *paRes 'hit with an object' > *pažasa* 'knock down fruit with a pole'. The nature of this change is puzzling. Since words that originally ended in *-Ca did not geminate the onset of the final syllable (*apa > *apa* 'what?', *lima > *lima* 'five', *mata > mata 'eye', *qasawa (> *sawa) > *saβa* 'spouse', *taumata > *taumata* 'person') *-a* epenthesis evidently could not have preceded gemination. But if gemination preceded *-a* epenthesis this change would have resulted in phonetically difficult or impossible final geminate obstruents *-pp*, *-tt*, etc. There seem to be two alternatives: either *-a* epenthesis occurred first and consonants preceding final *-a* were geminated only in words of more than two syllables, or final consonant gemination and *-a* epenthesis in Talaud were part of a single complex sound change.

The decision whether geminates should be treated as unit phonemes or as clusters can sometimes be difficult. Since the Indic-based Batak syllabary still represents geminates as nasal + stop, speakers of Toba Batak are likely to perceive *-pp-*, *-tt-* and *-kk-* as clusters. However, it is possible that cross-linguistically geminates which reflect consonant clusters preserve the syllable boundary, while those that reflect single stops are perceived as units.

Geminate, or long vowels, have developed in a few AN languages as a result of 1) the loss of intervening consonants, or 2) the reduction of longer morphemes to monosyllables, which are obligatorily bimoraic in many daughter languages. Either or both processes are operative in e.g. POC *layaR > PPN *laa 'sail'(with one syllable peak), but only the former can be implicated in e.g. POC *apaRat 'west monsoon' > PPN *afaa 'storm, hurricane', next to POC *apa > PPN *afa 'what?'. Potentially confusable with geminate or long vowels, but distinct from them are sequences of heterosyllabic like vowels that have developed through the loss of an intervening consonant as with PMP *daRaq > Mukah Melanau *daaʔ* 'blood' (with two syllable peaks; cp. *gaʔ* 'generic marker of location').

9.1.9 Innovations affecting nasals

Two types of change affecting nasal consonants are found in western Indonesia and the Chamic languages of mainland Southeast Asia, with a major concentration of cases on the island of Borneo. In the first type a number of languages have developed what Court (1967) called 'preploded' nasals, a typologically unusual category that was discussed briefly in Chapter 4. Impressionistically these are heard as nasals preceded by a short homorganic stop, an effect produced by postponing velic lowering until an oral closure is

achieved. To illustrate, whereas Malay has *tajam* 'sharp', *hujan* 'rain', and *tulaŋ* 'bone', the corresponding forms in Selako, a Malayic Dayak language of southwest Borneo, are *taja^{p}m*, *uja^{t}n* and *tua^{k}n*. Similar final nasals are widespread in the Land Dayak languages of southwest Borneo. In some languages, as Selako, the preploded portion of the nasal is heard as a fleeting voiceless stop. In others, as Kuap Land Dayak, it is heard as a fleeting voiced stop (PMP *dalem > Kuap *dərə^{b}m* 'in, inside', *quzan > *uje^{d}n* 'rain'). The most striking features of preploded nasals in AN languages are: 1) they occur only word-finally, 2) preplosion is suspended in syllables that begin with a nasal consonant, 3) over time preploded nasals tend to merge with corresponding voiceless stops, 4) nasal preplosion has an areal character, yet is found in languages that are geographically discontinuous.

As noted in Chapter 4, primary nasal consonants trigger rightward nasal spreading in most AN languages for which adequate phonetic information is available, and reverse nasal spreading is normally very slight. For this reason preplosion never affects final nasals if the last syllable has a nasal onset, as in *malem > Lundu *maəm* 'night' *diŋin > *diŋin* 'cold', or *kuniŋ* 'yellow' (Malay *kuniŋ*). When nasality spreads rightward a syllable of the form -NVN presents a conflict between nasal spreading and final preplosion. In such situations nasal spreading predominates, nasalising the final vowel and thus preventing oral closure before the velic is lowered. Where the onset of a final syllable is non-nasal, however, the vowel that follows it is oral. Nasal preplosion, then, can be seen as a reactive tendency to protect the vowel before a nasal coda from reverse nasal spreading (Blust 1997c). In some languages preploded nasals have been further simplified to voiceless stops, as in Keninjal, a Malayic Dayak language spoken in the southern Sarawak-Kalimantan border region: *tazem > *tajap* 'sharp', *quzan > *ujat* 'rain', *tuqelaŋ > *tulak* 'bone', but *haŋin > *aŋin* 'wind'. Although nasal preplosion is perhaps best-known from the Land Dayak region of southwest Borneo (Court 1967), it has a wide and discontinuous distribution in western Indonesia from north of Sabah (Bonggi), most Land Dayak languages, and some contiguous Malayic Dayak communities, as Selako, in southwest Borneo, Tunjung in east Kalimantan, Lom, on the islands of Bangka and Belitung between Sumatra and Borneo, Rejang, in southern Sumatra, Mentawai, spoken in several dialects in the Mentawai Islands west of Sumatra, Urak Lawoi', a phonologically aberrant Malay dialect spoken in southwestern peninsular Thailand, and two Chamic languages, Roglai and Tsat. Some of these languages, as Keninjal, no longer have preploded nasals, but once did, as shown by reflexes of final nasals as the homorganic voiceless stop unless the last syllable begins with a nasal consonant. Closely similar forms of the same phenomenon are found in some of the Mon-Khmer languages of mainland Southeast Asia, and in somewhat different forms, in other parts of the world (Blust 1997c).

Nasal postplosion begins with medial clusters of nasal + voiced stop. It has been shown that the stop portion of nasal-stop clusters is shorter with voiced stops than with voiceless stops in many languages (Cohn 1990). In a number of languages of western Indonesia including Narum of northern Sarawak, Iban of southern Sarawak and adjacent parts of Kalimantan, Acehnese of northern Sumatra, Rejang of southwest Sumatra, and Balinese, this universal phonetic tendency has been carried further, so that the voiced stop is barely perceptible, as in Narum: *kambiŋ > *am^{b}iŋ* 'goat', or *p-inzam > *pin^{j}am* 'borrow'. In effect, nasal postplosion is a change whereby voiced stops following medial nasals are shortened to the point of near-imperceptibility, leaving an oral vowel after the medial nasal as the principal indication of a former or underlying cluster. As noted in Blust (1997c), there is a strong tendency for languages in this area to have both preploded final nasals and

postploded medial nasals, as seen in the native pronunciation of the Kendayan ([kən^dája^tn]) Dayak of southwest Kalimantan, bordering the Land Dayak region of Sarawak.

Preploded final nasals are easily recognised even in phonetically unsophisticated sources, but this is not true of postploded medial nasals, which in published sources may be written simply as *-mb-*, *-nd-*, *-nj-*, *-ngg-*, or as the corresponding simple nasals. If etymological information shows that an apparent simple medial nasal reflects a prenasalised voiced obstruent, it is likely that the simple nasal is a mistranscription for a postploded nasal, or was a postploded nasal at an earlier time. Finally, most languages with postploded medial nasals appear to have resyllabified the nasal from coda to onset position, since the morpheme structure of forms that once had prenasalised voiced obstruents has changed from CVCCVC (hence CVC.CVC) to CVCVC (hence CV.CVC).

The second region in which nasals show significant innovations is the Pacific, where many Oceanic languages reflect *m (and *p) as labiovelars. The definition of 'labiovelar' in Oceanic languages is variable. In many languages labiovelar consonants appear to be simple labials with rounded offglides. In Nuclear Micronesian languages, however, labiovelar consonants such as *m^w* reportedly show lip spreading rather than lip rounding. As noted in 4.1.12, still other languages, especially in northern Vanuatu, have labiovelar consonants that appear to be produced by co-articulation: [ŋm^w], [kp^w]. The antiquity and source of these consonants is a matter of some debate. Blust (1981a) pointed out that although labiovelars are found in many Oceanic languages, and often in cognate morphemes, cross-linguistic disagreements between labial vs. labiovelar are rather common. To illustrate, the great majority of Oceanic languages reflect *lima 'hand, arm; five', yet Mota *lim^wa* 'hand; five', and Fijian *liŋa* 'forearm and hand' reflect *lim^wa. Likewise, most Oceanic languages reflect *Rum^waq 'house', yet a few languages that have labiovelar consonants instead reflect *Rumaq, as Wogeo, Arosi *ruma*, Lau *luma*, Sa'a *nume* 'house', Nauna *yum* 'house', *yuma-n* 'nest, web'. In forms such as PMP *Rumaq > POC *Rum^waq, the development of a labiovelar can be seen as conditioned by an adjacent rounded vowel. In others, however, as *lima > Mota *lim^wa*, Fijian *liŋa*, this explanation is not available. Lynch (2002) has revisited this issue in great detail, and concludes that most labiovelars in Oceanic languages are products of conditioned change or borrowing. This may be true, but it leaves a significant residue of cases like Mota *lim^wa*, Fijian *liŋa* unexplained. Apart from these changes little else need be said about the evolution of nasal consonants in AN languages apart from the merger of all word-final nasals as *-ŋ* in a number of languages.

9.1.10 Vocalic change

PAN had just four vowels: *a, *i, *u and *e (schwa). Of these the schwa is historically the least stable, showing a wide range of reflexes. The vowels of the classic vowel triangle, on the other hand, are preserved in at least some environments in all known languages. Where *a changes, it is generally to schwa or *o* in word-final position, or to *i* or *e* through low vowel dissimilation in Oceanic languages, or low vowel fronting in languages of northern Sarawak and northeast Luzon. Low vowel dissimilation, which affects the first of two low vowels in successive syllables, has already been discussed, and low vowel fronting, which affects low vowels that follow a voiced obstruent, will be addressed under 'Bizarre sound change'. Other types of vocalic change are treated separately below.

9.1.10.1 Expansions of the vowel inventory

Various types of change have led to expansions of the original inventory of four vowels. In some cases these expansions are moderate, but in others they are extreme. Suprasegmental changes such as the addition of tonal distinctions or an oral/nasal contrast are treated separately below.

High vowels are lowered in certain environments in some languages, and the loss of the conditioning environment has sometimes given rise to new mid-vowels *e* and *o*. An example is the Long Labid dialect of Penan in northern Sarawak, where *R and *s > *h*, high vowels lowered before final *h* and *ʔ*, and *h* then disappeared: PMP *ikuR (> ikuh > ikoh) > *iko* 'tail', *Ratus (> hatus > atuh > atoh) > *ato* 'hundred', *sendiR (> səndih > səndeh > səreh) > *səre* 'lean against', *beties (bətih > bəteh) > *bəte* 'calf of the leg'. As a result of height harmony some phonemic mid-vowels were also created in penultimate position, as in *nipis > *nepe* 'thin, of materials'. Far more dramatic expansions are seen in some Oceanic languages, as Chuukese, which has developed nine contrasting vowels as a result of various assimilatory processes and subsequent loss of the conditioning factors (Dyen 1949), or some of the languages of northern Vanuatu which, according to François (2005) have developed as many as 16 vowel phonemes (this counts long vowels separately, rather than treating them as containing a prosody of length).

Other major changes in the vowel inventory have created numerous new diphthongs in addition to new vowels. Among the most spectacular elaborations of the original vowel system into new systems of vowels and diphthongs are those of the Musi dialect of Rejang in south Sumatra, which shows some 27 splits and 21 mergers of the PAN vowels (Blust 1984b, McGinn 2005), and the rich processes of vowel breaking in the Melanau languages of coastal Sarawak. As noted in §4.3.2.10, in a number of languages in coastal Sarawak, and some further inland high vowels have undergone two patterns of breaking or diphthongisation. The first pattern is found in closed final syllables, where high vowels developed a mid-central offglide before final *k* and *ŋ*. The second is found in open final syllables, where high vowels developed a mid-central onglide. Offgliding has left a residue in the synchronic grammar of most languages, but ongliding has led to restructuring, as seen in Mukah Melanau, where *titik > *titik* ([títijəʔ]) 'speck, dot', *manuk > *manuk* ([mánuwəʔ]) 'bird', *pusiŋ > *pusiŋ* ([púsijəŋ]) 'to turn', *buŋbuŋ > *bubuŋ* ([búbuwəŋ]) 'ridge of the roof', next to *qubi > *ubəy* 'yam', *telu > *tələw* 'three'. The details of vowel breaking differ across languages. Mukah, for example, is unusual in that *a participated in this change, raising and offgliding to [ejə], as in *anak > *anak* ([ánejəʔ]) 'child', or *bintaŋ > *bitaŋ* ([bítejəŋ]) 'star'. It is also unusual in that final *k lenited to -*ʔ* after breaking occurred, leaving the distinction between underlying -*ik* and -*iʔ* as [ijəʔ] vs. [eʔ]. The Uma Juman dialect of Kayan, on the other hand, shows breaking of *i before both *-k and *-ŋ, but *u lowers in this environment. Table 9.15 plots some of the variability in patterns of vowel breaking found in languages of coastal Sarawak and adjacent areas (KT = Kampung Teh, KK = Kampung Kekan). The data is given in phonetic transcription:

Table 9.15 Patterns of vowel breaking in languages of Sarawak
(predictable glides are omitted)

	*-ik	*-iŋ	*-uk	*-uŋ	*-ak	*-aŋ	*-i	*-u
Kiput	-iəʔ	-iə	-uəʔ	-uə	-ak	-aŋ	-əy	-əw
Balingian Melanau	-iək	-iəŋ	-uk	-u(ə)ŋ	-ak	-aŋ	-əy	-əw
Mukah Melanau	-iəʔ	-iəŋ	-uəʔ	-uəŋ	-eəʔ	-eəŋ	-əy	-əw
Dalat Melanau KT	-iək	-iəŋ	-uək	-uəŋ	-ak	-aŋ	-əy	-əw
Dalat Melanau KK	-eək	-eəŋ	-oʔ	-u	-iʔ	-i	-əy	-əw
Matu Melanau	-it	-in	-ok	-oŋ	-ak	-aŋ	-əy	-əw
Sarikei Melanau	-iəʔ	-in	-uəʔ	-uəŋ	-ak	-aŋ	-ay	-aw
Uma Juman Kayan	-iək	-iəŋ	-ok	-oŋ	-ak	-aŋ	-eʔ	-oʔ
Uma Bawang Kayan	-ik	-iŋ	-uk	-uŋ	-eək	-eəŋ	-eʔ	-oʔ
Long Wat Kenyah	-iək	-iəŋ	-uək	-uəŋ	-ak	-aŋ	-əy	-əw

Table 9.15 provides only a limited picture of the complex variability of vowel breaking in the languages of Sarawak. Even with this restricted set of languages and dialects, however, it is clear that vowel breaking in Sarawak represents an exuberant expression of variations on a theme. In Mukah Melanau, for example, breaking of the low vowel also occurs before *-h* from *R > r: *sandaR > *sadar* ([sádejəh]) 'lean on', *gagar* ([gágejəh]) 'kind of raised platform', in Long Wat Kenyah the common environment for breaking with offglides (before final *-k and *-ŋ) has been extended to include -ut and -un in at least some forms (but apparently not -it and -in), in Dalat Melanau KT, the pattern of mid-central onglides that led to the restructuring of final high vowels is also found before final glottal stop and h: *puluq > *puləuʔ* 'group of ten', *betis > *bətəih* 'calf of the leg', and in Sarikei Melanau offgliding of high vowels also occurs before final *h* and *t*. The most elaborate histories of vowel breaking in Sarawak are found in the Melanau languages and to a lesser extent in such Lowland Kenyah languages as Long Wat. From a coastal site of origin this innovation then evidently spread up the river systems, where it reached some of the Kayan dialects in an attenuated form. While vowel breaking is well known from the history of the Romance languages, the data from Sarawak differs from these in two respects. First, breaking is conditioned by final velars, but in those languages that distinguish -*k* from -*g*, as Mukah or Dalat Melanau, breaking does not take place before -*g*: Mukah Melanau [hig] 'budge, move slightly', [dúhig] 'mythical forest monster', [tug] 'ball of the heel', [pádʒug] 'foot', [tátag] 'patch, repair', [típag] 'stamp feet', Dalat Melanau KT [lílig] 'tree resin', [ríbig] 'pinch' (cp. the disyllabic [píjəg] 'shivering'), [múug] 'rub, scrub, as a floor', [tug] 'heel'. Second, Mukah stress is penultimate, and the vowel that undergoes breaking—unlike vowel breaking in the history of the Romance languages—is thus found in an unstressed syllable.

The one feature that is more or less constant throughout the variability seen in Table 9.15 is the breaking of final high vowels as *-əy* and *-əw* (only Kayan, which has almost certainly acquired vowel breaking by diffusion from the Melanau languages, differs). The same feature is found in southern Sumatra, as in Rejang, where *mi-Sepi (> mipi) > *mipəy* 'dream', *hisi > *isəy* 'contents', *waRi > *biləy* 'day', *bulu > *buləw* 'body hair', feathers', *sapu > *supəw* 'broom' or *qulu > *uləw* 'head', and in the Chamic languages, as in PMP *beli > Jarai *bləy* 'buy', *duRi > Jarai *drəy* 'thorn', *bulu > Jarai *bləw* 'body hair, feather', or *telu > Jarai *kləw* 'three'. Minangkabau of southern Sumatra lacks ongliding of final high vowels (*tali > *tali* 'rope', *kutu > *kutu* 'head louse'), but shows offgliding of high vowels before certain final consonants. Adelaar (1992:42ff) notes that earlier *-i and *-u

are not offglided in Minangkabau 1) before -*ʔ* from *p or *t, 2) before -*n* from *m and *n, 3) before -*h* from *s, or 4) when final in the reconstructed form. They are, however, offglided as -*iə*, -*uə* (1) before -*ʔ* from *k, 2) before -*ŋ*, 3) before -*h* from PMP *q, and 4) before zero from earlier *l or *r. What is intriguing about this pattern is that final velars are again among the core elements that trigger offgliding of high vowels. Like several other changes considered in this chapter, the geographical distribution of vowel breaking in AN languages is puzzling: it is confined to parts of northwest Borneo, where it seems clearly to be an areal phenomenon, but skips the Land Dayak area that is closest to Sumatra, only to reappear in parts of southern Sumatra and mainland Southeast Asia. Whether this marks an earlier pattern of contact remains to be seen, but it is striking that vowel breaking of a very specific form is found in just this discontinuous yet geographically confined region and apparently nowhere else in the AN language family.

One other change to high vowels is noteworthy, but irregular. In a number of Oceanic languages, and some languages outside the Oceanic group, *u is occasionally reflected as *i*, without clear evidence of phonological conditioning, as in POC *tusuq > PPN *tusi 'to point, point out', POC *turu > PPN *turi 'knee', PPN *tupu-na > Maori *tipuna* 'ancestor', PPN *qumu > Hawaiian *imu* 'earth oven', or PPN *tamaŋu > Hawaiian *kamani* 'a tree: *Calophyllum inophyllum*'. This peculiar drift-like tendency in a number of Oceanic languages was noted in Blust (1970b), but has not since been revisited, or explained.

This leaves the schwa, which is the vowel most likely to change, and the vowel with the widest range of change paths. PAN *e apparently was a mid-central vowel, although its reflexes in many Philippine languages are high-central (written *i*). In Sundanese the reflex of *e in native words is *i* and the reflex of *e in Javanese loans is *ə*, making this one of the few AN language in island Southeast Asia to contrast high-central and mid-central vowels.

In conditioned change the schwa is reflected as *i, u, o, ə* or *a*. Some languages reflect it as *i* or *u*, depending upon the environment. In Thao of central Taiwan, for example, *e usually was lost in penultimate syllables but preserved before final consonants, where it split into *i* before coronals and velars and *u* before labials: PAN *bukeS > *fukish* 'head hair', *keRet > *klhit* 'cut, sever', *RameC > *lhamic* 'root of tree or grass', *ŋipen > *nipin* 'tooth', *Sajek > *shazik* 'smell, odor', *lemlem > *ma-rumrum* 'dim, unlit', *qaRem > *qalhum* 'pangolin', *dakep > *sakup*/*sapuk* 'catch, seize', *Sulem > *urum* 'cloud, mist'. In Tagalog *e became *u/o* if *u occurred in an adjacent syllable, but otherwise became *i*: PMP *buhek > *buhók* 'head hair', *tebuh > *tubó* 'sugarcane', but *beRas > *bigás* 'husked rice', *deŋeR > *diníg/diŋíg* 'to hear', *tanem > *taním* 'to plant'. As noted already, in Sri Lankan Malay *e > *i* if the next vowel was front, but *e > *u* if the next vowel was back (*u or *a). Some languages show divergent reflexes of *e that are conditioned by place in the morpheme rather than by neighbouring vowels. Most dialects of Malay, for example, reflect *e as *a* in the final syllable, and as [ə] (written *e*) elsewhere, as in *enem > *ənam* 'six'. Buli and other South Halmahera-West New Guinea languages, on the other hand, reflect *e as *o* in the penult but *a* in the final syllable, as in PMP *depa > Buli *lof* 'fathom', *kuden > *ulan* 'clay cooking pot', or *enem > *wonam* 'six'. No examples are known of *e > [e] in a conditioned sound change.

In unconditioned changes the schwa has become every vowel except *i*. The change *e > *u* is attested in Bunun of central Taiwan, Cebuano and other languages of the central Philippines, and Chamorro (with *u/o* allophony), *e > *e* in the northern dialect of Ilokano, and Ma'anyan, and *e > *a* in Isneg and some other languages of northern Luzon, in Kapampangan (and Kapampangan loanwords in Tagalog), in the Murut languages of Sabah, Minangkabau of southwest Sumatra, and Makasarese of southwest Sulawesi. By far

the most common unconditioned change that affected schwa is *e > *o,* which is found in Ifugaw of northern Luzon, Ata Manobo of Mindanao, some of the Samalan languages of the southern Philippines, the Dusunic languages of Sabah, Toba Batak of northern Sumatra, some dialects of Lampung in southern Sumatra, Bolaang Mongondow and its close relatives in northern Sulawesi, Proto Eastern Celebic (Mead 2003b), Wolio, Soboyo, and Proto Oceanic, among others.

9.1.10.2 Monophthongisation

There is some disagreement whether *-ay, *-aw, *-uy and *-iw should be called 'diphthongs' (Clynes 1997, 1999), but they clearly behave differently from other -VC sequences in undergoing mutual assimilation and contraction, usually as *-ay > *-e*, *-aw > *o*, and *-uy and *-iw > *i* in many of the modern languages. Although *-ay > *-e* and *-aw > *o* usually co-occur, as do *-uy > *i* and *-iw > *i*, the latter monophthongisation has taken place in some languages without corresponding changes to *-ay or *-aw.

9.1.10.3 Tonogenesis

Phonemic tone is rare and historically secondary in AN languages. Remijsen (2001:3) points out that it is found in the following areas: 1) some Chamic languages of mainland Southeast Asia, 2) Mayá, Matbat, Moor and other South Halmahera-West New Guinea languages of western New Guinea and neighbouring islands, 3) Kara, Barok and Patpatar of New Ireland, 4) Yabem and Bukawa of the Huon Gulf region of New Guinea, and 5) in five languages of New Caledonia. Most AN languages that have become tonal have also shifted from a predominant disyllabic word canon to predominant monosyllabism. In some of these cases tonogenesis appears to have been initiated by contact with tone languages, but in others no such causal factor can be identified.

The Chamic languages represent a continuum of areal adaptations. In addition to various segmental features that have been acquired from contact with their Mon-Khmer neighbours Western Cham, which is in closest contact with Khmer, has developed register (breathy vs.clear voice) and Eastern, or Phan Rang Cham, which is in closest contact with Vietnamese, has developed incipient tone, although there is some disagreement about how large a role contact has played in this development (Thurgood 1999, Brunelle 2009). Tsat, spoken by a Muslim population on Hainan Island in southern China, has developed a full tonal system with five contrasts (Ouyang and Zheng 1983, Haudricourt 1984, Benedict 1984, Maddieson and Pang 1993).

Thurgood (1999:179) notes that 'Register itself constitutes a complex of features that tend to occur together: voice quality (phonation type), vowel length, pitch, and voice quality induced vowel gliding … individual languages may emphasise one or another of those features, suppressing the other features.' In Western Cham, breathy phonation developed after the voiced obstruents, whereas a modal voice quality was found after other consonants. The difference between breathy and modal voice resulted in two complementary sets of vowels, one associated with each phonation type. In time the breathy phonation type was extended to vowels after sonorants (this is atypical for Chamic), and the transparency of the original conditioning was lost through merger of voiced and voiceless stops and the spreading of register from the pretonic first syllable to the stressed main syllable. In Eastern, or Phan Rang Cham of Vietnam, the changes set in motion by contact have led to an incipient tone system with high and low contrasts. Final glottal stop has further conditioned contours. In Jarai, for example, vowels are predictably

pronounced with a short rising pitch before final glottal stop from *t, as in PMP *epat > Jarai *paʔ* ([paʔ²⁴]) 'four'.

Tsat, spoken on Hainan Island for roughly the past millennium, has been in contact with Tai-Kadai languages and Chinese for many generations, during which time it has become almost completely monosyllabic, and has developed a system of five contrasting tones that Thurgood (1999:215) describes as high (55), mid (33), low (11), high or mid falling (42), and low or mid rising (24). Proto Chamic voiced obstruents as the onset either to the word or to the final syllable produced a breathy voice register that gave rise to 55 before *-h, 42 before *-ʔ, and 11 before other finals, including zero. Other Proto Chamic initial consonants produced a modal voice register that gave rise to 55 before *-h, 24 before *-ʔ, and 33 before other finals, as shown in Table 9.16:

Table 9.16: Development of a five-tone system in Tsat of Hainan Island

Proto Chamic	Tsat	Gloss
*babah	pha⁵⁵	mouth
*mamah	ma⁵⁵	chew
*batuk	tuʔ⁴²	cough
*manuk	nuʔ²⁴	chicken
*habəw[109]	phə¹¹	ash
*dua	thua¹¹	two
*ʔular	la³³	snake
*lima	ma³³	five

Proto Chamic probably had already begun to acquire distinct modal and breathy phonation types from its initial contact with typical Mon-Khmer languages. From this point syllable endings determined contours, and where exposure to tone languages allowed, a full tonal system arose from an atonal parent language that was very much like Malay. The results closely mimic the typology of neighbouring languages belonging to other families, and it is clear that the divergent typological traits associated with these AN languages are the product of generations of language contact.

Tone has also developed in Mayá, Matbat, Moor and some other South Halmahera-West New Guinea languages. Remijsen (2001) has made the remarkable discovery that Mayá, spoken on the islands of Waigeo, Salawati and Misool in the Raja Ampat Archipelago just to the north and west of the Bird's Head Peninsula of New Guinea, has an apparently unique hybrid prosodic system characterised by both phonemic stress and phonemic tone. Most Mayá lexemes are monosyllabic or disyllabic, and syllables may carry one of three tones: high (symbolised /³/), and rising (symbolised /¹²/), which occur only in the final syllable of content words, and something he writes as [+stress] and describes as an 'unmarked toneme.' In polysyllables stress may fall on either the penult or the ultima without respect to tonal values. Matbat, spoken on Misool Island, is said to have five tonemes: extra high falling (/⁴¹/), high level (/³/), low rising (/¹²/), low level (/¹/ and falling (/²¹/). Although he discusses the history of tonal development in these languages in general terms, Remijsen does not provide a set of diachronic rules that enables the tonal values of Mayá or Matbat to be fully predicted from reconstructed forms, as seen in the comparison of etymologies such as PMP *bunuq > Mayá *buᶾn* 'kill' vs. *penuq > *foⁱ²n* 'full', *salaq > *saᶾl* 'error' vs. *qateluR (> *teluR) > *toⁱ²l* 'egg', or *matay > *maⁱ²t* 'die' vs.

---

[109] Thurgood (1999:183) writes Proto-Chamic diphthongs *-ɛy, *-ɔw, for what I interpret to be *-əy, *-əw.

*kutu > u³t 'louse'. The conditions for tonal development in these languages thus remain unclear, as does the possible role of contact with Papuan languages in stimultating tonogenesis.

Capell (1971:264) reported that Kara of northern New Ireland, and Barok of central New Ireland are two-tone languages (high vs. low). Beaumont (1976:391) adds that this is also true of the Sokirik dialect of Patpatar, spoken immediately south of Barok. Given their geographical contiguity, the presence of tone in Barok and one dialect of Patpatar may be due to diffusion, but this is unlikely for Kara, which is separated from Barok and Patpatar by Nalik, Notsi, Madak and the non-AN language Kuot or Panaras, all of which appear to be non-tonal. This matter was taken up again by Hajek (1995), who includes a vocabulary of 141 lexical items for Kara, two dialects of Barok, and the Sokirik dialect of Patpatar. Unfortunately, this material is of limited use for diachronic purposes, and the history of tone in these languages consequently remains obscure.

Yabem and Bukawa of the Huon Gulf region on the north coast of New Guinea have acquired a simple contrast of high and low tones. The development of phonemic tone in these languages followed three steps: 1) a system of obstruent harmony developed in which all stops or fricatives within a word had to agree in voicing; where a conflict occurred voiceless obstruents became voiced, 2) vowels after voiceless obstruents acquired high tone and vowels after voiced obstruents acquired low tone, 3) voicing distinctions for obstruents were lost (Bradshaw 1979). By comparison with some other areas, then, tonogenesis in Yabem and Bukawa is relatively well-understood.

Rivierre (1993) reports that five of the 28 AN languages spoken in New Caledonia are tonal. These are Cèmuhî and Paicî of north-central New Caledonia, Drubea and Numèè of extreme southern New Caledonia, and Kwenyii of the Isle of Pines to the southeast of the main island. In a pioneering diachronic study Haudricourt (1968) pointed out that the aspirated stops of some northern New Caledonian languages and the voiceless fricatives of others correspond in cognate forms to the high tones of Cèmuhî and Paicî. He further related these correspondences to processes that have been observed in a number of other Oceanic languages, particularly in the Polynesian Outliers, where CV- reduplication followed by syncope has given rise to initial geminates, which in some languages have further evolved into aspirated stops. If this suggestion is correct tonogenesis in New Caledonian languages may represent the end-point, although not necessarily the inevitable end-point, of a line of development that is seen in less advanced form in other parts of the Oceanic subgroup of AN: 1) bases are reduplicated, 2) syncope of the reduplicative vowel produces initial geminates, 3) geminates become aspirates, 4) aspirates produce high tone on a following vowel, non-aspirates low tone. Unlike most other known cases, the New Caledonian languages that have developed tone could not have done so as a result of contact with non-Austronesian tonal languages.

In addition to tonogenesis some AN languages have developed a historically secondary phonemic stress. The case of Mayá has already been noted. Other cases include Thao of central Taiwan, Rukai of south-central Taiwan, Pangasinan of central Luzon, Malagasy, and apparently Lindrou of western Manus. In Thao about 98% of all lexical bases have penultimate stress, but roughly 2% are oxytone: *bakóŋ* 'soup bowl', *baksán* 'sticky rice confection', *dadú* 'leader', *falhán* 'rib', *tufúsh* 'sugarcane', *ushán* 'menstruation', etc. Some words with final stress are loans, but others appear to be native, and the conditions under which final stress developed are not always clear. Several Rukai dialects have contrastive stress, but they often disagree, and for this reason Li (1977b) omitted stress from his reconstructed Proto Rukai. In disyllables the Tanan and Tona dialects generally

agree, and differ from Budai: Tanan, Tona *abó*, but Budai *ábo* 'ash', Tanan, Tona *comáy*, but Budai *cómay* 'bear', Tanan, Tona *cakí*, but Budai *cáki* 'excrement', etc. However, in polysyllables Budai often agrees with the other dialects, or with one of them at the expense of the other: Tanan, Budai, Tona *mabosóko*, Maga *mabusúku* 'drunk', Tanan, Budai, Tona *kisísi* 'goat', Budai *ʃikólaw*, Tona *ikólaw*, but Tanan *ʃikoláw* 'clouded leopard'. The history of stress in Rukai remains to be worked out. Several Philippine languages, including Casiguran Dumagat, Ibanag, and Pangasinan in the Philippines, and Ratahan in northern Sulawesi, lost the inherited stress of Proto Philippines and then developed new patterns of contrastive stress (Zorc 1979). In Malagasy stress contrasts arose through the development of extrametrical supporting vowels: PMP *taliŋa > *tadíny* 'ear', *taŋan 'thumb' > *tánana* 'hand'.

Lindrou stress contrasts were transcribed by the writer in a corpus of about 750 words and 30 sentences collected in a little over nine hours of fieldwork. It is likely that the perception of stress in some cases was influenced by contrastive gemination, which was recorded inconsistently. Where etymologies are available the basis for an innovative system of stress contrasts (or gemination) is unclear: POC *onom > *ono-h* 'six', *bʷatu 'head' > *battu-k* 'my head', *kiajo > *kies* 'outrigger boom', *lima > *lime-h* 'five', with penultimate stress, but *tolu > *talóh* 'three', *mata 'eye' > *madá-k* 'my eye', *kanase > *kanás* 'mullet', *talise > *dralís* 'a shore tree: *Terminalia catappa*', *tina 'mother' > *tiné-k* 'my mother', with final stress. It is possible that stress was penultimate before the loss of final vowels, and remained in place after this change, giving rise to a predominant oxytonality, but this cannot explain all forms.

9.1.10.4 Vowel nasalisation

Phonemically nasalised vowels are rare in AN languages. They occur in a number of the languages of New Caledonia, where—in at least the northern languages—they reportedly have evolved from earlier postnasalised stops through a complex sequence of changes of the general form *CVNV > CⁿV > CʰṼ > CṼ (Ozanne-Rivierre and Rivierre 1989). In Miri of northern Sarawak vowel nasality is phonemic, as seen in the minimal pair *hããw* 'you' (PMP *kahu) : *haaw* 'rafter' (PMP *kasaw). However, contrastive nasality appears to be confined to this single form, and the basis for nasalisation is obscure.

In Seimat, spoken in the extensive Ninigo Lagoon west of Manus in the Admiralty Islands, apparent nasal vowel phonemes occur only after *w* and *h* (Blust 1998a, Wozna and Wilson 2005:6). Like most AN languages Seimat has onset-driven nasal harmony: nasality originates in a nasal consonant and spreads rightward to subsequent vowels. The Proto Oceanic nasals *m, *n, *ñ and *ŋ are reflected as Seimat *m, n, n* and *ŋ* respectively, but Proto Oceanic/Proto Admiralty *mʷ is weakened to a glide. When *mʷ lenited the allophonic nasality on a following vowel became contrastive, since some vowels following *w* now were nasalised, while others were not, as seen in Table 9.17:

Table 9.17 Sources of Seimat nasalised vowels after *w*

POC/PADM	Seimat	English
*mʷata	wãt	snake/earthworm
*mʷaqane	wã-wãn	man, male
*mʷalutV (PADM)	wãlut	dove sp.
*dramʷa	kaw(ã)-	forehead
*watiRi (PADM)	wat	monitor lizard
*waka	wa	boat
*qawa	aw(a)-	mouth

Nasalised vowels also arose in Seimat following *h* from POC *r. This condition for vowel nasalisation—which Matisoff (1975) has facetiously called 'rhinoglottophilia'—is somewhat less familiar than nasalisation following a nasal consonant, but is now well-established as a source of vowel nasality in several language families. Allophonic rhinoglottophilia is seen, for example in the Polynesian Outlier language Rennellese, where *a* is strongly nasalised (and *i* less so) in the word *hahine* 'woman'. Seimat rhinoglottophilia is complicated by the fact that *h* has two historical sources, POC *r and *p, and only *h* from *r triggered nasality of the following vowel:

Table 9.18: Sources of Seimat nasalised vowels after *h*

POC/PADM	Seimat	English
*roŋoR	hõŋ	to hear
*rua	hũ-hũa/hũo-hũ	two
*matiruR	mati(hũ)	sleep
*maqurip	moĩh	living, alive
*wara- (PADM)	wah(ã)	root
*panek	han	to climb
*pija	hil	how much/how many?
*poñu	hon	green turtle
*puqaya	hua	crocodile
*qutup	utuhi	draw water

Why did vowels nasalise after *h* from POC *r, but not after *h* from POC *p?. The most likely scenario is that there were two phonetically distinct glottal fricatives that merged only after the development of postglottal nasality. Closer inspection of the Seimat data suggests that at least for vowel nasality after *h* the locus of nasality is the consonant rather than the vowel. This is seen most clearly in the allomorphy of the transitive suffix *-i*, which surfaces as a nasal vowel after *h* from *r, but as an oral vowel after *h* from *p: POC *tuqur 'stand' > *tu* 'stand up', *tu-a* 'stand up! (imper.)', *ha-tuh-ĩ* 'to erect'make something stand', *qutup* 'submerge a vessel to fill it' > *utuh-i* 'fill a vessel with water', *utu-a* 'fill it! (imper.)'. Since this is the same morpheme with both bases the nasality in *ha-tuh-ĩ* must be derived by nasal spreading. Although similar data do not exist for *w*, the pattern presumably would be the same, and the general conclusion must be reached that Seimat has a phonemically nasalised glottal fricative and labiovelar glide (Blust 1998a).

9.1.11   *Other types of normal sound change*

Two other types of 'normal sound change' can be mentioned here, although the consequences of one of these are typologically unusual.

The first of these changes is a type of restructuring. PMP permitted consonant clusters only medially. The most common of these were nasal-homorganic stop sequences, as in *tumbuq 'grow', or *punti 'banana'. In insular Southeast Asia these sequences clearly are clusters, but in many Oceanic languages and in Proto Oceanic they are reflected as voiced stops *b, *d, *g with automatic prenasalisation. In Malay, for example, *tumbuh* 'to grow' contains a consonant cluster that contrasts with clusters such as that in *kampuŋ* 'village'. In Fijian *tubu* 'grow' contains a unit phoneme *b* that is phonetically similar to the Malay cluster ([mb]), but is structurally different, since both prenasalised voiceless stops and plain voiced stops are impossible. In effect, the voicing and prenasalisation of stops have come to imply one another in many Oceanic languages, and this change has introduced a new type of phoneme (predictably prenasalised stops).

A second set of sound changes has introduced a novel type of morphology, namely verbal ablaut. This phenomenon is discussed under §6.5, and will not be repeated here.

9.2   Bizarre sound change

Bizarre sound changes are inferred from reflexes that cannot easily be attributed to an accumulation of normal changes. In a language family the size of AN it should come as no surprise that a number of bizarre changes have occurred. Some of these are described and discussed in Blust (2005e).

As a general classificatory schema we can distinguish changes that are bizarre on any of three grounds: 1) transition, 2) condition, 3) result. Changes that are bizarre on the grounds of transition show an unexpectedly large number of feature values that differ between a reconstructed phoneme and its reflex, for which evidence of transitional stages is unconvincing, as with *t > *k*. Changes that are bizarre on the grounds of condition may have ordinary transitional properties, but occur under unexpected conditions. Changes that are bizarre on the grounds of result may be ordinary in other respects, but produce segments that are typologically remarkable. The following discussion takes note of five sound changes that are bizarre on the grounds of unexpected transitions, five that are bizarre on the grounds of unexpected conditions, and three that are bizarre on the grounds of unexpected results. It does not pretend to be exhaustive.

9.2.1 Bizarre transitions

Noteworthy among sound changes that involve bizarre transitions are: 1) *t > *k,* 2) *l > ŋg, 3) *w/y > -p, 4) *p > y and 5) *w > c-, -nc-.*

9.2.1.1   **t > k*

Perhaps the best-known bizarre sound change in AN languages is *t > *k*. Although this change is commonly cited only from Hawaiian and Samoan, it is known to be reflected in 43 languages that represent at least 20 historically independent innovations (Blust 2004b). It will be represented here by data from just five languages: 1) Hawaiian, Samoan, Luangiua (also called 'Ontong Java'), Bipi and Likum.

When the Hawaiian orthography was established by Boston missionaries in the 1820s there was some vacillation in the choice between *l* and *r,* and *t* and *k,* suggesting that these sounds were in free variation at that time. The change of *t to *k* in fact has not yet reached all dialects, since *t is preserved under certain phonological conditions in the socially isolated dialect of Ni'ihau Island. In Samoan this change is still in progress, and can be said to be sociolinguistically conditioned: *k* variants are commonly found in casual or informal speech, as in normal conversational settings among native speakers, while *t* variants are preferred in formal settings, as in religious sermons, official speeches, and conversations with outsiders (Mayer 2001). Less information is available about the history or sociolinguistics of the *t > *k* shift in Luangiua, a Polynesian Outlier language spoken in the Solomon Islands, although a reasonably good vocabulary of the language is available (Salmond 1975). The data in Table 9.19 show that the change *t > *k* took place in these three Polynesian languages, and that it was unconditioned. Samoan forms before and after a slash represent formal and informal pronunciations respectively:

Table 9.19 Examples of POC *t > *k* in Hawaiian, Samoan and Luangiua

PPN	Hawaiian	Samoan	Luangiua	English
*taʔe	kū-kae	tae/kae	kae	feces
*tolu	kolu	tolu/kolu	kolu	three
*turi	kuli	tuli/kuli	kuli	knee
*katafa	ʔākaha	ʔātafa/ʔākafa	akaha	frigate bird
*ʔate	ake	ate/ake	ake	liver
*ʔatu	aku	atu/aku	aku	a fish, the bonito
*kutu	ʔuku	ʔutu/ʔuku	uku	hair louse

Theoreticians who confronted this change in Hawaiian using classical feature theory during the 1970s found it perplexing because of the number of feature values that must be altered in passing from *t* to *k.* Typologists have found it equally perplexing because of its rarity. What is remarkable about the AN examples is that the same change has happened repeatedly in languages that have been separated from a common ancestor and out of contact with one another for centuries or millennia. This is true with regard to all AN languages that show the change, and it is also true on a smaller scale of those that belong to a well-defined subgroup, as Hawaiian, Samoan and Luangiua, which are only distantly related within the Polynesian subgroup. In Samoan, where *t* and *k* alternate sociolinguistically the *t* is a voiceless unaspirated alveolar stop and the *k* a voiceless velar stop, thus providing no phonetic clue for why such a change would occur.

Although some details remain obscure, as first argued for Hawaiian by Wise and Hervey (1952), and made more generally known by Schütz (1994:215), the best explanation for the recurrence of the *t > *k* change in AN languages appears to be that once *k was lost, leaving a stop system *p, t, ʔ,* or just *p, t,* the *t* was free to vary widely in its phonetic realisation. Where this resulted in a more frequently backed stop it amounted to a sound change. Where there was a range of variation from [t] to [k] it created a situation like that encountered by the Boston missionaries who were confronted early in the twentieth century with the task of deciding on a Hawaiian orthography and could in principle have chosen either *t* or *k* to represent the phoneme (Blust 2004b).

9.2.1.2 *l > ŋg

Unlike the change *t > *k,* which is recurrent in the AN family, the change *l > *ŋg* is restricted to one language. Proto Oceanic contrasted *r and *l, as in *rua 'two' and *luaq 'vomit'. This contrast was retained in Proto Polynesian, but was lost after the split into Proto Tongic and Proto Nuclear Polynesian. Fijian, which is closely related to the Polynesian languages, retains this distinction as *r : l* (*rua* 'two', *lua* 'vomit'), Tongan, which forms the core of Tongic, one of the two primary branches of the Polynesian subgroup, retains the distinction as zero : *l* (*ua* 'two', *lua* 'vomit'), and Nuclear Polynesian languages have merged the two as *l* (Samoan *lua* 'two', *lua-i* 'spit out'), or *r* (Maori *rua* 'two' : *rua-ki* 'vomit'). The Polynesian Outlier language of Rennell Island, south of the main Solomons chain, reflects Proto Nuclear Polynesian *l as *ŋg* (written *g* in the standard orthography), and the Bellona dialect reflects it as *ŋ*, probably a reduction of earlier *ŋg*. This change is well-attested, as seen in Table 9.20:

Table 9.20 Examples of PPN *l and *r > ŋg in Rennellese

PPN	Rennellese	English
*hala	aŋga	path, road
*ali	aŋgi	flatfish, flounder
*ʔaro	ʔaŋgo	front'
*walu	baŋgu	eight
*laa	ŋgaa	sail
*leʔo	ŋgeʔo	voice/clear voice
*lima	ŋgima	hand, arm; five
*rua	ŋgua	two
*luaq	ŋgua	spit out/vomit
*fale	haŋge	house, building
*rano	ŋgano	lake
*roŋo	ŋgoŋo	hear/news, report
*tolu	toŋgu	three

Although both PPN *r and *l underwent this change, it is clear that they merged as *l or *r before it took place. In either case it is surprising that a simple alveolar liquid would change to a prenasalised velar stop. One could speculate that the immediate antecedent was an alveolar rhotic that was backed to a uvular *r*, which then strengthened to a stop that became prenasalised in accordance with the Melanesian areal pattern in which voiced stops are automatically prenasalised. Such phonetic speculation about possible intermediate stages in sound change is to some extent testable, or at least plausible when such stages are preserved in closely related languages. However, this is not the case with Rennellese, and the main function of such speculation is to buttress the position that all sound change is phonetically motivated.

Rennellese has an *l* phoneme in demonstrable loanwords, and it is clear that this Polynesian Outlier language was in sustained contact with a Southeast Solomonic language during some period of its history (Blust 1987a). If Rennellese *ŋg* directly continues a lateral phoneme these Southeast Solomonic loans must have entered the language after the change *l > *ŋg*. On the other hand, if Rennellese *ŋg* directly continues a rhotic phoneme the Southeast Solomonic loanwords could have entered the language before the change *r

> *ŋg*. In either case it seems clear that the change took place before the separation of the Rennell and Bellona dialects, presumably a millennium or more in the past.

9.2.1.3   **w, *y > -p*

As seen earlier, the fortition of *w and *y to obstruents is not uncommon in AN languages. However, such fortitions generally produce voiced stops that are homorganic with the glide, and almost always produce a different stop for each semivowel. By contrast, in Levei and Drehet of western Manus, both *w and *y are strengthened to *p* word-finally, a change that is fundamentally different from the usual process of glide fortition. To increase the quantity of comparative material available, Levei and Drehet forms are compared directly with cognates in other languages of the eastern Admiralties that have preserved *w and *y. When dealing with a group of languages that has not yet received thorough comparative treatment this procedure could be considered circular, but in several cases POC etyma confirm the inferred direction of change (*p^wiyey* 'crocodile' and *p^wiley* 'rat' are from Likum, not Lindrou, and Levei *solay* 'swordfish' is assumed to be a loan from Sori):

Table 9.21 Examples of POC *w and *y > *p* in Levei and Drehet

POC	Levei	Drehet	Lindrou	English
*pakiwak	peʔep	peʔep	beʔew,	shark
*qayawan	ep	ep	ew	Banyan
*pati	ha-hup	ha-hup	ha-huw	Four
	isop	asap	asiw	House
	kamop	kamop	kam^wew	left side
*kanawe	kanap		kanaw	Seagull
	kop	kop	kow	Fence
	nasop	nosop	lasow	bandicoot
	nelip	nelip	ñalew	canarium nut
*boRok	pup	pup	bow	pig
*pitaquR		p^wisip	besew	a tree: *Calophyllum inophyllum*
	usip	isip	osew	Rattan
	kenep	kenep	kaney	mangrove crab
*kayu	kep	kep	key	Tree
*laqia	lip	lip	ley	ginger
*waiwai	owip	owip	ewey	Mango
*paRi	pep		bey	stingray
*puqaya	puep	puip	bua, p^wiyey	crocodile
	p^wilip	p^wilip	p^wiley	rat
*sakulayaR	(solay)	solap	solay	swordfish

The change *w > *p* is sufficiently similar to some types of glide fortition not to merit notice in a discussion of bizarre sound changes, but this cannot be said for *y > *p*. So far as the evidence permits us to infer, *w and *y were altered in a single change: historically secondary final glides became *p*. Most of the languages of western Manus form a dialect network and so, despite rapid and ongoing sound change, are closely related. The change of *w and *y to *p* must, therefore, have happened rather quickly and in the not too distant past. It is assumed to have been a single change in the immediate common ancestor of

Levei and Drehet. All recorded examples of this change affected historically secondary word-final glides, some of which derive ultimately from *i (POC *laqia > *laya > *lia [lija] > liy > *lip*) or *R (POC *paRi > *payi > *pay > *pey > *pep*). No examples of initial *y are available, but initial *w did not undergo the change.

9.2.1.4   *p > y

The surprising change *p > y is found in two widely separated Oceanic languages, Marshallese of eastern Micronesia, and Sakao of northeast Malakula in Vanuatu. PMP *b and *p merged as Proto Oceanic *p, but PMP *mb and *mp merged as POC *b, a voiced bilabial stop that was automatically prenasalised. POC *b became Proto Micronesian (PM) *p, and POC *p became PM *f (Bender, et. al. 2003a,b). Since Marshallese consistently reflects PM *p as *p*, and almost always reflects *f as *y* it seems reasonably clear that the change in this language was from a voiceless labiodental fricative to a voiced palatal glide (because the standard orthography obscures the point at issue, Marshallese forms are cited in phonemic transcription): PM *faka- > MSH *ya(k)-* 'causative' (fossilised), *afara > *hayeray* 'shoulder', *fai- > *yayi-* 'reciprocal prefix' (fossilised), *faa > *ya-* 'four' (in *ya-biqiy* '400'), *fai > *yayi-biqiy* 'stingray', *fanifani > *yanyen* 'to bail', *fanua > *yaney* 'land, island', *faŋi > *yag(i)-* 'north', *fara > *yar* 'core of pandanus', *fara > *yar* 'lungs', *farafa > *yerey* 'outrigger platform', *fasu > *yat(i)-* 'eyebrow'.

In Sakao POC *p reportedly became *y* in stressed syllables, but disappeared elsewhere. Since most languages in Vanuatu reflect *p as *v* or some lenited continuation of *v*, it appears that Sakao also developed a voiced palatal glide directly from a labiodental fricative. Guy (1978) maintains that Proto North New Hebridean *v became a semivowel which then fronted before *a, *e and *i: PNH *vano > *yan* 'to go', *vati > *yed* 'four', *vili > *yil* 'to hit', *sava > *aya* 'what?', *vatu > *e-yed* 'stone'. Sakao historical phonology presents major challenges, and many features of change in this language remain unclear. It is possible, therefore, that this change was less bizarre than appears at first sight, proceeding from *v to *w and then fronting before front vowels and *a.

9.2.1.5   *w > c-, -nc-

In Sundanese, spoken over much of the western third of the island of Java, *w became *c* in initial position, and *-nc-* word medially. Some examples of *b underwent the same change, probably after first leniting to *w*. Examples of this change appear in Table 9.22. Cognates from Malay are included where no PMP reconstruction is known.

Like other changes considered in this section, the direct transition from *w to a voiceless palatal affricate seems phonetically improbable. The search for intermediate steps in this case, however, is not likely to be fruitful. Since *b did not normally become Sundanese *w*, and became Malay *w* only in *-aba- > *-awa-*, Sundanese forms that show the change *b > *c-, -nc-* probably have been borrowed from Javanese, where *b usually became *w*. The antiquity of Javanese loans in Sundanese is unknown, but recorded history suggests that most Javanese influence on Sundanese happened within the past millennium. In any case it is clear that *w > *c-, -nc-* could not have occurred in Sundanese until PMP *b became Javanese *w*, and Javanese forms that reflect this change were borrowed. Moreover, forms such as *katuncar*, which Nothofer (1975:298) derives from Sanskrit *kutumburi/kustumbari* 'coriander seed', and *lincaŋ* suggest that *-mb- became *-nc- without intermediate steps. All of this points to a relatively recent and rapid change in Sundanese, and one for which a phonetic motivation is not at all apparent.

Table 9.22 Examples of PMP *w or *b > Sundanese *c-* or *-nc-*

PMP	Sundanese	Malay	English
(1) *w > c-/-nc-			
*wahiR	cai[110]	air	water, river
	kancah	kawah	vat, cauldron
	karancaŋ	kərawaŋ	openwork; à jour design
*lawaq	lancah		spider
	ranca	rawa	swamp, morass
*sawa	sanca		python
(2) *b > c-/-nc-			
*bahaq	caʔah	bah	floodwaters
*badas	cadas		gravelly ground
	canir	banir	buttress root
*baŋkudu	caŋkudu	bəŋkudu	*Morinda* spp.
	cariŋin	bəriŋin	banyan, fig tree'
	cauŋ	bauŋ	catfish
*bayaR	caya	bayar	give in compensation
*laban	lancan	lawan	opponent
	katuncar	kətumbar	coriander seed
	lɨncaŋ	ləmbaŋ	swollen with water

9.2.2 Bizarre conditions

Among sound changes that involve unexpected conditions are 1) intervocalic devoicing, 2) postnasal devoicing, 3) gemination of the consonantal onset to open final syllables, 4) rounding of final *a, and 5) low vowel fronting after voiced obstruents.

9.2.2.1 *Intervocalic devoicing*

Many of the languages of northern Sarawak have complex historical phonologies that include a variety of bizarre sound changes. The Berawan dialects are among the most noteworthy of these. In initial position *b normally did not change and *R, which probably was a uvular trill in pre-Berawan, became *g* (examples from the dialect of Long Terawan unless noted otherwise): *balu > *billoh* 'widow(er)', *bana > *binnəh* 'husband', *batu > *bittoh* 'stone', *bulu > *bulloh* 'body hair, feathers', *buku > Long Jegan *bukkyəw* 'node, joint'; *Ratus > *gitoh* 'hundred', *Ribu > *gikkuh* 'thousand', *Ramut > *gimauʔ* 'root'. In intervocalic position, however, both *b and *R are reflected as *k*: PMP *qabu > *akkuh* 'ash', *balabaw > Long Jegan *bəlikiw* 'rat', *babuy > Long Jegan *bikuy* 'pig', *bubu > Long Jegan *bukkəw* 'bamboo basket trap for fish', *Ribu > *gikkuh* 'thousand', *tuba > *tukkih* 'fish poison', *qubi > *ukkih* 'yam'; *beRas > *bəkəh* 'husked rice', *beRat > *bəkəiʔ* 'heavy', *qabaRa (> *baRa) > *bikkih* 'shoulder', *hadiRi > *dəkih* 'post', *duRi > *dukkih* 'thorn', *kaRaw > *kikiw* 'scratch an itch', *paRa > Long Jegan *pakkyəy* 'firewood shelf', *tageRaŋ > *takiŋ* 'rib'. In two known cases intervocalic devoicing took place prior to the loss of the initial syllable of a base: *biRuaŋ > *kəbiŋ* 'the Malayan sun bear', *duRian >

---

[110]   Shortened to *ci-* in numerous place names throughout Sundanese-speaking region of western Java, as in the towns of Ciamis, Cianjur, Cikalong, Cilacap, Cirebon, etc.

kəjin 'durian.'   Other examples of intervocalic devoicing followed by the loss of a *prefix* can be inferred from the fact that *b > *b* and *R > *g* word-initially in nouns and dynamic verbs, but both became *k* in stative verbs which originally took the prefix *ma-: *ma-buat > *kəbəiʔ* 'long', *ma-Raya > *kijjih* 'big', *ma-Raqen > *kiʔən* 'light in weight'. These changes provide evidence that at some point in their history the Berawan dialects added a rule of intervocalic devoicing. Since *d became *-r-* devoicing applied only to *b and *g, and since *b is reflected as intervocalic *k* it is likely that it first changed to *-g-* in intervocalic position, and that intervocalic devoicing then applied solely to the velar stop.

Glide fortition in Long Terawan Berawan produced *b* from *w and *j* from *y, as in PMP *tawa > *tabəh* 'laugh' or *ma-Raya > *kijih* 'big'. Since neither of these devoice intervocalically it must be assumed that they arose after intervocalic devoicing applied. By contrast, in neighbouring Kiput the inherited intervocalic stops *b and *d generally did not devoice, but *g did, as well as v (< *w) and j (< *y): PMP *qabu > *abəw* 'ash', *babuy > *babuy* 'pig', *nibuŋ > *nibuŋ* 'nibung palm', *tuba > *tubih* 'derris root fish poison', *hajek (> *adek) > *m-adək* 'sniff, kiss', *ŋajan (> *ŋadan) > *adin* 'name', *pajay (> *paday) > *padəy* 'riceplant', *t-aji (> *tadi) > *tadəy* 'younger sibling', but but *tugal > *tukin* 'dibble stick', *jaway (> *javay) > *dafiəy* 'face', *duha (> *dua > *duva) > *dufih* 'two', *kahiw (> *kayu) > *kacəw* 'wood, tree', *qasawa (> *sava) > *safəh* 'wife', *tian (> *tijan) > *ticin* 'belly, abdomen', *quay (> *uvay) > *ufiəy* 'rattan'. In Kiput, then, intervocalic devoicing applied only to *g* and to fricatives and affricates which derive from glides, while in Berawan it applied only to *g* from *b, *R, and presumably *g.

9.2.2.2 *Postnasal devoicing*

Postnasal devoicing has been noted in three geographically and genetically separated languages: Murik in northwest Borneo, the Bengoh dialect of Land Dayak in southwest Borneo, and Buginese in south Sulawesi. In Murik (Blust 1974c) the rule is partly phonemicised and partly allophonic: Proto Kayan-Murik *kelembit > *kələmpit* 'shield', *umbuŋ > *umpuŋ* 'ridge of the roof', *lindem > *lintəm* 'dark', *mandaŋ > *mantaŋ* 'to fly', *tundek > *tuntuk* 'beak of a bird', *lindiŋ > *lintiŋ* 'wall of a house', *undik > *untik* 'upper course of a river', *tandab > *tantap* 'dive to catch something', *andeŋ > *antəŋ* 'deaf', *pindaŋ > *pintaŋ* 'blossom', *pendan > *pəntan* 'small fruit bat', *nji > *nji* [nʧi] 'one', *menjat > *mənjat* [mənʧat] 'pull', *anjat > *anjat* [anʧat] 'rattan tote bag', *tunjuq > *tunjuq* [tunʧuʔ] 'to point.'   For Bengoh Land Dayak there is little information apart from the statement that "In Bengoh nasal + voiced obstruent clusters are in a sense preserved because voiced obstruents have become devoiced, thereby strengthening the clusters" (Rensch, Rensch, Noeb and Ridu 2006:69, fn. 40). In Buginese and Mandar, where voiced stops are unchanged except that PMP *d and *j often became Buginese *-r-*, postnasal devoicing appears to have invariably led to restructuring (I write schwa where Mills (1975) writes a high central vowel): Proto South Sulawesi *aŋgəp > Buginese *aŋkəʔ* 'price', *anjap > Buginese *ancəʔ* 'offerings to spirits hung on banyan trees', *bemba > *bempa* 'water jar', *lambuk > *lampuʔ* 'pound rice', *limboŋ 'deep water' > Buginese *lempoŋ* 'pond', *rambu > *rampu* 'fringe', *rumbia > *rumpia* 'sago palm', *tambiŋ 'addition to a house' > *tampiŋ* 'outhouse', *barumbun 'a colour pattern' > Buginese *warumpuŋ* 'bluish-white colour (of chickens)', *bumbun > *wumpuŋ* 'heap up'.

Since both intervocalic and postnasal positions are *voicing* environments both of the preceding changes can be considered dissimilations, but this does little to explain why a change of this type would occur. Alternatively, to adopt a position similar to that used in the explanation of *t > *k, postnasal devoicing may have taken place because the voicing

contrast for obstruents was first lost after nasals, but not elsewhere, as a result of voicing assimilation (hence nasal + voiceless stop became nasal + voiced stop). In this environment, then, voice was free to vary. If the voiceless variant of postnasal obstruents prevailed over time postnasal devoicing took place; otherwise voicing assimilation after nasals remained as a 'natural change'.

9.2.2.3   *Gemination of the onset to open final syllables*

As noted already, AN languages have several well-attested sources for geminate consonants, including assimilation in consonant clusters, and compensatory lengthening after schwa. In the Berawan dialects of northern Sarawak, however, geminate consonants developed under very different conditions: if the final syllable began with a consonant and was open at the time of the change the syllable onset was geminated. Final *q and *R were subsequently lost, and final -*h* was subsequently added In Long Terawan and Batu Belah, but not in Long Jegan, thereby obscuring the original conditions: PNS *mata > Long Terawan *mattəh*, Long Jegan *matta* 'eye', *mataq > LJ *mata* 'raw, uncooked', *bulu > LT *bulloh* 'body hair, feathers', *buluq > LT *bulu* 'bamboo sp.', *bana > LT *binnəh* 'husband', *tanaq > LT *tana* 'earth', *aku > BB *akkoh* 'I', *ikuR > BB *iko* 'tail', *Ribu > LT, BB *gikkuh* 'thousand', *babuy > BB *bikuy* 'pig', *anipa > LT *lippəh* 'snake', *tapan > LT *tapan* 'winnow', *tama > LT *tamməh* 'father', *tumid > LT *tumin* 'heel'. There are numerous examples of this change, and no doubt about the conditions under which it occurred, but the phonetic motivation behind these conditions, if any, remains puzzling.

9.2.2.4   *Rounding of final *a*

Many languages have rounded a low vowel if a high back vowel is found in the next syllable. Such a change occurs, for example, in Rejang of southern Sumatra, where *a > *o* if the next syllable is closed, but *u* if it is open (McGinn 2005): *zaRum > *dolom* 'needle', *manuk > *monoʔ* 'chicken, fowl', *Ratus > *otos* 'hundred', *batu > *butəw* 'stone', *sapu > *supəw* 'broom'). Unexpectedly, however, a number of AN languages in island Southeast Asia round *a only word-finally. Languages in which this change occurs include 1) Kadazan of western Sabah, 2) Timugon Murut of western Sabah, 3) Kedayan of Brunei, 4) Bekatan of southern Sarawak, 5) Ida'an of eastern Sabah, 6) Òma Lóngh Kenyah of central Kalimantan, 7) Minangkabau of southwest Sumatra, 8) Lampung of southern Sumatra, 9) Javanese, and 10) Gorontalo of northern Sulawesi. Differences of detail in some of these changes should be noted.

Although Kadazan and Timugon Murut are rather closely related, the rounding of final *a in these languages appears to result from independent historical changes, as each language has closer relatives that have not participated in this innovation. Examples from Kadazan include PMP *duha > *duvo* 'two', *lima > *himo* 'five', *ina > *ino* 'that', *mata > *mato* 'eye', *taliŋa > *tohiŋo* 'ear', but *salaq > *hasaʔ* (met.) 'mistake, error', *ma-etaq > *mataʔ* 'raw, uncooked', *natad > *natad* 'courtyard', etc. without rounding.

In Javanese, which has seven phonemic vowels (*i u e o ɛ ə a*), final *a was rounded to *ɔ*, which triggered regressive assimilation of a preceding *a if a single consonant intervened. The change remains allophonic: *lima > *lima* [limɔ] 'five', *mata > *mata* [mɔtɔ] 'eye', *ñawa 'breath; soul' > *ñawa* [ɲɔwɔ] 'soul, spirit, life', etc.

Other languages, including Bukat of Borneo, Gayō of northern Sumatra, and some dialects of Balinese, Sasak and Malay, have weakened *a to schwa only in word-final position. Since Kadazan, Timugon Murut, Kadayan, Bekatan and Gorontalo all reflect

PMP *e (schwa) as *o*, it is possible that the change *a > -o* in these languages passed through the stages *a > -ə > -o*. However, this is not the case in Òma Lóngh Kenyah, Minangkabau, Lampung, Javanese, or Ida'an where, so far as the evidence allows us to say, apparently occurred without an intermediate step.

9.2.2.5 Low vowel fronting

A particularly striking sound change that is characterised as bizarre on account of its conditions is low vowel fronting after voiced obstruents. This innovation is found in at least nine languages of northern Sarawak, and a number of languages of northeast Luzon. Although many of the languages in Sarawak which share it belong to a well-defined subgroup and neighbour one another, differences of form suggest that the fronting of low vowels after a voiced obstruent is a product of independent historical changes. In Miri of northern Sarawak a low vowel was fronted in the final syllable if and only if there was a voiced obstruent earlier in the word. As a result, fronting occurred immediately after a voiced obstruent, or at a distance from it: Proto North Sarawak *adan > *aden* 'name', *paday > *fadih* 'riceplant', *tugal > *tugel* 'dibble stick', *ujan> *ujen* 'rain'; *busak > *buek* 'flower', *dua > *dəbeh* 'two', *dilaq > *jaleʔ* 'tongue'. If a word beginning with a voiced obstruent contained two low vowels, fronting skipped the first vowel and targeted the second: *baRa > *bare* 'ember, hot ash', *daqan > *daʔen* 'branch', *jalan > *jalen* 'path, road'. In Miri fronting followed PNS *ə > a, and so affected forms with either vowel under the stated condition (Miri e = [ɛ] ~ [e]) : *agəm > *agem* 'hold', *daləm > *dalem* 'in, inside', *bəbʰər 'fan' > *fer* 'blow on.' Fronting also followed the introduction of many Malay loanwords, as these generally undergo the change: *badek* 'rhinoceros' (Malay *badak*), *barjeh* 'work' (Malay *bəkərja*, ultimately from Sanskrit), *fajer* 'dawn' (Malay *fajar*, ultimately from Arabic), *karbew* 'water buffalo, carabao' (Malay *kərbau*, ultimately from a Mon-Khmer source), *tabageh* 'copper' (Malay *təmbaga*, ultimately from a Middle Indic source). Finally, low vowel fronting was blocked by an intervening voiceless stop or nasal : *bakaw > *bakaaw* 'mangrove', *butan > *butaan* 'coconut', *dəpa > *dəpa* 'fathom', *gatəl > *gatal* 'itch'; *bana > *banah* 'husband', *baŋaR > *baŋaar* 'stench of stagnant river mud', *danaw > *danaaw* 'lake', *jaməq > *jamaʔ* 'dirty', *bənaŋ* 'thread' (Malay *bənaŋ*). Narum, a closely related language spoken some twenty five miles up the Baram River, shows a similar change, but without any blocking consonants, as seen in the forms *bana > *baneəh* 'husband', *bataŋ > *batiəŋ* 'tree trunk, log', *buŋa > *buŋeəh* 'betel nut', *dəpa > *dəpeəh* 'fathom', *dətaq > *dətiəʔ* 'boiling', *dipaR > *dipir* 'opposite bank', *gatəl > *gətel* 'itch', *jameq > *jameəʔ* 'dirty', and the loanword *səgupin* 'tobacco pipe' (Brunei Malay *səgupan*). These and other differences suggest that, despite their geographical proximity and nearness of relationship, Miri and Narum acquired low vowel fronting after their separation from an immediate common ancestor.

The Berawan dialects, in the middle course of the Baram river, differ from these cases in targeting not the last low vowel in a word, but the first low vowel that follows a voiced obstruent. In these languages low vowel fronting may occur at a distance, but never skips a segment to which it could apply, hence PNS *baRa > Long Terawan *bikkih* 'shoulder', *bulan > *bulin* 'moon', *jalan > *ilan* 'path, road', *dua > *ləbih* 'two'. Given this difference in the targeted vowel, Berawan dialects sometimes front vowels in morphemes that are passed over by Miri and Narum because they are not found in the last syllable, or that are passed over by Miri (because of the operation of a blocking consonant), but not by Narum: *batu > Long Terawan *bittoh*, Miri *batauh*, Narum *bateəw* 'stone', *bana > Long Terawan *binnəh*, Miri *banah*, Narum *baneəh* 'husband'.

Sa'ban, a divergent Kelabit dialect that is geographically and genetically separated from other languages of northern Sarawak that share this sound change, also targets the last low vowel in a word whether the conditioning segment immediately precedes the fronted vowel or not. However, in Sa'ban 'erosion from the left' has subsequently deleted or altered many of the voiced stops that triggered fronting: PNS *adaq > *adəiʔ* 'shadow, ghost', *dalan > *alin* 'path, road', *daRaq > *areaʔ* 'blood', *labaw > *labiəw* 'rat', *daqan > *laʔin* 'branch', *buaq > *wəiʔ* 'fruit'.

A strikingly similar pattern of low vowel fronting "runs throughout Black Filipino languages from Dupaningan Agta in the far north of Luzon … to as far south as Manide and Inagta Alabat, and including Umirey Dumaget, Northern Alta, and Southern Alta in the middle" (Lobel 2013:253), and other, apparently related variations on consonant-conditioned vowel change (low vowel backing, back vowel fronting) are also found in Manide and Inagta Alabat (Lobel 2010). In addition to the heavy attestation of low vowel fronting in the languages spoken by the Negrito populations of northeast Luzon, the same type of historical change also appears sporadically in southern Gaddang and Isinay, which are spoken by non-Negrito groups. In Casiguran Dumagat low vowels fronted to *i* if they were in the first syllable, and to *e* if they were in the last syllable immediately following a voiced obstruent: PMP *balay > *bile* 'house', *danum > *dinom* 'water', *Ramut > *gimot* 'root', *quzan > *uden* 'rain', *kaRat > *kaget* 'bite', *daRaq > *dige* 'blood'. The targeted vowel is therefore the same as in the Berawan dialects, but the result differs in that the degree of raising is determined by syllable position.

Low vowel fronting after a voiced obstruent is perplexing for two quite different reasons. First, it is not at all clear what phonetic motivation might lie behind it. In many Mon-Khmer languages and in some of the Chamic languages that have been in contact with them voiced stops produce a prosodic feature of breathy voice which carries rightward unless interrupted by an intervening voiceless stop. The principal vocalic effect of such breathy prosody is to raise vowels. We might, therefore, consider low vowel fronting in northern Sarawak and northeastern Luzon to be examples of the same phenomenon. However, the facts in Borneo and Luzon point to fronting as the primary effect of a voiced obstruent on a following low vowel, with raising an incidental by-product. The schwa, for example, undergoes this change only if it has first shifted to *a*, but could in principle be raised to a high central vowel if the phenomenon were fundamentally the same as that in Mon-Khmer languages. Moreover, phonation types should affect all relevant segments in their path, but as noted, low vowel fronting skips some relevant vowels to target more distant ones in languages such as Miri, Narum or Sa'ban. Second, this change shows a clear geographical bias, occurring in two fairly restricted regions. This distributional fact suggests that contact is implicated in the shared development within each of these two regions, yet differences of detail in the individual languages argue against such an explanation. The notion of 'stimulus diffusion', whereby the *idea* of an alien cultural product is borrowed and reworked, has a respectable history in cultural anthropology, but has hardly affected linguistics, since cultural products such as sound changes are difficult to borrow without borrowing the vocabulary in which they are embedded. Some such notion may need to be invoked in order to account for the pervasiveness of this change in North Sarawak languages that appear to have innovated it independently. On the other hand, the idea that contact played a role of any kind in creating the distributional pattern for low vowel fronting is difficult to maintain for Sa'ban, which is geographically separated from the other North Sarawak languages that show this change,

and is virtually impossible to maintain in accounting for the appearance of a strikingly similar bizarre change in northern Sarawak and northeastern Luzon.

9.2.3 Bizarre results

Sound changes that produced bizarre results may or may not be ordinary in themselves. What is noteworthy about them is their interaction with typology: before the change a structural feature was typologically nondescript; after the change it was transformed into a feature that attracts special notice. Sound changes that are noteworthy primarily for the unusual results they produced include: 1) bilabial stops to bilabial trills in various languages of Manus and Vanuatu, 2) bilabial stops, fricatives and nasals to linguo-labials in various languages of Vanuatu, 3) plain voiced obstruents to voiced aspirates in Kelabit-Lun Dayeh and some other languages of northwest and northern Borneo. In addition we might add the development of preploded final nasals in western Indonesia and mainland Southeast Asia, or of glottalised consonants in Yapese of western Micronesia, and Waimoa of East Timor, although both of these features are found elsewhere in the world.

9.2.3.1   *Bilabial trills*

Bilabial trills are so rare that they were once used as the type case of sounds that any child can produce, but which are absent from natural languages. However, these rare speech sounds have been reported from three distinct areas in the AN language family, where they generally reflect *mb/__u. About one third of the roughly 30 languages of Manus in the Admiralty Islands have at least one prenasalised trill, *br* (bilabial) or *dr* (alveolar). Some languages have both trills, and at least one of these (Leipon) permits both to occur in the same morpheme, as in POC *na pudi (> mpudi > budi) > *brudr* 'banana'. Similar segments have been reported for several languages of Vanuatu (Crowley 2006a), and in Nias, spoken in the Barrier Islands west of Sumatra (Catford 1988). The discussion here will be limited to the languages of Manus.

At least seven languages on Manus have a bilabial trill *br* (Blust 2007a). These include 1) Nali, 2) Titan, 3) Papitalai, 4) Kuruti, 5) Leipon, 6) Lele, and 7) Kele. Several languages have an alveolar trill, but no bilabial trill. These include Ere, in eastern Manus, which has [mb], Lindrou in western Manus, which has [b], and Bipi, a migrant from eastern to western Manus, in which [p] corresponds to the bilabial trill of other languages. Apart from their rarity the trills of Manus are noteworthy for two reasons: 1) they are always prenasalised, and 2) the bilabial trill almost always occurs before *u*. Maddieson (1989a) has described the phonetic properties of these segments, and has suggested the mechanism through which they arose from earlier sequences of [mbu]:

1.   In pronouncing *mbu-* the nasal port is initially open, with the vocal cords vibrating. Intraoral pressure is low during the nasal portion of this sequence, and increases very little during the stop portion, which is always very brief in prenasalised voiced stops.

2.   At the time of labial release the pressure acting to separate the lips is weak, contributing to a slow initial rate of labial opening.

3.   Because the consonant is released into a high rounded vowel the labial aperture remains constricted for a time after release.

4.   The lips are also relatively slack at the time of release due to rounding and protrusion for the vowel.

5.   Bernoulli forces created by accelerated airflow through the narrow lip aperture then may result in involuntary and possibly repeated reclosure of the lips during the stop-vowel transition, analogous to the mechanism of voicing by vibration of the vocal cords.

In the languages of Manus an important precondition for the rise of both *br* and *dr* was the presence of the common noun article *na, which tended to lose its unstressed vowel and then merge with a following noun as prenasalisation. As a result of the 'capture' of this article most nouns have what Ross (1988) called a 'secondary nasal grade' reflex, and where this preceded *pu- the result is generally the development of a bilabial trill. Examples from Nali of eastern Manus include: POC *na puaq kayu > *brua key* 'fruit', *na puqaya > *bruay* 'crocodile', *na puki > *brui-* 'vagina', *na punti > *brun* 'banana', and *na putun > *brut* 'a shore tree: *Barringtonia asiatica*'. Bilabial trills developed only before rounded vowels, mostly *u. The present pattern of geographical distribution is a product of selective retention, with the alveolar trill tending to be the more stable of the two. Both trills have disappeared completely in the southeastern Admiralties (Baluan, Lou, Rambutyo, Nauna), and in the western islands (Wuvulu-Aua, Ninigo lagoon).

9.2.3.2   *Linguo-labials*

A number of the languages of central Vanuatu have developed a series of consonants that Maddieson (1989b) calls 'linguo-labials'. These are articulated with the tongue tip against the upper lip, much like apico-dentals, but with a more forward point of articulation. In their maximum extension they include stops, fricatives, and nasals. Languages that have such a series, in whole or in part, are: 1) Mafea, 2) Aore, 3) Tangoa, 4) Araki, 5) Vao, 6) Mpotovoro, 7) Leviamp, and 8) Unmet. Tryon (1976) writes these segments as /p̃, ṽ, m̃/. Other languages clearly had linguo-labials earlier, but unmarked /p̃/ to /t/, /ṽ/ to /ð/ or /θ/, and /m̃/ to /n/. These include 9) Tolomako, 10) Roria, 11) Tur, 12) Tambotalo, 13) Sakao, 14) Lorediakarkar, 15) Shark Bay I, 16) Shark Bay II, 17) Orap, and 18) Mae in Tryon (1976). In effect, linguo-labials developed by substituting the tongue for the lower lip as an active articulator. Before rounded vowels (*u, *o) labials were unchanged (Table 9.23, part 1), but before unrounded vowels (*i, *e, *a) they fronted to linguo-labials, which in some languages, as Tolomako, subsequently unmarked to apico-dentals (Table 9.23, part 2).

Since nothing about the collection of vowels *i, *e, *a appears to favor the substitution of the tongue for the lower lip as an active articulator, and since a narrowed labial aperture would disfavor protrusion of the tongue, lip rounding evidently blocked the development of linguo-labials. This is true not only of simple labial consonants before rounded vowels, but also of labiovelar consonants that in some languages have subsequently unrounded: *bʷatu > Tangoa *patu-* 'head', *bʷoe > Tangoa *poi*, Tolomako *poe* 'pig', *Rumaq (> *Rimʷa) > *ima* 'house', *mʷata > Tangoa, Tolomako *mata* 'snake' (cp. 'eye' in Table 9.23). In some languages the fronting of labials to linguo-labials is sporadic, as seen in non-altered forms such as POC *pisa > Tangoa /moβisa/ (expected **mo p̃isa), Tolomako /βisa/ (expected **tisa) 'how much/how many?'. The distribution of this phenomenon and its apparent irregularity in some languages suggests that it began in one relatively restricted geographical area, either in eastern Santo or northern Malakula, and diffused from there into outlying languages.

Table 9.23 The development and unmarking of linguo-labials in languages of Vanuatu

POC	Tangoa	Tolomako	Gloss
(1)			
puaq	βua	—	fruit
pulu	βulu	βulu	hair
saŋapuluq	saŋaβulu	—	ten
putos	puto	pito	navel
pose	e-βose	—	canoe paddle
	mohi	moɣi	mosquito
(2)			
piri	p̈irip̈iri	—	squeeze
bebe	p̈ep̈e	tete	butterfly
pakiwak	p̈aheu	—	shark
pano	v̈ano	βano	go
kamiu	kam̈im	kaniu	2pl
meme	m̈em̈e	nene	tongue
mata	m̈ata	nata	eye
kamali	ham̈ali	ɣanali	men's house
lima	lim̈a	lina	hand

9.2.3.3 *Voiced aspirates*

As noted in previous chapters, several dialects of Kelabit in northern Sarawak have a typologically rare set of phonemic voiced aspirates b^h, d^h, g^h that contrast with *b*, *d*, *g*, and are distinct from the more familiar 'murmured stops' of languages such as Hindi. The Kelabit voiced aspirates begin voiced and end voiceless, with the voiceless coda continuing into the onset of a following vowel for some speakers, hence producing 'aspiration'. Although these segments are about twice the length of *b*, *d*, *g* they are not clusters (Blust 1974a, 2006a). The voiced aspirates b^h : d^h : g^h in Bario Kelabit derive from Proto North Sarawak phonemes *b^h, *d^h, *j^h and *g^h, with the alveolar and palatal aspirates merging. The corresponding segments in some other Kelabit dialects and some other languages in northern Sarawak are shown in Table 9.24.

Even where they appear typologically ordinary, as with Long Wat, Long Merigam *b*, *d*, *j*, *g*, at least *b* and *d* usually show greater constriction that reflexes of PNS *b and *d. In Blust (1969, 1974b) the voiced aspirates were thought to reflect PNS clusters *bS, *dS, *jS, *gS which arose by syncope, an explanation called the 'Proto North Sarawak vowel deletion hypothesis'. Although other observations appeared to support this reconstruction, the principal basis for it was the observation that Bario Kelabit b^h corresponds to Kiput *s*. This interpretation required that many pre-PNS reconstructions be revised through the interpolation of an extra syllable with *S, and assumed that these were the only languages outside Taiwan in which PAN *S was retained as a sibilant. Thus Bario Kelabit *təb^huh*, Kiput *təsəw*, Bintulu *təɓəw* 'sugarcane', which had previously been attributed to PAN *tebuS were attributed to *tebuSu, Bario Kelabit *pəd^huh*, Kiput *pəsəuʔ*, Bintulu *lə-pədəw* 'gall, gall bladder', which had previously been attributed to *qapeju, were attributed to *qapejuSu, and so on. In some cases it was not necessary to add a new syllable, as with PAN *bukeS (then written *buSek) > Bario Kelabit *əb^huk*, Kiput *suaʔ*, Bintulu *ɓuk* 'head hair', where there appeared to be direct evidence that a vowel had been lost between a voiced obstruent and a following *S.

Table 9.24 Reflexes of the Proto North Sarawak voiced aspirates

Proto North Sarawak	b^h	d^h	j^h	g^h
Kelabit, Bario	b^h	d^h	d^h	g^h
Kelabit, Pa' Mada	p	t	t	k
Kelabit, Pa' Dalih	p	s	s	k
Kelabit, Long Napir	f	s	s	k
Kenyah, Long Anap	p	t	c	k
Kenyah, Long Dunin	ɓ	ɗ	s	ɠ
Kenyah, Long Wat	b	d	j	g
Penan, Long Merigam	b	d	j	g
Kenyah, Long San	ɓ	ɗ	d^y	ɠ
Berawan (all dialects)	p	c	c	k
Kiput	s	s	s	k
Narum	f	t	c	k
Miri	f	s	s	k
Bintulu	ɓ	ɗ	j	g

Evidence from Formosan languages, which generally preserve *S as a sibilant, and could be expected to retain the hypothesised extra syllable, sometimes contradicted the vowel deletion hypothesis, as with Pazeh *apuzu* (not **apuzuu), or Paiwan *qapǝdu* (not **qapǝdusu) 'gall'.

Moreover, over time it became apparent that there was a skewing of forms that contain a reflex of PNS voiced aspirates: almost all reflexes of PMP reduplicated monosyllables, and almost all reflexes of voiced obstruents after PMP *e (schwa) became PNS voiced aspirates. This skewing provided the critical clue to a reinterpretation. The reflex of PAN *e in many of the languages of insular Southeast Asia is a mid-central vowel that is extra short. Because its duration is less than that of other short vowels consonants it either deflects stress one syllable to the right, or holds stress by geminating a following consonant. Voiceless stops are still allophonically geminated after schwa in Bario Kelabit, but plain voiced stops almost never occur in this environment. Geminate consonants also often arise from inherited or derived consonant clusters. It now appears likely that the PNS voiced aspirates began as voiced geminates that arose primarily in two environments: 1) through allophonic lengthening after *e (schwa), and 2) from the complete assimilation of the first of two abutting consonants in a reduplicated monosyllable. Examples of the first type occurred with a directly inherited schwa, as with PAN *tebuS, PMP *tebuh (> *tǝbbu) > PNS *tǝb^huh 'sugarcane', or with a schwa that was acquired to reconstitute a lost disyllabism after a content morpheme had reduced to a monosyllable through regular sound change, as with PAN *bukeS > PMP *buhek (> *buk > *ǝbuk > ǝbbuk) > PNS *ǝb^huk 'head hair'. Examples of the second type occurred in forms such as *butbut (> *bubbut) > Bario Kelabit *bub^hut* 'pluck, pull out'. Some etymologies allow either interpretation, as PAN, PMP *qalejaw (> *ǝlǝdaw > *ǝldaw > *ǝddaw) > PNS *ǝd^haw 'day', where gemination has two possible sources.

It is known that consonant gemination favors voiceless segments, since aerodynamically it is difficult to maintain voicing over longer closure durations. For this reason universal phonetic factors would have disfavored voicing throughout earlier *-bb-, *-dd- and *-gg-. It is possible that the historical stage immediately preceding Proto North Sarawak had both voiced and voiceless geminates. If so, the voiceless geminates remained allophonic, or were reduced to simple voiceless stops, but the voiced geminates were partially devoiced,

giving rise to true voiced aspirates. These in turn evolved into the phonetically diverse and rather surprising set of sound correspondences seen in Table 9.24. In Lower Baram languages such as Miri, Dali', Lelak, Kiput or Narum PNS *b^h evidently became *f. Then, as first proposed by Dahl (1981a:60), pre-Kiput *f became *s*, an unusual sound change, but one that must be assumed, given the absence of any other evidence that PAN *S survived as a sibilant outside Taiwan.

9.3   Quantitative aspects of sound change

This section briefly addresses two issues affecting sound change in AN languages that can be stated in quantitative terms. The first of these arises from the general statement in Lehmann (1992:191) that 'it may be true that no change affects all occurrences of a phoneme. Completely unconditioned changes may, then, be impossible to document in languages.' The second arises from the impression that vocalic change is more common in Indo-European than in AN languages, while the reverse is true of consonants.

Lehmann illustrates his statement with Proto Indo-European *o > Proto Germanic *a, and notes that this change 'is called unconditioned, even though the *o* of unstressed syllables was lost.' This could mean 1. PIE *o > PGmc *a, and then *a was lost, or 2) PIE *o was lost, and then all remaining examples of this phoneme became PGmc *a. Under either interpretation the change was unconditioned. To say otherwise is to confuse reflexes with changes. Many changes in AN languages are unconditioned, but to see this it is necessary to distinguish change from reflex. To cite only a few of many possible examples, Proto Polynesian *ʔ disappeared unconditionally in Hawaiian: *ʔate > *ake* 'liver', *ʔone > *one* 'sand', *ʔila > *ila* 'birthmark', *ʔuha > *ua* 'rain', PPN *matuʔa > *makua* 'parent', *taʔu > *kau* 'year, season'. Following this change, all examples of *k became ʔ, as in *kutu > *ʔuku* 'louse' or *ika > *iʔa* 'fish', and then all examples of *t became k, as in *tolu > *kolu* 'three', or *mata > *maka* 'eye'. In addition, *n and *ŋ merged unconditionally, as in PPN *laŋi > *lani* 'sky', next to *ono > *ono* 'six'. To argue that *t > *k* is a conditioned change because PPN *laŋi derives from PMP *laŋit, is pointless, since final consonants had been lost in Proto Polynesian long before the changes that are peculiar to Hawaiian.

This discussion shows that unconditioned sound changes occur in AN languages, but says nothing about their frequency. To gain some idea of relative frequency Table 9.25 summarises the major sound changes in six languages, Thao, Tagalog, Chamorro, Tetun, Motu and Hawaiian (UC = unconditioned change, CC = conditioned change, US = unconditioned phonemic split):

Table 9.25 Relative frequency of conditioned and unconditioned
sound changes in six languages

Language	UC	CC	US
Thao	10	2	1
Tagalog	5	3	4
Chamorro	10	4	1
Tetun	12	6	1
Motu	7	4	2
Hawaiian	9	2	2
Total	53	21	11

In Thao at least ten changes are best treated as unconditioned, and only two as conditioned. The unconditioned changes are: 1. *C > c (from probable voiceless alveolar affricate to voiceless interdental fricative), 2. *b > f, 3. *d/z > s (treated as one change, though possibly two), 4. *s > t, 5. *j > z (from probable palatalised voiced velar stop to voiced interdental fricative), 6. *N/ñ > z, 7. *ŋ > n, 8. *l > r, 9. *R > lh (from probable voiced alveolar trill to voiceless alveolar lateral), 10. *h > Ø. *S shows an unconditioned development > *sh* or zero: *Suni > *shma-shuni* 'chirp, twitter (birds)', *luSeq > *rushaq* 'tears', *tapeS > *tapish* 'winnow', but *kaSiw > *kawi* 'wood, stick; tree', *kuSkuS > *kuku* 'fingernail, claw', *SuReNa > *ulhza* 'snow, ice'. All changes that regularly produced *c*, *s*, *sh*, or *z* were later subject to 'sibilant assimilation', as in *CaqiS > *shaqish* (expected **caqish) 'to sew', *daRa > *lhalha* (expected **salha) 'the Formosan maple', or *Sidi > *sisi* (expected **shisi) 'wild goat, serow' (Blust 1995b). Each of these splits (*C > *c/sh*, *d > *s/lh*, *S > *sh/s*) could be treated as conditioned, but this would fail to capture the underlying unity of sibilant assimilation. It is thus simpler and more realistic to posit unconditioned changes *C > *c*, *d/z > *s*, and *S > *sh* and a single later process of sibilant assimilation that created new developments from the uniform outcome of these changes. The clearly conditioned changes in Thao are 1. *e > Ø in the penult, and to *u* before syllable-final labials, but *i* before syllable-final coronals or velars, 2. sibilant assimilation.

In Tagalog at least five changes are best treated as unconditioned, three as conditioned, and four as showing an unconditioned phonemic split. The unconditioned changes are: 1. *c > *s*, 2. *q > *ʔ*, 3. *z > *d*, 4. *ñ > *n*, and 5. *R > *g*. Dempwolff (1934) recognised both *g* and *y* as reflexes of *R in Tagalog, but it is now clear that *R > *y* is due to early borrowing from Kapampangan. The uncontroversially conditioned changes in Tagalog are: 1. *d > *d-*, *-r/l-*, *-d*, as in *deŋeR > *diŋíg* 'to hear', *bukid > *búkid* 'hill; forested mountain region', but *dadaŋ > *daráŋ* 'expose to a fire', or *qudaŋ > *uláŋ* 'lobster', 2. *e > *u/o* if *u is found in an adjacent syllable, or *i* elsewhere, as in *buhek > *buhók* 'head hair', or *tebuh > *tubó* 'sugarcane', but *tenek > *tiník* 'thorn', 3. *u to *o* in a closed syllable (allophonic until the introduction of Spanish loanwords). Four changes show apparently unconditioned phonemic splits: 1. *-d- to *-r-* or *-l-*, 2. *l to *l* or zero, 3. *g to *g/k* or *k to *k/g* (considered a single type of change), 4. *S > *h* or zero.

In Chamorro at least ten changes are best treated as unconditioned, and only four as conditioned. The unconditioned changes are: 1. *z > *ch* (voiceled palatal affricate to voiceless palatal affricate), 2. *p > *f*, 3. *b > *p*, 4. *j > *ʔ*, 5. *q > *ʔ*, 6. *S > Ø, 7. *R > *g*, 8. *d > *h*, 9. *k > *h*, and 10. *e > *u*. Changes 1-6 are noncontroversially unconditioned. Changes 7-10 are superficially conditioned, since *R is reflected as *g-*, *-g-*, *-k*, *d and *k are reflected as *h-*, *-h-*, *-Ø*, and *e became *u* in open syllables, but *o* in closed syllables. To treat these divergent reflexes as conditioned, however, would confuse the distinction between reflex and change; reflexes of *R are far more likely to be products of an unconditioned change *R > *g* followed by final devoicing than products of a conditioned change *R > *g-*, *-g-*, *-k*. Similarly, reflexes of *d and *k are more likely to be products of unconditioned changes *d > *h* and *k > *h* followed by loss of *-h*, than of single conditioned changes that yielded zero word-finally and *h* elsewhere, and *e > *u/o* is clearly the result of *e > *u*, followed by lowering of high vowels in closed syllables, since it also affected reflexes of *i and *u.

The four changes that are uncontroversially conditioned are 1. *w > g^w and *y > dz/__V (treated as a single change), 2. vowel syncope, 3. vowel lowering in closed syllables, and 4. *l > *t* when not preceding a vowel. The first of these is seen in *walu > $g^w alu$ 'eight', *qasawa > $asag^w a$ 'spouse', *paya > *fadza* 'sardine', or *niuR > *nidzok* 'coconut', as

against *buRaw > *pugaw* 'chase away', or *qazay > *achay* 'chin'. The second is seen in *tuqelaŋ > *toʔlaŋ* 'bone', *peRes-i > *foks-e* 'squeeze out, express', or *aRemaŋ > *h-akmaŋ* 'moray eel'. The third change is seen in *toʔlaŋ*, *foks-e*, *bukbuk > *poppo* 'powder from decay', *quRut > *ugot* 'to massage', and many other forms. The last uncontroversially conditioned change is seen in e.g. *laki > *lahi* 'man, male', *kali > *hali* 'dig up tubers', but *qipil > *ifet* 'a tree: *Intsia bijuga*', or *qalejaw > *atdaw* 'day; sun'.

In Tetun at least twelve changes appear to be unconditioned, and only six conditioned. The unconditioned changes are: 1.*p > *h*, 2. *c > *s*, 3. *q > Ø, 4. *z > *d*, 5. *d/j > *r*, 6. *g > *k*, 7. *ñ > *n*, 8. *ŋ > *n*, 9. *R > Ø, 10. *S > Ø, 11. *y > Ø, 12. *w > Ø. Conditioned changes are: 1. *-ay/aw > -*e*/*o* (one change), as in PMP *qatay > *ate* 'liver', or *lakaw > *laʔo* 'to walk', 2. *-uy/iw > -*i*, as in PMP *hapuy > *ahi* 'fire', or *kahiw > *ai* 'tree', 3. *-m > -*n*, as in PMP *enem > *neen* 'six' or *zaRum > *daun* 'needle', 4. *b- > *f*, but *-b- > *h*, as in PMP *batu > *fatu* 'stone', *beRas > *fos* 'husked rice', *babuy > *fahi* 'pig', *bukbuk > *fuhuk* 'wood weevil', *labaw > *laho* 'rat', or *tuba > *tuha* 'derris root fish poison', 5. *-mp- > *b* and *-nt- > *d* (counted as one change), as in PMP *kempuŋ > *kabun* 'stomach', *punti > *hudi* 'banana', 6. *e > *o* in the penult and *a* in the final, as in PMP *depa > *roha* 'fathom', or *deŋeR > *rona* 'hear, listen'. One change shows an apparent unconditioned phonemic split, namely *k > *k*, *ʔ*, or zero, as in PMP *kima 'giant clam' > *kima* 'sea shell', *hikan > *ikan* 'fish', but *takut > *taʔu*-k 'afraid', *kutu > *utu* 'louse'.

Of changes that can be described with confidence in Motu at least seven are unconditioned, while only three are clearly sensitive to phonological context. Unconditioned changes from Proto Oceanic are: 1. *q > Ø, 2. *y > *l*, 3. *ñ > *n*, 4. *ŋ > Ø, 5. *s > *d*, 6. *l > Ø, 7. *w > *v*. None of these are controversial. Conditioned changes are: 1. –C > Ø, 2. *t > *s* before *i or *e, as in *tina > *sina* 'mother', or *qate > *ase* 'liver', 3. Ø > y/#__a (followed by *y > *l*, and *q > Ø), as in POC *apaRat > *lahara* 'northwest wind and season', or *aŋin > *lai* 'wind' (but not e.g. *qapuR > *ahu* 'lime, quicklime'), 4. *-lu > *i*, as in *tolu > *toi* 'three', or *qatoluR (> *qatolu) > *atoi* 'egg'. There are two apparently unconditioned phonemic splits: 1. *p > *p* or *h*, and 2. *k > *k*, ɣ or zero.

Hawaiian shows at least nine unconditioned changes, two conditioned changes, and two apparently unconditioned splits. The unconditioned changes from Proto Oceanic are: 1. *t > *k*, 2. *k > *ʔ*, 3. *q > Ø, 4. *b > *p*, 5. *dr/r > *l*, 6. *ñ > *n*, 7. *ŋ > *n*, 8. *R > Ø, and 9. *y > Ø. None of these are controversial. The conditioned changes are loss of final consonants (already complete in Proto Central Pacific), and *pap > *wah* (Eastern Polynesian labial dissimilation), as in POC *papine > Hawaiian *wahine* 'woman'. There are two apparently unconditioned phonemic splits: *u > *u*, but occasionally *i*, and *s > *s*/Ø.

A sample of six languages to represent the AN family clearly is inadequate, but it is worth noting that all six languages show the same pattern, namely that unconditioned changes are the norm, and conditioned changes the exception. It is thus likely that a much larger data sample would reveal similar results. The claim that truly unconditioned sound changes may not exist is thus not only false, but denies the most common type of sound change known in AN languages. In addition, it appears that conditioned changes are more likely to affect vowels than consonants, an observation that assumes greater interest in light of the second issue taken up in this section.

The second issue arises from the impression that vowels tend to be much more stable than consonants in AN, as compared with e.g. Germanic languages. This can be seen at a glance with AN subgroups such as Polynesian, where no regular changes have taken place in the vowel system of any daughter language, although a number of consonant changes have occurred. The changes considered in Table 9.25 can also be used to evaluate the

relative frequency of vocalic change vs. consonantal change. These are summarised in Table 9.26 (UC = unconditioned change, CC = conditioned change, US = unconditioned phonemic split):

Table 9.26 Relative frequency of consonant change vs. vowel change in six languages

	UC	CC	US	Total
Thao:				
Consonant	10	1	1	12
Vowel		1		1
Tagalog:				
Consonant	5	1	4	10
Vowel		2		2
Chamorro:				
Consonant	9	2		11
Vowel	1	2	1	4
Tetun:				
Consonant	12	3	1	16
Vowel		3		3
Motu:				
Consonant	7	4	2	13
Vowel				0
Hawaiian:				
Consonant	9	2	1	12
Vowel		1		1

Again, the sample size is small, and some distortion is introduced by the use of different proto languages (PAN for Thao, PMP for Tagalog, Chamorro and Tetun, and POC for Motu and Hawaiian). Despite these problems the pattern is clear: consonant change in these languages (and presumably AN languages in general) is far more common than vowel change (52 to 1 for unconditioned changes, 13 to 8 for conditioned changes, 9 to 2 for unconditioned splits). While it can be argued that this pattern reflects the fact that consonants outnumber vowels in the phoneme inventory, this does not appear to provide a complete explanation of the difference, since the same can be said for, e.g. the Germanic languages, where vowel change appears to be much more extensive relative to consonant change than is true of AN languages.

9.4   The Regularity Hypothesis

The mention of unconditioned phonemic splits in the preceding section raises questions about the Regularity Hypothesis. The regularity of sound change is an issue that has engendered much controversy, and has sometimes led to entrenched opposing positions. While there is no question that most sound changes in AN languages are regular, some are merely recurrent, and the causes of this lack of complete regularity remain a challenge to the theory of sound change. Only a small sample of data can be reviewed here, but this will show that there are some apparently unconditioned phonemic splits that are peculiar to

individual languages, and others (pandemic irregularities) that are common to entire populations of languages.

In Tagalog PMP *l normally became *l*: *(la)-laki > *laláki* 'man, male', *leŋa > *liŋá* 'sesame', *lima > *limá* 'five', *luheq > *lúhaʔ* 'tears', *qilaw > *ílaw* 'torch, light', *bileR > *bilíg* 'ocular cataract', *beli > *bilí* 'buy', *walu > *waló* 'eight'. However, in some etymologies it became *h*, or was lost (with automatic insertion of glottal stop between a resulting sequence of like vowels or unlike vowels the first of which is low): *luslus > *hushós* 'slip down', *balay 'public building' > *báhay* 'house', *buluq > *búhoʔ* 'bamboo sp.', *balu > *bálo ~ báʔo* 'widow(er)', *zalan > *daʔán* 'path, road', *bulan > *buwán* 'moon, month', *uliq 'return home' > *ulíʔ* 'again, once more' ~ *uwíʔ* 'return home', *selsel > *sísi* 'regret', Proto Philippines *habel 'weave cloth' > *hábi* 'woven pattern on fabric'. No condition can be stated for this split, which may be a product of early borrowing from a Bisayan language in which *l has lenited.

Several AN languages show an apparently unconditioned split of the voiced stops *b and *d. Dempwolff (1934) documented this in considerable detail for Javanese, where *b > *b* or *w* and *d > *d, ḍ* or *r* without stateable conditions, as in *batu > *watu* 'stone', but *balay > *bale* 'hall, public building'. In some cases *b > *b* may indicate a loanword from Malay, but there are many etymologies in which this seems unlikely, and some words show variation between *b* and *w*, as with *balik/walik* 'on the contrary, the other way'. Similar problems appear in some other languages, as Maranao and Tiruray of Mindanao, and Timugon Murut of Sabah: PMP *bulan > Maranao *olan* 'moon, month', *bihaR > *wiag* 'alive', *babuy > *bəboy* 'pig', *buaq 'fruit' > *boaʔ* 'endosperm of germinating coconut', *baRah > *baga/waga* 'ember', *batu > *batu/watu* 'stone'. In some languages such irregularities may result from sound change in progress, but if so it is unclear why this would result in phonemic splitting much more commonly for *b and *d than for *p and *t.

Apparently unconditioned sound changes, such as the splitting of *b and *d without statable conditions, are problematic for the Regularity Hypothesis because they suggest that sound change may proceed one lexical item at a time, as argued by Wang (1969). While lexical diffusion is a plausible explanation for phonemic split in a single language or set of closely related languages, it is much more difficult to apply to changes that independently affect large numbers of related languages. These kinds of recurrent changes, which apparently have not been recognised outside the AN language family, have been called 'pandemic irregularities' (Blust 1996d).

Following earlier writers, Dempwolff (1934-1938) recognised that the obstruent phonemes he reconstructed often have two reflexes in Oceanic languages, as with *puluq > Samoan *se-fulu* 'ten', but *pulut > *pulu* 'breadfruit sap'. He attributed these differences to prenasalisation, and in Oceanic linguistics the terms 'oral grade' and 'nasal grade' came to be used as designations of the more lenis and more fortis reflexes of an obstruent. By the 1970s it had become apparent that appeal to an oral grade reflex deriving from a simple stop, and a nasal grade reflex deriving from a prenasalised stop was insufficient to account for the full range of reflexes in many languages (Blust 1976a, Geraghty 1983). The first attempt to account for these anomalies within a comprehensive Proto Oceanic phonological system was that of Ross (1988), who recognised a three-way split in the obstruent reflexes of many Oceanic languages: oral lenis grade, oral fortis grade, and nasal grade. While the distinction between oral grade and nasal grade is largely predictable by evidence of prenasalisation in non-Oceanic AN languages, the lenis : fortis distinctions have developed by unconditioned phonemic split. What is peculiar about this split is that it affects the reflexes of *p, *k and *s (but not *t) in hundreds of languages, and since lenis

and fortis grade reflexes often disagree across languages in cognate morphemes, it must be assumed that this split is a product of independent historical changes. It is impossible to reconcile the treatment that Ross proposes with classical Neogrammarian theory, but it is equally impossible to account for the data without accepting a set of assumptions very much like those he adopts. Why lexically-specific splits of the inherited obstruents would affect hundreds of languages over a wide geographical area within a single subgroup of AN in a drift-like fashion remains a complete mystery.

A problem that is similar, but distinct from this, occurs in many AN languages of insular Southeast Asia. In addition to simple and homorganically prenasalised medial obstruents, as in *putiq 'white', *punti 'banana', Dempwolff (1934-1938) proposed nearly 200 reconstructions that contain a 'facultative nasal', such as *tu(m)buq 'to grow'. While it is easy to interpret the use of parentheses in such etyma as marking an ambiguity, this convention actually marks a cognate set that is internally contradictory, some languages pointing unambiguously to the simple stop and others to a prenasalised stop, as with Tagalog *túboʔ* 'sprout, shoot of a plant', Javanese *tuwuh* 'to grow' (< *tubuq), but Malay *tumbuh*, Fijian *tubu* 'to grow' (< *tumbuq). The problem of variable medial prenasalisation is both frequent in terms of the number of forms affected, and almost universal in languages of the Philippines and western Indonesia, where no obvious morphological or phonological solution is available (Blust 1996d). To cite only a few of many possible examples, Dempwolff reconstructed *hasaq 'whet, sharpen', a form that is reflected with –s- in hundreds of languages, yet Maloh of western Borneo has *ansaʔ*, and Javanese has *asah/aŋsah* 'whet, sharpen', Dempwolff reconstructed *betuŋ 'k.o. large bamboo', but Western Bukidnon Manobo has *bəntuŋ*, and Dempwolff reconstructed *ribu (now written *Ribu and reflected with a simple medial stop in many languages that permit medial prenasalisation), but Banggai of eastern Sulawesi has *imbu* 'thousand'. In some cases prenasalisation could not possibly have been original, as in PMP *busbus > Iban *bumbus* 'perforated', where a heterorganic consonant cluster had to reduce before prenasalisation was possible, or *i kahu > Maloh *iŋko* '2sg', where prenasalisation has occurred across an original morpheme boundary that has since been lost.

Even more fundamentally, Grace (1996) has suggested based on data from New Caledonia, that traditional models of language and language community are inaccurate, and should be given a far more fluid interpretation than the usual one in which languages are conceived as socially and spatially well-bounded entities. The revolutionary implications of this view in accounting for apparent irregularity in sound change are yet to be fully explored.

These cases are not meant to give the impression that sound change is chaotic in AN languages; it is not. By and large the broad patterns of change show an overwhelming tendency to regularity. But the key word here is 'tendency': there are many irregularities in individual languages, as well as broad patterns of irregularity across large populations of languages, as with the fortis/lenis reflexes of Proto Oceanic *p, *k and *s in many Oceanic languages, and the rampant pattern of sporadic medial prenasalisation in the AN languages of the Philippines and western Indonesia. Anomalies such as these should not be taken as the norm, but neither should they be swept under the carpet with facile appeals to borrowing, analogy, or lexical diffusion, since history shows that an honest acknowledgement of theoretically nonconforming data often provides the stimulus for the next great advance in understanding.

9.5   Drift

As used by Sapir (1921) the term 'drift' refers to the tendency for languages that have separated from a common ancestor to undergo parallel changes independently. In illustration, Sapir noted that English and High German (but not Low German) have developed a system of pluralising umlaut (*mouse* : *mice, Maus* : *Mäuse*) from an earlier system in which this class of nouns was marked by a suffix *-i*. Despite a chilly initial reception, the concept of drift has since been vindicated repeatedly in a number of different language families, including Austronesian. Three examples of widespread drift in AN languages, will be used to illustrate.

9.5.1 The drift to open final syllables

Because final consonant loss is considered a natural evolutionary path in sound change, one would expect it to occur randomly in a language family, or to be concentrated in a single subgroup due to innovation in an immediate common ancestor. What makes the loss of final consonants intriguing in AN is the clear evidence of geographical bias: despite their genetic diversity none of the Formosan languages show this change, nor do any of the more than 100 languages of the Philippines, or any of the scores of languages of Borneo or most other parts of western Indonesia. In stark contrast, many of the languages of Sulawesi south of the northernmost peninsula have drastically reduced word-final consonant contrasts either by merging or losing final consonants, or by adding supporting vowels, and about 90% of the more than 450 Oceanic languages have lost final consonants, even though every PMP final consonant except *h was retained in POC.

Several writers (Sneddon 1993, Mead 1996) have shown that at least some final consonants were found in the common ancestors of most Sulawesian microgroups, but that these are lost or reduced in number in some or all daughter languages. In other words, there has been a common tendency to erode word endings in these languages through independent, but parallel changes. Sneddon (1993) recognised ten indigenous microgroups, three isolates, and Manado Malay as distinct genetic units on the island. Of these, three groups show almost no tendency to reduce final consonants: 1. Minahasan, 2. Saluan, 3. Banggai. Two others show moderate reduction of final consonant contrasts (mostly merger of final nasals as *-ŋ*): 4. Tomini-Tolitoli, 5. Balaesang. A sixth group (Gorontalo-Mongondow) is split, the Gorontalic languages reflecting *-C as *-Co, and the Mongondowic languages retaining virtually all Proto-Philippine consonants in final position. Each of the remaining eight groups shows a strong tendency to reduce final consonant contrasts, including complete loss of final consonants in many languages. Since this is true even of Manado Malay, which is intrusive into the area, it suggests that contact has played a part in these changes. Although this may be a factor in some cases, however, it is difficult to apply to the Sangiric languages, which are separated from most other final-eroding languages of Sulawesi by a block of Minahasan and Mongondowic languages that show no trace of final consonant erosion.

In the Oceanic subgroup the loss of final consonants is even more pervasive, affecting all but a few dozen languages. Most of the languages that have not lost final consonants have preserved them through the addition of an echo vowel, or a supporting vowel *-a*. Why final consonant loss should be so much more common in Oceanic languages than other AN languages remains a major unsolved riddle. Unlike Sulawesi, which is relatively small and compact, the Oceanic languages are dispersed over a very wide area, and evidently acquired their attested distribution as the result of a rapid expansion of Neolithic farmers

into the Pacific. Those languages that retain final consonants are scattered widely over this area (New Ireland, the French Islands, southeast New Guinea, southern Vanuatu, New Caledonia), and the evidence suggests that final consonant loss was rapid, abrupt and independent in many parts of the Pacific.

9.5.2 Disyllabic canonical targets

Chrétien (1965) showed that about 94% of word-bases reconstructed by Dempwolff (1934-1938) contain two syllables. While affixed words typically are longer than this, the overwhelming predominance of a disyllabic canonical shape has had enduring consequences for sound change during the five and one half to six millennia since the breakup of the PAN speech community on Taiwan. Throughout the AN language family a variety of processes operate either on derived monosyllables or on derived trisyllables to restore a lost disyllabism, but are otherwise inoperative. Collectively these processes can be called 'canonical restoration'.

One type of canonical restoration is seen in Javanese, where loss of *h or *R caused many words to reduce to monosyllables, as with PMP *zuRuq > Old Javanese *duh* 'juice, sap', *paRih > OJ *pe* 'stingray', *tuRut > OJ *tut, tūt* 'follow', or *baRah > OJ *wa, wā* 'ember'. In modern Javanese all of these words have been restored to disyllables by CV-reduplication: MJ *duduh, pe/pepe, tutut, wawa*. That this process was persistent over many generations of speakers is apparent from the fact that some examples of restorative reduplication already existed in OJ (ninth to fifteenth century): *daRa (> *ra/rā) > OJ *rara, rarā* 'young girl, virgin', *duha > *ro, roro* 'two', *uRat > *otwat/otot* 'vein, tendon'. Notably, reduplication has no identifiable morphological value in these forms, and apparently functioned only to satisfy a disyllabic canonical target.

Old Javanese (Zoetmulder 1982) had a stative verb prefix *a-*, seen in *bwat* 'weight, burden' : *a-bwat* 'heavy, burdening' < *beReqat, *doh* 'a distance, afar' : *a-doh* 'far away, distant' < *zauq, *lon* 'slowness' : *a-lon* 'slow' < *laun, or *twas* 'hard core of a tree' : *a-twas* 'hard' < *teRas. Many of the modern Javanese equivalents have weakened or lost the morpheme boundary in such forms. Horne (1974), for example, lists *abot* 'heavy, hard, weighty', *adoh* 'far, distant', *alon* 'slow', *atos* 'hard, tough', and *awor* 'mixed, mingled' as dictionary entries (with a few forms listed separately under *bot*, *doh* and *wor*, but no entries for *lon* or *tos*). Superficially, loss of a morpheme boundary has no connection with reparative reduplication, yet the outcome of each is canonical restoration.

One of the most common processes used to restore a lost disyllabism in content morphemes is epenthesis. In most cases the epenthetic vowel is schwa. This type of change is relatively rare in the Philippines, where most intervocalic consonants are retained. Isneg, however, shows it in etymologies such as PMP *bahaR > *abāg* 'loincloth', *buhek > *aboʔ* 'head hair', *baquR 'trigger of a tension-set trap' > *abóg* 'fishing pole', *paqet > *apāt* 'chisel', or *tuhud > *atúd* 'knee'. In the southern Philippines Tiruray shows a similar change in *nahik > *ənik* 'climb, ascend', *tau > *ətəw* 'person', *tahep > *ətof* 'winnow', *tuhud > *ətur* 'knee', or *buhek > *əbuk* 'head hair', and even in such loanwords as *əsah* 'tea' (from northern Chinese). In Borneo several languages show the same type of change, as in *bahaR > Kelabit *əbʰar* 'loincloth', *buhek > *əbʰuk* 'head hair', or *bahaq 'floodwaters' > *əbʰaʔ* 'water, river'. Again, loanwords are treated in the same way so as to eliminate monosyllabic content morphemes: *əbol* 'ball', *əpəm* 'pump', *əti* 'tea' (all from English). Finally, Javanese shows apparent cases of restorative epenthesis with schwa in *duRi > OJ *rwi* > MJ *əri* 'thorn', *paRih > MJ *əpe* 'stingray' (variant of *pepe*), *əbuk*

'notebook' (Dutch *boek*), *əgaŋ* 'passage' (Dutch *gang*), or *əsop* 'soup' (Dutch *soep*). What is noteworthy in all of these cases is that schwa epenthesis happens only with historically secondary monosyllables, never affecting the numerically predominant disyllabic bases.

Tagalog reduplicates disyllables without change, as in *baʔit* 'good' : *magpaka-baʔit-baʔit* 'try to be very kind/good' (Schachter and Otanes 1972:339). However, reduplicated bases with glottal stop or *h* between like vowels often show laryngeal deletion and syllable contraction, as in *búhos* 'spilling' : *b<al>usbós* 'grain spilled from package', *laʔáb* 'spreading flame' : *l<ag>abláb* 'noisy conflagration', *láhad* 'opening of the hand' : *ladlád* 'opened, unfolded', or *súhol* 'bribe' : *sulsól* 'instigation to do evil'. This type of change may not appear to be connected with CV- reduplication, loss of morpheme boundary or epenthesis, but it never affects simple bases such as *búhos* or *laʔáb*, and the result once again is the satisfaction of a disyllabic canonical target.

One of the most surprising mechanisms used to satisfy a disyllabic canonical target is antiantigemination (AAG), a process whereby vowel syncope takes place only between identical consonants (Blust 2007b). This type of change is common in the Polynesian Outliers, but is also found in a similar form in Nuclear Micronesian languages, Mussau in western Melanesia, Iban in southwest Borneo, dialects of Eastern Peninsular Malay, and the Tomini-Tolitoli languages of northern Sulawesi. AAG can be illustrated with data from Tuvaluan of northwest Polynesia, and Iban of southwest Borneo. Besnier (2000:618) lists a number of Tuvaluan base forms and their syncopated derivatives. The base forms (marked with an asterisk) are described as 'both a synchronic abstraction and a historical reconstruction,' and generally correspond to CV- reduplications in other Polynesian languages, such as Samoan: **vavae* : *vvae* 'divide' (Samoan *vavae* 'divide, separate'), **mamao* : *mmao* 'far' (Samoan *mamao* 'far, distant'), **totolo* : *ttolo* 'crawl' (Samoan *tolo* 'crawl, swarm', *totolo* 'crawl, creep'). In Iban what probably was a historical process of CV- reduplication, together with prepenultimate neutralisation of vocalic oppositions as schwa, has produced a number of trisyllabic word bases that begin with the sequence $C_1 \partial C_1$-. In nearly every case these are given with alternative pronunciations, one with the schwa and one lacking it: *bəbadi/bbadi* 'have an accident, suffer', *gəgudi/ggudi* 'a kite (toy)', *jəjage/jjage* 'kind of water insect', *ləlaki/llaki* 'male', etc.

Together, these types of historical change can be seen as motivated by a single inherited structural pressure: the predominance of a disyllabic canonical target. Collectively they function much like a conspiracy in synchronic phonology, and constitute another form that drift can take.

9.5.3 The reduplication-transitivity correlation in Oceanic languages

One last example of drift is particularly instructive, since it is driven by the same canonical pressure just examined, but has had fundamental consequences for the syntactic typology of Oceanic languages. It is, in effect, a phonologically-driven syntactic drift, and understanding how it worked sheds important light on the interconnectedness of all parts of language in the process of change.

Many Oceanic languages share a morphological pattern in which intransitive counterparts of transitive verbs are reduplicated. This is highly transparent in Tok Pisin, which superimposes an English lexicon on a structure typical of Oceanic languages, as in *wasim* 'to wash something' : *waswas* 'to bathe', *tingim* 'to remember, think of' : *tingting* 'to think, ponder', *lukim* 'to see' : *lukluk* 'to look', or *tokim* 'to say something, speak to' : *toktok* 'to talk' (cp. *iu* 'to wash' : *iuiu* 'to bathe', *tumu* 'to write down' : *tutumu* 'to write',

or *kal* 'to dig up' : *kakal* 'to dig' in Tolai, which has provided much of the non-English vocabulary and structure of Tok Pisin). So widespread is this pattern that some writers have attributed it to Proto Oceanic. However, comparative evidence suggests that the reduplication-transitivity correlation has arisen repeatedly as a result of satisfying the same disyllabic canonical target described in 9.5.2.

Proto Malayo-Polynesian had a substantial number of unanalyzable reduplicated monosyllables such as *butbut 'pluck, pull out', *gemgem 'hold in the fist', or *tuktuk 'knock, pound, beat'. As a result of regular phonological change these were simplified to CVCVC in Proto Oceanic: *puput, *kokom, *tutuk, etc. Some Oceanic languages reflect these forms regularly, as with Lau *fufu* 'pick fruit' (< *puput), Tongan *koko* 'press, squeeze' (< *kokom), or Samoan *tutu* 'beat mulberry bark on a special anvil' (< *tutuk). However, many more languages appear to reflect forms like POC *puti 'pluck, pull out', *komi 'hold in the fist', or *tuki 'knock, pound, beat', as with Nggela *vuti* 'pull up, out, off or away; root up', Rennellese *komi* 'clasp firmly', or Nukuoro *dugi* 'punch, hit, strike, pound'. If only a few examples of this kind were known they might be dismissed, but because a similar pattern has been recorded in 37 forms it cannot be considered accidental (Blust 1977b). What, then, is the relationship of pairs such as *puput : *puti, *kokom : *komi, or *tutuk : *tuki, and what is their connection with the pattern of transitive marking shown above in Tok Pisin?

Pawley (1973) reconstructed POC *-i 'close transitive'. This is a very widespread affix in Oceanic languages, and its presence in POC is not in dispute. Given POC *-i 'close transitive' POC *puti 'to pluck', *komi 'to grasp', or *tuki 'to pound' may be interpreted as transitive verbs (*put-i, etc.) that in their intransitive forms were unsuffixed: *puput, *kokom, *tutuk. There there still remains one discrepancy between between these two sets of reconstructed forms, namely the absence of the first syllable in *put-i, etc. In a small number of other Oceanic languages the cognates of Nggela *vuti*, Rennellese *komi* or Nukuoro *dugi* are trisyllabic, as with Lau *fufusi*, Sa'a *huhusi* 'pluck, pull out' (with regular *t > s/__i), Tongan *kokomi* 'press or squeeze', or Manam *tútuki* (transitive singular), *tutúki* (transitive plural) 'crush, smash (food, betelnut)'. It appears, then, that earlier reconstructions were fragmented, showing parts of a morphological paradigm without an awareness of their internal coherence. To correct this problem it is necessary to reconstruct both simple and suffixed bases, POC *puput : *puput-i 'pluck, pull out', *kokom : *kokom-i 'hold in the fist', *tutuk : *tutuk-i 'knock, pound, beat', etc.

This brings us back to the disyllabic canonical target in AN languages, and shows yet another way that this target could be satisfied, namely by haplology. Alternatively, it is possible that forms like Nggela *vuti* arose by antiantigemination, with subsequent geminate reduction, although this appears unlikely in view of the lenis reflex of POC *p. Naturally, POC *-i was also suffixed to non-reduplicated bases, as with *qunap 'fish scale, turtle shell' : *qunap-i 'to scale fish', and here no allomorphy developed initially. Because the reduplication-transitivity correlation would have developed as an automatic by-product of the tendency to satisfy a disyllabic canonical target in suffixed reflexes of reduplicated monosyllables, this pattern would have been confined at first to this restricted set of forms. The comparative evidence suggests that Oceanic languages recurrently underwent parallel reductions of the type *puput-i > *put-i, with eventual loss of the morpheme boundary in the shorter forms. The number of forms that exemplify the pattern *puput: *put-i (> *pupu : *puti) evidently was sufficiently large that it was then generalised to the reflexes of non-reduplicated bases. Thus, based on the pattern *pu-ti]$_t$: pu-pu]$_i$ 'pluck, pull out', paradigm-mates such as *una-pi]$_t$: una]$_i$ 'to scale fish' were analogised to *una-pi]$_t$: una-una]$_i$ so as

to produce a reduplication-transitivity correlation in forms such as Mokilese *wina* (trans.) : *winaun* (intrans.) 'to pluck, to scale' (Harrison 1973). As noted at the outset, given the fact that a similar general correlation of form and function is found in many other Oceanic languages it is tempting to attribute this feature to POC. The evidence from reflexes of PMP reduplicated monosyllables, however, suggests that this grammatical feature was not present in POC, but arose repeatedly in the histories of diverse and widely separated Oceanic languages as a consequence of the blind operation of sound change.

10 *Classification*

10.0 Introduction

The problem of classifying languages in accordance with their histories falls into two broad categories: the establishment of genetic relationship and subgrouping. The first of these research areas has seen a surge of interest in recent years, as scholars who generally (but not always) operate outside the mainstream of historical linguistics have made a number of bold proposals that attempt to connect established language families into larger phyla or superfamilies. Before addressing some of the more fully developed of these proposals as they relate to AN, however, it will be useful to consider the general problem of establishing genetic relationship, and the boundaries of the AN language family itself, as this is normally understood.

10.1 The establishment of genetic relationship

To establish that languages are related it is necessary to weigh the relative merits of four competing explanations for similarity: 1. chance, 2. universals, 3. borrowing, 4. divergent descent from a common ancestor. It is impossible to show that languages are not related. When explanation (4) is not available, all that can be said with confidence is that sufficient evidence has not yet been presented to demonstrate relationship. This naturally leaves open the possibility that currently unrecognized relationships may be demonstrated in the future. However, it is important to stress that the demonstration of genetic relationship among languages depends on the prior elimination of chance, universals, and borrowing, no matter how distant the proposed relationship is.

It is widely recognized that the surest way to eliminate chance as a cause of cross-linguistic similarity is through lexical comparison. As Greenberg (1957) has noted, since the lexicon consists almost entirely of arbitrary sound-meaning pairings, each lexical item is independent of all others, something that is not true of structural features in phonology, morphology, or syntax. Moreover, although vocabulary can be borrowed, loanwords can often be detected by irregular phonological developments, and are not likely to dominate the basic vocabulary of a language. The general arbitrariness of the form-meaning relationship does not mean, however, that historically unrelated lexical resemblances do not occur. Indeed, since the vocabulary of any natural language contains thousands of forms it is only to be expected that chance will sometimes produce striking results. To give an idea of how easy it is to find spurious comparisons, Table 10.1 lists a few chance resemblances between lexical items in AN and various other language families that the writer was able to collect in less than four hours of casual browsing through published dictionaries (*Austronesian*: Malay, Seraway, Erai, Ifugaw, Lindrou, Kayan, Gane, Kanakanabu, Saaroa, Rotinese, Malagasy; *Indo-European*: Sanskrit; *Arawakan*: Yavitero; *Jivaroan*: Aguaruna, *Niger-Congo*: Swahili; *Eskaleutian*: Yupik; *isolate*: Ainu):

694

Table 10.1 Chance lexical similarities between Austronesian and other languages

Austronesian language	Other language	
Malay dua 'two'	Sanskrit dva	'two'
Seraway axi 'day'	Yavitero axi	'day'
Erai ani 'honeybee'	Yavitero ani	'wasp'
Thao apu 'grandparent, ancestor'	Aguaruna ápu	'chief'
Ifugaw búuk 'head hair'	Aguaruna buúk	'head'
Malay ikat 'tie, fasten'	Aguaruna íkat	'tie, fasten'
Seimat mut 'vomit'	Aguaruna ímut	'vomit
Lindrou babu 'grandfather'	Swahili babu	'grandfather'
Kayan bua 'fruit'	Swahili bua	'stalk, stem'
Lamaholot bobũ 'stupid'	Swahili bubu	'dumb, mute'
Pazeh damu 'blood'	Swahili damu	'blood'
Malay kaka 'elder sibling'	Swahili kaka	'elder brother'
Lamaholot əmaʔ 'mother; older woman'	Yupik ema-	'grandmother'
Saaroa naani 'here'	Yupik maani	'here'
Gane manik 'bird'	Yupik manik	'bird's egg'
Kanakanabu nanu 'where?'	Yupik nani	'where?'
Rotinese tamu 'smack lips while eating'	Yupik tamu	'chew once'
Malay api 'fire'	Ainu abe, api	'fire'
Malagasy nunu 'breast'	Ainu nunnu	'suckle'

To scholars of a certain temperament some of these resemblances will appear stunning, particularly where the forms are essentially identical in both shape and meaning, as with Seraway (southeast Sumatra) and Yavitero (Upper Orinoco) *axi* 'day', Erai (Wetar Island, eastern Indonesia) *ani* 'honeybee', and Yavitero *ani* 'wasp', Lindrou (Manus Island, western Melanesia) and Swahili *babu* 'grandfather', Pazeh (central Taiwan) and Swahili *damu* 'blood', Malay *kaka* 'elder sibling' and Swahili *kaka* 'elder brother', or Malay and Ainu *api* 'fire'.[111] The similarity of Malay *dua* to Sanskrit *dva* 'two' was noted more than 160 years ago by Bopp (1841), who misconstrued this and other random resemblances or loanwords as evidence for a genetic relationship between Indo-European and AN languages. Many hypotheses of distant genetic relationship appear to begin in this way—a scholar working outside the normal paradigm of his/her field observes one or two striking lexical similarities between languages not previously regarded as related, and becomes convinced that these resemblances could only be due to common origin. Once such a mindset becomes fixed there is no turning back: a search begins in earnest for additional

---

[111]   Ruhlen (1994), and Bancel and Matthey de l'Etang (2002) have shown that forms similar to Malay, Swahili *kaka* are globally distributed. They consequently attribute *kaka 'elder brother, uncle' to Proto Human. If, for the sake of argument we accept the monogenesis of human language, this would require us to assign *kaka to a language that almost certainly was spoken at least 100,000 years ago. Yet, with very few exceptions the forms cited as support for this global etymology have velar stops, a conservatism that strains all credibility. A number of the AN forms that Bancel and Matthey de l'Etang (2002) cite to support this argument are unrelated to PAN *kaka (Ontong Java *kainga*, Rhade *awa*, Tubetube *kaukaua*, Nggela *kukua*, Fijian *tuaka*), and the same is therefore likely to be true of data from other language families they cite. This is not to deny that kVkV occurs with surprising frequency as a term for an elder consanguineal relative, but the proper explanation for this association is yet to be determined.

support, and many more forms are adduced for the relationship, generally of poor quality and lacking systematic correspondence.

Chance similarity can be described as a type of convergence: languages that begin different come to resemble one another through lexical and phonological change. Malay *dua* : Sanskrit *dva* is a more attractive comparison, for example, than PAN *duSa : PIE *dwo- 'two', as are Malay *api* : Ainu *abe, api*, as against PAN *Sapuy : Ainu *abe, api* 'fire', or Gane *manik* : Yupik *manik* as against PMP *manuk 'chicken' : Yupik *manik* 'bird's egg'. Since finding chance resemblances between widely separated languages of different families is not difficult, it is only to be expected that even the comparison of proto languages may produce results of similar quality.

Convergence may also be motivated by language universals. Probably the best-known example of motivated convergence is the similarity of the words *mama* 'mother' and *papa* 'father' in many of the world's languages, as with Samoan *mama* 'mother' (child language), *papa* 'father' (child language), Spanish *mamá* 'mother', *papá* 'father', or Mandarin, Swahili *mama* 'mother', *baba* 'father'. Thanks to Jakobson (1960) the universal nature of this set of terms is well-established, and unlikely to lead to false hypotheses of genetic relationship. Some semantic universals or universals of metaphor, however, are less well-known, and could give rise to misguided notions of historical relationship, as shown in Table 10.2:

Table 10.2 Similarity between AN and non-AN languages due to semantic universals

Austronesian language	Other language
Malay mata hari (eye-day) 'sun'	Irish suil an lae (eye of day) 'sunrise'
Malay mata ikan (eye-fish) 'callus'	German Hühnerauge (chicken-eye)
	Mandarin jīyǎn (chicken-eye)
	Thai taapla (eye-fish) 'callus'
Malay mata jala (eye-net) 'mesh of net'	Hebrew ayin ha rɛshɛt (eye of net)
	Japanese ami no me (net **gen** eye)
	Vietnamese mat luoi (eye-net)
	Thai taakhàay (eye-net) 'mesh of net'

Although the similarity here is structural rather than lexical, failure to recognise the universal tendency to use the morpheme for 'eye' in the abstract sense of 'center, focal point' could encourage spurious hypotheses of genetic relationship in much the same way as the material in Table 10.1. Other types of structural convergence motivated by language universals are word order typology and canonical shape, both of which appear to have played a role in the genesis of some hypotheses of genetic relationship.

Similarity due to borrowing is generally well understood, and does not require extensive discussion. On the whole it is more likely to create problems in subgrouping than in the establishment of genetic relationship. Although any feature of language structure or content can diffuse through contact, the probability of diffusion is not equally distributed over all parts of a language. In particular, basic vocabulary is less likely to be borrowed than cultural vocabulary, and this distinction is often used to determine whether lexical similarity between languages is due to genetic relationship or to borrowing.

Once chance, universals and borrowing have been safely eliminated as plausible bases for similarity, the only remaining alternative is common origin, and this must be established on the basis of recurrent sound correspondences. Recurrence in sound correspondences can be demonstrated on the basis of a surprisingly small set of data. What

matters is not so much the number of cognates as the number of distinct and non-controversial instances of a given sound correspondence. Needless to say, these two measures of quantity tend to be interrelated, since each cognate identification necessarily provides either additional examples of established sound correspondences, or new sound correspondences. Nonetheless, a body of say, 50 cognates would convincingly establish the genetic relationship of two languages or language groups provided that each proposed sound correspondence is exemplified in at least two, and preferably three forms. Table 10.3 illustrates this point in a limited comparison of Malay and Hawaiian:

Table 10.3 Recurrent sound correspondences between Malay and Hawaiian

Malay	Hawaiian	English
mata	maka	eye
kutu	ʔuku	louse
ikan	iʔa	fish
laŋit	lani	sky
taŋis	kani	weep, cry

If we agree that it is necessary to demonstrate recurrence only for sound correspondences that involve a phonemic *difference* (e.g. *t* : *k*, but not *m* : *m*), then the cognation of the forms in Table 10.3 can be established beyond reasonable doubt by the following evidence: 1. *t* : *k* (eye, louse, weep/cry), 2. *k* : *ʔ* (louse, fish), 3. -C : zero (fish, sky, weep/cry), 4) *ŋ* :*n* (sky, weep/cry). With these five forms, we have three examples of Malay *t* : Hawaiian *k* and of Malay final consonant : Hawaiian zero, as well as two examples of Malay *k* : Hawaiian *ʔ*, and Malay *ŋ* : Hawaiian *n*. This is not to imply that more data would not be desirable (and much more exists); rather, it shows that even with limited data chance can effectively be eliminated by clear evidence of recurrence. Since universals and borrowing are also implausible explanations for the above similarities, the only remaining alternative is genetic relationship: these languages (along with many others) are continuations of an ancestral language that was spoken in the prehistoric past.

10.1.1 *Problems in the demarcation of the Austronesian language family*

Although the classification of most AN languages is non-controversial, this has not always been true, and for some languages questions still remain. Perhaps the first serious conceptual problem that Westerners faced in explaining the similarity and diversity of AN languages was how to classify the languages of Melanesia. As noted in Chapter 1, in 1778 Förster stated that the Polynesian languages exhibit similarities to Malay, but that the languages of Melanesia constitute a variety of independent groups. We now know that (with minor qualifications) Förster was wrong, but it may be instructive to examine the reasons for his error, since it persisted in scholarly discussions for over a century.

Two factors appear to have made the classification of AN languages more problematic in Melanesia than elsewhere. The first of these was the tendency of European observers to correlate language and race. Whereas AN speakers in insular Southeast Asia share a general phenotypic similarity with AN speakers in Micronesia or Polynesia, the speakers of AN languages in Melanesia often resemble speakers of Papuan languages more closely than they resemble speakers of AN languages outside Melanesia. The reason for this skewing of language affiliation and physical type clearly is gene flow, extending in some cases over a period of more than three millennia. Speakers of Papuan languages have been

in the western Pacific for upwards of 30,000 years, and the AN speakers who arrived there some 3,500 years ago interbred with the older population to varying degrees over time, leading in some cases to a radical change in phenotype.

The second factor is the undeniable lexical deviance of some AN languages in Melanesia. This becomes clear if we take the five lexical items in Table 10.3, add ten more, and extend the comparison to languages in Melanesia, as seen in Table 10.4:

Table 10.4 Patterns of lexical conservatism in five Austronesian languages

Malay	**Kaulong**	**Nengone**	**Fijian**	**Hawaiian**	**English**
mata	mara	warowo	mata	maka	eye
kutu	əmut	ote	kutu	ʔuku	louse
ikan	ili	pashawa	ika	iʔa	fish
laŋit	hiŋis	gulaʔawe	lomā-laŋi	lani	sky
taŋis	hau	mane(o)	taŋi	kani	weep, cry
kulit	po	nenun	kuli	ʔili	skin
susu	susu	mimi	suðu	ū	breast
hati	əran	tareat	yate	ake	liver
api	yau	ciʔiei	bukawaŋga	ahi	fire
air	eki	tin(i)	wai	wai	(fresh) water
dua	ponwal	rewe	rua	lua	two
tiga	miuk	tini	tolu	kolu	three
əmpat	mnal	ece	va	ha	four
lima	eip	doŋ	lima	lima	five
ia	yaŋ (m.)	bon/ic(e)	koya	ia	3sg.
	wut (f.)				

Although this sample is selective, it is sufficiently representative to be sure that were it increased to 100 or 200 items the same pattern would be visible: Malay, Fijian and Hawaiian show clear evidence of relationship to one another and to other AN languages, Kaulong shows limited evidence of relationship to these, and Nengone shows almost no evidence of relationship to these or to most other AN languages. Since Malay and Hawaiian show the greatest geographical separation, it is clear that divergence does not correlate with geographical distance in any straightforward manner. Rather, if anything, linguistic divergence tends to correlate with the physical type of the speakers. In other words, most lexically 'aberrant' AN languages are found in Melanesia, where they are spoken by populations that in general do not differ markedly in physical type from many Papuan-speakers. Such a statement admittedly is an oversimplification, since: 1. skin color in Melanesia (to choose only one measure of phenotypic variation) ranges from types that are almost Southeast Asian or Micronesian (Wuvulu and Aua Islands and the extinct population of the Kaniet Islands in the Admiralties, Tenis, east of Mussau) to the jet-black peoples of the western Solomons, and 2. within Melanesia the average language shows significantly lower lexical retention rates than in most parts of the AN world, but individual languages are sharply at variance with this trend, as with Motu, several of the languages of the central and southeast Solomons (Nggela, Arosi, Sa'a), Raga in northeast Vanuatu, or Fijian, all of which are lexically about as conservative as a typical Polynesian language. Nonetheless, apart from Enggano in western Indonesia, Yapese in western Micronesia, and possibly Nauruan in southeast Micronesia no AN language that might be called 'lexically aberrant' is spoken by a population that is phenotypically southern

Mongoloid, while many lexically aberrant AN languages are spoken by populations that are phenotypically 'Melanesian'. The overall picture thus suggests that—despite striking and still unexplained exceptions—contact and gene flow has played a significant part in accelerating lexical replacement rates in many of the AN languages of Melanesia.

Returning to the list in Table 10.4, which includes some of the most stable items of basic vocabulary in AN languages generally, Kaulong (southwest New Britain) shows only two clearly related forms (*mara* 'eye', *susu* 'breast'). Altogether, only 10 of 194 words that could be supplied for Kaulong on a variant of the Swadesh 200-word lexicostatistical test-list, or 5.2% show clear AN affinity. Not only is the rest of the vocabulary unfamiliar, but Kaulong distinguishes gender in the 3sg pronoun, a structural feature that is very rare in AN languages, but common in Papuan languages. At least six Papuan languages are spoken in New Britain: Anem, Baining, Kol, Pele-Ata, Sulka, and Taulil (Gaktai may be a seventh, or a divergent dialect of Baining). Given this array of observations—'Melanesian' physical type, little shared vocabulary, divergent structural properties, neighboring Papuan languages—it is natural to wonder whether the few recognizable items of basic vocabulary in a language such as Kaulong might not be AN loans in a non-AN language. Two considerations have traditionally persuaded scholars to classify Kaulong as AN. The first of these is the presence of a few clearly AN elements in very basic parts of the lexicon. In addition to *mara* 'eye' and *susu* 'breast', for example, Kaulong has *it* < *kita '1pl incl.', although virtually nothing else in the pronoun system is familiar, and only a handful of other items of basic vocabulary can be identified as AN. The second consideration is that Kaulong subgroups with Miu and Asengseng in the 'Western Whiteman Family', and a few additional AN elements surface in these languages (although not many). Since basic vocabulary is far less likely to be borrowed than non-basic vocabulary, the argument for classifying these languages as AN has depended on the presence of some clearly AN forms in basic vocabulary, together with the absence of any clear relationship to an uncontroversially Papuan language.

To see how shaky these arguments are one only needs to find an uncontroversially Papuan language that has at least as much AN basic vocabulary as Kaulong (or other members of the 'Western Whiteman Family'). This condition is met by Mailu, the largest of six languages in the non-AN Mailuan family in southeast New Guinea (Dutton 1975). On a modified form of the 200-item Swadesh list, Mailu has19 of 179 items, or 10.6%, that are just as clearly AN and just as basic as the material used to call Kaulong an AN language: 1. hand: *ima* (< POC *lima), 2. intestines: *sinae* (< *tinaqi, 3. blood: *lala* (< *raRaq, 4. dream: *nivi* (< *nipi), 5. thatch/roof: *ato* (< *qatop), 6. alive: *auauri* 'alive', *mauri* 'life' (< *maqurip), 7. buy: *woiwoi* (< *poli), 8. bird: *manu* (< *manuk), 9. fat/grease: *mona* (< *moñak), 10. louse: *tuma* (< *tuma), 11. mosquito: *nemo* (< *ñamuk), 12. sand: *one* (< *qone), 13. fresh water: *mami* (< *mamis 'sweet'), 14. flow: *aruaru* (< *qaRus), 15. star: *vitiu* (< *pituqon), 16. wind: *ani* (< *aŋin), 17. sick/painful: *marai* (< *ma-sakit), 18. shy/ashamed: *mai* (< *mayaq), 19. above: *ata-na* (< *atas; cp. the similar suffix in *au-na*, *goda-na* 'below'). In addition to these forms Mailu naturally has a number of AN loanwords in non-basic vocabulary as well (*tua* < *tupa 'derris root fish poison', *uaea* < *puqaya 'crocodile', etc.). The classification of Mailu is regarded as uncontroversial because there is complete agreement that it is genetically related to the other Mailuan languages (Binahari, Morawa, Domu, Bauwaki, and Labu), which have little or no AN lexical content. These other languages are spoken in inland areas, while Mailu is spoken on Mailu Island off the coast of southeast New Guinea, where it received much greater exposure to AN contact influence than the interior languages. If the basic

vocabulary of Mailu contains about twice the amount of unambiguous AN material as Kaulong, surely the basis for classifying the former as AN and the latter as non-AN must be questioned. In the final analysis it is clear that Kaulong (and other members of the Western Whiteman Family) could be classified as Papuan if some unambiguously Papuan relative or relatives were located in New Britain. No such relatives have been found, but this may be little more than an accident of history. Given the presence of other Papuan languages in New Britain, the Western Whiteman languages may well have had earlier relatives in interior New Britain that lacked AN features, much like the interior Mailuan languages in relation to Mailu. Looked at another way, if the interior Mailuan languages were to become extinct, Mailu could easily be misclassified as an aberrant AN language much like Kaulong or other Western Whiteman languages. This is not far-fetched, since these languages are all quite small: Dutton (1975:614) gives population figures as: Binahari 770, Morawa 755, Domu 482, Bauwaki 378, Labu 51?; Mailu 4,662. Problems such as this have driven some scholars to the radical conclusion that the family tree model (regarded by nearly all historical linguists as an idealisation of the processes of language split) is *never* appropriate in modeling genetic relationship. As Thurston (1987:4) puts it, "I would like to suggest that ultimately, all languages owe their earliest forms to processes such as pidginisation, and that after generations of use among intimates, these languages acquire the complexity that obscures their former origin." Few historical linguists would go this far, but in areas like New Britain generations of contact between unrelated languages has significantly blurred the boundaries between language families that must have been considerably clearer in the past.

Difficult as Kaulong and other Western Whiteman languages are to classify, Nengone presents even greater problems. Geraghty (1989) and others have claimed that the languages of the Loyalty Islands belong with those of New Caledonia, and this classification is accepted without reservation by Lynch, Ross and Crowley (2002:888). However, the evidence for this claim is tenuous. Even Nengone and Dehu, which are spoken on neighboring islands, show little lexical similarity. A casual check of the first 50 items on a modified Swadesh 200-word list turned up only seven likely cognates between these two languages (Dehu *aj* : Nengone *al* 'swim', Dehu *jun* : Nengone *dun* 'bone', Dehu *madra* : Nengone *dra* 'blood', Dehu *mano* : Nengone *nono* 'breathe', Dehu *thinem* : Nengone *gutinen* 'tongue', Dehu *xen* : Nengone *kān* 'eat', and Dehu *deŋ* : Nengone *dredreŋ* 'hear'), and the evidence linking them to languages of the New Caledonian mainland is even more limited. In reality, the classification of the languages of the Loyalty Islands as AN is heavily dependent upon their geographical position. Since the nearest uncontroversially Papuan languages are spoken in the Solomon Islands some 1,700 km. to the northwest, and the archaeological record to date does not clearly indicate a pre-Lapita population in southern Melanesia, there seems to be little alternative but to automatically classify all of the languages of this region as Austronesian.

Although there currently is universal agreement that the Western Whiteman languages, Dehu and Nengone are AN, the classification of a few other languages in Melanesia has given rise to real controversy. The first of these controversies began in 1911, when W.M. Strong and Sidney H. Ray, in back-to-back articles made opposite claims about the classification of Maisin, spoken at the southeastern tip of New Guinea. Strong believed that Maisin was a Papuanized AN language, while Ray believed that it was an Austronesianized Papuan language. For more than 65 years the classification of Maisin (sometimes called a 'mixed language') remained in limbo. Then Lynch (1977b) and Ross (1996b:192ff) both sided with Strong's original conclusion of 1911.

The second controversy concerns the Reef-Santa Cruz languages, spoken in the Santa Cruz Archipelago, east of the Solomons chain. These are Reef (or Gnivo), spoken on a few small islands in the group, Lödäi, a chain of dialects around Santa Cruz Island, and Nagu (or Nanggu), on the southeast corner of Santa Cruz Island. Lincoln (1978:930) argued, largely on the basis of evidence from numerals, pronominal affixes and tense markers, that these languages are AN. The numerals, however, show little resemblance to Proto Oceanic, as seen in the following data from Nemboi and Nagu:

Table 10.5 Numerals in Nemboi and Nagu of the Reef-Santa Cruz Islands

POC	Nemboi	Nagu	
*tasa	tüöte	töti	one
*rua	ali	tüli	two
*tolu	atü	tütü	three
*pat, pati	awä	tup^{w}a	four
*lima	nöwlün	mööp^{w}m	five
*onom	pötäŋimö	temũũ	six
*pitu	itumütü	tũtüü	seven
*walu	itumüli	tumulii	eight
*siwa	itumöte	tumatee	nine
*sa-ŋa-puluq	nöpnũ	napnũ	ten

Wurm (1978:971), on the other hand, adopted the view that "the Reef-Santa Cruzans originally spoke a non-Austronesian language or languages and that they have incompletely taken over an Austronesian language." Until recently most scholars sided with Wurm on this issue (Lynch 1998:217, Lynch, Ross and Crowley 2002). However, Næss (2006) has argued that the evidence for a Papuan affiliation is no stronger than that for an AN affiliation, and Ross and Naess (2007) have finally shown convincingly through a detailed analysis of the historical phonology, that these languages are indeed Oceanic, albeit Oceanic languages with highly atypical surface characteristics, and ones that apparently form a primary branch of the Oceanic group.

A third controversy concerns Magori and related languages of southeast New Guinea, which were classified as Papuan by Ray (1938). However, further descriptive and comparative work by Dutton (1976) has shown unambiguously that these languages are members of the Central Papuan Family, along with such better-known languages as Motu and Mekeo, and that their deviant appearance is a result of prolonged influence from the Austronesianized Papuan language Mailu.

Outside Melanesia the only real controversy that has ever arisen regarding the boundaries of the AN language family concerns Chamic. Here again the problems are a product of contact and convergence, but in this case they are due to convergence in language typology rather than convergence in phenotype, or rapid replacement of basic vocabulary. Well before the end of the nineteenth century the Chamic languages were recognized as Malayo-Polynesian. Shortly thereafter they were mentioned as such in a well-known publication by Kern (1889), and in 1891 Niemann argued that they are most closely related to Acehnese of northern Sumatra. Finally, those French scholars who contributed the most substantial descriptions of the languages (especially Aymonier and Cabaton 1906) unambiguously asserted their Malayo-Polynesian origin. Some German-speaking writers, especially the influential Austrian ethnologist and linguist Wilhelm Schmidt, however, differed with this interpretation. Largely because he relied heavily on

typological resemblance as a basis for genetic classification, Schmidt (1906) labeled the Chamic languages 'Austroasiatic mixed languages', an error that persisted in some parts of the English speaking world as late as the 1950s. Pittman (1959) set the record straight (although Francophone scholars surely wondered why this was necessary), and there has been no disagreement since. Thurgood (1999) provides the most complete account of the history of the Chamic languages, which are quite closely related to Malay and its closest relatives in western Indonesia. This study not only documents in great detail the steps which led from a proto language with a western Indonesian typology to daughter languages with typical Mon-Khmer typological features, but it also argues that Acehnese is a descendant of Proto Chamic rather than a sister language of the Chamic group, a claim that is challenged by Sidwell (2005).

10.2 The external relationships of the Austronesian languages

A number of attempts have been made to link AN to other language families. Some of these, as the claim that the now extinct Beothuk of Newfoundland was related to AN because of a common practice of headhunting (Campbell 1892) seem too frivolous to merit discussion. The following proposals, listed in the order that they were first advanced, will be briefly reviewed: 1) Austronesian-Indo-European, 2) Austronesian-Austroasiatic = Austric, 3) Austronesian-Semitic, 4) Austro-Japanese, 5) Austronesian-Amerind), 6) Austronesian-Tai-Kadai = Austro-Tai, 7) Austronesian-Chinese = Sino-Austronesian, and 8) Ongan-Austronesian.

10.2.1 Austronesian-Indo-European

In his pioneering comparative study of Austronesian Humboldt (1836-1839) examined Kawi, the poetic language of the Old Javanese texts, in some detail. He noted the obvious fact that much of the vocabulary has been borrowed from Sanskrit, but suspected that under this layer of relatively recent borrowing was a deeper layer indicative of genetic relationship. Humboldt's rather programmatic remarks were greatly amplified by Bopp (1841), who opened his treatise with the remarkable statement that Humboldt's materials and his own studies had led him to conclude (1841:1) "that the Malayo-Polynesian family is a descendant of Sanskrit, that stands to it in the relationship of a daughter language, whereas the majority of European languages are sister languages to Sanskrit." In modern terms this amounts to a claim that all AN languages are members of the Indic sub-branch of the Indo-Iranian branch of Indo-European languages. Bopp built his case on phonetic resemblances of an unsystematic kind, sometimes choosing forms that are confined to a single language for comparison with Sanskrit. Thus Tahitian *pae* : Sanskrit *pañća* 'five' are said to be related even though it is clear from Humboldt's tables that *pae* is not widespread (it appears, in fact, to be peculiar to Tahitian), and that the original AN word for 'five' must have been something like Malay, Hawaiian *lima*.

Despite his explicit assertion that the AN languages are descendants of Sanskrit, Bopp did not hesitate to cite evidence of relationship with other Indo-European languages when Sanskrit appeared wanting. The widespread AN form *lima*, for example, was compared with Irish, Scots Gaelic *lamh* 'five'. In other cases he considered only the syllable that suited his purpose, without bothering to justify the segmentation, as with Kawi *pitu* : Sanskrit *sa-pta* 'seven'. The great grammatical differences between Sanskrit and such AN languages as Malay were attributed to devolution. In Bopp's view (1841:2) the Romance

languages in comparison with Latin exhibit a state of deterioration ('eines verfallenen Sprach-Organismus'), and in much the same way he saw the AN languages as having arisen "from the ruins of Sanskrit."

These extraordinary claims are all the more remarkable in coming from one of the most famous figures in nineteenth century Indo-European linguistics. To compound the irony, after lying neglected for nearly a century the Indo-European-AN hypothesis was resurrected by Renward Brandstetter (1937), one of the leading figures in early twentieth century Austronesian linguistics. In contrast to Bopp, Brandstetter did not claim a special relationship between the AN languages and any single Indo-European language. Again, the case is built on phonetic resemblances of an unsystematic kind. It is, moreover, assisted by a theory of monosyllabic roots that is meritorious in its own right (Blust 1988a), but which loosened the criteria for comparison even further than would otherwise be the case, as seen in the comparison Kapampangan *dakáp* 'catch, arrest', Toba Batak *taŋkap* 'seize, grasp, comprehend', Karo Batak *akap* 'think', Latin *capere* (common root: *kap*) 'seize, grasp, comprehend.'

10.2.2 Austric

Shortly after the middle of the nineteenth century some British colonial writers began to note similarities between the Munda languages of eastern India and the Mon-Khmer languages of mainland Southeast Asia. In 1906 these observations were placed on a firmer footing by the Austrian anthropologist and linguist Wilhelm Schmidt, who proposed a language family spanning mainland Southeast Asia from Vietnam to central India. He called this family 'Austroasiatic' (AA), and divided it into seven primary branches: 1) a group of 'mixed' languages including Chamic, together with Sedang and other languages now generally regarded as Mon-Khmer, 2) Mon-Khmer, consisting of the two languages of empire after which it is named, along with scores of languages spoken by hill tribes in Vietnam, Laos and Cambodia, 3) the aboriginal (pre Malay) languages of the Malay Peninsula, 4) Palaung-Wa and Riang of the Burma-China frontier, 5) Khasi of Assam, 6) the languages of the Nicobar Islands, and 7) the Munda languages of India. It is now universally recognized that Schmidt's inclusion of the Chamic languages in the AA family was mistaken, but the rest of AA is generally accepted.

In the fourth chapter of his book *Die Mon-Khmer Völker* Schmidt, following up the preliminary remarks of earlier scholars, argued that the AA and AN languages are part of a larger group that he called 'Austric'. Schmidt's evidence included several formally similar affixes, and a set of 215 lexical comparisons. The morphological evidence is particularly interesting, and at first sight compelling. As has been seen, the major syntactic relations of PAN and PMP evidently were expressed through an elaborate system of verbal affixes that probably worked in conjunction with a smaller set of particles preposed to the noun phrase. These affixes include a variety of prefixes, the infixes *-um- and *-in-, and the suffixes *-an, *-en and *-i. In some Formosan and Philippine languages reflexes of *-um- and *-in- co-occur as *-inum-* or *-umin-*. Schmidt noted that Khmer, and to a lesser extent other AA languages, have infixes *-m-*, *-n-*, and that the double infix *-mn-* is found in Khmer and Nicobarese. He maintained further that although the functions of these infixes differ in the two language families, they can plausibly be regarded as related.

Given the rarity of infixes in human languages generally these agreements, down to specific details of co-occurrence, clearly are impressive. However, as Schmidt himself pointed out, in some AA languages almost any sonorant may be infixed: *m, n, ñ, ŋ, l, r*, and

close attention to the glosses does not suggest semantic similarity between *-um-* and *-m-* or *-in-* and *-n-*. The striking resemblance between AN and AA languages in the use of infixes thus suffers considerably when the comparisons are examined more closely. In other cases Schmidt simply misrepresented the facts. The PAN suffixes **-an* 'locative' and **-en* 'patient', for example, were treated as variants of the same morpheme. The adjectival suffix *-a* in Polynesian languages was then taken to reflect **-an*, and used to gloss the original as **-an* 'adjectival', thereby permitting comparison with a similar suffix in Mundari of the Munda group.

Schmidt's lexical comparisons, while sometimes quite striking, do not display systematic sound correspondences. Where the formal similarity is greatest he sometimes permits the meaning to vary considerably, as with Malay *lumut*, Tagalog *lumót*, Mota *lumuta* 'moss' next to Khmer *ləmuot* 'sticky, viscous, slippery'. Where the meaning is identical the phonetic agreement—still not based on recurrent sound correspondences—need not be close, as with Malay *ñamuk*, 'Munda' *gamit* 'mosquito'. In a few cases Schmidt appears to have identified forms that are genuinely related, but these can be explained as loans, as with Malay *danau*, Jarai *dənaw*, next to *dönau* 'lake' in the Mon-Khmer language Bahnar, which borders Jarai on the north.

Shortly after Schmidt's monograph appeared the Austric hypothesis was endorsed by several prominent Austronesianists, most notably Kern (1908), and Brandstetter (1916:25). Other scholars, as Blagden (in Skeat and Blagden 1906:2:444) were less receptive, and the idea slipped into dormancy for several decades. Kuiper (1948) attempted to bolster the evidence for Austric with typological resemblances between Munda and what he called 'Indonesian' languages. His argument, however, suffers from the same weaknesses as Schmidt's (failure to establish recurrent sound correspondences), and moreover contains serious misrepresentations of the AN facts. It had no impact on scholarship in the AN field, and no prominent Austronesianist after Brandstetter has accepted Schmidt's argument until quite recently.[112]

The most serious attempt to resurrect the Austric hypothesis is Reid (1994b). Despite errors in his use of the data, the strongest part of Schmidt's argument clearly was morphological, and Reid has tried to bring that aspect of the Austric hypothesis up to date by drawing on fuller descriptive sources than were available at the beginning of the twentieth century. He does this mainly by using data from Nancowry, a language of the Nicobar Islands, and from Katu, spoken in the highlands of central Vietnam. Until recent decades both languages were comparatively isolated, and perhaps for this reason are morphologically and phonologically conservative in crucial respects. Table 10.6 lists the most important affixes or clitics that appear to agree between the two language families:

---

[112]   This was not true of outsiders to Austronesian, however, as with Benedict (1942, fn. 55) or Shorto (1976).

Table 10.6: Morphological resemblances between Austroasiatic and Austronesian

Katu	Nancowry	PAN
pa-	ha-	*pa- 'causative'
pa-ka- 'double causative'		*pa-ka- 'causative (statives)'
	-um- 'causative'	*-um- 'actor focus'
	-an-/-in- 'nominalizer'	*-in- 'perfective/nominalizer'
	ma-/-am- 'agentive'	*ma- 'stative'
ta- 'involuntary action'		*ta(R)- 'involuntary action'
	-a 'object nominalizer'	*-a '3sg object'
	na 'ligature'	*na 'ligature'
	i 'locative'	*i 'locative'

From the beginning the shared morphological feature that caught the attention of most researchers was the use of infixes of similar shape. As Diffloth (1994:310) has noted, in assessing these similarities one must keep in mind "the great rarity, worldwide, of true infixes such as we find in both Austroasiatic and Austronesian." Schmidt's morphological data was compromised by the variability of the vowel, allowing him to posit only *-m- and *-n- or *-Vm-, *-Vn-. It was, moreover, also compromised by functional differences between these affixes in the two language families.

What is important about the data from Nancowry is that it evidently allows the Proto Austroasiatic forms to be reconstructed as *-um- and *-in-. This may seem like a minor change, but it greatly reduces the probability that the resemblance of AA and AN infixes is a product of convergence. On the debit side, Reid notes that the Nancowry nominalizing infix has allomorphs -an- (in monosyllables) and -in- (in disyllables). Since PAN *-in- can be reconstructed in only one shape, a corresponding Proto Austric form would have to be *-in-. Data from other branches of Mon-Khmer, however, points to *-an-, as with Katu -an- 'nominalizer', suggesting that if PAN *-in- is cognate with the nominalizing infix of AA languages -in/an- allomorphy already existed in Proto Mon-Khmer. Although the great bulk of evidence supports the view that PAN *-um- and *-in- were verbal affixes that marked actor focus/agent voice and perfective aspect respectively, reflexes of both appear in some daughter languages as nominalizers. As seen in Chapter 6, this is so common with reflexes of *-in- that a secondary nominalizing function can reasonably be assigned to this morpheme in PAN. The nominalizing function of *-um-, however, is far less well-supported, and generally is dependent on a larger syntactic context. As additional support for a genetic connection of these AA and AN infixes, however, Reid (following Schmidt) notes that the two can be infixed together as -mn-, a process that closely parallels the perfective form of actor focus verbs in PAN, marked by *-umin-.

While infixation with similar forms is especially striking, it is only part of the evidence for a possible distant genetic relationship between AA and AN. A causative prefix *pa- is common to both language families, but also to Tibeto-Burman. More strikingly, Costello (1966) reports several affix combinations in Katu, including *pa-ha-* 'causative passive' and *pa-ka-* 'double causative', composed of *pa-* 'causative' + *ha* 'causative passive' and *pa-* 'causative' + *ka-* 'causative'. The second of these is remarkably parallel to PAN *pa-ka- 'causative of stative verbs'. Reid (1994b:327ff) tries to make a case for PAN *ka- 'causative', but the evidence much more clearly supports *pa- 'causative' + *ka- 'stative' (Zeitoun and Huang 2000, Blust 2003c:443). Despite these misalignments of function the formal agreement of affix sequences in both language families cannot fail to make a strong impression. Finally, Costello (1966) gives Katu *ta-* 'involuntary action', a form that

resembles PAN *ta/taR- 'prefix of spontaneous or involuntary action', but was not included in Reid's evidence for Austric.

In the same issue of the journal in which Reid attempted to revive the Austric hypothesis based on new evidence of shared morphology, Diffloth (1994) surveyed the lexical evidence for Austric, and concluded that this is 'not impressive.' Of 41 items considered, most are easily dismissed as products of chance or borrowing. As Diffloth (1994:312) notes, however, "Ironically, it is the relative poverty of shared vocabulary between Austroasiatic and Austronesian, combined with evident agreement in the morphology, that argues for a genetic, and against a contact relationship between the two families, provided we allow for a great time depth in order to avoid the obvious paradox."

A very different view of AA-AN lexical relationships has been promulgated by Hayes (1992, 1997, 1999a, 1999b), who presents what he calls "irrefutable evidence for the Austric hypothesis." The last of these studies focuses on a limited set of putative sound correspondences, as shown in Figure 10.1:

Proto Austric (PAS)		PAA	PAN
1. *s	>	*s	*s > *h
2. *z	>	*z > *s	*z > *D
3. *z	>	*z > *nz > d, ʔd, ḍ	*z > *D
4. *z	>	*z > *s > *ns > t, ʔt, ṭ	*z > *D

Figure 10.1 Putative sound correspondences linking Proto Austroasiatic
and Proto Austronesian (after Hayes 1999b)

Expanding this compressed format into its full expression, Hayes allows the following PAA : PAN 'denti-alveolar sibilant' correspondences: 1. *s : *s, 2. *s : *h, 3. *z : *z, 4. *z : *D, 5. *s : *z, 6. *s : *D, 7. *nz : *z, 8. *nz : *D, 9. *d : *z, 10. *d : *D, 11. *ʔd : *z, 12. *ʔd : *D, 13. *ḍ : z, 14. ḍ : *D, 15. *ns : *z, 16. *ns : *D, 17. *t : z, 18. *t : *D, 19. *ʔt : *z, 20. *ʔt : *D, 21. *ṭ : *z, 22. *ṭ : *D. No fewer than twenty-two putative sound correspondences thus map onto PAS *s and *z. Needless to say, without evidence of conditioning this allows great latitude to the play of chance. Although space will not allow a consideration of more than a small part of the data presented, to avoid selection bias in the following critique all PAA and PAN examples given in support of PAS *s are cited in Table 10.7 as they are presented by Hayes:

Table 10.7 Evidence for Proto Austric *s, according to Hayes (1999b)

No.	PAA	Austronesian
1.	*ɣasi	*biRah 'Alocasia sp.'
2.	*sarʔom	*ha[r]um 'aroma, scent'
3.	*sa[ʔ]ak	*hawak 'body'
4.	*caḷus	*[s]ilu[h] 'fingernail'
5.	*saqi	*bahaq 'flood(ed)'
6.	*suk	*buhuk 'hair'
7.	*tunqas	*tuqah 'old'
8.	*saḷu]	*haluh 'pestle, pounder
9.	*ganosi	*sa(ŋ)guh 'pith, sago'
10.	*sa(n)qaɣ	*haka[r] 'root'
11.	*suɣup	*hiRup 'sip'
12.	*rawasi	*ñawah 'spirit, soul'
13.	*buɣasi	*buRah 'spray, sprinkle'
14.	*(m)bus	*sebuh 'develop steam'
15.	*sarut	*hurut 'stroke'
16.	*g[a]nis	*gigih 'tooth'
17.	*(kən)[ḷ,r]us	*peñuh 'turtle'
18.	*ŋkasi	*zaŋkah 'unit of measure'
19.	*(tam)pis	*ta(m)pih 'winnow'

The first general observation about this list is that the reconstructed AN forms are not assigned to a specific proto language, a point that will be clarified below. Second, the PAA forms are not glossed. This might lead some readers to conclude that the glosses given for the AN forms also apply to PAA, but this does not appear to be the case. Instead, Hayes cites forms from individual AA languages or the proto languages of lower-level subgroups to the left of his PAA reconstructions, and only these are glossed (e.g. Katu *saak* 'corpse' : PAA *sa[ʔ]ak : AN *hawak 'body'). Third, although most of the PAA reconstructions proposed by Hayes are disyllabic, in some cases the PAA form is evidently thought to match just the final syllable of the AN form, while in others it is thought to match the entire AN disyllable. We turn now to the individual comparisons and a detailed account of the problems with them.

1. On the AA side Hayes cites Khmer *ṛas* 'root', evidently meant to match the last syllable of *biRah, but the AN reconstruction should read: PAN *biRaq 'elephant ear taro: *Alocasia* spp.'. **Formal problems**: 1) The loss of the first syllable in PAA is unexplained, 2) PAA *s : PAN *q is not a stated sound correspondence, 3) the final vowel of the PAA form is unexplained. **Semantic problems**: Nowhere in the AN language family is a semantic change found connecting *Alocasia* taro with the meaning 'root'. Rather, *Alocasia* taros, which cause itchiness when eaten, are metaphorically associated with sexual lust (Blust and Trussel, ongoing). In many Formosan languages apparent reflexes of *biRaq mean 'leaf', as this is easily the most salient part of the plant.

2. On the AA side Hayes cites Vietnamese *thɤm* 'fragrant'. Dempwolff's *ha[r]um 'aroma' is restricted to Malay and a few other languages in western Indonesia that have borrowed heavily from Malay, as Javanese and Ngaju Dayak. There is thus no evidence that this form is old in AN. Moreover, the correspondences between these languages point to a form with initial *q. **Formal problems**: 1) PAA *s : PAN *q is not a stated sound correspondence.

3.   On the AA side Hayes cites Katu *saak* 'corpse'. The right column, however, should read PAN *Sawak 'waist; back of the waist'. ***Formal problems***: 1) PAA *[ʔ] : PAN *w is otherwise unattested, and evidently is not assignable to earlier *ʔ or *w. ***Semantic problems***: In most AN languages this is a body part term that lacks an exact equivalent in English: Kavalan *sawaq* 'waist, back of the waist', Hanunóo *háwak* 'back of the waist' (no general term for 'waist'), Cebuano *háwak* 'waist'; *háwak-un* 'tend to suffer from backaches', Simalungun Batak *awak* 'waist', Manggarai *awak* 'hips, waist'. The evidence suggests that PAN *Sawak and PMP *hawak referred to the unprotected space between the rib cage and the pelvic bone which is not covered by the muscles of the diaphragm, hence a part of the body corresponding roughly to the English concept 'waist', but applying only to the sides and back.

4.   On the AA side Hayes cites Semelai *cəruus* 'claw (nail)'. The correct AN form is *silu, a word that is found in a number of the languages of Borneo and Sumatra, and apparently in the Manobo languages of the southern Philippines, where it shows assimilation of the first vowel, as in Ilianen Manobo *sulu* 'fingernail'. The PAN form for 'fingernail, claw', however, was *kuSkuS. ***Formal problems***: 1) PAA *c : PAN *s is not a stated sound correspondence, 2) PAA *a : PAN *i is irregular, 3) whatever its antiquity in AN, *silu had no final consonant, an error that Hayes inherited from his sources, which transcribed Dempwolff's 'spiritus asper' automatically as -*h*.

5.   On the AA side Hayes cites Bahnar *kəsayʔ* 'sprinkle'. The AN form is PMP *bahaq 'flood; floodwaters', but no PAN form is known. ***Formal problems***: 1) The loss of the first syllable in PAA is unexplained, 2) The palatal glide of the PAA form is unexplained. ***Semantic problems***: Sprinkling and flooding are not closely associated concepts in any known AN language. The former is associated either with small amounts of water, or with the sowing of seeds.

6.   On the AA side Hayes cites Old Mon *sok* 'hair', a form that is widespread in AA languages. ***Formal problems***: 1) The loss of the first syllable in PAA is unexplained, 2) the *h in *buhuk derives from PAN *S, but the development was PAN *bukeS > PMP *buhek with medial and final consonants in the desired order only after the change *S > *h (Blust 1993c), 3) PAA *u : PAN *e is not a regular sound correspondence.

7.   On the AA side Hayes cites Katu *takɔh* 'old man or woman'. The right column should read *tuqaS 'elder; mature'. ***Formal problems***: 1) Given the Katu form the basis for the order of the vowels in the PAA reconstruction is not explained, 2) the implicit claim that Katu -*k*- derives from *q is contradicted by other forms cited by Hayes (1999b). In many AN languages this form, often with the stative prefix *ma- refers to elder relatives, but also to ripened fruits.

8.   On the AA side Hayes cites Katu *saal* 'pound rice'. The right column should read PAN *qaSelu 'pestle for pounding grains'. ***Formal problems***: 1) The loss of the first syllable in PAA is unexplained, 2) the use of square brackets in the PAA form evidently is intended to avoid the problem of accounting for the absence of a final vowel in AA forms, 3) the correspondence PAA *a : PAN *e is irregular.

9.   On the AA side Hayes cites Semelai *gnɔs* 'heart'. The AN form can be reconstructed as *sagu, but has a limited distribution and in most cases appears to be a loanword from Malay. ***Formal problems***: 1) The loss of the first syllable in PAA is unexplained, 2) the correspondence of PAA *-no- to PAN *-gu is unparalleled, 3) the loss of the last syllable of the PAA form is unexplained. ***Semantic problems***: Reflexes of *sagu refer to processed sago flour (not to the sago palm). The semantic connection of 'sago flour' with 'heart' is unexplained.

10. On the AA side Hayes cites Soui *səŋkaal* 'skin'. The right column should read PMP *akaR/wakaR 'root' (both forms can be reconstructed). ***Formal problems***: 1) PAA *s : PAN *w or zero is not a stated sound correspondence, 2) PAA *q : PAN *k is not a regular sound correspondence. ***Semantic problems***: The notions 'skin' and 'root' are not connected in the semantic histories of AN languages, and the basis for the proposed connection here is obscure.

11. On the AA side Hayes cites Khmer *sruup* 'swallow, sip'. The phonetic resemblance to PAN *SiRup 'sip' is close, but no more striking or systematic than many of the resemblances in Table 10.1.

12. On the AA side Hayes cites Vietnamese *[h]wa:ś > *vay* 'ancestor'. The AN form should read PAN *niSawa 'breath; breath soul'. ***Formal problems***: 1) PAA *r : PAN *ñ is additionally attested only in cp. (17), which is problematic for several other reasons, 2) the loss of the last syllable of the PAA form is unexplained. ***Semantic problems***: In the semantic history of AN languages the notions 'breath' or 'breath soul' have no semantic connection with the notion 'ancestor'.

13. On the AA side Hayes cites Proto Monic *pruus 'squirt'. The AN form is PAN *buReS 'spray liquid from the mouth'. ***Formal problems***: 1) Loss of the last vowel of the PAA form is unexplained. ***Semantic problems***: In AN languages squirting (produced by external pressure on an object that contains liquid) and spraying from the mouth are very distinct notions. Reflexes of *buReS often refer to shamanistic healing in which medicinal substances are chewed and spewed on the afflicted part of a patient.

14. On the AA side Hayes cites Proto Monic *ʔbuh 'boil'. The AN form is PMP *sebu 'extinguish a fire'. ***Formal problems***: 1) The loss of the first syllable in the PAA form is unexplained, 2) PAA *s : PAN zero is not a stated sound correspondence. ***Semantic problems***: In various attested AN languages reflexes of *sebu refer to the use of water to extinguish a fire, or to temper metals. The basic idea thus appears to be the sudden meeting of water and heat. The notions of boiling and extinguishing are distinct, and unconnected in any known AN language.

15. On the AA side Hayes cites Pearic *sro*(*o:*)*t* 'undress oneself'. The AN form is PMP *quRut 'massage'. ***Formal problems***: 1) PAA *s : PAN *q is not a stated sound correspondence, 2) PAA *a : PAN *u is irregular. ***Semantic problems***: The notions of undressing and massaging are not known to be connected in any AN language.

16. On the AA side Hayes cites Proto Monic *gnis 'canine tooth'. The AN form is *gigi, but is a likely Malayo-Chamic innovation (cp. PAN *nipen 'tooth', and various doublet forms with the same meaning). ***Formal +problems***: 1) PAA *a : PAN *i is irregular, 2) PAA *n : PAN *g is otherwise unattested, 3) PAA *s : PAN zero is not a stated sound correspondence.

17. On the AA side Hayes cites Proto Waic *rɨs 'turtle'. The AN form is PAN *peñu 'the green turtle: *Chelonia mydas*'. The basis for this comparison is completely obscure. ***Formal problems***: 1) The correspondence of PAA *(kən)- to PAN *pe- is unparalleled, 2) PAA *r : PAN *ñ is additionally attested only in cp. (12), which is problematic for several other reasons, 3) PAA *s : PAN zero is not a stated sound correspondence. ***Semantic problems***: PAN distinguished at least two types of turtles: *qaCipa 'freshwater turtle' and *peñu 'the green turtle'. All reflexes of *peñu refer consistently and exclusively to the green turtle, a large marine chelonian which presumably would have been unknown to speakers of PAA.

18. On the AA side Hayes cites Pacoh *ŋeaih* 'count'. The AN form is *zaŋka, with known reflexes restricted to languages of western Indonesia, and considerable variation in

meaning: Toba Batak *jaŋka* 'measure, weigh', Javanese *jaŋka* 'term, length of time', Malay *jaŋka* 'measuring off, esp. with calipers or compasses; unit of land measurement (in Negri Sembilan)'. ***Formal problems***: 1) The loss of the first syllable in PAA is unexplained, 2) PAA *s : PAN zero is not a stated sound correspondence, 3) the final vowel of the PAA form is unexplained. ***Semantic problems***: The notions of counting and measuring are distinct in AN languages, and terms relating to one domain rarely if ever undergo semantic changes that intersect the other domain.

19. On the AA side Hayes cites Proto Waic *pes 'sweep'. The AN form is PAN *tapeS 'winnow'. ***Formal problems***: 1) The correspondence PAA *i : PAN *e is irregular. ***Semantic problems***: The AN form referred to winnowing grains by tossing in the air on a winnowing basket. In AN languages the ideas of sweeping and winnowing are very distinct, and are unconnected in any known linguistic comparison

One need go no further to see that this type of "irrefutable evidence" is little more than a collection of random, and sometimes quite desperate attempts to match individual syllables in languages belonging to different families. A number of the attempts to match PAA *s with PAN *h are based on erroneous shapes of AN reconstructions. In other cases final vowels or entire syllables are unaccounted for, and the semantics of more than half of the comparisons are severely strained. Advocates of long-range comparison have learned from past criticism that evidence of genetic relationship will be accepted by most historical linguists only if it is based on recurrent sound correspondences. Every effort is therefore made to create the appearance of meeting this basic standard. Where such correspondences are not available they are manufactured by contrivance, but rarely without formal or semantic problems or both. Reid (2005) has reexamined the lexical evidence for Austric by assembling a set of 78 comparisons. He distinguishes these (very generously) as probable (17), possible (22), weak (7), or unacceptable (32), and concludes somewhat confusingly on an optimistic note, while at the same time acknowledging that the lexical evidence for Austric "is not as convincing as one would like."

The Austric hypothesis is unique, then, in the radical disjunction of morphological and lexical evidence: the morphological agreements between AA and AN are deserving of serious consideration as possible evidence for a distant genetic relationship (one that has led to the almost total elimination of common lexicon while still preserving a handful of core grammatical morphemes), but the lexical evidence for Austric, as Diffloth (1994) has correctly pointed out, remains elusive.

10.2.3 Austronesian-Semitic

On the grounds of geographical distribution alone, one of the most bizarre proposals concerning the external relationships of the AN languages is that of Macdonald (1907). This book, which follows up earlier publications on the same theme, contains a grammar of 81 pages and a dictionary of 220 pages for the non-Polynesian language of Efate Island in what was then the New Hebrides, where Macdonald spent many years in missionary work. Although the dictionary is useful in itself, both it and the grammar are offered as support for a theory that the AN (called 'Oceanic') languages are related to Semitic. As evidence for his position Macdonald makes numerous grammatical comparisons between various AN and Semitic languages, and appends a 36 page index of Efate words that he claims to be of Semitic origin. Like some of the proposals based on lexical evidence that have already been discussed above, the argument is distinguished by its almost exuberant disdain for method. In an example of a type that could be multiplied many times over, it is

said (1907:25) that the Semitic collective, abstract or feminine suffix -t 'often becomes *k* (or *h*) in Malagasy and Malay,' as in Malay *goso-k* 'rub, scour'. Since Macdonald is not concerned with the justification of morpheme divisions on either formal or semantic grounds one could argue with equal plausibility that the final consonant of English *hoo-k* is cognate with the Semitic suffix. Needless to say, no attempt is made to demonstrate recurrent sound correspondences between AN and Semitic languages, either where there is some degree of phonetic similarity (Efate *sumi* 'to kiss' : Arabic *s'amma* 'to smell'), or where there is none (Efate *uota* 'chief, lord; husband' : Hebrew *ba'al* 'lord; husband').

10.2.4 Austro-Japanese

At a time when some Austronesianists still gave serious consideration to the Austric hypothesis van Hinloopen Labberton (1924) put forth the novel proposition that the AN (called 'Malay-Polynesian') languages are genetically related to Japanese. At the same time he accepted Schmidt's argument, and so implied the existence of an Austric-Japanese superfamily. The evidence offered for an AN-Japanese relationship is partly typological and partly lexical (e.g. Malay *ikan* : Japanese *sakana* 'fish', Malay *nasi* : Japanese *meshi* 'cooked rice'), but is very limited. A second article published the following year offered considerably more 'evidence', but nothing that further strengthened the argument.

More than half a century later arguments for a genetic relationship of AN and Japanese were advanced again, most vigorously by Takao Kawamoto who, beginning in 1977, produced a long series of publications offering mostly lexical evidence for a historical connection of these languages. In some publications, as Kawamoto (1977:23) he has stated that 'Japanese is cognate with the Austronesian family of languages,' while in others, as Kawamoto (1984:31) he has said instead that 'Japanese was at least twice Austronesianized,' thus suggesting that the observed similarities are due to creolisation rather than genetic relationship. At another point in the same publication, however, he states (1984:32) that the Oceanic element in Japanese is directly inherited, while the non-Oceanic AN element is due to contact, implying that Japanese is an Oceanic language that has been overlaid with non-Oceanic AN loanwords. Although he presents hundreds of lexical items along with tables of reportedly recurrent sound correspondences to support his case, an analysis of his data supports none of these claims. Rather, virtually every etymology is problematic in one or more ways, and the lexical similarities are no more systematic or semantically convincing than those given by Hayes as evidence for Austric.

During the period that Kawamoto was occupied in publishing hundreds of proposed Japanese-AN etymologies, the American linguist Paul K. Benedict (1990) expanded his Austro-Tai hypothesis (see below) to include Japanese in a larger superfamily with three major nodes: 1. Proto Austro-Tai (PAT) splits into Miao-Yao (Hmong-Mien) and the rest, 2. Proto Austro-Kadai (PAK) splits into Kadai and the rest, 3. Proto Austro-Japanese (PAJ) splits into AN vs. Japanese-Ryukyuan. Benedict based his argument on both morphological and lexical evidence. He noted that proposals linking Japanese with languages other than AN have received more attention in the literature, and holds that 'The failure of the Austronesian hypothesis to capture a larger share of the audience lies primarily in the completely inadequate presentation it has suffered. Kawamoto … presents a host of comparabilia or 'look-alikes' rather than cognate sets per se and, in fact, states (1977:25) that his comparisons involve 'cognates or possible cognates'. What is more, the phonological correspondences that he suggests read in either direction/all directions (!), hardly enhancing their credibility.'

Given this critical commentary one would expect to see stronger arguments given for the proposed Japanese-AN connection, but Benedict's (valid) remarks prove to be just another example of the skillet calling the kettle black. Among hundreds of 'look-alikes' in his 104 page glossary of 'Austro-Japanese roots' are the following: 1) 'belly': PAJ *ba[r]aŋ : PAN *ba[rɣ]aŋ, Proto Rukai *baraŋ, Japanese *Fara*, 2) 'bird (of prey)': PAJ *taka : PAN *taka-, based on Paiwan *tjakaŋa* 'black kite (bird sp.)', Saisiyat *takako* 'falcon', Japanese *taka* 'hawk, falcon', 3) 'dog' : PAT *ʔa(ŋ)klu : PAN *asu/wasu, Old Japanese *winu*, and 4) 'feather/arrow' : PAJ, PAN *lawi : Japanese *ya* 'arrow'. In the first of these there is no internal AN evidence for PAN *ba[rɣ]aŋ. Rather than noting the internal evidence for PAN *tiaN 'abdomen' Benedict compares Proto Rukai *baraŋə 'belly' with the Old Japanese form. Reconstruction 'from the top down' is a valid and often useful procedure in historical linguistics, provided that the relationship of the languages has already been securely established. In the present case, however, it implies that *tiaN replaced *baraŋ in a proto language ancestral to all AN languages except Rukai, and hence makes the otherwise unsupported claim that at the highest levels the AN language family splits into two subgroups: 1. Rukai and 2. the rest. Although this view was defended by Starosta (1995:691), the arguments for it are flawed, and Benedict himself never explicitly adopted such a hypothesis. In the second comparison there is no evidence for PAN *taka-, which is arbitrarily extracted from the first two syllables of the Paiwan and Saisiyat forms, neither of which contains a known synchronic or historical morpheme boundary. In the third comparison Benedict posits a unique medial consonant cluster, giving him *carte blanche* to describe its development.[113] Comparison 4) involves a form meaning 'tail feathers' in AN languages. No known reflexes are associated with the feathering of arrow shafts, and the comparison with Japanese *ya* is both semantically and phonologically problematic, since Benedict's statements of correspondence lead us to expect PAJ *w to be reflected as Japanese -*w*- (1990:107), and the loss of the final vowel in Japanese is likewise unexplained. The arbitrariness of these claims is apparent from the observation that, while van Hinloopen Labberton offered the comparison Malay *ikan* : Japanese *sakana* 'fish', Benedict compares the same AN form (PAN *Sikan) with Japanese *ika* 'squid'.

Vovin (1994) dismisses all of the foregoing arguments as ill-founded. These multiple attempts to relate Japanese to AN nonetheless serve a useful purpose in relation to method, since on the one hand they show that the same potential relationship has appeared plausible to various scholars at different times, but on the other they highlight the complete arbitrariness of most cognate decisions.

10.2.5 Austronesian-American Indian connections

Various proposals, none of them convincing, have been made for a genetic relationship of AN with one or another language family of the Americas. The earliest and most prolific proponent of this view was the French linguist Paul Rivet, who is well-known for his many contributions to the comparative linguistics of South America. In a brief report Rivet (1925) argued for the common historical origin of what he called 'Melano-Polynesian' and 'Hoka' languages. The terminology is strange, and one is initially led to suspect that Rivet was speaking of similarity due to borrowing (from languages in Melanesia and Polynesia to the New World) rather than genetic relationship. But his position is stated clearly

---

[113]   Benedict reconstructs one example of *k₁ in initial position, where it reportedly produced Japanese *k*-, in *kusi* 'spit, skewer'.

(1925:51): 'Un groupe linguistique nord-américain et un groupe sud-américain peuvent être rattachés respectivement à la famille mélano-polynésienne et à la famille australienne.' Rivet inexplicably alters the term 'Hokan', first proposed by Dixon and Kroeber (1912) and subsequently used by Sapir to refer to a putative language phylum reaching from Oregon to Nicaragua, and the term 'Malayo-Polynesian'. More disturbingly, he cites no material from identified languages. Rather, he gives generalized 'Melano-Polynesian' and 'Hoka' words, with no indication of language, as with 'MP' *wahine* : 'Hoka' *huagen* 'woman', where the first term is taken from Hawaiian, but 'MP' *tasi* 'sea' : 'Hoka' *tasi* 'water', where the first term is taken from some unidentified non-Polynesian Oceanic source. Needless to say, this kind of random fishing for similarities between widely separated languages with no controlled evidence of recurrence is no different than the material in Table 10.1. Rivet (1926) is a far more ambitious attempt to demonstrate genetic relationship between AN and American Indian languages, but contains the same methodological errors as the earlier publication.

More recent proposals for a genetic relationship between AN and various languages of the Americas include Key (1984, 1998), and Foster (1998). According to Key (1984:1) "The languages of Polynesia contain elements also found in North and South American Indian languages that suggest distant historical connections. In this preliminary study it is not yet possible to determine whether the resemblances are due to borrowings, or whether the common structural characteristics go back to the same genetic origins." She claims (1984:5) to base cognate identifications on "the correspondences and reflexes that result from language change," but an inspection of her 35 pages of proposed cognate sets reveals no evidence that she has adhered to this ideal. Rather, the material she has assembled shows no more systematic coherence than that of e.g. Macdonald (1907). The first five items in her appendix of 'Word Sets' are sufficient to make the point: 1) PAN *buka, PPN *fuke, Aztec *tapowa, Quechua *phaska-, Tacana *pohoke* 'to open', 2) Hawaiian *tūtū/kūkū*, Nggela *kukua*, Proto Uto-Aztecan *kaku, Mapuche *kuku* 'grandparent', 3) PAN *halaŋ 'lie down', Easter Island *taha* 'incline', Proto Uto-Aztecan *ka, Proto Panoan *raka-, Cavineña (Tacanan) *ha'ra-* 'lie down', 4) Qae *caro*, Nggela *saro*, Proto Quechua *siri- 'lie down', 5) Proto Polynesian *loto, Hawaiian *i-loko*, Proto Quechua *ukhu, Cavineña (Tacanan) *-doko-*, Chama (Tacanan) *-doxo-*, Amahuaca (Panoan) *ʔokə* 'inside'.

Comparison 1) is based on the occurrence of an initial labial stop and a medial -*k*- in most witnesses. There is, however, no attempt to demonstrate recurrent sound correspondences. The second example shows that even within AN the claims of cognation are based purely on sporadic phonetic similarity (regular correspondences are Hawaiian *k* : Nggela *t*, or Hawaiian *ʔ* : Nggela *k/g*). The third comparison takes an AN form *halaŋ 'lie athwart, block the way' that is not assignable to PAN or even PMP on family-internal evidence (Blust and Trussel ongoing), and proposes South American connections. The inclusion of Easter Island (Rapanui) *taha* shows quite clearly again that—despite her claims to the contrary—recurrent sound correspondences are irrelevant to Key's proposals of cognation, since the regular Rapanui reflex of *halaŋ would be **ala. In the fourth comparison the AN forms are chosen from individual languages, and have no known antiquity in AN. If the claim is that these are loanwords a case must be made for migration from the central Solomons to the high Andes (or vice versa). Moreover, Fox (1955) gives *saro* with the meaning 'to spread, of a mat; spread out or over', not 'lie down'. Comparison 5) takes Proto Polynesian *loto 'inside, pool, lake, lagoon' (Walsh and Biggs 1966) and compares it with a random collection of forms that contain a medial -*k*- in various American Indian languages, all apparently meaning 'inside'. The semantics of this

form in Polynesian languages suggests a development from the meaning 'lagoon' to the meaning 'inside' (from concrete noun to relational term). Key makes it clear (1984:8) that she is struck by the parallel change *t > *k* in Hawaiian and in Chama, a Tacanan language of Bolivia. While this is of interest to diachronic typology, it hardly constitutes evidence for historical contact, yet this is the way she seems inclined to use it. In the case at hand the exclusive citation of American Indian forms with -*k*- rules out genetic relationship, since a change *t > *k* has been demonstrated only for Chama. Because she does not consider chance, this would seem to leave borrowing as the only explanation Key is willing to accept. But loanwords must have sources, and the only Eastern Polynesian languages that have undergone a *t > *k* change are Hawaiian and Tahitian (Blust 2004b). In both languages this is a very late change. In Hawaiian there is documentary historical evidence that change *t > *k* probably was spreading from the island of Hawai'i northward during the late eighteenth century, with the conquests of Kamehameha I. In Tahiti it is confined to dialects of the Leeward Islands, and to rapid speech styles in the standard language, where it is dissimilatory. Any hypothesis of borrowing, then, must assume contact between speakers of (southern) Hawaiian and various peoples of the high Andes within the past two and one half centuries.

Key (1998) continues in the same vein, citing vaguely similar forms in AN and South American Indian languages as diverse as Guarani of Paraguay (Tupian), Selknam of Tierra del Fuego (Chon), Cayapa of Colombia (Paezan), Waiwai of the Guianas (Cariban), and Quechua (Quechuan) and Aymara (Aymaran) of the high Andes. The resemblances are not qualitatively different from the similarities noted in Table 10.1, and are generally far less striking. Foster (1998) chides Key (1998) for her inconsistency in advocating use of the comparative method to determine linguistic relationship, while abandoning it in practice. With reference to her own proposals she notes that 'The comparative evidence discussed here follows the sound methodological principles that she recommended, demonstrating a surprisingly close, though geographically extensive, genetic relationship between Quechua (Qu) of South America, Afroasiatic (AA) of Africa and Mesopotamia, and the Pacific-island languages of the Austronesian (AN) family.' She calls this unpromising collection of putatively related languages the 'Proto Pelagian language phylum,' sets up tables of asserted sound correspondences, and then proposes a number of cognate sets that repeatedly violate the very method she advocates.

10.2.6 Austro-Tai

Two new proposals concerning the external relationships of AN, the first a book by the Danish linguist Kurt Wulff, the second an article by the American linguist Paul Benedict, appeared in 1942. Although they were completely independent, both proposals pointed to a superfamily that includes AN and the Tai languages. Wulff's book was published three years after his death. In it he assumes the validity of the traditional view that the Tai languages are related to Chinese, but argues further for a Chinese-Tai-AN language phylum. Wulff's evidence consists of some 145 lexical comparisons. As the argument for an AN-Tai connection has been developed in considerably greater detail by Benedict it will be considered here.

Benedict's view of an AN-Tai relationship—generally known as the 'Austro-Tai hypothesis'—differs from that of Wulff in attributing the resemblances between Thai and Chinese to borrowing. In his first publication on the topic, Benedict (1942) noted striking resemblances between the numerals of Dempwolff's Uraustronesisch (called 'Indonesian')

and various minority languages of southern China and northern Vietnam, collectively called 'Kadai'. To show how tantalizing the evidence is, these comparisons are reproduced here, omitting only the data from Northern Li and Southern Kelao, which adds nothing to his case, and modernizing the shapes of Dempwolff reconstructions:

Table 10.8 Austronesian-Thai-Kadai numerals (after Benedict 1942)

	PAN	**Laqua**	**S.Li**	**N. Kelao**	**Lati**
one	isa	tiă	ku	si	tsam
two	duSa	ðe	dau	so	fu
three	telu	tău	su	da	si
four	Sepat	pe	sau	bu	pu
five	lima	mö	ma	mbu	ŋ
six	enem	nam	nom	naŋ	nă
seven	pitu	mö tău	t'u	ši	ti
eight	walu	mö du	du	vleu	be
nine	Siwa	mö diă	pöü	su	lu
ten	sa-puluq	păt	p'uot	beu	pa

The resemblances seen between the reconstructed AN numerals and the Southern Li (now called 'Hlai') numerals 5-8 in particular, seem to push the limits of chance as a plausible explanation. Instead, they suggest that originally longer words were reduced to monosyllables by loss of all but the last syllable in the Kadai languages. In this first publication Benedict was confident, but restrained—he did not propose reconstructions for the proposed proto language ancestral to AN and Thai-Kadai, and he suggested (1942:597) that 'most of the important lexical correspondences have been uncovered.'[114]

Benedict did not publish on this topic again for another 25 years, and when he returned to it he had extended the Austro-Tai hypothesis to include the Miao-Yao (Hmong-Mien) languages of southern China, and abandoned all methodological restraint with regard to sound correspondences. In a number of publications beginning in 1967 and culminating in Benedict (1975), he proposed hundreds of new etymologies together with protoforms. To give an idea how these protoforms relate to the data used to support them we need only consider the words for 'buffalo', 'citrus' and 'dog'. For 'buffalo' Benedict (1975) reconstructs PAT *k[a]R[ ]baw; *k[a]R[]b/l/aw to account for forms such as Malay *kərbau*, Proto Tai *grwaay. This word, however, has no antiquity in AN, being in all likelihood a Mon-Khmer loanword into Malay, whence it spread into a few other coastal languages of western Indonesia and the Philippines. The Tai form is said to derive from *g[]rbaay < *k[]rbaay < *k(a)Rbay < *k(a)Rb/al[/aw] through a series of changes that are posited purely for *ad hoc* convenience. The word for 'citrus' is given as PAT *m[i]law, based on Malay *limau* and similar forms in other languages of western Indonesia, Proto Polynesian *moli, and Proto Tai *naaw, said to be from *[ml]aaw < *m(a)law < *m[i]law. But clear reflexes of *limaw 'citrus fruit' are restricted to Old Javanese *limo*, Malay *limau* and a few other languages of western Indonesia, and there is no compelling reason to relate it to Oceanic forms reflecting *molis. Although a PAN term is unknown, a better PMP candidate for the meaning 'citrus fruit' is *muntay, with reflexes reaching from at

---

[114] In his earlier work Benedict used 'Thai' to refer to a language group. After 1975 he began to use 'Tai' to refer to the Tai languages as a group, and 'Thai' to refer to Siamese proper, hence the terminological inconsistencies.

least the southern Philippines to Mentawai west of Sumatra, and eastward to the Moluccas (Blust and Trussel ongoing). Finally, the word for 'dog' was mentioned above in connection with the Austro-Japanese hypothesis, where it was reconstructed as *ʔa(ŋ)klu. Benedict (1975) instead posited *[wa]kləwm[a], based on reflexes of *asu in most AN languages (*wasu in some Formosan languages), *hma in Proto Tai and *klu in Proto Miao-Yao. The liberal use of square brackets and the reconstruction of unique consonant clusters appear calculated to maximise the role of chance in comparing phonetically dissimilar forms with no evidence of recurrent sound correspondences.[115] It is unfortunate that Benedict's later work does such violence to sound method, since the Austro-Tai hypothesis, like the Austric hypothesis, may very well be a valid construct that has been unjustly discredited by undisciplined etymologizing.

Table 10.9 Buyang : Austronesian etymologies (after Sagart 2004)[116]

PAN	PMP	Buyang	
*maCay	*matay	ma°tɛ⁵⁴	'die'
*maCa	*mata	ma° ta⁵⁴	'eye'
*qayam	*manuk	ma°nuk¹¹	'bird'
*quluh	*qulu	qa°ðu³¹²	'head'
*kuCu	*kutu	qa°tu⁵⁴	'louse'
*qetut	*qetut	qa°tut⁵⁴	'fart'
*qudip	*qudip	qa°ʔdip⁵⁴	'raw'
*Cumay	—	ta°mɛ³¹²	'bear'
*-ku	*-ku	ku⁵⁴	'I'
*-Su	*-mu	ma³¹²	'thou'
*duSa	*duha	ca⁵⁴	'two'
*telu	*telu	tu⁵⁴	'three'
*Sepat	*epat	pa⁵⁴	'four'
	*lima	ma³¹²	'five'
	*enem	nam⁵⁴	'six'
	*pitu	tu³¹²	'seven'
	*walu	ma°ðu³¹²	'eight'
	*siwa	va¹¹	'nine'
	*sa-ŋa-puluq	put⁵⁴	'ten'

More recent work that observes stricter methodological controls has demonstrated a very likely historical connection between Tai-Kadai and AN, but the nature of this connection remains in dispute. Ostapirat (2000, 2005) has assembled a core set of comparisons linking AN and Tai-Kadai (called 'Kra-Dai') that are reminiscent of the

---

[115] Matisoff (1990) describes the use of 'reconstructions' that essentially combine unrelated morphs into a single conglomerate form as 'protoform stuffing'. Under a similar procedure English 'dog' and Tagalog *asó* might be derived from an *ad hoc* 'protoform' *dogaso, with appropriate rules of deletion, and so on with any randomly chosen forms from languages not known to be related.

[116] Sagart also includes Buyang ta°qup¹¹ : Western Malayo-Polynesian *ta(ŋ)kup 'to cover', but this is a much weaker comparison, as the monosyllabic root *-kup 'enclose, cover' occurs with many different initial syllables in AN languages, and a form compatible with this reconstruction is currently known only in Ilokano *takúp* 'patch, mend by patching' (Blust 1988a:116). In Sagart's interpretation the PAN numerals above 'four' involve special complications that will be discussed below.

etymologies in Benedict (1942), but that go beyond them in several ways. Examples include PAN *nipen : Tai *fan* : Gelao *pan* 'tooth', PAN *qetut : Tai *tot* : Buyang *tut* 'fart', PAN *kuCu : Kam-Sui *tu* : Hlai *tshou* : Laha *tou* 'louse', PAN *maCa : Tai *taa* : Hlai *tsha* : Laha *taa* 'eye', and PAN *aku : Tai *kuu* : Buyang *kuu* 'I'. His proposed sound correspondences differ in several respects from those of Benedict, and avoid many of the methodological problems that plagued Benedict's sometimes extravagant claims. Similarly, Sagart (2004:432-433) has drawn attention to Buyang, a Kra (Kadai) language spoken near the China-Vietnam border, which has apparent cognates of the AN numerals 1-10, as well as a dozen or more elements of apparently cognate basic vocabulary. What is striking about this language is that it is not strictly monosyllabic, and both syllables show recurrent correspondences with AN forms, as shown in Table 10.9.

Buyang preserves penultimate syllables, but these are restricted in form, as the vowel is always *a* with a neutral tone, and the initial consonant must belong to the set *m, q* or *t*. What is important about the data in Table 10.9 is that it greatly strengthens the evidence offered by Benedict (1942) for a historical connection between AN and Tai-Kadai. Few linguists viewing these comparisons would feel comfortable attributing them to chance. Whether the historical connection is due to contact or divergent descent, however, remains a point of contention. Although the Buyang evidence was not yet available to him, Thurgood (1994) argued on the basis of claimed irregularities in tone correspondences that AN and Tai-Kadai have a historical connection, but one of early borrowing before AN speakers left the Asian mainland rather than one of genetic relationship. Ostapirat (2005) claims that the tone correspondences noted by Thurgood (1994) are in fact regular, and point to distant genetic relationship. As will be seen in connection with the Sino-Austronesian hypothesis, Sagart, argues for a genetic relationship, but a much closer one than envisioned by either Benedict or Ostapirat.

10.2.7 Sino-Austronesian

Following earlier, less systematic proposals by Wulff (1942), and Peyros and Starostin (1984), the French Sinologist Laurent Sagart commenced an ambitious attempt to demonstrate the genetic relationship of Sino Tibetan to AN. In its initial formulation (Sagart 1990, 1993), the proposal linked AN to Chinese, but not to Sino Tibetan as a whole. Shortly thereafter, in response to criticism from Southeast Asian specialists Sagart (1994) expanded his proposed language family to include Austric + Sino Tibetan, with a suggestion that it may ultimately turn out to include Miao-Yao (Hmong-Mien) and Tai-Kadai as well. Sagart (1995) modified this again by suggesting that although Chinese and Tibeto-Burman may be genetically related, the relationship is very distant, is obscured by early Chinese loanwords in Tibeto-Burman languages, and is not exclusive (i.e. Sino Tibetan may be valid, but this does not exclude the possibility of a closer Chinese-AN relationship). Table 10.10 lists selected comparisons between PAN, Old Chinese and Proto Tibeto-Burman as given by Sagart (1994), who calls these 'potential cognates':

Table 10.10 Preliminary evidence for Sino-Austronesian (after Sagart 1994)

PAN	OC	PTB	
*quluH$_1$	hl<j>uʔ	*lu	head
*punuq	nuʔ	*nuk	brain
*nunuH$_1$	n<j>oʔ	*nuw	breast
*-luR	hl<j>uj	*twiy = twəy, lwi(y)	flow
*qiCeluR	prefix+lonʔ	*twiy = twəy	egg
*tuktuk	tok, t<r>ok	*tuk 'cut, knock, pound'	peck
*-kuk	kh<(r)j>ok	*kuk	bent, crooked
*a(n)Dak	t<r>ik	*l-tak	ascend
*kuSkuS	k<r>ot	*kut	scrape
*-lus	lot, hlot	*g-lwat 'free, release'	slip off, loose
*-sep	tsɨp	*dzo:p	suck
*tutuH$_1$	tuʔ	*tow/dow	beat, pound
*imay	mijʔ	*moy (Bodo-Garo)	rice
*D$_2$amaR	hmɨjʔ	*mey	torch/fire

On the whole it is difficult to find this evidence persuasive. In four of the fourteen comparisons an AN monosyllabic root is taken in isolation as a comparandum, even though in AN languages these roots occur only in combination with an initial syllable. The treatment of Old Chinese postconsonantal glides and liquids as infixes is a matter for Sinologists to determine, but if it is disallowed it will have clearly deleterious consequences for many of the remaining comparisons. Even if the infixal analysis of postconsonantal glides and liquids is allowed, the material in Table 10.10 hardly constitutes convincing evidence of a genetic relationship between AN and Sino Tibetan. The Tibeto-Burman homophones *twiy = twəy, lwi(y) 'flow' and *twiy = twəy 'egg' in comparison with the homophonous final syllables of PAN *-luR 'flow' and *qiCeluR 'egg' may appear striking, but similar phonetically dissimilar rhyming sets can be found between other languages that are not conventionally regarded as related, as with Dutch *een*, *steen*, Malay *satu*, *batu* 'one; stone'.

Most recently Sagart has presented what he considers an improved argument for Sino-Austronesian. In response to criticisms by (Li 1995b) that few Old Chinese-AN comparisons offered in earlier publications involve basic vocabulary Sagart (2005:161) proposes "sixty-one basic vocabulary comparisons between AN, Chinese and TB." This change is at once welcome and a cause of concern. It is welcome because it fully acknowledges the genetic relationship of Chinese and Tibeto-Burman, in accord with the view of most specialists working on these languages. At the same time it is a cause for concern, since with over 250 Tibeto-Burman languages to match with data from PAN or other proposed AN proto languages the probability of chance agreements is greatly increased. It is difficult to see any qualitative difference between the comparisons offered here and those that Sagart has offered in earlier papers; the overwhelming impression one gets is that he has convinced himself of the relationship, and is searching desperately (if skillfully) for any scrap of data that can be taken as evidence to support his views. Table 10.11 lists the first ten comparisons that he gives minus the Chinese characters; B. = Benedict (1972), PECL = Proto East Coast Linkage, a putative first-order descendant of PAN on the east coast of Taiwan; raised letters mark tones:

Table 10.11 Improved evidence for Sino-Austronesian (after Sagart 2005)

		PAN or PECL	OC	TB
1.	body hair	gumuN	[b]mu[r] (eyebrow)	B. mul
				(Moshang kemul)
2.	bone	kukut	[a]kut	
3.	brain	punuq	[a]nuʔ	B. (s-)nuk
4.	elbow	siku(H$_2$)	[b]t<r>kuʔ	Gyarong tkru
5.	female breast	nunuH$_1$	[b]noʔ	B. nuw
6.	foot	kakay		B. kriy
7.	head	quluH$_1$	[b]hluʔ	Lushai lu
8.	palm of hand	dapa	[b]pa	B. pa
9.	pus	nanaq		Tib. rnag
10.	mother	ina(-q)	[b]nraʔ (woman)	B. m-na

In general Sagart assumes that forms in Old Chinese and many attested Tibeto-Burman languages reflect only the last syllable of a PAN reconstruction. Comparison (1) does not appeal to a known PAN etymon (the usual reconstruction is *bulu 'body hair; downy feather; plant floss'). Rather it cites a form from a proposed language called the 'Proto East Coast Linkage'. There is no published support for such a reconstruction, and its status is obscure. Among attested languages I have been able to find a related form only in Tamalakaw Puyuma, which has *gumul* 'body hair; feather.' Needless to say, this comparison is distributionally, formally and semantically unconvincing, and the addition of a disyllabic form from the Baric language Moshang simply reinforces the impression that this kind of etymologizing is grasping at straws. Comparison (2) fares no better. The generally accepted PAN word for 'bone' is *CuqelaN, while the source of *kukut is unpublished and obscure. Comparison (3) faces fewer objections, since *punuq 'brain' can be reconstructed for PAN based on Formosan-only comparisons. Comparison (4) apparently requires both syllables of PAN *sikuh 'elbow' to be retained, but since this is not normally the case it raises unanswered questions. Comparison (5) raises another issue, since the vowels do not match; inspection of Sagart's list shows that he allows both OC *o and *u to correspond to PAN *u, without conditions. Comparison (6) contains an error; the PAN word for 'foot/leg' is *qaqay, not *kakay. In either case it is unclear why the Benedict reconstruction *kriy is thought to show anything more than a chance resemblance to the AN form. Comparison (7) can be considered one of the triumphs of Sagart's efforts to impose a historical interpretation on random resemblances: *q appears to yield OC *h*- here and in PAN *qaluR 'to flow' > OC *hlu[r]ʔ* 'water, river'. One can argue that this is evidence for recurrent sound correspondences, but as noted already two examples of an apparent sound correspondence in forms of similar or identical meaning can arise by chance. Comparison (8) takes a 'Proto Western Malayo-Polynesian' form proposed in Blust (1980b) based only on Cebuano *lapalapa* 'sole, bottom surface of the foot' and Sasak *dampa* 'palm, sole' and treats it as though it were the PAN reconstruction for 'palm of the hand'. In fact, no PAN form has yet been established, and the most widely distributed word with this meaning is PMP *palaj 'palm of hand; sole of foot'. Comparison (9) should be *naNaq, and the comparison with Tibetan *rnag* is irregular. Comparison (10) fails to account for the liquid in OC, and assumes gratuitously that the final glottal stop derives from the PAN vocative suffix *-q (Blust 1979).

In addition, 16 of Sagart's 61 basic vocabulary comparisons attempt to match a monosyllabic root in AN (*-qem 'cloud(y)', *-taq 'earth', *-kut 'dig', etc.) with a word in

OC or TB, even though these roots are invariably bound in AN languages. All-in-all, then, this most recent attempt does nothing to strengthen the case for Sino-Austronesian. In a separate table Sagart adds 14 items of 'cultural vocabulary comparisons' between AN, OC and TB, including words for *Setaria* and *Panicum* millet, husked rice, paddy, chicken, cage or enclosure, net, broom, stopper or plug, to bury or tomb, loincloth or robe, to plait or braid, to shoot, and to hunt. Some of these are interesting, as PAN *panaq 'aim at a target; flight of an arrow', OC ᵃnaʔ 'crossbow', but for the most part they appear far-fetched (e.g. PAN *beCeŋ : OC ᵇtsĭk 'foxtail millet: *Setaria italica*', where both OC -ŋ and -k are allowed to correspond to PAN *-ŋ without conditions).

Far more than most proposals that have attempted to link AN to other language families, Sagart's work is methodologically rigorous and meticulous in its attention to detail. Despite these virtues most scholars have been left with the impression that the Sino-Austronesian hypothesis is the product of an *idée fixe*: once the claim was made that Chinese and AN are genetically related, no stone has been left unturned in trying to find further support for it. The reason that Sagart's conclusions diverge so radically from those of most scholars in both Austronesian and Chinese historical linguistics (Wang 1995), is that he regards comparisons such as those associated with PAN *-luR, *qiCeluR and *D₂amaR in Table 10.9 as evidence for recurrent sound correspondences (in this case PAN *R : OC j : PTB *y), although no single comparison is of high quality. At least part of the material that Sagart uses to buttress his thesis apparently is extracted from Old Chinese texts that abound in obscure and rarely used words and, as noted by Li (1995b), in many cases the proposed comparisons are semantically rather loose.

10.2.8 Ongan-Austronesian

Most recently Blevins (2007) has advanced the startling claim that Jarawa and Onge of the Andaman Islands form an 'Ongan' family that is a sister of Austronesian, the two forming an 'Ongan-Austronesian' superfamily. This is surprising for several reasons. First, by all indications the Andaman Islands probably have been separated from all other landmasses since at least the end of the Pleistocene epoch, some 10,000 years ago. At the time of Western contact the Andamanese lacked seagoing canoes, and so probably reached the islands by foot when they were part of the Asian mainland, or by sea when distances between the islands and mainland were much shorter than they have been during the Holocene period. Second, all Andamanese are physically very distinct from the southern Mongoloid type of most Austronesian speaking peoples in stature, skin color, hair form and various other bodily features. They appear, moreover, to form part of a once more widely-attested population of Southeast Asian Negritos that includes surviving (generally mixed) groups in the Malay Peninsula and scattered portions of the Philippines. Third, they show radical cultural differences from most modern AN-speaking groups and from speakers of PAN. Comparative linguistic evidence shows clearly that speakers of PAN lived in permanent pile dwellings, had rice and millet agriculture, cultivated non-grain crops such as bananas and sugarcane, used the simple back loom to weave textiles, made pottery and domesticated such animals as the dog, pig, and probably chicken. By contrast the Andamanese were hunter-gatherers at the time of initial Western contact, and hence lacked all of the traits just mentioned.

This background information suggests that AN speakers and the Andamanese represent fundamentally distinct divisions of humanity that have been separated for tens of thousands of years. A priori, then, the physical and cultural evidence does not present an

encouraging picture for a hypothesis of genetic relationship between AN and any of the languages of the Andamans. In fact, as noted by Blevins (2007:158) and other researchers before her, it has not been possible to show that Jarawa and Onge are genetically related to the languages of Great Andaman—a clear indication in itself of how long these islands have been inhabited by a Negrito population of hunter-gatherers. Blevins argues that Jarawa and Onge (the 'Ongan family') are genetically related to AN, but is non-committal on their possible relationship to the languages of Great Andaman. Given the close physical and cultural traits shared by all Andamanese groups, it would be very odd to find that Jarawa and Onge formed a single community with speakers of AN languages at a time when they were already separated from other Andamanese. Despite these formidable obstacles Blevins has proposed regular sound correspondences between Proto Ongan and PAN, and supported these with over 100 proposed etymologies and several types of grammatical data. The results are not impressive. Although she claims to use the comparative method in arriving at her results, the semantic comparisons are often very loose, morpheme boundaries are supplied where they are convenient rather than where they are justified, and *ad hoc* hypotheses of phonological change are invoked to make many comparisons 'work'.

Most of the above proposals appeal to unsystematic or far-fetched resemblances between AN and other language families. The clearest exceptions to this statement are Austroasiatic (based on morphological correspondences) and Tai-Kadai/Kra-Dai (based on lexical correspondences). Apart from Tai-Kadai, none of the lexical data offered so far is qualitatively superior to the material in Table 10.1, yet advocates of these views have not infrequently accused their rivals of violating sound method. Thus, Macdonald (1907:6-7) criticizes Bopp for methodological lapses of a sort that he himself is guilty of to a far greater degree, and van Hinloopen Labberton (1924:250-257) takes both Bopp and Macdonald to task for errors of a type that he seems blind to in his own work. Similarly, Benedict (1990) criticizes Kawamoto for proposing a Japanese-Austronesian relationship on the basis of 'look-alikes', yet his own comparisons are equally impressionistic. Finally, Benedict has attributed all resemblance between AA and AN languages to contact, although at least one scholar has proposed the same explanation for the resemblances between AN and Tai-Kadai languages (Thurgood 1994). Any one of the preceding propositions or some combination of them may, of course, ultimately prove to be correct (a historical connection of some type now appears virtually certain for Tai-Kadai and AN). But, on currently available evidence confident assertions about the external relationships of the AN languages remain the province of the true believer.

10.3 Subgrouping

The problem of linguistic subgrouping can be subdivided into three areas: 1. models of subgrouping, 2. methods of subgrouping, and 3. results of subgrouping. Before the results of subgrouping in the AN language family can be properly understood it is necessary to examine briefly the kinds of models and methods that have been used to attain them.

10.3.1 Models of subgrouping

Although the general literature of historical linguistics recognizes just two models for subgrouping languages (family tree model, wave model), several other alternatives that do not correspond exactly to either of these have been proposed in the AN literature.

10.3.1.1 The family tree and wave models

Nineteenth century linguists understood the basis for establishing genetic relationship before they understood the basis for subgrouping, or were able to develop conceptual tools for dealing with it. The first such conceptual tool was the family free model, which was developed by the German Indo-Europeanist August Schleicher in 1861-1862. Schleicher's family tree for Indo-European contained some branches that are generally rejected today, as 'Slavo-Germanic', 'Aryan-Greco-Italo-Celtic', 'Greco-Italo-Celtic', or 'Albanian-Greek'. A decade later Schleicher's former student, Johannes Schmidt countered with the wave model. To explain the distribution of linguistic features that had led Schleicher to propose e.g. a 'Slavo-Germanic' branch for Indo-European, Schmidt (1872) argued that many linguistic innovations are propagated from centers of prestige or influence, and spread out like ripples on a pond into which a stone has been dropped. In time it was shown that both models have their utility under given sets of circumstances. The Wave Model tends to work best with dialect networks, where a high degree of mutual intelligibility permits linguistic changes to diffuse easily from one community to another, while the family tree model tends to work best with languages that have been separated for some time. The reasons for this are partly linguistic and partly geographical. With increasing separation time language communities have greater opportunities to move apart, and are more likely to be separated by considerable distances, while the opposite is true for language communities that have been distinct for a relatively short time.

In practice many historical linguists find the family tree model more convenient than the wave model as a general diagrammatic tool, and there has been a tendency to use family trees almost exclusively in representing subgrouping relationships. When this is done it is sometimes noted that the model is an idealisation of the processes of change, since it assumes sharp cleavages between communities that have no further contact. In fact, however, most language splits probably begin with a partial separation, meaning that innovations continue to diffuse between the daughter communities for some generations before complete separation is achieved (if ever). This point is sometimes misunderstood by non-linguists who read family tree diagrams as literal claims about the process of language split, and assume wrongly that historical linguists fail to take account of diffusional influence (sometimes called 'reticulation', in contrast to 'fission').

10.3.1.2 The shifting subgroup model

The first alternative to the family tree model proposed anywhere in the AN language family was that of Geraghty (1983), who found it impossible to apply strict family tree principles to subgroup the languages of Fiji. In particular, the dialects of the Lau Islands in eastern Fiji (called 'Tokalau Fijian') share features exclusively with Polynesian languages that do not occur elsewhere in Fiji. These shared innovations are not products of recent borrowing from Tongan, but are said to be shared "by Proto Polynesian and Eastern Fijian (especially Tokalau Fiji) exclusively of Western Fijian" (1983:366-367). Geraghty argued that the simplest explanation of the facts is one that assumes language separation followed by convergence: the dialects of the Lau Islands were part of the Proto Polynesian speech community that were later reincorporated into the Fijian dialect network. In a family tree model such relationships cannot easily be represented, since branches diverge, but never converge. As Geraghty puts it (1983:381), "The genetic model, therefore, is supplemented to explain the relationships observed, by allowing a language to change its subgroup membership over time." Geraghty's model for Fiji does not correspond exactly to the classical wave model developed in Indo-European linguistics, since the wave model allows

overlapping isoglosses as a result of gradual or imperfect separation during language split, but does not allow convergence after split. It might therefore be called a 'convergence model'. However, this name could be misleading, since Geraghty also recognizes the normal process of linguistic fission. For now I call it 'the shifting subgroup model.'

10.3.1.3 The network-breaking model

Pawley and Green (1984) have proposed a model of language split that is different from both the wave model and the shifting subgroup model of Geraghty. They characterise the standard family tree as a '*radiation* model,' the essential feature of which is that a localized, homogeneous language community expands by hiving off daughter communities that are widely dispersed, and therefore isolated from one another in their ongoing development. Pawley and Green oppose this model to what they call the 'network-breaking model', in which the territory of a proto language may not expand at all, but daughter communities arise through a process of gradual contraction in the spheres of communication within a dialect network. In this model a dialect complex that is unified by communication networks based on trade, intermarriage, etc. evolves into a number of semi-independent descendants as communicative links gradually weaken and spheres of linguistic unity contract. While the classic wave model stresses the spread of innovations across space, then, the network-breaking model stresses the weakening of communicative bonds over time, and the consequent divergence of the daughter communities. In this sense both the family tree and wave models can be seen as dynamic models of language split (since both involve outward movement of language communities or of innovations), whereas the network-breaking model is a static model of language split.

10.3.1.4 Innovation-defined and innovation-linked subgroups

In what is surely the most thorough discussion of subgrouping models in the AN field, Ross (1997) has surveyed the literature on processes of fission and fusion both in linguistics, and in the sister disciplines of social/cultural anthropology and archaeology. He argues that cladistic (schismatic) and rhizotic (fusional) processes can be identified in the history of languages and cultures, but whereas both types of process are common in ethnogenesis, rhizotic processes are rare in glottogenesis, occurring only in situations of pidginisation and creolisation. Ross touches on a number of important points, the most significant of which can be summarized in the contrast between what he calls 'innovation-defined' and 'innovation-linked' subgroups. The first of these corresponds to the type of subgroup represented in family tree diagrams, or the radiation model of Pawley and Green. The second corresponds roughly to the wave model, but with a difference of emphasis. While the wave model was proposed to show that two otherwise well-defined subgroups share certain features as a result of contact (as Slavic and Germanic, or High German and Low German in the region of the Rhenish Fan), an innovation-linked subgroup proposes a genetic unit in which no innovation encompasses the entire population of languages. Rather, Innovation (1) may be found in Languages A-E, Innovation (2) in Languages B-G, Innovation (3) in Languages D-K, etc., thus allowing the broad subgroup to be established by linked innovations. Another noteworthy point raised by Ross (1997:222) is that residual categories may be represented by single descent lines in a family tree diagram and so give the impression that these represent innovation-defined or innovation-linked subgroups rather than terminologically unified collections of multiple primary branches. This is the case, for example, with the Western Malayo-Polynesian group first proposed in Blust

(1977a), and reasserted in many subsequent publications despite the absence of a clear body of exclusively shared innovations shared only by these languages.

10.3.2 Methods of subgrouping

Because of their number, great geographical range, and apparently rapid spread (punctuated by two noteworthy 'pauses', one in Taiwan and the other in western Polynesia), the AN languages present major subgrouping challenges. In this respect AN is very different from some language families, such as Indo-European, where there is virtually universal agreement about the assignment of languages to subgroups. In part because of these problems and in part because of unique features of the zoogeography of the AN world, the classification of AN languages has produced some of the most original subgrouping methods to emerge in historical linguistics since the nineteenth century.

10.3.2.1 Exclusively shared innovations

As late as the 1930s the AN languages were classified by geographical region: those of insular Southeast Asia were 'Indonesian', those of Melanesia were 'Melanesian', those of Micronesia were 'Micronesian' and those of Polynesia were Polynesian. Dempwolff (1934-1938) adopted this schema in vol. 1 of his *Vergleichende Lautlehre*, but then found evidence for a large 'melanesisch' (= Oceanic) subgroup that includes all AN languages of the Pacific except Palauan and Chamorro. His evidence for this group consisted primarily of phonological mergers, in particular the highly distinctive merger of *p and *b, which is relatively easy to demonstrate in most languages. Other mergers that confirmed this grouping were the merger of the palatal obstruents *c, *s, *z and *j, and of *e (schwa) and *-aw as *o*. The merger of *j with the other palatal obstruents is now known to have taken place after the separation of the languages of the Admiralty Islands from other members of the Oceanic group, and the merger of *e and *-aw occurs in some other AN languages, but the merger of *s and *z is nearly as distinctive as that of *p and *b.

Dempwolff's method of establishing the Oceanic subgroup agrees closely with the type of evidence that Brugmann (1884) had recommended more than half a century earlier in subgrouping the Indo-European languages on the basis of exclusively shared innovations. All generally accepted AN subgroups have been established on this basis, but as will be seen below, the way in which these innovations have been demonstrated has sometimes been quite innovative. Other fairly standard uses of the same procedure differ from the kind of evidence Dempwolff used to establish the Oceanic subgroup in appealing mostly to lexical rather than phonological innovations. Eastern Malayo-Polynesian, for example, is defined almost entirely in terms of innovative lexicon, a critical part of which consists of clear replacement innovations (PMP *anak replaced by PEMP *natu 'child', PMP *nunuk replaced by PEMP *qayawan 'banyan tree', PMP *tuRun replaced by PEMP *sobu 'go down, descend', etc.).

10.3.2.2 Lexicostatistics

Subgrouping by lexicostatistics was pioneered in the AN language family by Dyen (1965a). In this massive comparison 245 AN languages, drawn from a set of 371 word lists, were compared on the basis of data from the Swadesh 200-word test list. Using impressionistic cognate decisions Dyen fed 'plus' and 'minus' values into an early-generation mainframe computer at Yale University to generate a matrix of cognate

percentages between the languages in the sample. The result was in need of interpretation and fine-tuning, but the major conclusions were stated with a good deal of confidence: the AN family divides into 40 primary branches, of which 34 are located in or near Melanesia, pointing to a primary center of dispersal in the area of New Guinea and the Bismarck Archipelago. Although this conclusion was accepted almost instantly by some anthropologists (as Murdock 1964), it is rejected today by virtually all linguists and by nearly all Pacific archaeologists. Dyen's conclusions followed from the type of data he used: cognate percentages without regard to the innovation/retention distinction. Since the rate of replacement of basic vocabulary evidently varies significantly between languages (Blust 2000a), the assumption of a constant replacement rate can easily lead to distortion in the inferences reached from raw matrices of cognate percentages.

10.3.2.3 Quantity vs. quality

The use of exclusively shared innovations in subgrouping is usually conceived in the purely quantitative terms originally expressed by Brugmann (1884): the security of a subgrouping inference varies directly with the number of exclusively shared innovations that support it. Brugmann was not specific about what constitutes an adequate number of innovations to permit secure subgrouping inferences, but he is clear about relative quantity. Other things being equal, few linguists would quarrel with this position. However, other things are not always equal, since the quality of phonological innovations may vary markedly. The changes *h > zero, *s > h, and devoicing of final obstruents in a collection of languages constitutes weak evidence for a subgroup, since all three changes are common in global perspective, and could easily arise independently. By contrast, a single highly unusual innovation may have greater subgrouping value than a collection of more mundane changes.

To illustrate, a North Sarawak group that may include all languages of northern Borneo has been proposed based on a single change, namely the split of the PMP voiced obstruents *b, *d/j, *z, and *g Blust (1969, 1974b). This change has great subgrouping value for two reasons. First, the conditions for splitting are complex, and in some cases apparently irregular. Second, the novel segment produced by splitting differs from the 'normal' reflex in being phonetically complex or in devoicing in non-final position:

Table 10.12 Double reflexes of PMP voiced obstruents in North Sarawak languages

PMP		b	d/j	z	g
Bario Kelabit	(1)	b	d	d	g
	(2)	b^h	d^h	d^h	g^h
Bintulu	(1)	b	d	j	g
	(2)	ɓ	ɗ	j	g
Highland Kenyah	(1)	b	d	j	g
	(2)	p	t	c	k
Berawan	(1)	b/k/m	r/n	s	g
	(2)	p	c	c	k
Kiput	(1)	b/p	d/t	d/j	g/k
	(2)	s	s	s	k

Comparisons that illustrate parts of this table include reflexes of PMP *qabu 'ashes' and *tebuh 'sugarcane' (Bario Kelabit *abuh* : *təbʰuh*, Bintulu *avəw* : *təɓəw*, Highland Kenyah

abu : *təpu*, Berawan *akkuh* : *təpuh*, Kiput *abəw* : *təsəw*), PMP *ŋajan 'name' and *qapeju 'gall/gall bladder' (Bario Kelabit ŋ*adan* : *pədʰuh*, Bintulu *ñaran* : *lə-pədəw*, Highland Kenyah ŋ*adan* : *pətu*, Berawan (ŋ)*adan* : *pəcuh*, Kiput *adin* : *pəsəwʔ*), *quzan 'rain' and *haRezan 'notched log ladder' (Bario Kelabit *udan* : *ədʰan*, Bintulu *ujan* : *kəjan*, Highland Kenyah *udan* : *can*, Berawan *usin* : *acin*, Kiput (*pəraaʔ*) : *asin*.

As noted earlier, in addition to these splits Lowland Kenyah dialects such as Long San or Long Sela'an have a single phonetically imploded reflex of PMP voiced obstruent at labial, alveolar, palatal and velar positions. Implosion is only now beginning to emerge as phonemic in these languages due to the reduction of earlier clusters *-mb-, *-nd-, etc. to nonimplosive voiced stops, but the fact that all originally simple voiced stops are imploded adds to the already considerable evidence for an earlier series of phonetically complex consonants. What is methodologically noteworthy is that the North Sarawak subgroup was proposed on the basis of just one sound change—the innovation that produced series (2) reflexes. Although additional evidence could be used to bolster this hypothesis, clear evidence of phonemic splitting in cognate morphemes is sufficient in itself to rule out chance or drift as likely causes of the shared agreements. As will be seen below, a similar change is attested in the Begak dialect of Ida'an, spoken in eastern Sabah, where it has resulted in what appear to be synchronic consonant clusters (Goudswaard 2005).

10.3.2.4 *The linguistic value of the Wallace Line*

The AN world is bisected by a major zoogeographical boundary first identified by the English naturalist Alfred Russel Wallace in 1869. Perhaps nowhere else can an artifact of geological history provide clues to linguistic subgrouping, but the non-correspondence of linguistic subgroups and biotic zones makes this possible with the Wallace Line. Given the Formosan/Malayo-Polynesian split a number of terms for placental mammals can be assigned to PAN and PMP, including PAN *luCuŋ, PMP *lutuŋ, PMP *ayuŋ 'monkey'; PAN *takeC 'barking deer'; PAN/PMP *saladeŋ 'stag'; PMP *qaNuaŋ, PMP *qanuaŋ 'wild buffalo'; PAN *Sidi 'serow, Taiwan mountain goat'; PAN/PMP *babuy 'pig'; PAN/PMP *qaRem 'pangolin'; and PAN *buhet, PMP *buet 'squirrel' (Blust 1982b, 2002b). Since placental mammals are not native to areas east of the Wallace Line, PAN and PMP must have been spoken west of that boundary. In crossing the Wallace Line AN speakers would then have encountered marsupials for the first time. If this happened through separate migrations into eastern Indonesia and the western Pacific there would be no basis for expecting that the terms for marsupial mammals would be cognate, since they would have been independently invented (or borrowed) at different times and in different places. As it happens, however, many languages of eastern Indonesia and the Pacific reflect *kandoRa 'cuscus', and *mansar/manser 'bandicoot' which could not have been present in PAN or PMP.

Table 10.13 Cognate marsupial terms in eastern Indonesia and the Pacific

	*kandoRa 'cuscus'	*mansar/manser 'bandicoot'
Moluccas		
Leti-Moa		mada/made 'bandicoot'
Damar		madar (?)
Yamdena		mande 'cuscus'
Ngaibor (Aru)		medar 'marsupial sp.'
Ujir (Aru)		meday 'cuscus'
Elat (Kei)		mender 'cuscus'
Asilulu (Ambon)		marel 'cuscus'
Kamarian (Seram)		maker 'bandicoot'
Kei		medar 'cuscus'
Watubela	kadola	
Kesui (Keldor)	udora	
Geser-Goram	kidor	
Ambelau		mate 'cuscus'
Misool (Fofanlap)	do:	
Buli	do 'small marsupial'	
Melanesia		
Lou		mʷas
Sori	ohay (met.)	
Seimat	koxa (met.)	
Penchal	kotay (met.)	
Nauna	kocay (met.)	mʷac
Mussau	aroa	
Vitu	hadora	
Manam	ʔodora	
Motu		mada
Takia		madal
Wogeo		mʷaja 'cuscus'
Duke of York		man
Lungga	ɣandora	
Nggela	kandora	
Vitu	hadora	

Schapper (2011) has correctly pointed out that most eastern Indonesian reflexes of PEMP *mansar/mansər 'bandicoot' actually mean 'cuscus', a fact that was not correctly reported in this table as it appeared in Blust (2009a). However, this observation has no effect on the subgrouping implications of the cognate sets for marsupials (Blust 2012a). Since convergent innovation cannot plausibly explain this distribution it follows that the Central Malayo-Polynesian and South Halmahera-West New Guinea languages of eastern Indonesia subgroup immediately with the Oceanic languages in a larger Central-Eastern Malayo-Polynesian subgroup. There are additional types of evidence for CEMP, but the distribution of cognate sets for marsupial mammals must be considered the jewel in the crown for this hypothesis. This argument is complex in that it requires acceptance of an antecedent subgrouping condition (the MP hypothesis) which at first may seem remote

from the facts to be explained. But if terms for placental mammals are assigned to PAN and PMP, terms for marsupial mammals must be innovative, and since no basis for convergence is apparent it is simplest to attribute these innovations to a single speech community that was ancestral to the CMP, SHWNG and OC languages.

10.3.2.5 Sibilant assimilation

One other unusual type of evidence has been used for subgrouping the AN languages. The numerals 1-10 commonly attributed to PMP are 1. *esa/isa, 2. *duha, 3. *telu, 4. *epat, 5. *lima, 6. *enem, 7. *pitu, 8. *walu, 9. *siwa, 10. *sa-ŋa-puluq. By including Formosan evidence this system must be modified in several ways. First, *duha and *epat must be written *duSa and *Sepat. Second, the numeral ligature in *sa-ŋa-puluq cannot be reconstructed, hence PAN *sa-puluq. Third, and most critical, while nearly all non-Formosan languages point unambiguously to *siwa, all diagnostic Formosan languages point with equal clarity to *Siwa 'nine'.

To explain this discrepancy Dyen (1971a:34) split Dempwolff's *s into $*s_1$ and $*s_2$ and assigned the latter segment to the word for 'nine', hence $*s_2$iwa. Scholars such as Dahl (1981b) saw this approach as simply labeling the data, so the problem was revisited in Blust (1995b), where it was observed that at least three Formosan languages show synchronic or historical evidence of 'sibilant assimilation', as already described under 4.3.1.2 and 9.1.2. Briefly, several Formosan languages, including Paiwan, Saisiyat and Thao show historical evidence that sibilants within the same word tended to assimilate, a process that continues synchronically in Thao and affects at least *c* ([θ]), *s*, *sh* ([ʃ]), and *lh* ([ɬ]). Although such a set of consonants is larger than the usual category 'sibilant', and has never been recognized as a natural class, it seems clear that the kind of articulatory interference seen in these examples is essentially the same as the interference exploited in English tongue-twisters such as 'she sells sea shells by the sea shore'. It seems clear, then, that the onsets of successive numerals show interference effects in many languages if they are similar but non-identical. If its shape had been determined only by regular sound change English 'four', for example, would have been **whour, but since it is immediately followed by 'five' the expected sequence 'whour, five' became 'four, five'. In the case at hand the PAN word for 'ten' was *sa-puluq, with an unambiguous initial *s. If PAN 'nine' was *siwa, as reconstructed by Dempwolff, there would be no basis for explaining why Formosan languages reflect *Siwa. If the PAN numeral sequence was *Siwa 'nine' : *sa-puluq 'ten', however, a type of sibilant assimilation very similar to that observed in the history of Thao, Saisiyat and Paiwan could have converted *Siwa to *siwa. Since this is an innovation that is universal outside Taiwan it is simplest to assume that it happened once in the common ancestor of the non-Formosan languages.

10.3.3 Results of subgrouping

The preceding section draws attention to novel methods for subgrouping that have been used in AN linguistics. In doing this a few proposed subgroups have already been discussed. The purpose of this section is to provide an overview of the subgrouping of the AN languages in greater detail. Because many of the most serious controversies affect the highest levels of the classification it will be desirable to begin at the bottom of the tree and to work up from reasonably well-established lower-level subgroups to higher ones.

10.3.3.1 Polynesian

The Polynesian subgroup was recognized very early, and in the popular imagination it is commonly taken to be the best exemplar of the AN language family as a whole. As noted in earlier chapters, a Polynesian subgroup was recognized in more-or-less its full sweep during the Cook expeditions of 1768-1779. The challenges that were left for linguists to confront were 1. the internal subgrouping of the Polynesian languages, and 2. a determination of the closest external relatives of Polynesian. Although an important internal classification of the Polynesian languages was published by Elbert (1953:169), who recognized the Tongic : Nuclear Polynesian (called 'Samoan-Outlier-Eastern') split, the split of Eastern Polynesian into Rapanui vs. the rest, and of the latter into Marquesic and Tahitic groups (with Hawaiian, however, assigned to Tahitic rather than Marquesic), the most influential subgrouping of the Polynesian languages in the second half of the twentieth century was proposed by Pawley (1966, 1967):

Table 10.14 Subgrouping of the Polynesian languages (after Pawley 1966, 1967)

I.   Tongic
　　　A. Tongan, Niue
II.   Nuclear Polynesian
　　　A.   Samoic-Outlier (Samoan, East Uvean,
　　　　　East Futunan, Outliers)
　　　B.   Eastern Polynesian
　　　　　1.   Rapanui
　　　　　2.   Central-Eastern
　　　　　　　a.   Marquesic (Hawaiian, Marquesan, Mangarevan)
　　　　　　　b.   Tahitic (Tahitian, Tuamotuan, Rarotongan, Maori)

Like Elbert, Pawley broke with previous views that Tongan and Samoan form a unit, an idea that had arisen largely through confusion of culture areas and linguistic subgroups. As with Elbert's classification, Pawley's classification drew attention to the crucial role of Tongic in the reconstruction of Proto Polynesian, and of Rapanui to the reconstruction of Proto Eastern Polynesian. One clear break with Elbert's earlier classification involved the position of Hawaiian, which Pawley assigned to a Marquesic group, with an apparent closest connection to Southeast Marquesan. Although questions were raised about the standard classification of Polynesian languages from time to time (especially by Wilson 1985), it stood until the close of the twentieth century.

More recently Marck (2000) has revisited the subgrouping of the Polynesian languages in great detail, and proposed several fundamental revisions. While the primary division into Tongic and Nuclear Polynesian is preserved, the Samoic-Outlier group is demolished. In its place Nuclear Polynesian divides into eleven primary branches. Ten of these are represented by single Outlier languages: 1. Pukapuka (not mentioned by Pawley), 2. East Uvea, 3. East Futuna, 4. West Uvea, 5. Futuna-Aniwa, 6. Emae, 7. Mele-Fila, 8. Tikopian, 9. Anuta, and 10. Rennell-Bellona. The eleventh division is Ellicean, which is said to have three coordinate branches: 1. Samoan-Tokelauan, 2. the Ellicean Outliers (Kapingamarangi, Nukuoro, Sikaiana, Ontong Java, Takuu, Tuvaluan), and 3. Eastern Polynesian. The internal subgrouping of Eastern Polynesian is retained intact. There are a number of problems in demonstrating the existence of a subgroup that includes the Ellicean Outliers and Tuvaluan, and it is unlikely that these will be resolved in any very definitive way in the near future if at all. Marck's subgrouping nonetheless has the

advantage both of accounting better for a range of linguistic observations, and of helping to close what had become a widening gap between the archaeological record for Outlier Polynesia and the earlier taxonomy of Polynesian languages (as noted by Kirch 1997:60-61, the radiocarbon chronology for Outlier Polynesia includes dates of 2680 ± 90 BP for Tikopia, and 2830 ± 90 BP for Anuta in the Solomon Islands, suggesting a very early separation of some Outliers from other Polynesian languages).

Otsuka (2006) addressed an issue raised by earlier linguists, who had noted that despite its generally non-controversial classification as Tongic, Niuean shows some phonological and syntactic features "that are characteristic of Eastern Polynesian" (2006:429). Essentially three explanations for this distribution are available: 1. the features in question were found in PPN but lost in Tongan, 2. the features were acquired by Niuean speakers through contact with Eastern Polynesian languages, or 3. Niuean and Eastern Polynesian show parallel innovations. After carefully considering the evidence from syntax she concludes that the last of these explanations is the most plausible, and that Niuean and Eastern Polynesian have consequently come to resemble one another more than would otherwise be the case as a result of syntactic drift.

Most recently a re-evaluation of the radiocarbon chronology for eastern Polynesia by a team of Pacific archaeologists (Wilmshurst et al. 2011) has dramatically shortened the settlement time that can plausibly be allowed for any part of this large area. In this carefully conducted study the earliest reliable date is about 1025 AD for the Society group, with archaeological evidence that virtually all other habitable islands in eastern Polynesia were settled within the next 100-150 years. Since this shortened chronology leaves little time for the Tahitic-Marquesic distinction to develop, and no obvious site for a Central-Eastern Polynesian language community apart from the Societies, which is generally regarded as the Proto Eastern Polynesian homeland, linguists have been forced to rethink both the internal subgrouping of Eastern Polynesian languages and the dynamics of language split in a part of the world where travel across hundreds or even thousands of miles of open sea was a viable possibility. Walworth (to appear) summarizes the new conclusions: the evidence for a Tahitic/Marquesic split presented by Green (1966) and accepted by virtually all subsequent Polynesianists does not stand up well under close examination. Rather, it appears that all languages of eastern Polynesia diverged from a common source within roughly a century of one another, and that many of the innovations they share were developed in an inter-archipelagic community of seafarers who remained in contact with one another for the first six to eight generations after separation, when kin ties and oral traditions of common origin continued to preserve a sense of unity, despite the great distances involved.[117] Because of its geographical isolation, Rapanui was settled just once, with little or no further contact, thus giving rise to the impression (in a strict family tree model) that it was settled when all other Eastern Polynesian languages remained a single speech community. In the new subgrouping model all Eastern Polynesian languages apart from Rapanui remained a single speech community only in the sense that they continued to participate in this long-distance network of communication for a century or more after Rapanui had been cut off from it.

This model helps to reconcile the shortened radiocarbon chronology demanded by the archaeological record with comparative linguistic evidence, but it still leaves the question of the Proto Eastern Polynesian homeland unsettled. Eastern Polynesian languages are

---

[117]   Essentially the same idea was expressed by Pawley (1996b:403), who noted that "Proto Eastern Polynesian probably developed as a dialect chain extending from the Marquesas through the Society Islands to the Cooks (Green 1988 makes a similar proposal)."

defined by a number of distinctive innovations, and if all of eastern Polynesia was settled within 100-150 years after the first landfall in the Societies, this leaves far too little time for these innovations to accumulate *in situ*, leaving the linguist to search for some location further west in which the Proto Eastern Polynesian language community could have developed before relocating to the Societies. The tacit assumption of most scholars has been that the Societies were settled from the Samoan region (Pawley and K. Green 1971), but this leads to a quandary, since there are no islands in the region of Samoa that are reasonable candidates for the immediate relatives of Eastern Polynesian languages. In addressing this problem Wilson (2012) has argued persuasively that the immediate relatives of Eastern Polynesian are Takuu and Luangiua, which he calls the Central Northern Outlier (CNO) languages, with Sikaiana being slightly more distantly related in what he calls a 'Northern Outlier-Eastern Polynesian' (NO-EPN) subgroup. If valid, this proposal implies that the Societies were settled from an immediately antecedent location further north, and much further west than the Samoan homeland that has been traditionally assumed as the staging site for Proto Eastern Polynesian. Needless to say, it also raises other questions, such as how such tiny atoll communities as those of Takuu and Luangiua could power such a major population expansion, leading to the settlement of an area larger than the continental United States in less than two centuries.

10.3.3.2 Central Pacific

Grace (1959) argued that the closest relatives of the Polynesian languages are Fijian, spoken by a physically non-Polynesian population in the Fiji Archipelago, and Rotuman, spoken by a physically Polynesian or Polynesian-like population on a remote island group some 320 miles NNW of the western tip of Vanua Levu. Grace (1967) named this proposed subgroup 'Central Pacific'.

The proposed Central Pacific subgroup raised several questions. First, given the physical differences between the modern populations of triangle Polynesia and Rotuma as against Fiji, what was the physical type of the speakers of Proto Central Pacific? Second, as noted by Biggs (1965) Rotuman has at least two 'speech strata', one of which consists of a heavy layer of Polynesian loanwords. Has borrowing in any way obscured the primary genetic connections of Rotuman to the Polynesian languages? While it is widely agreed today that the answer to the second question is 'no', the first question touches on issues that have never been satisfactorily addressed. Although an archaeological signature for multiple settlements may be invisible, it is likely that Fiji was settled at least twice in rapid succession. The first settlement would have been accomplished by what Bellwood (1979) calls a 'southern Mongoloid' population (in this context ancestral Polynesians), moving out from Vanuatu soon after contact with that area. The second, and perhaps subsequent settlements would have been accomplished by a physically mixed population that resulted from several generations of interbreeding between southern Mongoloids and a clearly distinct 'Melanesian' population encountered either in Vanuatu or somewhere further west in the Pacific. The first wave of migration pushed on to Tonga, Samoa, and ultimately all of Polynesia, while the second remained in Fiji, continuing to mix with the southern Mongoloid population that remained *in situ*. Under this interpretation Rotuman presumably would be slightly more closely related to the Polynesian languages than to Fijian, since it would be a product of the first settlement of Fiji and neighboring regions.

The next major step in classifying Central Pacific languages was taken by Geraghty (1983), who showed that the evidence for a Central Pacific group is less robust than previously thought, depending on a single phonological innovation (loss of POC *R), four

shared grammatical morphemes, and two shared lexical items. The loss of *R occurs in other Oceanic languages, and there are plausible external cognates for some of the grammatical morphemes, further weakening this subgrouping hypothesis. Second, the forms that are uniquely shared by Fijian and Polynesian are almost entirely shared by Eastern Fijian and Polynesian exclusively of Western Fijian. This includes not only the innovations originally used to justify Central Pacific, but two additional phonological innovations and 113 lexical innovations, of which 27 involve functors. As noted already, for this reason Geraghty concluded that Eastern Fijian (called 'Tokalau Fijian') had originally been part of a dialect network that gave rise to the Polynesian languages, but was subsequently reincorporated into a greater Fijian dialect complex. Geraghty (1986) took up the issue of Central Pacific again, reiterating his doubts about the validity of the larger grouping, but at the same time proposing a 'Proto Central Pacific' sound system.

The Central Pacific hypothesis remains in a somewhat unsettled state. Lynch, Ross and Crowley (2002) accept it as a 'linkage', not a family, claiming in effect that it is not a discrete group. Geraghty's results are welcome in helping to confront the dilemma of determining the physical type of Proto Central Pacific speakers, since they suggest multiple settlements. The wider genetic connections of 'Central Pacific' languages, which will be addressed briefly below, remain very much in the realm of speculation.

10.3.3.3 Nuclear Micronesian

Unlike the languages of Polynesia, which form a well-defined genetic unit that was recognized as early as the eighteenth century, the languages of Micronesia are heterogeneous. They include a group of Oceanic languages that Bender (1971) called 'Nuclear Micronesian', Yapese, which is Oceanic, but does not subgroup closely with any other Oceanic language in Micronesia, and Palauan and Chamorro, which are only distantly related to the Oceanic languages and to one another. Again, in contrast with the Polynesian languages, the Nuclear Micronesian languages were poorly known until after the Second World War. While our knowledge of Polynesian languages has accumulated slowly over time through the efforts of a great many diverse scholars working in different parts of the world, our knowledge of Nuclear Micronesian languages owes almost everything to the concerted efforts of a small group of scholars at the University of Hawaii working on the Pacific Languages Development Project, or 'PALI Project', in the 1970s (Rehg 2004). Once adequate descriptions were available it became clear that most of the Oceanic languages of Micronesia form a well-defined genetic unit. The Nuclear Micronesian group is generally divided into the following branches (Bender et al 2003a, adapted from Jackson 1983):

Table 10.15 Subgrouping of the Nuclear Micronesian languages

1.	Kosraean
2.	Central Micronesian
2.1.	Gilbertese
2.2.	Western Micronesian
2.2.1.	Marshallese
2.2.2.	Proto Pohnpeic-Chuukic
2.2.2.1.	Pohnpeic (Pohnpei, Mokil, Ngatik, Pingelap)
2.2.2.2.	Chuukic (the Chuukic dialect continuum)

Missing from this diagram is Nauruan (and its geographically disconnected dialect Banaban), a language that clearly is Oceanic and that probably subgroups with the Nuclear Micronesian languages, perhaps as a primary branch in relation to everything else in the table, but which still is imperfectly known (Nathan 1973, Rensch 1993).

10.3.3.4 Southeast Solomonic

Pawley (1972) posited a Southeast Solomonic (SES) group with two primary branches: 1. Guadalcanal-Nggelic (GN), and 2. Cristobal-Malaitan (CM). The integrity of each of these groups is obvious on inspection, and their articulation into a larger SES group is supported by a considerable body of evidence. However, as noted by Blust (1984a), there is a countervailing body of evidence which suggests that the CM and Nuclear Micronesian languages shared a brief period of common history apart from all other AN languages. Since there is a much larger body of evidence supporting an immediate subgrouping connection between the CM and GN languages, the history of CM-Nuclear Micronesian may be similar to what Geraghty described for 'Tokalau Fijian': an early but brief period of exclusively shared innovations was later largely obliterated by the reintegration of CM into SES, just as Tokalau Fijian was reintegrated into the larger Fijian dialect complex.

10.3.3.5 North and Central Vanuatu

Tryon (1976) proposed a subgrouping of the languages of Vanuatu (at that time the 'New Hebrides'). Based on shared innovations in phonology he concluded (1976:51ff) that there is a 'New Hebrides' group with two primary subdivisions: 1. the languages of the Banks and Torres Islands in the far north, and 2. the rest**.** Within his residual group he recognized a 'Central New Hebrides' group with a number of subgroups (Malakula, Ambrym, Paama, Epi, Tanna). Based on lexicostatistical evidence he recognized a somewhat different classification, with six coordinate divisions: 1. North and Central New Hebrides (West Santo, Malakula Coastal, East New Hebrides, Central New Hebrides, Epi), 2. East Santo, 3. Malakula Interior, 4. Erromanga, 5. Tanna, and 6. Anejom.[118] Subsequent work by Lynch (1978b, 2001), Guy (1978) and Clark (1985) modified this classification in a number of ways, leading to a basic separation between the languages of the far south (Erromango, Tanna and Anejom) and those of the center and north.

10.3.3.6 The Southern Vanuatu Family

Since the nineteenth century the languages of southern Vanuatu have been considered aberrant, and their position was long in doubt. Kern (1906a) had to make a special effort to demonstrate that Anejom is AN, and although this issue was long settled by the middle of the twentieth century, the languages of southern Vanuatu were still regarded as among the most divergent and poorly described languages in the Pacific by Capell (1962:383). This situation began to change with the work of Lynch (1978b), who identified a number of exclusively shared innovations that define 'Proto South Hebridean,' including the languages of Erromanga, Tanna and Anejom. This early effort has been considerably refined and extended in Lynch (2001), where he calls the ancestral language 'Proto Southern Vanuatu'.

---

[118]   Tryon wrote 'Malekula' and 'Aneityum'. I follow current practice in converting these to 'Malakula' and 'Anejom'.

10.3.3.7 New Caledonia and the Loyalties

The languages of New Caledonia and the Loyalty Islands present some of the most recalcitrant comparative problems of any Oceanic languages, sharing low cognate percentages with other AN languages and even with each other, and having complex and often apparently irregular phonological histories (Grace 1990). Given these characteristics their position within Oceanic has not been easy to determine. Nonetheless, Geraghty (1989) has identified a number of likely innovations that collectively appear to be shared exclusively by these languages, and has called their hypothetical immediate common ancestor 'Proto Southern Oceanic'. Further evidence for this grouping is given by Ozanne-Rivierre (1986, 1992), and the proposal is accepted by Lynch, Ross and Crowley (2002:887), who suggest a larger grouping called the 'South Efate/Southern Melanesian Linkage', with two branches: 1. South Efate, and 2. the Southern Melanesian Family. The latter divides in turn into 2.1. the Southern Vanuatu Family, and 2.2. the New Caledonian Family, with a further division of the latter into the New Caledonian Family and the Loyalty Islands Family.

10.3.3.8   Wider groupings in the southern and eastern Pacific

Given the similarities of Polynesian and Micronesian cultures and physical types it has sometimes been suspected that the Polynesian and Nuclear Micronesian languages belong to a subgroup within the larger Oceanic group, but no evidence of an exclusive connection between these languages has ever been found. Rather, both groups have been assigned in some classifications to a larger collection of languages that includes them together with various languages of eastern Melanesia or eastern and southern Melanesia. One of the earliest of these proposals is that of Biggs (1965:383), who posited a subgroup called 'Eastern Oceanic', which was said to encompass 'Fijian, Polynesian, Rotuman and certain languages of the Solomons-New Hebrides chain, including probably Arosi of San Cristoval, Ulawa of Contrariété Island, Sa'a, Lau and Kwara'ae of Malaita, Nggela of Florida, Kerebuto and Vaturanga of Guadalcanal, Mota of the Banks Islands, and Efate of the New Hebrides.' This claim was elaborated by Pawley (1972), who added Nuclear Micronesian languages and tried to put the hypothesis on a firmer foundation by reconstructing part of the Proto Eastern Oceanic lexicon, including a number of grammatical morphemes. While some of these forms may prove to be exclusively shared, others have clear cognates outside the proposed Eastern Oceanic group.

Somewhat later Lynch and Tryon (1985) proposed a larger 'Central-Eastern Oceanic' group with three branches: 1. Eastern Oceanic (Southeast Solomonic, North and Central Vanuatu, Central Pacific, Micronesian), 2. Southern Vanuatu, and 3. Utupua and Vanikoro (southern Santa Cruz Islands). The evidence for this grouping is shaky, and in any case, as noted above, it has now been replaced by a 'South Efate/Southern Melanesian Linkage' ('linkage' indicating that it is not a discrete group).

10.3.3.9 The North New Guinea Cluster

Following a preliminary survey by Capell (1943), little detailed work was done on classifying the languages of western Melanesia until fairly recently. Grace (1955) had posited nineteen primary subgroups of Oceanic languages, fourteen of them in New Guinea, the Bismarck Archipelago and the western Solomons, and Dyen (1965a), who did not recognise Oceanic, had posited a large concentration of lexicostatistically-defined primary subgroups of the AN family in western Melanesia. This somewhat chaotic and

implausible situation changed abruptly with Ross (1988), whose work has revolutionized the comparative study of this large collection of AN languages. Within the area reaching from New Guinea to the western Solomons Ross recognized four 'clusters' of languages, which he called the 'North New Guinea Cluster', the 'Papuan Tip Cluster', the 'Meso-Melanesian Cluster', and the 'Admiralties Cluster', and one isolate or mini-cluster, the St. Matthias Group.

The North New Guinea Cluster (NNGC) "includes all the AN languages of the north coast of Papua New Guinea, the coast of the Huon Peninsula and the Huon Gulf with its hinterland (including the Markham Valley), and most of the island of New Britain west of (but not including) the Willaumez Peninsula, as well as its south coast as far as a point just beyond Jacquinot Bay … The cluster also includes the AN languages of all the offshore islands of the region except the French Islands to the north of western New Britain" (Ross 1988:120). Ross's NNGC includes three main branches: 1. the Schouten Chain, 2. the Huon Gulf Family, and 3. the Ngero/Vitiaz Family. Ross (1988:120, 183) indicates that he has considerably more confidence in the genetic unity of the Huon Gulf and Ngero/Vitiaz Families than he does in North New Guinea, which is defined by loose interlinkages between the constituent groups rather than by a set of exclusively shared innovations found throughout the proposed group. Stated differently, "Proto North New Guinea was not a single communalect which diffused or split but a linkage of communalects which, so to speak, became Proto North New Guinea at the time it became independent of other Oceanic communalects" (1988:120).

10.3.3.10   The Papuan Tip Cluster

The second large grouping of AN languages that Ross recognized in western Melanesia is the Papuan Tip Cluster (PTC). The PTC is better established than North New Guinea or Meso-Melanesian, and so was recognized early in the linguistics literature on this region. It includes all AN languages of southeast New Guinea and the Kilivila/Louisiades Network, including Goodenough, Ferguson and Normanby Islands, the Trobriands, Woodlark Island, and Misima, together with the many smaller islands that lie around or between these. There appears to be general agreement that this is a valid group, one that evidently originated in the islands to the southeast of the tail of New Guinea, and subsequently spread onto the mainland, reaching the Central Papuan area (Mekeo, Roro, Motu, Sinaugoro, etc.) by about 2,000 BP.

10.3.3.11   The Meso-Melanesian Cluster

Ross's third large group is the Meso-Melanesian Cluster (MMC). This is geographically the most extensive of the four groups, covering the whole of New Ireland (with New Hanover), portions of the north coast of New Britain (together with the French Islands), Nissan, and the Solomons chain from Buka through New Georgia and Santa Isabel. Ross (1988:261) compares the taxonomic status of the MMC with that of the North New Guinea Cluster, seeing both as linkages rather than as descendants of a linguistically uniform speech community. In his view early Oceanic speakers may have formed a chain or network of dialects stretching from the north coast of New Guinea to the western Solomons, with a significant break occurring around the Willaumez Peninsula of New Britain that created two extensive sub-linkages, the ancestral North New Guinea Linkage on the one side and the ancestral Meso-Melanesian Linkage on the other. Ross (1988:384, 2010) also points out that Southeast Solomonic lexical elements in the Meso-Melanesian

languages that now occupy the western Solomons suggest two layers of migration into the Solomons by AN speakers, a point also made on the basis of a somewhat different argument by Pawley (2009b). The first of these was by speakers ancestral to the Southeast Solomonic languages. The second was by speakers of Meso-Melanesian languages who overlaid the earlier population in the western Solomons, but did not extend beyond New Georgia and Santa Isabel. The result is a sharp linguistic break between the western Solomons and the rest of the island chain, but clear evidence of Southeast Solomonic substratum in the west.

10.3.3.12   The Admiralties Cluster

Like the Papuan Tip Cluster, the Admiralties Cluster (called the 'Admiralties Family' in Lynch, Ross and Crowley 2002) is a relatively discrete unit. Grace (1955) recognized two groups within this collection of languages as primary branches of Oceanic: Group 14, 'Admiralty and Western Islands' and Group 15, 'Wuvulu and Aua'. Blust (1978a:34) united these in a single Admiralty Islands group, but without supporting evidence. Evidence for an Admiralty Islands group was first presented by Ross (1988:315-345), and modified by Blust (1996f:31ff). Superficially, languages such as Wuvulu-Aua (dialects of a single language), Kaniet (probably two languages, now extinct) and Seimat appear to be rather different from the languages of Manus and its immediate satellites, but they share a number of lexical and morphological innovations that are not known in other AN languages. Replacement innovations in lexicon include POC *siku > PAdm *kusu 'elbow' (Wuvulu/Aua *utu*, Lindrou *kusuʔu-*, Titan *kusu-*), POC *qapaRa > PAdm *pose 'shoulder' (Wuvulu *foka*, Aua *fore*, Bipi *pose-*, Levei *pose/pʷese-*, Nali *pʷese-*), and POC *ikan > PAdm *nika 'fish' (Wuvulu, Aua *nia*, Bipi, Lindrou, Likum, Kuruti, Leipon, Nali, Loniu, Pak, Nauna *ni*, Lou, Baluan, Lenkau *nik*; but also Yapese *niig*). Morphological evidence includes reflexes of the numeral suffix *-pu in Kaniet *-fu*, Seimat *-hu*, Bipi, Lindrou, Likum, Kele, Lele, Kuruti, Leipon, Ere, Nali, Loniu, Pak, Nauna *-h*, Sori, Lou, Baluan, Penchal, Lenkau *-p*, and the sporadic reduplication seen in forms such as POC *panako > PAdm *papanako 'to steal' (Aua *fafanao*, Kuruti *pahna*, Ere *panna*, Nali *pahana*.

10.3.3.13   The St. Matthias Family

The St. Matthias Family contains just two languages, Mussau and Tenis or Tench (the latter almost completely unknown). Ross (1988) suggested that the Admiralties Cluster and St. Matthias Family may form a larger subgroup, but the evidence offered for this is not convincing, and the two are separated in Lynch, Ross and Crowley (2002:878).

10.3.3.14   Western Oceanic

At the conclusion of his pathbreaking study Ross (1988) tentatively proposed a larger grouping with the North New Guinea Cluster, Papuan Tip Cluster, and Meso-Melanesian Cluster as primary branches. He called this 'Western Oceanic'. In Lynch, Ross and Crowley (2002:879) this is described as the 'Western Oceanic Linkage', and the Sarmi/Jayapura Family of Irian Jaya is added to it.

10.3.3.15   Oceanic

Oceanic is the largest well-defined AN subgroup short of Malayo-Polynesian, with an estimated 466 languages (Lynch, Ross and Crowley 2002:878ff). It was first posited by

Dempwolff (1927, 1937) on the basis of exclusively shared phonological mergers, and further evidence for it was added by Milke (1958, 1961, 1968), Grace (1969), and Pawley (1973). Today it is accepted by virtually all historical linguists who work on the AN languages of the Pacific. The Oceanic subgroup is defined primarily by a number of phonological mergers, as shown in Table 10.16:

Table 10.16 Phonological evidence for an Oceanic subgroup

PMP	POC
*b/p	*p
*mb/mp	*b
*c/s/z/j	*s
*nc/ns/nz/nj	*j
*g/k	*k
*ŋg/ŋk	*g
*d/r	*r
*e/-aw	*o
*-i/uy/iw	*i

Some of these mergers occur in non-Oceanic languages, but the overall pattern is unknown west of the Mamberamo River in Irian Jaya, and some individual mergers, such as that of *b and *p, are extremely rare outside the Oceanic group.

The principal point of contention in the higher-level structure of the Oceanic subgroup is the position of the languages of the Admiralty Islands. Lynch, Ross and Crowley (2002) posit five primary branches of Oceanic: 1. the St. Matthias Family, 2. Yapese, 3. the Admiralties Family, 4. the Western Oceanic Linkage, and 5. the Central-Eastern Oceanic Grouping. Blust (1978a, 1998d) pointed out that the languages of the eastern Admiralties merge PMP *c, *s and *z like other Oceanic languages, but unlike other Oceanic languages merge *j with *d and *r. This distinctive difference in phonological history is most simply explained by a hypothesis that Oceanic contains two primary branches: 1. the Admiralties Family, and 2. the rest. Under this interpretation terminological issues arise: should 'Proto Oceanic' be reserved for the ancestor of all languages previously included in this category, or only for those that have merged PMP *j with *s? In either case, a different name will be required for the other group.

Finally, the position of Yapese within Oceanic remains unsettled. For many years it was unclear whether Yapese is Oceanic, like most other AN languages of the Pacific, or whether—like Palauan and Chamorro of western Micronesia—it is non-Oceanic. Ross (1996c) demonstrated the Oceanic affinity of Yapese beyond reasonable doubt, and suggested that it may subgroup distantly with the Admiralties Family. In Lynch, Ross and Crowley (2002), however Yapese is treated as a primary branch of Oceanic.

10.3.3.16   *South Halmahera-West New Guinea*

Dempwolff identified an Oceanic ('melanesisch') subgroup as early as 1927, but he was unclear about its western boundary, and this indeterminacy persisted for decades. Held (1942:7) suggested that the boundary between what he called 'Indonesian' and 'Melanesian' languages ran through the middle of Geelvink (now Cenderawasih) Bay, separating a West Geelvink Bay Group of 'Indonesian' languages (Numfor-Biak and Wandamen-Windesi), from an East Geelvink Bay Group of 'Melanesian' languages (the

languages of Yapen and Kurudu Islands, and Waropen). Grace (1955) placed the western boundary of Oceanic around the border between what was then Netherlands New Guinea and the Australian Trust Territory of New Guinea. Milke (1958:58), on the other hand, included Numfor within the Oceanic group. Dyen (1965c:304) pointed out that 'The merging of *p and *b suggests that nearly all of the languages east of Biak and Palau perhaps constitute a single subgroup, though not necessarily against the languages to the west taken as a group.' This issue was finally resolved in Blust (1978b), where evidence was given for a major linguistic boundary near the mouth of the Mamberamo River in Irian: east of this line the AN languages are Oceanic; west of the line they belong to a group that Esser (1938) called 'South Halmahera-West New Guinea' (SHWNG).

This group of about thirty languages, many known only from a comparative vocabulary of about 250 words in the preliminary survey of Anceaux (1961), spreads from southern Halmahera through the Raja Ampat Islands to Waropen, spoken along the east coast of Cenderawasih Bay as far as (and a bit beyond) the mouth of the Mamberamo River. It is defined by a number of phonological and lexical innovations. The evidence for SHWNG holds special interest in that certain sound changes shared by some South Halmahera languages such as Buli, and by some West New Guinea languages such as Numfor, occur in opposite orders. This can be seen in Table 10.17:

Table 10.17 Shared sound changes that differ in order in Buli and Numfor

PMP	Buli	Numfor	Taba	Munggui	
*kutu	ut	uk	kut	utu	louse
*telu	tol	kor	p-tol	bo-toru	three
*susu	sus	sus	susu-	susi	breast
*m-atay	mat	—	-mot	—	to die
*manuk	mani	man	manik	—	bird
*Rebek	opa	rob	-opa	yoba	to fly
*Rusuk	usi	—	—	—	rib
*hajek	—	yas	—	—	sniff, kiss
*paniki	fni	—	nhik (met.)	—	flying fox

The material in Table 10.17 shows that both Buli and Numfor 1) lost *k and 2) final vowels, but in different orders (1,2 in Numfor, 2,1 in Buli). Buli is now located some 800 km west of Numfor. Languages to the west of Buli, such as Taba (= East Makian), show change 2), but not change 1). Languages to the east of Numfor, such as Munggui, show change 1), but not change 2). Stated more generally, change 1) covers SHWNG languages from the easternmost part of their territory through Buli in eastern Halmahera, but misses those further west, while change 2) covers SHWNG languages from the westernmost part of their territory through Numfor, but misses those further east. In more central areas both changes occur, but in opposite orders. This is thus a classic example of sound change propagated by bidirectional diffusion at similar times and rates, with overlap in the central zone. Since change 1) began in the east it reached Numfor before it reached Buli, and since change 2) began in the west it reached Buli before it reached Numfor. Sound changes rarely spread among distinct languages, and this overlap thus suggests that SHWNG began as a dialect chain in which innovations diffused between then-adjacent communities that evolved over time into a chain of mutually unintelligible languages, with increasing geographical separation.

10.3.3.17   *Eastern Malayo-Polynesian*

The case for EMP is based mainly on 56 proposed lexical innovations (Blust 1978b). Some of these have since been found to have cognates in CMP languages, but others are clear replacement innovations, and so carry considerable subgrouping weight. Among the more significant are: 1) PMP *anak > PEMP *natu 'child' (Buli *ntu*, Waropen *ku*, *ku-ku*, POC *natu), 2) PMP *bahuq > PEMP *boi/bui 'smell, stench' (Buli *pu-pui* 'stench, stink', Fijian *boi* 'have a smell'), 3) PMP *nunuk > PEMP *qayawan 'banyan, strangler fig' (Buli *yawan*, POC *qayawan; a reflex of *nunuk also was retained in POC), 4) PEMP *ka(dR)a 'cockatoo; parrot' (Ansus, Ambai *kara* 'cockatoo', Roviana *kara* 'the general name for parrots'; no PMP term), 5) PEMP *sakaRu 'reef' (Buli *sa*, POC *sakaRu; no PMP term) and the irregular change in *besuR > PEMP *masuR/mosuR 'satiated, full after eating' (Buli *mose*, Windesi *mos*, Wuvulu *maku*, Aua *maru*, Label *masur*, Raluana *maur*, Fijian *macu*). EMP is now widely accepted, but much basic documentation remains to be done on its member languages, particularly those in the SHWNG branch.

10.3.3.18   *Central Malayo-Polynesian*

Central Malayo-Polynesian (CMP) includes some 120 languages in eastern Indonesia, from Bimanese in eastern Sumbawa through the Lesser Sundas to the southern and Central Moluccas**.** The internal cohesion of this proposed group is less clearly marked than SHWNG. As noted in Blust (1993b), CMP languages share a number of phonological, lexical and morphosyntactic innovations, that overlap without covering all members of the group. In the terms employed by Ross (1988, 1997) it is thus best seen as a 'linkage'.

A number of smaller genetic units have been identified within CMP. Following the pioneering work of Stresemann (1927), Collins (1982, 1983a) recognized a large Central Maluku (CM) group of languages centered on the island of Seram. Central Maluku divides into West CM (languages of Buru, the Sula Archipelago, and Ambelau Island), and East CM (languages of Seram, Ambon, the Seran Laut Islands, and Banda, plus Kayeli on Buru). Taber (1993) has presented a lexicostatistical classification of the languages of southwest Maluku, recognizing Wetar, Kisar-Roma, Luang, East Damar, Teun-Nila-Serua and Babar groups, and a West Damar isolate. It seems likely that Kisar-Roma and Luang share a common node apart from other languages of the southwest Moluccas. In addition, Blust (1993b:276ff) presented evidence for a geographically discontinuous Yamdena-North Bomberai subgroup within CMP that includes at least Yamdena of the Tanimbar Archipelago in the southern Moluccas, together with Sekar, Onin, Uruangnirin and perhaps some other languages spoken on the northern side of the Bomberai Peninsula of New Guinea, located some 500 km. to the northeast.

Despite this progress much work remains to be done on the internal subgrouping of CMP languages. Most larger groupings that try to reach beyond individual islands remain elusive. The 'Bima-Sumba' group proposed by Esser (1938), for example, is illusory, and the linguistic position of Bimanese remains very much an unsettled matter (Blust 2008a). The subgrouping of the other languages of the Lesser Sundas is sketchy, although it is generally assumed that there are Timor and Flores groups, the former including all AN languages of Timor plus Rotinese, and the latter including at least the languages of western and central Flores. The languages of Sumba are closely related and form a noncontroversial genetic unit that groups with Hawu-Dhao. Questions still remain concerning the position of other languages, including West Damar in the eastern Lesser Sundas, and the languages of the Aru Islands.

10.3.3.19   *Central-Eastern Malayo-Polynesian*

Central-Eastern Malayo-Polynesian (CEMP) includes all CMP and EMP languages, hence all AN languages of eastern Indonesia and the Pacific except Palauan and Chamorro of western Micronesia. Some of the evidence for this large grouping was presented earlier. In addition to the unique perspective provided by cognate marsupial terms (Table 10.13), many languages distributed over the same subgroups show sporadic lowering of high vowels in cognate morphemes, as in PMP *uliq > PCEMP *oliq 'return, go back', PMP *tudan > PCEMP *todan 'sit', or PMP *ma-qitem > PCEMP *ma-qetəm 'black', assimilation of *aCi > eCi, as in PMP *i-sai > PCEMP *i-sei 'who?' or PMP *kali > PCEMP *keli 'dig up tubers', replacement innovations in lexicon, as with PMP *dilaq but PCEMP *maya 'tongue', or PMP *qabaŋ but PCEMP *waŋka 'canoe', as well the apparent innovation of PCEMP *kanzupay next to the inherited *ka-labaw 'rat', and of PCEMP *keRa(ŋŋ) 'hawksbill turtle' (Blust 1983/84a). Given the high quality of evidence for CEMP it probably is safe to say that this subgroup is as well-established as any large AN subgroup apart from Oceanic.

Table 10.18 Major AN subgroups in eastern Indonesia and the Pacific

No.	Group	Source	Type
1.	Central-Eastern MP	Blust 1983/84a, 1993b	F
2.	Central MP	Blust 1983/84a, 1993b	L
3.	Central Maluku	Stresemann 1927, Collins 1982	F?
4.	Yamdena-North Bomberai	Blust 1993b	F?
5.	Eastern MP	Blust 1978b	F
6.	South Halmahera-W. New Guinea	Esser (1938), Blust 1978b	F
7.	Oceanic	Dempwolff 1927, 1937	F
8.	Western Oceanic	Ross 1988	L
9.	St. Matthias Family	Ross 1988	F
10.	Admiralties Family	Blust 1978a, Ross 1988	F
11.	Meso-Melanesian Cluster	Ross 1988	L
12.	Papuan Tip Cluster	Ross 1988	L
13.	North New Guinea Cluster	Ross 1988	L
14.	New Caledonian Family[119]	Haudricourt 1971	F
15.	Southern Vanuatu Family	Lynch 1978b	F
16.	Northern and Central Vanuatu	Grace 1955	L
17.	Southeast Solomonic	Grace 1955	F?
18.	Nuclear Micronesian	Grace 1955	F
19.	Central Pacific	Grace 1955	F?
20.	Polynesian	Förster 1778	F

Table 10.18 summarizes the subgroups discussed so far, showing where each was proposed, and classifying it as follows: F = family: any group that is highly discrete—that is, defined by a large collection of exclusively shared innovations, L = linkage: a collection of languages united by overlapping innovations, R = residue: the languages left after subgrouping others. I include only groups that are widely accepted, and for which

---

[119]   Grace (1955) recognized separate New Caledonia and Loyalty Island subgroups. Haudricourt (1971) appears to have been the first to explicitly claim that all languages of this region descend from an immediate common ancestor, for which he proposed a number of lexical reconstructions.

supporting evidence is given either in the primary source or in later publications. Hence, Central-Eastern Oceanic and Eastern Oceanic are omitted on the grounds that they have rarely been mentioned in the literature since they were proposed, and several groups proposed without supporting evidence in Grace (1955) are included as they have since been justified on the basis of exclusively shared innovations. Where Ross (1988) and Lynch, Ross and Crowley (2002) differ in subgroup name or type I follow the later work.

Pawley and Ross (1993:437) describe CEMP and CMP as "much more problematic" than EMP, SHWNG and OC, in the latter case presumably because the group does not appear to be discrete. However, the overlapping nature of the isoglosses that define CMP was noted explicitly by Blust (1993b:263, 270), who observed that the interconnections of these languages fits the description of what Ross (1988), in subgrouping the AN languages of western Melanesia, called a 'linkage'. CMP is thus defined in much the same way as Ross's Western Oceanic, Meso-Melanesian Cluster and North New Guinea Cluster. However, no reason is given for the reservations expressed about CEMP, which is by any normal method of subgrouping, a discrete unit.

Adelaar (2005b:24-26), on the other hand, expresses doubts about CEMP, CMP and even EMP. He questions the morphosyntactic evidence for CEMP, which is marginal to the argument, but ignores the subgrouping implications of the marsupial terms, the clear innovations in lexicon and the multiple examples of lexically-specific lowering of *i and *u noted above. With regard to EMP he raises no specific objections, but claims (2005b:26) that "Ross (1995) also doubts the evidence for EMP." Ross (1995c:84-85), however, notes that the evidence for EMP consists of 56 putative lexical innovations, and that 'There are no convincing phonological innovations shared in common by the two member groups,' simply rephrasing the statement in Blust (1978b) that the only clear evidence for EMP is lexical. The grounds for his questioning these groups are thus unclear. Similarly, if CMP is dismissed on the grounds that it is a linkage rather than a discrete group most proposed Oceanic subgroups short of Polynesian and POC itself would have to be abandoned on similar grounds.

A more concerted attack on both CMP and CEMP appears in Donohue and Grimes (2008), an article that has been cited by some writers (e.g. Spriggs 2011) as though it is a devastating critique, when in fact the argument is methodologically flawed in a number of respects (Blust 2009c).

Despite the exemplary discussion in Ross (1997) the classification of subgroups into types is not always clear. As noted in Blust (1978b), South Halmahera-West New Guinea appears to have evolved from a dialect network in which innovations were able to diffuse in opposite directions so as to be acquired in different orders in some languages. The SHWNG group nonetheless appears to be discrete in the sense that there is little question as to whether a given language belongs to it. In other words, the immediate ancestor of this group evidently evolved in sufficient isolation to acquire a distinct identity, and subsequently developed a dialect chain along which innovations spread. For this reason I have classified it as a 'family'. Northern and central Vanuatu in Grace (1955) seems to correspond roughly to the 'Northern Vanuatu Linkage' (Torres and Banks Islands, Santo, Ambae-Maewo) in Lynch, Ross and Crowley (2002), but was somewhat more inclusive. Southeast Solomonic has many characteristics of a discrete group, but as noted in Blust (1984a) there is some evidence that the Cristobal-Malaitan languages were once part of a dialect network with Nuclear Micronesian, thus raising questions as to whether the larger Southeast Solomonic group (which also includes Guadalcanal-Nggelic) should be considered a 'family'. Finally, as Geraghty (1983, 1986) has noted, efforts to assemble a

body of exclusively shared innovations that define a 'Central Pacific' group have met with considerable frustration, although it is widely assumed that this collection of languages forms a valid subgroup.

This survey of major AN subgroups has included everything from CEMP to Polynesian so far (hence eastern Indonesia and the Pacific apart from Palauan and Chamorro). It will now continue with lower-level subgroups in western Indonesia, the Philippines and Taiwan, linking these into more inclusive groupings up to PAN.

10.3.3.20   *Celebic*

Sneddon (1993) recognized ten Sulawesian microgroups, small collections of languages that form uncontroversial genetic units. These are 1) Sangiric: four languages in the Sangir and Talaud Islands and on the northern peninsula of Sulawesi, 2) Minahasan: five languages on the northern peninsula, 3) Gorontalo-Mongondowic: nine languages on the northern peninsula, 4) Tomini-Tolitoli: eleven languages in the neck of the northern peninsula between Palu and Gorontalo province (Himmelmann 2001), 5) Saluan: four languages in the eastern peninsula of Sulawesi and on the Togian Islands in Tomini Gulf, 6) Banggai: one language in the Banggai Islands off the eastern tip of the central Sulawesi peninsula, 7) Kaili-Pamona: five or more languages (depending on how the language/dialect distinction is drawn) in the mountain massif of central Sulawesi, 8) Bungku-Tolaki: 15 languages in the southeastern peninsula (Mead 1999), 9) Muna-Buton: seven languages on islands to the southeast of Sulawesi (van den Berg 1991a), 10) South Sulawesi: 27 languages in the southwestern peninsula of Sulawesi and on the island of Salayar. In addition he regarded Lemolang, Wotu and Balaesang as isolates, and Manado Malay as a non-indigenous introduction to Sulawesi. More recently Himmelmann (2001) has included Balaesang in the Tomini-Tolitoli group, and raised questions about whether Totoli and Boano belong in it. Donohue (2004) has questioned whether Tukang Besi belongs in the Muna-Buton group proposed by van den Berg (1991a,b), but van den Berg has presented a strong case for its inclusion (2003).

The most important subgrouping proposals for the languages of Sulawesi in recent years are those of van den Berg (1996b), who combined Sneddon's groups 7, 8 and 9 into a single 'Celebic' group, and Mead (2003b), who extended the proposed Celebic group to include all of Sneddon's groups 4-9 (counting Banggai and Saluan as one) plus Lemolang, Wotu and Balaesang. This proposal has important implications regarding the settlement history of this large island. Since both variants of the Celebic hypothesis exclude the extensive South Sulawesi group, it appears that the southwestern peninsula of the island and most of the center, east and southeast have had rather different linguistic histories. Given the greater territorial extent of Celebic the South Sulawesi languages may represent the holdouts of an earlier AN-speaking population that was largely replaced by the expansion of Proto Celebic in other parts of the island south of the northern peninsula.

10.3.3.21   *Greater South Sulawesi*

Following an earlier suggestion by the anthropologist Alfred Hudson, Adelaar (1994a) has argued that the Tamanic languages (Embaloh, Taman, Kalis) of the upper Kapuas River in West Kalimantan belong to the South Sulawesi group, with closest ties to Buginese. This is surprising, since these languages are located far inland, and the proposed subgrouping relationship implies a migration of some antiquity from south Sulawesi, with a movement up the Kapuas River past many other linguistic groups that must have been *in*

situ at the time. The specific association with Buginese, moreover, places the time of the migration after the separation of Buginese from other South Sulawesi languages, a time period that could hardly have been longer than 1,500-2,000 years ago.

10.3.3.22 *(Greater) Barito*

Although the relationship of Malagasy to Malay (and hence other AN languages) was recognized as early as 1603 by de Houtman, the position of Malagasy within the AN language family only began to attract interest in the first half of the nineteenth century. On the basis of typological similarity in the verb system von Humboldt (1836-1839) concluded that Malagasy belongs with the Philippine languages, and three decades later H.N. van der Tuuk (1865) proposed a grouping of Malagasy and the Batak languages of northern Sumatra. However, these efforts to show where this geographically displaced AN language fits were premature, and ultimately unsuccessful.

Dahl (1951) solved a longstanding mystery when he was able to show that Malagasy is most closely related to Ma'anyan of southeast Kalimantan. The material available at the time did not allow him to identify a larger grouping to which both of these languages belong. This was done by Hudson (1967), who coined the term 'Barito Family' for a collection of 5-15 languages (called 'isolects'), spoken in the basin of the Barito River of southeast Borneo. He divided the Barito Family into three coordinate branches: 1. Barito-Mahakam, with a single language (Tunjung), 2. West Barito (the groups traditionally known as 'Ot Danum' and 'Ngaju Dayak'), and 3. East Barito (Taboyan, Lawangan, Dusun Deyah, Dusun Malang, Dusun Witu, Paku, Ma'anyan and Samihim). Within the Barito Family Malagasy appears to group most closely with the East Barito languages, which include Ma'anyan. More recently Blust (2005c) has shown that the Sama-Bajaw languages of the Philippines have either borrowed from Barito languages at least a millennium ago, or belong to a 'Greater Barito' grouping that includes Sama-Bajaw and Barito as primary branches.

10.3.3.23 *Malayo-Chamic and beyond*

The close relationship of the Chamic languages to Acehnese of northern Sumatra was first recognized by Niemann (1891). Periodic remarks in the literature, as by Marrison (1975) also suggested a prior Chamic-Malay unity. These observations were brought together in a more systematic way in Blust (1981b, 1994a), where a 'Malayo-Chamic' group was proposed, and some lexical items reconstructed. More recently Adelaar (2005c) has proposed a larger grouping called 'Malayo-Sumbawan' with three primary branches 1. Madurese, 2. Sundanese, and 3. Proto Malayo-Chamic-BSS (where BSS = Balinese, Sasak, Sumbawanese). In his view this third branch divides in turn into three branches: 1. Proto BSS, 2. Chamic, and 3. Malayic. Several of these languages are comparatively well-studied, and Adelaar suggests that this group might have been recognized earlier if longstanding Javanese cultural and linguistic influence on Balinese had not obscured the closer relationship of this language to Malay. The evidence for Malayo-Sumbawan is suggestive, but falls short of being fully convincing, and the deconstruction of a Malayo-Chamic subgroup leaves certain types of evidence recognized in Blust (1994a) unexplained. For this reason it is rejected in Blust (2010).

10.3.3.24   *Barrier Islands-North Sumatra*

Nothofer (1986) presents evidence that the languages of the Barrier Islands subgroup with the Batak languages of northern Sumatra. Shared phonological innovations are limited. The strongest of these is perhaps the reflex of PAN *j as a velar stop or fricative, but a similar change occurs in some dialects of Rejang in south Sumatra, and in other AN languages. In addition, however, Nothofer lists six phonological irregularities shared by Barrier Island languages, and three others shared by Barrier Island languages with the Batak languages. Not all of these are found throughout the proposed subgroup, but a good case is made for linking Simalur, Sichule, Nias and Mentawai with one another and the Batak languages. Sixty nine lexical innovations are proposed in support of a Barrier Islands group, and sixty six others in support of a Barrier Islands-Batak group. Nothofer places Sichule-Nias under the lowest node, these plus Mentawai under the next node, then Simalur, Enggano, and the Batak languages under three progressively higher nodes, with a question mark after Enggano. However, he also includes some data from Gayō in his proposed lexical innovations without explicitly indicating that this language belongs in the proposed group. Tentatively, then, I will call this the 'Barrier Islands-North Sumatra' group. The greatest surprise in this proposal is the evidence for including Enggano, one of the most aberrant of all AN languages, as part of the Barrier Islands-Batak group. If valid it would follow that this language, for whatever reasons, has undergone extraordinarily rapid change on all levels (phonological, lexical, morphological, syntactic).

10.3.3.25   *North Sarawak*

Some evidence was given earlier for a North Sarawak subgroup that includes at least four major subdivisions: 1. Kelabitic (Kelabit, Tring, Lun Bawang/Lun Dayeh, Sa'ban), 2. Kenyah (a number of fairly closely related language communities in Sarawak and Kalimantan, including Sarawak Penan), 3. Berawan-Lower Baram (four dialects of Berawan, together with Kiput, Narum, Belait, Miri, Dali', Lelak and Lemeting), 4. Bintulu. The critical evidence for this group, presented in Blust (1969, 1974b) is the split of PMP voiced obstruents *b, *d/j, *z and *g into two series, one of which is unmarked and the other marked, or phonetically complex. Lexical evidence suggests that some other languages of central and western Borneo, as Melanau and Kayan, also belong to this group, but the phonological evidence is ambiguous, and the possible inclusion of these languages in North Sarawak remains to be clarified.

10.3.3.26   *North Borneo*

Although the evidence presented to date is limited and preliminary, there are indications of a wider 'North Borneo' group that includes North Sarawak and the indigenous languages of Sabah (Blust 1998e). One of the clearest indications of a likely North Borneo subgroup is the presence of phonetically complex stops or clusters, *bp*, *dt*, *gk* that have historically resulted from the split of plain voiced stops in both North Sarawak languages and members of the 'Ida'an Subfamily' of eastern Sabah (King and King 1984:9ff). The facts, however, are complicated by the clearly secondary nature of these stops in some lexical items in Ida'an. Table 10.19 gives complex stop reflexes of PMP *b, *d/j, *z and *g in the Bario dialect of Kelabit in northern Sarawak, and from the Begak dialect of Ida'an in eastern Sabah as described by Goudswaard (2005):

Table 10.19 Complex stops in Bario Kelabit and Ida'an Begak

PMP	Bario Kelabit	Ida'an Begak	
*baqbaq		babpaʔ	mouth
*baqeRu	bəruh	bagku	new; again
*beRas	bəra	bəgkas	husked rice
*beRat	bərat	bəgkat	heavy
*bahuq	bəw-an	bpow	smell; stink
*buhek	əbʰuk	bpuk	head hair
*ma-buhek	mabuk	bpuk	dizzy; drunk
*qalejaw	ədʰo	dtow	day; sun
*Rabiqi		gabpi	night
		gobpi	afternoon
*Rebaq		gəbpaʔ	collapse
*haRezan	ədʰan	gədtan	ladder
*peRaq		gkaʔ	squeeze out
*beRay		gkay	give
*heReŋ	ərəŋ	gkaŋ	roar, shout
*eguŋ		gkuŋ	large gong
*hizam		idtam	borrow
*qijuŋ	idʰuŋ	iruŋ	nose
*kiskis		kigkis	scrape clean
*lebuR		ləbpog	muddy water
*lebeŋ		ləbpoŋ	grave
*hapejes		pədtos	painful; ill
*sijem		sidtom	ant
*tebek	təbʰək	təbpok	pierce; inject
*tebeŋ	təbʰəŋ	təbpoŋ	fell a tree
*tebuh	təbʰuh	təbpu	sugarcane
*teduŋ		tədtuŋ	cover the head
*teRas		təgkas	hard wood
*teguk	təguk 'throat'	təgkuk	gulp; drink

Certain clarifications are needed in order to understand this material. First, stress in Kelabit is penultimate regardless of the quality of the vowel, and remains penultimate in the word under suffixation as a result of rightward stress shift, as in *taban* ([tában]) 'kidnapping, elopement' : *təban-ən* ([təbánən]) 'be kidnapped, be eloped with'. Second, as noted in Blust (1974a, 1993a, 2006a) the Kelabit segments *bʰ*, *dʰ*, *gʰ* are true voiced aspirates as defined by Ladefoged (1971:9): they begin voiced, end voiceless, and the voiceless terminus of the consonant may carry into the onset of a following vowel. Third, these are unit phonemes, since 1. if they were analyzed as clusters they would be the only tautomorphemic consonant clusters, 2. *bʰ*, *dʰ*, *gʰ* alternate with *b*, *d*, *g*, being almost completely predictable following a stressed schwa, as in [kə́dtʰa] 'able to withstand pain' : [kədáan] 'to suffer', or [ə́l:əg] 'cessation; divorce' : [lə́gkʰən] 'be divorced by someone', 3. at least for some speakers *bʰ*, *dʰ*, *gʰ* are pronounced with noticeable aspiration, to the point that *dʰ* may be affricated as [dʧʰ]; by contrast *p*, *t* and *k* are always unaspirated, and 4. high vowels have lowered or laxed allophones in closed syllables, but preceding a voiced aspirate there is no lowering or laxing, showing that the voiced portion of the voiced aspirates is not the coda of a preceding syllable (cf. *idʰuŋ* ([ʔídtʰʊŋ]) 'nose', where the last-

syllable vowel lowers, but the first-syllable vowel does not). By contrast, Ida'an Begak (IB) allows unambiguous consonant clusters (*ndow* 'child ghost', *agbod* 'small rice basket', *ləkpud* 'broken'), and has no suffixes, thus removing the possibility of determining whether *-bp-*, *-dt-* and *-gk-* alternate with their plain voiced counterparts under suffixation. None of the reasons for analyzing Kelabit *b^h*, *d^h*, *g^h* as unit phonemes are known to be present in IB, and it thus appears correct to treat *bp*, *dt* and *gk* as consonant clusters, even though it is clear that they originated from single voiced stops, some of which reflect earlier non-obstruent consonants (as *R).

There is an interesting implication in the above data. Where Kelabit has a voiced aspirate Ida'an usually has a homorganic voiced-voiceless cluster (six of seven cases, with no exceptions following schwa). The same relationship sometimes holds where Kelabit has no cognate but other North Sarawak languages indicate a PNS voiced aspirate (PMP *baqbaq > Long Anap Kenyah *paʔ*, Long Dunin Kenyah *ɓaʔ*, IB *babpaʔ* 'mouth'), or where a PMP etymon is unknown but North Sarawak languages reflect a voiced aspirate, as with Kelabit *r-əd^{h}an*, Bintulu *məɗan*, Long Anap Kenyah *mətan* : IB *m-adtan* 'fainted; passed out'. The reverse implication, however, is often violated, since IB developed *gk* following a stressed schwa after the sound change *R > g (*bəRas > [bə́ggas] > *bəgkas*, etc.). This might be interpreted as showing that the changes in NS and IB are independent, but this seems unlikely given the rarity of complex consonants or consonant clusters evolving from single voiced obstruents, and the consistent pattern whereby NS voiced aspirates imply IB voiced-voiceless clusters. The most likely scenario, then, is that *b^h, *d^h, *g^h were found in the immediate common ancestor of the North Sarawak and Sabahan languages. Since pre-IB did not allow plain voiced stops after a stressed schwa, the change *R > *g* was accommodated to the existing pattern in this environment (transcriptions such as *babpaʔ*, *bagku* or *gabpi* are suspect, and may contain penultimate schwas). The reconstruction of *b^h, *d^h, *g^h for a language ancestral both to North Sarawak languages and to Sabahan languages is further supported by Sabahan languages that show a corresponding split of earlier *b, *d/j, *z and *g, but lack a complex consonant series: PMP *qabu 'ashes', *tebuh 'sugarcane' > PNB *abu, *təb^{h}u > Kadazan *avu, toɓu*, Tombonuwo *awu, tobu*, PMP *bulan 'moon', *buhek 'head hair' > PNB *bulan, *əb^{h}uk > Kadazan *vuhan, toɓuk*, Tombonuwo *wulan, obuk*.[120]

10.3.3.27   *Greater North Borneo*

Blust (2010) shows there is a growing body of lexical evidence for an even larger 'Greater North Borneo' group that includes all languages of Borneo except Greater Barito. For some comparisons it is difficult to determine whether a form in a Bornean language is native or a Malay loan, but for others borrowing can safely be ruled out. The most striking lexical innovation defining this group probably is *tuzuq, replacing PAN *pitu 'seven', but others with a similar distribution can be cited, including *tikus 'rat, mouse', *labi 'soft-shelled turtle', *lamin 'room of a house', and *sakay 'visitor; stranger'. Other lexical innovations have a somewhat wider distribution that covers all languages of western Indonesia apart from Sulawesi, as *baRuaŋ 'the Malayan sun bear', *beduk 'coconut

---

[120]   Kadazan has homorganic voiced-voiceless consonant clusters, as in *abpai* 'cross legs, place legs on or across someone when sitting on the floor', *bodtuŋ* 'calf of the leg', or *higkaŋ* 'healthy and strong', but these sequences are confined almost exclusively to bases with no known etymology, and their source consequently remains obscure.

macaque', *duRian 'a tropical fruit, the durian', *pilanuk 'the mousedeer: Tragulus kanchil', *pulaŋ 'return to one's starting place', and *suŋay 'river'.

10.3.3.28 Philippines

The existence of a Philippine group of AN languages was questioned by Reid (1982), and more recently by Ross (2005b), but has been defended by Zorc (1986), and Blust (1991a, 2005c). As noted by Charles (1974), there is little phonological evidence for this group. Moreover, most Philippine languages are morphosyntactically conservative, further eroding the possibilities of finding useful innovations. There is, however, a substantial body of lexical innovations shared exclusively by these languages. Blust (2005c) presents a list of 241 proposed lexical innovations that appear to be restricted to Philippine languages, some of them clear replacement innovations. Together with material presented by Zorc (1986) this list can be extended to 327 forms. Since the 241 proposed lexical innovations are extracted from (Blust and Trussel ongoing), which is currently only about 33% complete, this suggests that fuller searching will turn up over 700 lexical innovations that define a Philippine subgroup.

The Philippine subgroup can be divided into 15 non-controversial microgroups: 1) Bashiic (Yami, Itbayaten, Ivatan), 2) Cordilleran (most languages of the Cordillera Central in northern Luzon, plus a few lowland languages such as Ilokano and Ibanag), 3) Central Luzon (Botolan Sambal, Tina Sambal, Bolinao, Kapampangan, various languages spoken by small Negrito populations, and possibly North Mangyan of Mindoro), 4) Inati (an isolate within the Philippine group, spoken by a small population of Negritos in the mountains of Panay), 5) Kalamian (Kalamian Tagbanwa and Agutaynen in the Calamian Islands between Palawan and Mindoro), 6) Bilic (Bilaan, Tboli, Tiruray and Giangan Bagobo of southern Mindanao), 7) South Mangyan (Buhid, Hanunóo), 8) Palawanic (Palawano, Aborlan Tagbanwa, Palawan Batak, Molbog), 9) Central Philippines (Tagalog, Bikol, the Bisayan dialect complex, Mamanwa, Mansaka, Mandaya, Kalagan, Tagakaulu, etc.), 10) Manobo (a number of languages spoken by hill tribes in the mountains of central and eastern Mindanao, Kinamigin on Camiguin Island in the Bohol Sea just north of Mindanao, and Kagayanen, spoken in the Cagayan Islands between the Bisayas and Palawan), 11) Danaw (Maranao, Iranun, Magindanao, spoken by a predominantly Muslim population in southwest Mindanao), 12) Subanen/Subanun (two or three closely related languages in the Zamboanga Peninsula of western Mindanao), 13) Sangiric (five languages spoken in the Sangir and Talaud Islands and on the northern peninsula of Sulawesi, plus Sangil, a dialect of Sangir spoken by fairly recent immigrants in the Sarangani Islands and southern Mindanao), 14) Minahasan (five languages in the general vicinity of Lake Tondano, in the northern peninsula of Sulawesi), 15) Gorontalic (nine languages spoken in the central and western portions of the northern peninsula of Sulawesi). As noted in Blust (2005c) members of the Philippine subgroup probably are descendants of a single proto language that expanded at the expense of others which existed around it during the early AN settlement of this area.

10.3.3.29 Greater Central Philippines

While the fifteen microgroups noted above are unlikely to provoke controversy, larger groupings within the Philippines have been more difficult to ascertain. There is, however, one major exception. Blust (1991a) presented evidence that microgroups 7, 8, 9, 10, 11, 12 and 15 form a larger genetic unit, called 'Greater Central Philippines' (GCP). The

distribution of GCP apparently resulted from the expansion of an ancestral language community located in northern Mindanao or the Bisayan Islands. One branch (Gorontalic) by-passed established Philippine languages in northern Sulawesi and gained a foothold further south. In addition, GCP loanwords are common in non-GCP languages in the Philippines and in the languages of Sabah, evidently a product of contact during the expansion process.

10.3.3.30   *Western Malayo-Polynesian*

Although a Western Malayo-Polynesian (WMP) subgroup has been recognized in many publications since Blust (1977a), there is no phonological evidence for such a group. A fairly large number of cognates appear to be shared exclusively by languages in the Philippines and western Indonesia, but it remains unclear whether these are innovations, or retentions from PMP. WMP languages in general, including languages from the northern Philippines through western Indonesia to Malagasy, Palauan and Chamorro, uniquely share the process of homorganic nasal substitution in active verb forms. Although this is not an active process in any Formosan or Central-Eastern Malayo-Polynesian language, there are some indications that it was once present in at least CMP and OC languages. It is therefore possible that WMP represents several primary branches of MP.

10.3.3.31   *Malayo-Polynesian*

Haudricourt (1965:315) proposed a tripartite division of AN languages into 1. Western, 2. Northern, and 3. Eastern. These correspond roughly to 1) Western Malayo-Polynesian, 2) all nine Formosan branches (= Northern) and 3) Oceanic in the present classification, although Haudricourt provided no indication of hierarchy among them, implying that the three are coordinate. The first statement of the Malayo-Polynesian hypothesis appeared independently in Mills (1975:581), Dahl (1976; original 1973) and Blust (1977a). Dahl (1976:123ff) proposed a bipartite division, arguing that "The Formosan languages seem to form one common subgroup of the family," and that "an evolution of the language after the separation from Formosan is common to the western languages and Eastern Austronesian. The two may thus be said to form a common subgroup different from the Formosan group." After discussing Dahl's suggestion that the Formosan languages were the first to split off, Mills (1975:581) added that 'The remainder—which can now properly be called Proto Malayo-Polynesian—merged *T and *C, then borrowed or somehow developed a new set of palatals *c j ñ.' At about the same time, ignorant of this terminological proposal, Blust (1977a) proposed a division into 1. Atayalic, 2. Tsouic, 3. Paiwanic, and 4. Malayo-Polynesian, with a further division of MP into 4a. Western Malayo-Polynesian, 4b. Central Malayo-Polynesian, and 4c. Eastern Malayo-Polynesian (the latter two later united under a common CEMP node). The Malayo-Polynesian hypothesis can thus be dated to 1973, and its acquisition of a name to 1975/1977.

The argument for a Malayo-Polynesian subgroup relies heavily on the quality of exclusively shared innovations. In this respect it resembles the argument for North Sarawak, but differs from the latter in appealing to a collection of irregular or narrowly conditioned changes rather than to a single regular phonological change that produced typologically rare results. One type of evidence is seen in PAN *Sepat > PMP *epat 'four', PAN *Si- > PMP *i- 'instrumental focus', PAN *Sipes > PMP *ipes 'cockroach'. These examples are valuable for two reasons: 1) the direction of change is unambiguous, hence permitting a clear distinction between innovation and retention, and 2) the change is

lexically-specific, since PAN *S- normally became PMP *h-. However, the value of this irregular change is compromised by the limited number of Malayo-Polynesian languages that distinguish *S- from zero (most Central Philippine languages, Kayan, Malay, Old Javanese and a few other languages in western Indonesia, Soboyo in eastern Indonesia). Nonetheless, where cognates are available there is general support for the view that the irregular loss of *S- took place in a language ancestral to all AN languages outside Taiwan (thus *Sepat > Tagalog *ápat*, Kayan *pat*, Old Javanese *pat*, *pāt* 'four', next to *SakuC > Tagalog *hákot* 'transport', *Suab > Kayan *huav* 'yawn', or *Sikan > Old Javanese *(h)ikan* 'fish'). A second example of irregular change is seen in PAN *paŋudaN > PMP *paŋdan 'pandanus'. Again, the direction of change is unambiguous, and syncope is lexically-specific (cp. PAN *Caliŋa > PMP *taliŋa 'ear', *qaNiCu > PMP *qanitu 'ghost', PAN *SapuSap > PMP *hapuhap 'feel, grope'). Irregular syncope in reflexes of PAN *paŋudaN is found from the northern Philippines (Ilokano *paŋdán*) to western Indonesia (Malay *pandan*), eastern Indonesia (Kambera *pàndaŋu*, Ngadha *pəda*, Tetun *hedan*, Wetan *edna*), and the Pacific (Wuvulu *paka*, Fijian *vadra*, Hawaiian *hala*). Finally, Formosan languages show the sequence *-CVS corresponding to *-hVC outside Taiwan (where C = any stop, V = any vowel, and S = PAN *S), as in PAN *bukeS, but PMP *buhek 'head hair', PAN *CaqiS, but PMP *tahiq 'to sew', or PAN *tapeS, but PMP *tahep 'winnow grains' (Blust 1993c). Although this change is recurrent, it is irregular. Moreover, in a few cases Formosan languages show the order *-CVS but non-Formosan languages show variation between *-hVC and -CVh, as with PAN *tuqaS > PMP tuqah/tuhaq 'old; mature', or PAN *liseqeS > PMP *liseqah/lisehaq 'nit, egg of a louse'. Since these cases require a PAN sequence *-CVS in any case, the simplest inference is that *-hVC forms in non-Formosan languages are innovative.

In addition to irregular sound changes several examples of apparently regular, but narrowly conditioned sound changes distinguish Formosan from MP languages. One of these is seen in PAN *CumeS > PMP *tumah 'clothes louse' and PAN *buReS > PMP *buRah 'spray water or chewed medicines from the mouth', where *ə merges with *a before a historically secondary final *h, but not before other final consonants. Again, the direction of change is unambiguous, and reflexes of *CumeS and *buReS are fairly common. A second change that appears to be narrowly conditioned is seen in PAN *baRuj : PMP *baluj 'a dove: *Ducula* sp.', and PAN *baRija : PMP *balija 'batten rod in weaving'. Given PAN, PMP *qulej 'maggot, caterpillar', or PAN *Sulij, PMP *hulij 'sleep next to', the simplest hypothesis (i.e. the one that assumes regular conditioned change rather than an unconditioned phonemic split that produced the unusual change *l > R in just two lexical items) is that the direction of change was *R > l/__Vj.

Blust (2001f) likens the development of the Malayo-Polynesian hypothesis to building a wall one stone at a time: a number of pieces of evidence, some of them small or obscure, need to be combined into a larger structure that shows increasing coherence with each new addition. However, the Malayo-Polynesian subgroup is also supported by three phonemic mergers (PAN *C/t > PMP *t, PAN *n/N > PMP *n, PAN *S/h > PMP *h), and by certain innovations in personal pronouns, most notably the change of PAN *-mu '2p genitive' to PMP *-mu '2sg genitive', without a similar change in the corresponding long-form pronoun (PAN, PMP *kamu '2p').

10.3.3.32   *Western Plains*

The subgrouping relationship of certain Formosan languages is sufficiently obvious that it was recognized without the need for extensive argumentation. This is true, for example,

of Atayalic, which was posited as a distinct group by Ferrell (1969), and confirmed by later research (Li 1980a, 1981, 1985). A second group that was considered well-established for many years is Tsouic (Li 1972, Tsuchida 1976). However, as noted earlier, the validity of Tsouic has been questioned by a number of scholars (Harvey 1982, Chang 2006, Ross 2012), and it now appears likely that while Kanakanabu and Saaroa subgroup together, Tsou is a primary branch of the Austronesian language family. Beyond these groups it has been difficult to establish more inclusive genetic units that have achieved general recognition. Tsuchida (1982) identified a group of four extinct Formosan languages (Taokas, Favorlang-Babuza, Papora, Hoanya), all of which were spoken on the western plains of Taiwan and were recorded by Japanese researchers before they disappeared early in the twentieth century. Blust (1996e) added Thao to this group. This addition is of some interest, since Thao is geographically separated from other Western Plains languages, and is the only surviving member of the group, although at the end of the twentieth century it was down to its last 15 known speakers (Blust 2003a).

10.3.3.33   *East Formosan*

East Formosan was posited by Blust (1999b) on the basis of a single sound change shared by Basay, Trobiawan, Kavalan and Amis of the east coast of Taiwan, and Siraya of the southwestern plain. The argument for this group takes us back to the issues raised in 10.3.2.4. Each of these five languages has merged PAN *j and *n, a change that is otherwise unknown in the AN language family. If this merger is treated as a product of convergence there is no obvious reason why the same change would not have affected some other members of a language family with a reported 1,262 members. This grouping has also been accepted by Li (2004a).

Apart from Atayalic, Tsouic, Western Plains and East Formosan there are no other widely accepted groupings of Formosan languages, and the next higher node in the AN family tree is thus PAN itself. Blust (1999b) found evidence for the following primary subgroups of AN: 1. Atayalic, 2. East Formosan, 3. Puyuma, 4. Paiwan, 5. Rukai, 6. Tsouic, 7. Bunun, 8. Western Plains, 9. Northwest Formosan, 10. Malayo-Polynesian. Of these Northwest Formosan (Saisiyat, Kulon, Pazeh) is perhaps the most problematic, as there is some evidence that Pazeh, like Thao, may form an independent branch of the Western Plains group. It follows that at least nine primary branches of the AN language family contain languages that are spoken only on Taiwan.

For convenience Table 10.20 summarizes the subgroups above CEMP, indicating where each was proposed, and classifying it in the same manner as in Table 10.18:

Table 10.20 Major Austronesian subgroups in western Indonesia,
the Philippines and Taiwan

No.	Group	Source	Type
1.	East Formosan	Blust 1999b	F
2.	Western Plains	Tsuchida 1982, Blust 1999b	F?
3.	Malayo-Polynesian	Dahl 1973, Mills 1975, Blust 1977b	F
4.	Western Malayo-Polynesian	Blust 1977b	R
5.	Greater Central Philippines	Blust 1991a	F
6.	Philippines	Zorc 1986	F
7.	Greater North Borneo	Blust 2010	F
8.	North Borneo	Blust 1998e	L?

No.	Group	Source	Type
9.	North Sarawak	Blust 1969, 1974b	F?
10.	Barrier Islands-North Sumatra	Nothofer 1986	F?
11.	Malayo-Chamic	Blust 1981b, 1994a, 2010	F?
12.	Barito	Dahl 1951, Hudson 1967	F?
13.	Greater Barito	Blust 2005c, 2007d	F?
14.	Greater South Sulawesi	Adelaar 1994a	F
15.	Celebic	van den Berg 1996b, Mead 2003b	L?

10.3.3.34   *Other proposals*

Other views on the classification of the Formosan languages can be found in the literature, but most of these are not well-supported. Ferrell (1969), for example, divided the Formosan languages into Atayalic, Tsouic, and 'Paiwanic' groups. The latter was in turn bifurcated into, 'Paiwanic I' (Rukai, Pazeh, Saisiyat, Luilang, Favorlang, Taokas, Papora, Hoanya and Thao), and 'Paiwanic II' (Bunun, Siraya, Amis, Kavalan, Yami). Continuing research has shown that Ferrell's Paiwanic group is untenable. With regard to Paiwanic II Yami is a Philippine language, and Bunun shares no known innovations exclusively with Siraya, Amis and Kavalan. Paiwanic I is likewise a diverse collection with no exclusively shared innovations, apart from members of the Western Plains group.

Harvey (1982) proposed a variant of the Malayo-Polynesian hypothesis, positing an Amis-Malayo-Polynesian branch of AN. Although this claim has often been repeated (sometimes under the name 'Amis-Extra Formosan') the evidence offered for it is weak. Moreover, it has since become clear that Amis is part of a larger East Formosan grouping that includes Kavalan, Basay, Trobiawan and Siraya (Blust 1999b).

Li (1985:259ff) accepts a contracted version of Ferrell's Paiwanic group, but also posits a Northern Formosan group containing Atayalic, Saisiyat, Pazeh, Taokas, Babuza, Papora and Hoanya. As evidence for this grouping he notes that 1) 'These languages generally do not distinguish the feature [+/- personal] in their construction markers,' 2) Dahl's (1981a) $*S_2$ and $*h$ merged as h, 3) $*N$ and $*ñ$ merged as l, 4) 'all languages in the Northern group retain the phonemic distinctions between PAN $*t$ and $*C$, PAN $*n$ and $*ł$,' and (5) PAN $*d$ and $*z$ have merged.

These types of evidence vary widely in their subgrouping value. Li acknowledges that 5) is shared with Tsouic, Bunun, Thao, Kavalan and Amis; its value as a marker of the proposed Northern Formosan group is thus compromised. Since 4) is a retention it has no value for subgrouping, and 1) could easily be a product of independent historical changes. More seriously, 1) is factually inaccurate, since the feature [+/- personal] is distinguished in the genitive construction markers of Atayal (Rau 1992:142). As for 3), Li states that this 'is found also in some other Formosan languages such as Rukai and Bunun.' Although this is not true of Bunun, which reflects $*N$ and $*ñ$ as n, lateral reflexes of these proto phonemes are found not only in Rukai, but also in Puyuma, Paiwan, Siraya, and Saaroa. Moreover, Thao, which belongs with Taokas, Babuza, Papora and Hoanya in a Western Plains subgroup, reflects $*N$ and $*ñ$ as z (voiced interdental fricative). These observations leave little choice but to conclude that the merger of $*N$ and $*ñ$ as some type of lateral (if, indeed, $*N$ was not a lateral already) happened independently in a number of languages. This leaves 2) as the sole evidence for a Northern Formosan subgroup. Li follows Dahl (1981a:38) in distinguishing $*S_1$ from $*S_2$, and argues that it is $*S_2$ that merged with $*h$. The problem with using this change as subgrouping evidence is that Dahl's $*S_2$ is reflected

as zero in a number of other languages, where it could have passed through an intermediate stage with *h*. Reflexes of *kaSiw (Dahl's *kaS₂iw) 'wood; tree', for example, show *h* in Atayalic, Saisiyat, and Pazeh, but zero in Babuza, Papora, Hoanya, Tsouic, Thao, Siraya and Bunun, and reflexes of PAN *SuReNa 'snow' show *h* in Atayalic, Saisiyat, and Pazeh, but zero in Hoanya, Tsouic, Rukai, Puyuma, and Thao. The reason that these lenited reflexes are shared by many Formosan languages in the same forms is interesting and important in itself, but the distribution of lenited reflexes of *S hardly supports the existence of a Northern Formosan subgroup. Other proposals for subgrouping the Formosan languages include Starosta (1995), which is methodologically flawed in various ways (Blust 1999b:62ff), and several claims for a Rukai-Tsouic group based on evidence that appears better explained as a product of diffusion.

At its apex the AN family tree splits into ten primary branches, nine of which are represented only by languages spoken in Taiwan. The lack of nesting in this structure has concerned some scholars. A commonly held view is that language communities split by bifurcation, and this happens recurrently, giving rise to successively nested structures. There are clear counterexamples to this view, as Indo-European, which is usually conceived as containing nine primary branches. Whether rake-like family tree structures arise from simultaneous multiple splits, or from sequential splits that are too closely spaced to allow them to be distinguished by the comparative method, they do appear to exist. Moreover, an AN family tree with ten primary branches is consistent with settlement patterns throughout the AN world. Like many of their descendants, PAN speakers had a maritime orientation, which initially included exploitation of marine resources in addition to farming rice and millet. In order to exploit marine resources it was necessary to settle coastal areas and to remain in his habitat until forced out of it under pressure from population increase or warfare. An initial coastal settlement of Taiwan, with its nearly 600 miles of coastline would have created a dialect chain encircling the island. Assuming that dialect areas occupied roughly 60 miles of coastline, this would have given rise over time to about ten language communities. Needless to say, this is an oversimplification, since periodic extinctions and expansions would have altered the picture in unpredictable ways, but the basic relationship of topography to language split is not incompatible with the language family splitting into roughly ten dialect areas very early in its history, and hence giving rise over time to about the same number of primary linguistic subgroups.

A second nesting issue has to do with the position of Malayo-Polynesian. If all MP languages are descended from an immediate common ancestor, it is reasonable to ask why no trace of that ancestor has been found in Taiwan. No non-forced migration that is historically documented has resulted in the complete removal of the migrating population or linguistic group from the homeland. Alternatively, even if all MP speakers were able to cross from Taiwan to the northern Philippines it must be asked why none crossed back to Taiwan to establish secondary settlements from south to north. There is no easy answer to these questions, but it is possible that the MP migration out of Taiwan was, like other migrations in history, a partial removal of a linguistic group, and that the remnant which stayed behind was later eliminated by the expansion of other language groups.

Undoubtedly the most challenging new proposal concerning the higher-level subgroups of AN is that of Sagart (2004), who has argued for much more nesting at the top of the AN family tree than is proposed here. According to Sagart the first split in AN separated Luilang, Pazeh and Saisiyat of Taiwan from all other AN languages. The main evidence for this claim consists of implicational relationships between the numerals 5-10. The numerals that are generally accepted for PAN are 1. *esa/isa, 2. *duSa, 3. *telu, 4. *Sepat,

5. *lima, 6. *enem, 7. *pitu, 8. *walu, 9. *Siwa and 10. *puluq. As Sagart (2004:413) has noted, "throughout Taiwan, a reflex of *puluq '10' implies the presence of a reflex of *Siwa '9', which implies the presence of *walu '8', which implies the presence of *enem '6', which implies the presence of *lima '5', which implies the presence of *pitu '7', while the reverse implications do not hold." In diagrammatic form:

puluq >> Siwa >> walu >> enem >> lima >> pitu

In other words, all languages that reflect *puluq are embedded in a larger set of languages that reflect *Siwa, this set of languages in turn is embedded in a larger set of languages that reflect *walu, and so on. Sagart interprets these implicational relations as evidence for successive language splits: PAN split into one language ancestral to Luilang, Pazeh and Saisiyat and another that he calls 'Pituish' (from *pitu 'seven'). 'Pituish' in turn split into a language ancestral to Atayalic, Thao, Favorlang, Taokas, Siraya, Papora and Hoanya, and another called 'Enemish' (from *enem 'six'), which is ancestral only to Siraya, and 'Walu-Siwaish' (from *walu 'eight' and *Siwa 'nine'). 'Walu-Siwaish' split into a language ancestral to Tsouic, Paiwan, Rukai, Puyuma, Amis and Bunun, and another called 'Muish', which was ancestral to all other AN languages. Under this interpretation the phylogeny of the Formosan languages shows a nesting structure that is not apparent from phonological change, or from any other proposed innovation in basic vocabulary. What is at issue is whether the implicational relationships that Sagart has identified in Formosan reflexes of the commonly accepted PAN numerals must be explained as a product of chronologically successive innovations, or whether they could have resulted instead from patterned losses, and this remains to be settled.

Sagart's 'Muish' group contains another surprise, one that fuses subgrouping issues with issues of distant genetic relationship. According to Sagart, 'Muish' has three branches: 1. Northeast Formosan (Kavalan + Ketagalan), 2. Tai-Kadai, and 3. Malayo-Polynesian. Not only is a genetic relationship between Tai-Kadai and AN claimed (in addition to Sino Tibetan), but Tai-Kadai is treated as a branch of AN and a sister of Malayo-Polynesian. The primary evidence that Tai-Kadai and AN are genetically related is drawn from Buyang, a recently described 'Kra' (Kadai) language of southern China (cf. Table 10.9). As noted earlier, there can be little doubt that the data from Buyang have established a historical connection with AN to the satisfaction of most scholars, although disagreements exist as to whether the observed similarities are due to common origin or early borrowing. But the claim that Tai-Kadai is a branch of the AN family rather than a sister family is utterly novel. In this view East Formosan languages such as Kavalan and Ketagalan/Basay are more closely related to Tai-Kadai than they are to Amis or Siraya, and Malayo-Polynesian languages such as Tagalog or Malay are more closely related to Tai-Kadai than they are to Paiwan, Puyuma, Amis, Bunun or Pazeh.

This proposal is based on Sagart's theory of serial innovation in the commonly accepted PAN numerals, and two asserted lexical or semantic innovations: PAN *-mu '2pl genitive' > *-mu '2sg genitive' and PAN *qayam replaced by PMP *manuk 'bird'. The first and most obvious problem with the claim that Tai-Kadai is a subgroup of AN is that languages such as Kavalan and Amis or Malay and Paiwan share hundreds of cognate forms, while there are perhaps two dozen plausible cognates linking Tai-Kadai with AN. To circumvent this anomaly Sagart (p.c.) assumes that the Tai-Kadai languages have undergone massive relexification from an unknown and now extinct language, leaving only a small core of original AN vocabulary.

The claimed non-numeral lexical innovations in 'Muish' are hardly more persuasive. First, the stated Buyang reflex of *-mu (*ma^{312}* 'thou') is too short to inspire confidence in a cognate judgement, and additionally has the wrong vowel reflex. Second, Sagart assumes that PAN had *qayam 'bird', and that Proto Kra (Ostapirat 2000) *¾ok 'bird' is related to PMP *manuk, pointing to 'Proto Muish' *manuk 'bird'. However, this interpretation is problematic. Sagart is forced to recognise reflexes of both *qayam and *manuk within his 'Northeast Formosan' group: Kavalan *alam* (Li and Tsuchida 2006), but Basay *manuk(ə)*, Trobiawan *manukka* 'bird' (Moriguchi 1991:233), and he assumes (2004:425) that "*manuk and *qayam then coexisted in Muish and PMP as 'wild bird' and 'domesticated bird' respectively." But, as noted in (Blust 2002b), PMP *manuk meant 'chicken', not 'bird'. A more plausible interpretation is that PAN *qayam meant 'bird' and *manuk meant 'chicken'. In most Formosan languages the latter term was replaced by onomatopoetic innovations (Kanakanabu *tarikuka*, Rukai *torokoku*, Favorlang *kokko*, Paiwan *kuka*, Rikavong Puyuma *torokok*, Amis *kokoq*, Bunun *tolkok*, etc). In PMP the same distinction was maintained but at an early date *qayam was replaced by *manu-manuk 'bird'. In short, the startling claim that Tai-Kadai is a subgroup of AN fails to bear close scrutiny, and we are forced to fall back on the more conservative view that these language families were in contact on the Asian mainland before AN speakers reached Taiwan, whether that contact involved separation from a common ancestor (the more likely hypothesis) or borrowing from AN into Tai-Kadai.

Most recently Ross (2009, 2012) has argued that a proposed syntactic innovation justifies a division of the AN languages into Puyuma, Tsou, and Rukai on the one hand, and the remainder, which he calls 'Nuclear Austronesian', on the other. What Ross proposes in effect, is a return to the claim first promulgated in Starosta, Pawley and Reid (1982) that the voice-marking morphology of attested Philippine-type languages originally had only nominalizing functions, and acquired its verbal properties through a subsequent reinterpretation. This view has been challenged by Sagart (2010), reasserted by Teng and Ross (2010) in a reply to Sagart's critique, and challenged again by Foley (2012a) in a careful syntactic analysis of Puyuma which concludes that their own data "casts doubt on Teng and Ross's (2010) claim that Puyuma has not undergone the reanalysis of nominalisations to verbs that they use to define Nuclear Austronesian, and from which they therefore exclude Puyuma" (Foley 2012a:38). In short, no attempt to reduce the number of primary branches of the AN family proposed in Blust (1999b) can be considered free from serious problems, and it appears safest for the present to maintain the more conservative view that the AN family divides into at least nine primary branches represented by the aboriginal languages of Taiwan and a single branch (Malayo-Polynesian) represented by all other members of the family.

10.3.4 The Austronesian family tree

For the convenience of the reader the higher-level subgroups of AN are given in table 10.21 together with the phonological correspondences between them. Voiced stops in POC were automatically prenasalized (*b = [mb], *d = [nd], *j = [ndʒ], *g = [ŋg]; c > s in POC indicates retention of earlier *j as *c in POC, but merger with *s in all OC languages outside the Admiralty Islands). In PAN and PMP *e = schwa:

Table 10.21 Phonological correspondences between
higher-level subgroups of Austronesian

PAN	PMP	PCEMP	PEMP	POC
p	p	p	p	p
	mp	mp	mp	b
t	t	t	t	t
	nt	nt	nt	d
C	t	t	t	t
	nt	nt	nt	d
c	c	s	s	s
	nc	ns	ns	j
k	k	k	k	k
	ŋk	ŋk	ŋk	g
q	q	q	q	q
b	b	b	b	p
	mb	mb	mb	b
d	d	d	r	r
	nd	nd	nd	dr
z	z	z	z	s
	nz	nz	nz	j
j	j	j	j	c>s
	nj	nj	nj	?
g	g	g	k	k
	ŋg	ŋg	ŋk	g
m	m	m	m	m
n	n	n	n	n
N	n	n	n	n
ñ	ñ	ñ	ñ	ñ
ŋ	ŋ	ŋ	ŋ	ŋ
s	s	s	s	s
	ns	ns	ns	j
S	h, Ø	h, Ø	Ø	Ø
h	h, Ø	Ø	Ø	Ø
l	l	l	l	l
r	r	r	?	?
R	R	R	R	R
	l/__j			
w	w	w	w	w
y	y	y	y	y
a	a	a	a	a
e	e	ə	o, ə	o
i	i	i/e	i/e	i/e
u	u	u/o	u/o	u/o
-ay	-ay	-ay	-e	-e
-aw	-aw	-aw	-o	-o
-uy	-uy	-uy	-i	-i
-iw	-iw	-iw?	-i	-i

It is clear from this table that POC is defined by phonological innovations, in particular regular mergers. By contrast, PEMP shows no phonological innovations that have subgrouping value (the monophthongisation of *-ay, *-aw, *-uy and *-iw is a recurrent change throughout the AN family, and could well have happened independently in Proto South Halmahera-West New Guinea, and Proto Oceanic). The phonological evidence for PCEMP, on the other hand, while not abundant or highly conspicuous, is of very high quality, since it involves the irregular lowering of *i and *u in the same cognate sets throughout the entire group, with more than 600 languages (and is distributed over roughly the same set of languages as the innovative terms for marsupial mammals). The phonological evidence for PMP is more subtle, requiring recognition of such easily overlooked conditioned changes as *R > l/_ j, the sporadic but recurrent metathesis of PAN *-CVS to *hVC (*bukeS > *buhek 'head hair', etc.), and other changes that are either sporadic or defined by narrow phonological conditions.

10.4 Migration theory

All inferences about centers of dispersal, or 'homelands' for language families or subgroup ancestors depend ultimately upon subgrouping (Dyen 1956b). As in determining the center of origin of cultivated plants, what is critical is not the geographical distribution of languages, but rather the distribution of major taxa. For reasons of parsimony the region where these are most heavily concentrated is the probable center of origin. In the case of AN the concentration of nine primary branches of the language family in Taiwan as against one large extra-Formosan branch (Malayo-Polynesian) presents a telling argument against the hypothesis of an extra-Formosan homeland, since this would require nine historically independent migrations of AN-speaking peoples into Taiwan rather than a single migration out of Taiwan into island and mainland Southeast Asia, Madagascar and the Pacific. Not only are the Formosan and non-Formosan homeland hypotheses asymmetrical in terms of what Dyen (1956b) called the 'principle of least moves', but there is no obvious reason why as many as nine migrations originating from some other locality would have an identical endpoint. The most parsimonious explanation for the attested distribution of primary AN subgroups, therefore, points unambiguously to Taiwan as the earliest region for which AN settlement can be established by linguistic means. Applying techniques derived from Bayesian inference that are currently revolutionising biological phylogenetics (Greenhill and Gray 2009), Gray, Drummond and Greenhill (2009) arrive at a very similar conclusion: the AN homeland almost certainly was in Taiwan, and from there the expansion of AN-speaking peoples was steadily southward and eastward into the Pacific.

Together with the emerging radiocarbon chronology for the introduction of Neolithic cultural assemblages, the linguistic evidence suggests that AN languages reached Taiwan from the adjacent mainland of China in the period 3,500-4,000 BCE. Once there AN speakers remained *in situ* for over a millennium before settling the northern Philippines, a time interval that can be called the 'first long pause' in the dramatic expansion of AN-speaking peoples over much of the island world of the tropics. There is no generally accepted explanation for this settlment pause, although it has been suggested that it was finally broken with the invention of the outrigger canoe complex (Blust 1999b), a cultural attribute that is richly attested outside Taiwan and in linguistic reconstructions for PMP, but not for PAN (Pawley and Pawley 1990). The unexpectedly close relationships of Philippine languages suggests that the current distribution of AN languages in the

Philippines is not a simple continuation of the original settlement pattern, with linguistic differentiation leading to a multiplication of languages over time. Rather, the clear evidence for a Philippine subgroup implies that after the initial AN settlement of the Philippine Islands, one language group (Proto Philippines) dramatically expanded at the expense of other early AN languages, which either became extinct or were dislocated, as through the migration of the ancestral Chamorros to the Mariana Islands around 1,500 BC (Blust 2000c).

Further south the migration history becomes cloudier: southward-expanding AN speakers split into at least two major streams, one leading into western Indonesia and the other into eastern Indonesia and the Pacific. However, it is possible that migration streams from the Philippines split three ways in their southward progress, with one leading into Borneo, and then to Sumatra, mainland Southeast Asia and Madagascar, a second into Sulawesi, and the third into eastern Indonesia and the Pacific. What has become increasingly clear in recent years is that the attested distribution of AN languages is due to a complex interaction of primary settlement and differentiation *in situ*, and subsequent language leveling not just in the Philippines, but in a number of areas. Long after the expansion of Proto Philippines precipitated a major episode of language leveling, the expansion of Proto Greater Central Philippines led to another reduction of linguistic diversity throughout most of the central Philippines, and in parts of northern Sulawesi. The recent establishment of a 'Celebic' supergroup that excludes the South Sulawesi languages (Mead 2003b) leaves the latter group as a clue to probable earlier diversity that was leveled by the expansion of Proto Celebic. Much the same is true of southern Sumatra, where a low level of diversity is found in precisely those areas that would have been settled earliest in almost any plausible scenario of AN migration (Blust 2005d). Other areas where the attested degree of linguistic diversity is radically at variance with expectation based on probable antiquity of settlement are Timor and southern Halmahera. The linguistic history of Timor is still poorly understood, but the AN languages of this island appear to be far less diverse than those of other parts of the Lesser Sundas, or in most of the Moluccas. Central and Central-Eastern Malayo-Polynesian almost certainly split from one another in the northern Moluccas, and the early diversity of AN languages in this area evidently was reduced by the expansion of South Halmahera-West New Guinea speakers out of Cenderawasih Bay in Irian into southern Halmahera and the smaller islands between Halmahera and New Guinea.

All of these inferences are based on language distribution: where diversity is high this implies a relatively long period of settlement, where it is low it implies a shorter period of settlement, but where it is low in areas where it is expected to be higher given the wider subgrouping picture, it implies a prehistoric episode of language leveling. It was through consideration of these general principles that Dyen (1965a) proposed an AN homeland in the area of New Guinea and the Bismarck Archipelago. The data that he used to support this inference was based on lexicostatistics. We now know that lexicostatistics, which fails to distinguish innovations from retentions, can yield seriously distorted subgrouping results if significant variation exists in the replacement rate of basic vocabulary across languages, and there is very good reason to believe that this is the case in the AN language family (Blust 2000a). Moreover, in 1965 little archaeology had been done in the Pacific, and it was possible to ignore the archaeological record in advancing such a proposal on the basis of linguistic evidence. In the 40 years since that time this picture has changed dramatically, since there is now a much more detailed (although still far from complete) record of Pacific prehistory derived from analyses of prehistoric material culture. What this record

shows is that the western Pacific at least as far east as Greater Bougainville, a large Pleistocene island that included much of what became the western and central Solomons in post-glacial times, was inhabited by speakers of 'Papuan' languages by at least 30,000 BP. Then, around 3,350 BP bearers of the Lapita culture complex appeared abruptly in the Bismarck Archipelago. Since they almost certainly entered the Pacific by passing over the north coast of the Bird's Head Peninsula of New Guinea, AN speakers must have first settled the north coast of New Guinea and small offshore islands before expanding into the Bismarck Archipelago. Most current scenarios see Proto Oceanic as spoken in the region of the Vitiaz Strait between New Guinea and New Britain. However, given the known voyaging capabilities of Proto Oceanic speakers there is no reason why some parts of this language community could not have sailed directly from the north coast of New Guinea to the Admiralty Islands or St. Mathias Archipelago before others, moving along the coast, had reached the Huon peninsula. Throughout at least New Guinea, the Bismarck Archipelago and the western Solomons AN speakers evidently encountered well-established populations of Papuan speakers who had settled western Melanesia tens of thousands of years before. Ongoing contact gave rise over time to syncretic populations and cultures (the 'Island Melanesians'), and to some shared features of linguistic typology. Within a century or two they had reached as far east as Fiji, Tonga and Samoa. At this point there was a second 'long pause' that gave rise to the distinctive features of Proto Polynesian, and then a late movement into eastern Polynesia, no part of which seems to have been settled before about 1,025 AD (Wilmshurst et al. 2011).

Although archaeological research in most parts of Micronesia is still relatively undeveloped, linguistic subgrouping points clearly to separate migration histories for Palau and the Marianas as against the Oceanic languages of this region. Among the Oceanic languages Yapese has had a very distinctive history, and may represent an early movement from the Admiralty Islands shortly after the settlement of the Bismarcks by AN speakers (Ross 1996c). Among other Oceanic languages of this area, subgrouping considerations suggest that speakers of Proto Micronesian entered the region in the east (Kiribati-Nauru), possibly from the southeast Solomons (Blust 1984a). From the initial point of entry they quickly settled the high islands of Kosrae, Pohnpei and Chuuk, and then gradually expanded westward into the atoll world of the western Carolines. In Vanuatu and southern Melanesia (New Caledonia and the Loyalties) the earliest archaeological evidence for human settlement is associated with the Lapita cultural complex, and hence with speakers of Oceanic languages. Various aspects of the physical anthropology, traditional cultures and linguistic typology of this part of Melanesia, however, are difficult to reconcile with the standard view, and suggest that much remains to be discovered in this area (Blust 2005a).

11 *The world of Austronesian scholarship*

11.0 Introduction

A survey of the AN languages would be incomplete if it failed to touch on the world of AN scholarship, including information on the size of the community of scholars actively engaged in work on these languages, the major academic centers for the study of AN languages, periodic meetings and publications, and the status of comparative scholarship on AN in relation to other major language families.

11.1 Size of the scholarly community, and major centers of Austronesian scholarship

To judge from recent conference presentations, and from articles or books that have been published within the past 10-15 years, some 400-450 persons worldwide appear to be actively engaged in the study of the AN languages. These individuals can be divided into two broad groups: those employed at universities and those employed as missionary-translators, primarily by the Summer Institute of Linguistics.

The following list of major centers for the study of AN languages or linguistics is presented with a full awareness of the hazards of attempting to compile such types of information. My aim here has been to highlight those institutions where a beginning student who is interested in gaining a broad knowledge of the AN language family or some branch of it, can most benefit from study. Important scholars, particularly in the formalist mold, are located at some other institutions, but often as lone individuals. Moreover, the primary research focus of many formalists who work on AN languages is on linguistic theory, with AN languages serving the role of data sources for theory-testing rather than being objects of general interest in themselves. Consequently, the published work of such scholars tends to have a deep, but narrow focus—one that is often too narrow to provide a useful introduction to the typology or historical relationships of the AN language family or any of its branches or geographically-defined areas.

Table 11.1 Major centers for the study of Austronesian languages/linguistics

Asia

Taiwan	(Hsin-chu) National Tsing-hua University, National Hsinchu University of Education
	(Hualien) National Dong Hua University
	(Chiayi) National Chung Cheng University
	(Pu-li) National Chi Nan University
	(Taichung) Providence University
	(Taoyuan) Yuan Ze University
	(Taipei) Academia Sinica, National Taiwan University, National Taiwan Normal University, National Taipei University of Technology,
Japan	(Nagoya) Aichi Prefectural University
	(Osaka) National Museum of Ethnology
	(Tokyo) Tokyo Woman's Christian University

Philippines	(Manila) University of the Philippines—Diliman, De la Salle University
Brunei	Universiti Brunei Darussalam
Malaysia	(Bangi, Selangor) Universiti Kebangsaan Malaysia

Australasia

Australia	(Canberra) The Australian National University
	(Newcastle) The University of Newcastle
	(Brisbane) Griffith University
	(Melbourne) The University of Melbourne, Monash University
	(Sydney) The University of Sydney
	(Perth) The University of Western Australia
New Zealand	(Auckland) The University of Auckland
	(Hamilton) The University of Waikato
Vanuatu	(Port Vila) The University of the South Pacific

North America

United States	(California) UC Santa Cruz, UC Santa Barbara, UCLA, California State University at Chico
	(Connecticut) Yale University
	(Delaware) The University of Delaware
	(Hawai'i) The University of Hawai'i
	(Iowa) The University of Iowa
	(Massachusetts) Harvard, MIT
	(New York) Cornell University
	(Texas) Rice University
Canada	(Montreal) McGill University
	(Ottawa) The University of Ottawa
	(Toronto) The University of Toronto

Europe

Russia	(Moscow) The Institute of Oriental Studies
	(St. Petersburg) The University of St. Petersburg
Netherlands	(Amsterdam) The Free University of Amsterdam
	(Leiden) The University of Leiden
	(Nijmegen) Max Planck Institute for Psycholinguistics
France	(Paris) Centre National de la Recherche Scientifique (CNRS)
Germany	(Cologne) The University of Cologne
	(Kiel) The University of Kiel
	(Leipzig) Max Planck Institute for Evolutionary Anthropology
United Kingdom	(London) School of Oriental and African Studies (SOAS), University of London
	(Manchester) The University of Manchester

Locally important centers for the study of the indigenous languages of the area also are found in the Philippines, Indonesia, Malaysia and Madagascar. The Summer Institute of Linguistics (SIL) has, or until recently had branches in the Philippines, Sabah, Sulawesi, the Moluccas, Papua/Irian Jaya, and Papua New Guinea. Until 1975 SIL also maintained a presence in Vietnam.

These centers vary greatly in the number of scholars engaged in research on AN languages, and in the range of languages represented by that research. Most Dutch Austronesianists have traditionally focused on the languages of Indonesia, although that focus is now broadening (largely through the efforts of Dutch nationals working outside the Netherlands). Similarly, French scholars have traditionally devoted their efforts largely to the AN languages of French possessions in mainland Southeast Asia and the Pacific. Almost all work on AN languages being conducted in Taiwan-based scholars is understandably concerned with the 15 remaining Formosan languages (including Yami), several of which are highly endangered. Work at some American universities is concentrated almost entirely on one or two languages, as with the tradition of syntactic research on Tagalog and Malagasy at UCLA. The programs that provide broadest coverage of the language family are the Department of Linguistics in the Research School of Pacific and Asian Studies at the Australian National University, which has, or until recently had, excellent coverage for western Melanesia, Polynesia-Fiji, and eastern Indonesia, as well as some strength in other areas, and the Department of Linguistics at the University of Hawai'i which has excellent coverage for Taiwan, the Philippines, Indonesia-Malaysia, western Melanesia, Micronesia, and Polynesia.

11.2 Periodic meetings

An important measure of the maturity of a scholarly discipline is the occurrence of periodic international gatherings of scholars to discuss the progress of research in the field. A number of regularly scheduled conferences have been taking place in the AN field since the early 1970s. The oldest and most inclusive of these are the International Conferences on Austronesian Linguistics (ICALs). For nearly a decade there was an attempt to maintain a separate series of general AN linguistics conferences which were situated in a location more convenient to scholars in the eastern United States. These were called the Eastern Conferences on Austronesian Linguistics, and only three were held. The first was organized at Yale University late in 1973, just months before the much larger First International Conference on Austronesian Linguistics was held in Honolulu. The second was held at the University of Michigan in Ann Arbor on May 4-5, 1976, and the Third Eastern Conference on Austronesian Linguistics (with acronym TECAL proposed for the first time) was hosted by Ohio University in Athens, Ohio, on May 6-7, 1983. Proceedings were published for each of the last two conferences (Naylor 1980, McGinn 1988).

In recent years a need has been felt in some quarters to create additional AN conferences with a more specialized orientation. These are defined either by a more limited geographical focus, or by a contrast between formal theoretical approaches and typological or historical approaches to language. Conferences motivated by the first demand are 1. the Conference on Oceanic Linguistics (initial acronym: ICOL; changed to COOL at the third meeting), 2. the International Symposium on Malay-Indonesian Linguistics (ISMIL), and 3. the East Nusantara Workshop/Conference on languages of eastern Indonesia (retrospective acronym: ENUS). The one conference that has been motivated by a desire to separate papers with a formal theoretical focus from others is the annual meeting of the Austronesian Formal Linguistics Association (AFLA). Most recently the United Kingdom Austronesian Research Group (UKARG) was organized at the University of Surrey to serve the interests of European scholars who are geographically far-removed from the AN world. A one-day conference under its sponsorship was held in 2005 with the acronym ALL (Conference on Austronesian Languages and Linguistics), a two-day meeting followed in June, 2006, and a third meeting in 2007. The founder of this conference (Bill Palmer) then relocated, and the site shifted to London for ALL 4 in 2009, with an expansion to include Papuan languages in 2012 under a new acronym APLL

(Conference on Austronesian and Papuan Languages and Linguistics). Table 11.2 profiles the accelerating pace of scholarship on AN languages since the early 1990s, as reflected in recurrent regular meetings (ECAL/TECAL = Eastern Conference on Austronesian Linguistics; FICCAL = First International Conference on Comparative Austronesian Linguistics; subsequent conferences dropped 'comparative', and were labeled SICAL, TICAL, FOCAL, VICAL, 6 ICAL, ICAL 7, 8 ICAL, 9 ICAL, ICAL 10, 11-ICAL and 12-ICAL; FICOL/SICOL = First/Second International Conferences on Oceanic Linguistics; acronym subsequently changed to COOL, AFLA = Austronesian Formal Linguistics Association, ISMIL = International Symposium on Malay-Indonesian Linguistics, ENUS = East Nusantara Workshop/Conference on languages of eastern Indonesia), ISLOJ = International Symposium on the languages of Java. No conferences on AN linguistics were held on a recurrent basis prior to 1973. From 1973 until 1993 only two conferences on AN linguistics were held on a periodic basis (one became defunct in 1983), but as of 2006 at least five conferences are held more-or-less regularly. The growth of conference-based scholarship on AN languages since the first meeting in 1973 has been nothing short of spectacular, as can be seen by comparing the 1980s, which witnessed just four conferences, with the 1990s, in which 17 were held, and the first decade of the 21st century, in which 36 conferences took place. Fully half of the individual years from 2000-2009 (2002, 2005, 2006, 2007, 2009) saw as many conferences on AN languages as the entire decade of the 1980s, and 2007 alone saw nearly as many as were held in the 15 year period from 1973 to 1987.

Table 11.2 Conferences on Austronesian linguistics held between 1973 and 2013

Year	Conference	Published Proceedings
1973	ECAL 1: New Haven, Connecticut, USA	no
1974	FICCAL: Honolulu, Hawai'i, USA	yes
1975		
1976	ECAL 2: Ann Arbor, Michigan, USA	yes
1977		
1978	SICAL: Canberra, Australia	yes
1979		
1980		
1981	TICAL: Den Pasar, Bali, Indonesia	yes
1982		
1983	TECAL: Athens, Ohio, USA	yes
1984	FOCAL: Suva, Fiji	yes
1985		
1986		
1987		
1988	VICAL: Auckland, New Zealand	yes
1989		
1990		
1991	6 ICAL: Honolulu, Hawai'i, USA	no
1992		
1993	FICOL: Port Vila, Vanuatu	yes
1994	AFLA 1: Montreal, Canada	no
	ICAL 7: Leiden, Netherlands	yes
1995	AFLA 2: Toronto, Canada	no
	SICOL: Suva, Fiji	yes

Year	Conference	Published Proceedings
1996	AFLA 3: Los Angeles, California, USA	no
1997	AFLA 4: Los Angeles, California, USA	no
	COOL 3: Waikato, New Zealand	no
	8 ICAL: Taipei, Taiwan	yes
	ISMIL 1: Pulau Pinang, Malaysia	no
1998	AFLA 5: Honolulu, Hawai'i, USA	no
	ISMIL 2: Ujung Pandang, Sulawesi, Indonesia	no
1999	AFLA 6: Toronto, Canada	no
	COOL 4: Niue	yes
	ENUS 1: Leiden, Netherlands	no
	ISMIL 3: Amsterdam, Netherlands	no
2000	AFLA 7: Amsterdam, Netherlands	no
	ENUS 2: Canberra, Australia	no
	ISMIL 4: Jakarta, Indonesia	no
2001	AFLA 8: Cambridge, Massachusetts, USA	no
	ENUS 3: Leiden, Netherlands	no
	ISMIL 5: Leipzig, Germany	no
2002	AFLA 9: Ithaca, New York, USA	no
	COOL 5: Canberra, Australia	no
	9 ICAL: Canberra, Australia	no
	ISMIL 6: Bintan island, Riau, Indonesia	no
2003	AFLA 10: Honolulu, Hawai'i, USA	no
	ISMIL 7: Nijmegen, Netherlands	no
2004	AFLA 11: Berlin, Germany	no
	COOL 6: Port Vila, Vanuatu	no
	ISMIL 8: Pulau Pinang, Malaysia	no
2005	AFLA 12: Los Angeles, California, USA	no
	ALL 1: Oxford, UK	no
	ENUS 4: Leiden, Netherlands	yes[121]
	ISMIL 9: Maninjau, West Sumatra, Indonesia	no
2006	AFLA 13: Hsin-chu, Taiwan	no
	ALL 2: Oxford, UK	no
	ICAL 10: Puerto Princesa, Philippines	no
	ISMIL 10: Newark, Delaware, USA	no
2007	AFLA 14: Montreal, Canada	no
	ALL 3: London, UK	no
	COOL 7: Noumea, New Caledonia	no
	ENUS 5: Kupang, Timor, Indonesia	no
	ISLOJ 1 : Semarang, Java, Indonesia	no
	ISMIL 11: Manokwari, Papua, Indonesia	no
2008	AFLA 15: Sydney, Australia	yes
	ISMIL 12: Leiden, Netherlands	no

---

[121]   See Ewing and Klamer (2010).

Year	Conference	Published Proceedings
2009	AFLA 16: Santa Cruz, California, USA	yes
	ALL 4: London, UK	no
	11 ICAL: Aussois, France	no
	ISLOJ 2: Senggigi, Lombok, Indonesia	no
	ISMIL 13: Senggigi, Lombok, Indonesia	no
2010	AFLA 17: Stony Brook, New York, USA	no
	COOL 8: Auckland, New Zealand	no
	ENUS 6: Kupang, Timor, Indonesia	no
	ISMIL 14: Minneapolis-St. Paul, USA	no
2011	AFLA 18: Cambridge, Massachusetts, USA	yes
	ISLOJ 3: Malang, Java, Indonesia	no
	ISMIL 15: Semarang, Java, Indonesia	no
2012	AFLA 19: Taipei, Taiwan	no
	APLL 5 (contination of ALL): London	no
	ENUS 6: Denpasar, Bali, Indonesia	yes[122]
	12-ICAL: Denpasar, Bali, Indonesia	yes
	ISMIL 16: Kelaniya, Sri Lanka	
2013	AFLA 20: Arlington, Texas, USA	
	COOL 9: Newcastle, Australia	
	ISLOJ 4: Padang, Sumatra, Indonesia	
	ISMIL 17: Padang, Sumatra, Indonesia	

In addition to these regularly scheduled conferences there have been several symposia that have shown signs of recurrence, either as independent entities, or as attachments to larger interdisciplinary meetings. The most important of these are the two International Symposia on Austronesian Studies Relating to Taiwan (ISART), held in December, 1992, and December, 2001, at the Academia Sinica in Taipei. The first of these meetings resulted in a large volume of published papers (Li et al. 1995). Other occasional meetings which have happened more than once are the occasional linguistics sessions at the meetings of the Borneo Research Council, at least one of which has resulted in a published volume (Martin 1992), and two meetings on the languages of western Borneo held in March, 2000, and January, 2005 at the Center for the study of the Malay world, Universiti Kebangsaan Malaysia in Bangi, Selangor.

As noted above, the basis for this proliferation appears to be the perceived need for smaller meetings with greater focus. One way to achieve this is by areal restriction, as with the new conferences on Oceanic languages, Malay/Indonesian, languages of Java, and languages of eastern Indonesia. The Austronesian Formal Linguistics Association, on the other hand, was organised by scholars working with formal models of syntax or phonology, some of whom evidently felt either that the International Conferences on Austronesian Linguistics were too much dominated by historical linguists, or that most work presented at these large, thematically diffuse conferences lacked a tightly articulated theoretical framework within which work in syntactic or phonological theory could be usefully discussed. There are signs that the AFLA conferences are now widening their scope to include historical linguistics, typology, and sociolinguistics, and in time the distinction between the AFLA and ICAL conferences may become blurred.

---

[122]   ENUS 6 was held as a two-day panel organized by Marian Klamer and Frantisek Kratochvil during 12-ICAL.

11.3 Periodic publications

The single most important publication outlet for AN languages undoubtedly is *Pacific Linguistics*, a series of books, monographs and collections of papers which began as the organ of the Linguistic Circle of Canberra in the early 1960s, but was published by the Department of Linguistics at the Research School of Pacific and Asian Studies, Australian National University for many years thereafter. As of this writing over 600 book-length publications have appeared in the PL series over a 50-year period (1962-2011), most of them concerned with AN languages. On average, *Pacific Linguistics* has published more than a book a month for half a century, and in the process has made an enormous contribution to the documentation of AN, Papuan and Australian aboriginal languages. At the beginning of 2012 the PL series moved to DeGruyter Mouton, who will continue to publish Pacific Linguistics monographs as part of a larger publication enterprise. In addition to this major publication outlet, several journals have a strong AN focus. Chief among these is *Oceanic Linguistics*, published twice a year since 1962 at the University of Hawai'i, and its associated set of Oceanic Linguistics Special Publications. Like *Pacific Linguistics*, *Oceanic Linguistics* focuses on AN, Papuan and Australian languages, although the bulk of the material that it publishes is concerned with AN languages. Other periodical publications which highlight AN languages are NUSA, a series of occasional papers that includes monographs and thematically or areally organized collections of papers on the languages of Indonesia and occasionally other regions, published by the Atma Jaya Catholic University in Jakarta since 1975, and averaging about two publications per year, *The Philippine Journal of Linguistics*, published twice a year since 1970 by the Linguistic Society of the Philippines, *Language and linguistics in Melanesia* (called *Kivung* from 1968-1980) published since 1968 by the Linguistic Society of Papua New Guinea, *Te Reo*, published since 1958 by the Linguistic Society of New Zealand, *Language and Linguistics*, published four times annually since 2000 by the Institute of Linguistics at the Academia Sinica in Taipei to further linguistic research on Chinese and the aboriginal languages of Taiwan, and *Dewan Bahasa*, a monthly journal published by the national language institute of Malaysia in Kuala Lumpur. A number of other journals that focus on the AN world are interdisciplinary, and only occasionally contain papers of linguistic interest. Among those that merit special mention are *Bijdragen tot de taal-, land- en volkenkunde*, published in the Netherlands since 1852, *The Journal of the Polynesian Society*, published by The Polynesian Society in Auckland, New Zealand since 1892, *The Sarawak Museum Journal*, published by The Sarawak Museum in Kuching, Sarawak since 1911, the *Bulletin of the School of Oriental and African Studies*, published by the University of London, from 1917 (as the *Bulletin of the School of Oriental Studies* until 1940), and *The Bulletin of the Institute of History and Philology*, published by the Institute of History and Philology, Academia Sinica since 1928.

11.4 Landmarks of scholarship vis-a-vis other language families

Austronesian is one of the best-studied of all language families. In some ways this is surprising, given its enormous territorial extent and the number of languages that it contains. This is not to deny that large gaps in documentation exist in some geographical regions and in some areas of linguistic analysis. What matters in comparing the development of scholarship on AN with that of other language families is that a great deal of descriptive and comparative work has been done, and that in many cases this work is of high quality. As a result we are in a position to say more about the nature and general theoretical relevance of many synchronic phenomena in AN languages, and about AN historical linguistics, than is true of many other language families. Tables 11.3 and 11.4 provide perspective on this statement. Table 11.3 lists

fifteen of the world's major language families, the number of languages reported for each, and the geographical spread of the family in approximate number of kilometers both East-West (EW) and North-South (NS):

Table 11.3 Major language families, with size in languages and territorial extent

		Languages	EW	NS
1.	Algonquian	38	4050	2740
2.	Arawakan	60	2200	5100
3.	Australian	258	3900	3725
4.	Austroasiatic	168	3690	3000
5.	Austronesian	1262	23000	10500
6.	Dravidian	75	1940	3000
7.	Indo-European	144	9140	7750
8.	Nadene	47	2425	5400
9.	Niger-Congo	1489	7050	6680
10.	Semitic	74	7500	4000
11.	Sino-Tibetan	365	4300	5280
12.	Tupi-Guarani	70	3800	3550
13.	Turkic	40	8270	4220
14.	Uralic	38	5000	3160
15.	Uto-Aztecan	41	2000	2700

Tables of this sort are inherently problematic, but they are nonetheless useful in attempting to approximate comparability of basic scholarship across language families. The number of known languages is given first. Needless to say, the language/dialect distinction is open to more than one interpretation. For this reason all data is taken from the *Ethnologue* (Lewis 2009). Language territories represent areas occupied prior to the European colonial expansions of the fifteenth to nineteenth centuries. Arabic in north Africa is relatively recent, but older than this (reaching coastal areas after 639 AD, with massive Bedouin migration after 1,000 AD), and the same can be said for AN languages in Madagascar and New Zealand. English, Spanish or French outside Europe are not counted in determining the territorial extent of Indo-European, then, but Arabic in north Africa and Malagasy and Maori are used for this purpose. Australian includes Tasmania, despite uncertainty regarding the classification of the extinct languages of this island. Another parameter that could be built into Table 11.3. is number of speakers, but this has proven difficult. Although the *Ethnologue* tries to distinguish speakers from those claiming a given ethnic identity, this is not always possible. Moreover, since language groups are listed by country, and some languages are spoken on both sides of a national border, there is a danger of duplicate counting that is sometimes difficult to avoid. For these reasons population figures have been left out, although some demographic data for AN languages are given in Chapter 2. Finally, although the *Ethnologue* numbers for languages are high compared to those of other sources (e.g. Ruhlen 1987) this is uniformly true, and does not appear to significantly distort their relative values.

Whereas Table 11.3 sketches the relative sizes of major language families in number of languages and territorial expanse at the time of European contact, Table 11.4 lists the year they were first recognized in print, the year that a systematic reconstruction of the phonology of the proto language was first proposed, and the approximate number of morphemes that have been reconstructed to date on various levels. In other words, it aims at an overview of major scholarly milestones with regard to the language family as a whole. In the interest of at least approaching true comparability the term 'language family' is restricted to collections of

languages for which a more-or-less complete sound system can be reconstructed, together with a minimum of several hundred proto morphemes (thus Algonquian rather than Algic, Semitic rather than Afroasiatic, Bantu rather than Niger-Congo, etc.). Except for Algonquian, Bantu, Dravidian, Tupi-Guarani and Turkic, which appear to have shallower time-depths, this corresponds to proto languages that probably began to differentiate on the order of 5,000-6,000 years ago. It can be argued, however, that a table of this kind is a procrustean bed, since at least Austroasiatic and Na-Dene press the limits of the comparative method in the reconstruction of phonology and lexicon. Another measure that might have been included is the year that significant sound correspondences were first proposed (1822 for Indo-European, with Grimm's formulation of the Germanic consonant shifts, 1861 for Austronesian, with van der Tuuk's formulation of the 'R-G-H' and 'R-L-D' laws, etc.). However, it was difficult to obtain this information for most language families (LFR = language family recognized, PR = phonology reconstructed, NRM = number of reconstructed morphemes):

Table 11.4 Highlights in the historical study of fifteen major language families

		LFR	PR	NRM
1.	Algonquian	1650/1703	1925	4,066
2.	Arawakan	1782	1972	203
3.	Australian	1841	1956?	1,561?
4.	Austroasiatic	1854?	1959	1,450
5.	Austronesian	1603/1708	1934-1938	6,370
6.	Bantu	1818	1899	1,000-4,000?
7.	Dravidian	1816	1961	1,496
8.	Indo-European	1786	1861-1862	4,200
9.	Na-Dene	1915	—	none
10.	Semitic	?/1702	1908-1913	500+
11.	Sino-Tibetan	1828/1896	1972	450-700+
12.	Tupi-Guarani	1806-1817?	1971	221
13.	Turkic	11th century	1949	3000?
14.	Uralic	1671/1770	1955	140?
15.	Uto-Aztecan	1859	1913-1914	2,622

Again, the first thing to note is that there are many problems in trying to construct a table of this kind. Does earliest recognition of a language family mean 1. recognition of all its primary branches, 2. recognition of at least two primary branches, or 3. recognition of the common origin of any two geographically widely separated languages? In most cases the last of these definitions appears to be the most practical, and is adopted here. Specific problems of interpretation are discussed below.

1. Algonquian: According to Goddard (1996:291ff) an Eastern Algonquian group was recognized by at least 1650, and the inclusion of the Algonquian languages of the Great Lakes region with it in a larger 'Algonkin' group was stated explicitly by Louis-Armand de Lom d'Arce, baron de Lahontan in 1703. Bloomfield's 1925 reconstruction of the sound system of what he called 'Central Algonquian' is generally regarded as the starting point for subsequent work on the comparative phonology of the Algonquian languages. Some 4,066 lexical reconstructions have been proposed for Proto Algonquian (Hewson 1993), but this figure reportedly is inflated by many morphologically related forms (Ives Goddard, p.c., March 17, 2004).

2. Arawakan: Noble (1965:1) states that a substantial portion of the Arawakan family 'was recognized as long ago as 1782 by Filippo Salvadore Gilij, an Italian missionary working

in Venezuela.' Gilij called this proposed grouping 'Maipuran', a name now used for the core group of languages which are indisputably related, and that form the nucleus of a more speculative 'Macro-Arawakan' group (Campbell 1997:178ff). Many Arawakan languages are now extinct, hindering attempts to determine the internal structure of the family and to reconstruct the phonology and lexicon of its earliest stages. The first reasonably systematic attempt to reconstruct Proto Arawakan phonology and lexicon apparently was that of Matteson (1972), who posited a phonological system with 22 consonants and 12 vowels (6 oral and six nasalized). Shortcomings in this work were noted by Payne (1991), who proposed 19 consonants and six vowels, plus 203 carefully justified lexical reconstructions for 'Proto Maipuran' which apparently was ancestral to all extant Arawakan languages.

   3.   Australian: In 1841 the English sea captain George Grey drew attention to widespread similarities in phonology, lexicon and pronoun systems among languages across the entire southern half of Australia (Dixon 1980:11). This is sometimes taken as the first recognition of an Australian language family, although most contemporary Australianists probably would object that the languages Grey compared belong to a single subgroup (Pama-Nyungan) which covers about seven-eighths of the Australian landmass. Capell (1956) spoke of 'Common Australian' phonology, noting that there is little variation in either consonant or vowel systems over much of the continent. Dixon (1980) discussed various features of 'Proto Australian' phonology, but without any clear use of the comparative method or subgrouping. Some Australianists therefore question whether a Proto Australian sound system has yet been reconstructed. O'Grady (1998:209) states that '1,561 putative cognate sets' are presented in an unpublished doctoral dissertation on Proto Pama-Nyungan defended by Susan Fitzgerald at the University of Victoria in British Columbia in 1997. These presumably are accompanied by reconstructions, as O'Grady himself presents 25 sample cognate sets together with etyma. Other Australianists, however, suggest that no more than about 300 reliable reconstructions have been proposed for Proto Pama-Nyungan to date, with even less done for Proto Australian (Nick Evans, p.c., March 4, 2004).

   4.   Austroasiatic: According to Ruhlen (1987:150ff) the Munda languages of eastern India were first recognized as a group distinct from Dravidian by Friederich Max Müller in 1854, a necessary precursor to establishing the Austroasiatic language family. Shortly thereafter J.R. Logan and F. Mason drew attention to lexical similarities between Munda and Mon-Khmer. After several decades of skepticism and reinterpretation the language family was firmly established—along with several major errors of omission and commission—by Wilhelm Schmidt in 1906. The reconstruction of a reasonably systematic sound system for Proto Mon-Khmer had to await the work of Pinnow (1959). Although the genetic relationship of Munda and Mon-Khmer is no longer an issue, little reconstruction has been done on the Proto Austroasiatic level. Gérard Diffloth (p.c., June 30, 2004) has compiled a computerized database with about 1,450 reconstructions for Proto Mon-Khmer, along with an appendix that contains about 300 Munda-Mon-Khmer comparisons. To date it has been made available only on an individual basis.[123]

   5.   Austronesian: As noted earlier, the first explicit recognition of a historical connection between two widely separated AN languages appears in de Houtman (1603). Somewhat over a century later Reland (1708) incorporated these languages with several from western Polynesia, into a larger, still unnamed grouping. Brandstetter (1906) compiled a preliminary comparative dictionary that was based only on material from the languages of insular Southeast Asia, and the

---

[123]   Hayes (1992 and sequels) has proposed a number of etymologies linking Austroasiatic to AN. In doing this he has proposed drastic revisions to the Proto Austroasiatic consonant system reconstructed by Pinnow (1959). However, his work contains serious methodological lapses and many of his comparisons are therefore unlikely to withstand close scrutiny by others within the Austroasiatic field.

comparative phonology of AN was placed on a firm foundation by Dempwolff (1934-1938), who reconstructed 2,213 lexical bases for 'Uraustronesisch'. About 600 of these are now rejected because they are restricted to languages in western Indonesia that are closely related to one another, or that have borrowed extensively from Malay, or both. Dempwolff's pioneering comparative dictionary has been superseded by Blust and Trussel (ongoing), an online comparative dictionary that currently contains 4,767 base entries that are not contained in a higher-level reconstruction, and over 13,000 morphological derivatives, but is less than 50% complete. Together with about 1,600 Dempwolff reconstructions that are yet to be updated and possibly upgraded, then, the number of reconstructed AN bases for which comparative evidence has been published to date, and that are not contained in a higher-level reconstruction is roughly 6,370. Some 1,290 of these are assigned to PAN (c. 5,500-6,000 BP), and 2,730 to PMP (c. 4,500-5,000 BP), as of April, 2013. However, numbers for nodes below PAN are deflated, since only the highest-level reconstructions are reported. A PAN reconstruction with reflexes in Formosan, Philippine and Oceanic languages, for example, automatically entails correlated reconstructions for PMP and all intermediate nodes through to Proto Oceanic, but is counted only as PAN, since the implied reconstructions for lower-order proto-languages are not headwords in a dictionary entry. All earlier reconstructions of the lexicon of Proto Oceanic, spoken c. 3,500 BP, have been superseded by the superlative work of Ross, Pawley and Osmond (1998, 2003, 2008, 2011, 2013, and volumes yet to appear).

6. Bantu: Although recognition of the Niger-Congo (or Niger-Kordofanian) family of which they are a part awaited the work of Greenberg in the 1950s, the genetic unity of a number of the western Bantu languages was explicitly noted as early as 1776 by the Catholic missionary Abbé Proyart, the relationship between southern and eastern Bantu languages was observed by Liechtenstein in 1808, and in 1818 William Marsden stated that the resemblances between western and eastern Bantu languages probably point to a prehistoric community of origin (Marten 2006). Meinhof (1899) posited a sound system for Proto Bantu, a proto language with a time-depth not likely to exceed 2,500 years. Schadeberg (2002) reports wide variation in the reported number of reconstructed Proto Bantu forms, much of it dependent upon subgrouping assumptions. In a personal communication of March 24, 2004 he suggested a figure of 1,000-1,500 reconstructed forms that go back to Proto Bantu, but the number of lexical reconstructions for ancestors of individual subgroups is as much as 4,000 in some cases. However, the time-depth of the multilayered Bantu subgroups and the numbers of reconstructed forms that represent interstage proto languages appear to be surrounded with uncertainty. Stewart (2002) has begun the reconstruction of Proto Niger Congo, but this remains in an incipient stage.

7. Dravidian: Krishnamurti (2003:16-17) credits the recognition of a Dravidian language family to the British civil servant Francis Ellis in 1816. Partial reconstructions of Proto Dravidian phonology were available earlier, but the first systematic account apparently is that of Krishnamurti (1961). Burrow and Emeneau (1984) propose 5,569 cognate sets, but suggest no etyma because of uncertainty concerning the reconstruction of vowels, and no more complete comparative dictionary has been compiled since.

8. Indo-European: By at least 1599 it was recognized that most languages of Europe share certain lexical similarities. However, these generally were attributed to borrowing, and the first clear recognition of an Indo-European language family was made by Sir William Jones in his famous address to the Asiatic Society of Calcutta in 1786. The reconstruction of Proto Indo-European phonology has undergone periodic revisions from the time of the Neogrammarians through the work of Saussure to the more recent 'glottalic theory'. Details aside, however, it seems fair to say that the main outlines of the phonology were in place by the time of Schleicher (1861-62), or at latest, by the time of Verner (1875). The most fully documented set of lexical

reconstructions for PIE continues to be Pokorny (1959) a work that contains around 2,215 base entries, but is considered by most Indo-Europeanists to be outdated. Many of Pokorny's bold-face entries reportedly have "no real existence (they involve old borrowings, or other very shadowy material based on one or two branches only, etc.), and conversely, many of the larger entries would now actually be split up into several different roots" (Brent Vine, p.c., January 4, 2005). More recent works propose a larger number of Proto Indo-European base forms, as with Mallory and Adams (1997:661-680), which contains about 4,200 entries, but without full supporting evidence.

   9. Na-Dene: According to Ruhlen (1987:197) "Gallatin, in his 1836 classification of North American Indian languages, recognized the Athabaskan family, though only its Northern varieties." The Na-Dene hypothesis, which connected Athapaskan, Tlingit and Haida was proposed by Edward Sapir in 1915. Eyak, spoken in southern Alaska and unknown in Sapir's time, was subsequently added as the next-of-kin to Athapaskan (Krauss 1979:845ff). The inclusion of Haida in this grouping has since been brought into serious question, and is generally rejected today (Krauss and Golla 1981:67). Sapir (1931) reconstructed a sound system only for Proto Athapaskan, a subgroup of Na-Dene with a time-depth of not much more than 2,000 years; his reconstruction was, moreover, limited to obstruent consonants. A full reconstruction of Proto Athapaskan-Eyak phonology was done by Krauss (1964). The reconstruction of Proto Na-Dene has yet to make significant progress.

   10. Semitic: The genetic relationship of the Semitic languages of the Near East has been known for an indeterminately long time. According to Ruhlen (1987:87) "A Frenchman, Guillaume Postel, reported the affinity of Hebrew, Arabic, and Aramaic in 1538, a connection long known to Jewish and Islamic scholars. In 1702 Hiob Ludolf extended this Semitic nucleus to include the Ethiopic Semitic languages of East Africa, and finally in 1781 von Schlözer proposed the name Semitic." Recognition of the full extent of the Semitic group can thus be dated to 1702, but recognition of the genetic relationship of at least some Semitic languages preceded this by centuries. The earliest reconstruction of Proto Semitic phonology and grammar is credited to Brockelmann (1908-13), but the number of reconstructed bases is more difficult to determine. Fonzaroli (1975) states that while the comparative phonology and morphology of Semitic languages has made great advances, there is no Semitic comparative dictionary, or large corpus of Proto Semitic reconstructions. He himself finds somewhat over 500 lexemes that are attested in at least one language or dialect from each of the three main geographical areas (East, Northwest and Southwest Semitic). Many of these (85) are concerned with human and sometimes animal anatomy and physiology. Other, more ambitious projects are promised on the internet, but surprisingly for such a well-studied group of languages, no published work to date appears to exceed the number cited by Fonzaroli.

   11. Sino-Tibetan: According to Ruhlen (1987:143) the large Tibeto-Burman group was recognized in 1828 by B.H. Hodgson, and in 1896 August Conrady proposed a larger family that included Chinese, Tai, and Tibeto-Burman. Since the earlier date concerns a part of the Sino-Tibetan family, and the later date a collection of languages larger than the Sino-Tibetan family it can be argued that the Sino-Tibetan family as such was not recognized until well into the twentieth century. However, priority is given to the earlier date, since Tibeto-Burman languages constitute some 95% of the Sino-Tibetan family, and the entire grouping is included by Conrady, even if it is excessively broad. The phonology of Proto Sino-Tibetan was not approached systematically until Benedict (1972). Benedict (1972) lists over 300 reconstructions for Proto Sino-Tibetan. Graham Thurgood (p.c.) states that these are still generally considered valid, and estimates that the total number of 'reasonably established' PST etyma is around 450, although not all of these have been brought together in a single place. Matisoff (2003) has over 700 reconstructions for Proto Tibeto-Burman, together with supporting evidence.

12. Tupi-Guarani: Tupi-Guarani is sometimes called the most extensive language family in South America, but Arawakan has a greater East-West spread (including the Caribbean), and a total territory that is nearly the same. A named Tupi-Guarani language family apparently was first explicitly noted in Adelung (1806-17). However, virtually all writers on the subject make the following points: 1. the 40-50 languages in this family are very closely related, despite a wide geographical dispersal, and 2. at the time of European contact speakers of these languages were expanding rapidly along the coast of Brazil, so that very similar forms of speech were found from the mouth of the Amazon to the Rio de la Plata. The Portuguese called this coastal language Tupinamba, or the Lingua Geral. While this name in itself shows that the Portuguese at the end of the fifteenth century recognized the relationship of widely dispersed Tupinamba communities, it does not show that there was a corresponding awareness of a larger language family. According to Lemle (1971:107) "It is evident from the most superficial inspection of word lists that there is a remarkable similarity among the nearly forty dialects that constitute the Tupi-Guarani family." Given the self-evident nature of this relationship it is possible that the relationship of distinct Tupian languages was recognized earlier than Adelung (1806-17), but evidence for this remains elusive. Some writers, as Rodrigues (1985) and Campbell (1997) treat Tupi-Guarani as a subgroup of a larger Tupi language family, but since reconstruction has so far been undertaken only on the Proto Tupi-Guarani level I focus on it here. Lemle (1971) proposes a 'tentative reconstruction of the phonological system and a restricted portion of the lexicon of Proto Tupi-Guarani.' The reconstructed lexicon contains just 221 items together with their reflexes in ten languages representing a wide geographical swath and both highest-order subgroups that she recognizes.

13. Turkic: According to Ruhlen (1987:128) the Turkic ('Tatar') family, was first recognized by Phillip Johann von Strahlenberg in 1730. However, Stefan Georg (p.c., Feb. 25, 2004) points out that, apart from Chuvash, "the Turkic language family is inspectionally obvious as a genetic grouping," and the entire Turkic grouping apart from the Siberian Turkic languages, which were unknown until recent times, was recognized in print by Mahmud al-Kashgari, a Turkic philologist who lived in the eleventh century. The Proto Turkic phonological system was first reconstructed by Räsänen (1949), and a Proto Turkic dictionary within the Russian Etymological Project reportedly contained about 2,100 entries early in 2004, with an expected total upon completion of about 3,000 entries (Alexander Vovin, p.c., Feb. 25, 2004).

14. Uralic: Ruhlen (1987:65) attributes the earliest recognition of a Finnish-Hungarian relationship to the Swedish scholar Georg Stiernhielm in 1671, and states that the full recognition of the Uralic family had been achieved by 1770. For Finno-Ugric Collinder (1955) proposes about 1,025 cognate sets, but no reconstructions. This set of lexical comparisons is not much larger than that of Budenz (1873-1881), who proposed some 996 cognate sets linking Finnic to Hungarian, although many of these are now considered to be erroneous. Janhunen (1981) has reconstructed some 140 Proto Uralic forms through a comparison of Proto Samoyedic and Proto Finnic, and has stated the sound correspondences linking the supporting cognate sets in the form of 58 rules for vocalism and 12 for consonantism (John Kupcik p.c., Sept. 4, 2004).

15. Uto-Aztecan: Aztecan languages were contacted by the Spanish early in the sixteenth century and led to some grammatical descriptions, but a comparative interest in this family did not begin until three centuries later. The recognition of the Uto-Aztecan family is generally credited to Buschmann (1859), although this is problematic, since he recognized a nonrandom similarity between the Aztecan languages of central Mexico and their relatives further north, but attributed it to contact. Credit for initial reconstruction of Proto Uto-Aztecan phonology is usually given to Sapir (1913-14). The first substantial set of lexical reconstructions was due to Miller (1967), who proposed about 500 starred forms, but suggested that these 'represent a

shorthand notation to enable the reader to see what phonemes have been compared' rather than being explicitly justified etyma (John C. McLoughlin p.c.). More recently, in a privately circulated study Stubbs (2008) has proposed some 2,622 lexical reconstructions, a number of which, however, represent a single branch of Uto-Aztecan (Lyle Campbell, p.c.).

Although these remarks take us outside the realm of AN linguistics, they provide perspective on the progress of comparative scholarship in the AN language family in relation to other major language families. It is clear from these tables that AN is not only one of the largest and most widely distributed language families, but was one of the first to be recognized. Phonological reconstruction was advanced earlier in AN than in most other large language families, and the quality and quantity of reconstructed vocabulary today probably is second to none.

11.5 Bibliographies of Austronesian linguistics

A number of bibliographies of AN linguistics have been compiled. More than with most publications, bibliographies rapidly become dated. For this reason I cite only books that have appeared since 1970, unless nothing else is available. The most important of these are as follows:

Taiwan: 1) Li (1992). This work contains a 38 page bibliography of Formosan linguistics (somewhat over 600 entries), and is reasonably complete through about 1990. It also contains a rather desultory list of references for comparative AN linguistics outside Taiwan. Not annotated.

Philippines: 1) Ward (1971). An exhaustive, partially annotated guide to the earlier literature from the Spanish colonial period, as well as much work in English up to about 1970. Contains around 3,300 entries for published works, and lists about 740 unpublished manuscripts on Philippine languages. 2) Asuncion-Landé (1971). A bibliography of Philippine linguistics with 1,977 entries. Lightly annotated. This work was eclipsed by Ward (1971) from the moment of its publication. 3) Makarenko (1981). An annotated bibliography of Tagalog/Pilipino, with 1,778 entries. The length of this list provides a clear indication of the extraordinary amount of attention that Tagalog has received in relation to many other AN languages. 4) Shinoda (1990). An annotated bibliography of both published work and unpublished manuscripts on Philippine languages by Japanese scholars between 1902 and 1989. 5) Hendrickson and Newell (1991). A thematically focused bibliography, concentrating on dictionaries and vocabularies of Philippine languages, with over 700 entries. This duplicates much material that is available in earlier bibliographies, but offers the convenience of exclusivity for those interested in Philippine lexicography. 6) Johnson (1996). A large compilation (3,908 entries), said (viii) to be 'a combination of several earlier bibliographic works.' This bibliography is, unfortunately, marred by very spotty coverage (e.g. some minor writers, as Laurance L. Wilson, are covered in great detail, while major contributors to Philippine linguistics such as R. David Paul Zorc are all but omitted). Not annotated. 7) Yamada (1997). This narrowly-focused but ambitious bibliography of the Bashiic-speaking peoples of the northern Philippines and Taiwan (primarily Ivatan, Itbayaten, and Yami) contains an astonishing 1,750-1,800 entries. It is, however, poorly constructed for its purpose, including some publications with little evident relationship to the topic, and numerous unpublished manuscripts down to the level of 'notes for a talk'. One of its more valuable features is a 7-page 'Bibliography of bibliographies' which precedes the main body of the text. 8) Johnson, Tan, and Goshert (2003). A bibliography celebrating the 50th anniversary of the work of the Summer Institute of Linguistics in the Philippines. Contains about 3,150 entries for both published and unpublished work by members of the Summer Institute of Linguistics, Philippines Branch.

Indonesia-Malaysia: 1) Voorhoeve (1955). A valuable guide to the older Dutch literature on the languages of Sumatra, with 205 entries, plus general reference works. Like other volumes in the same series, this book is annotated, contains short (5-6 page) sketches of individual languages or language groups, a language map, and photographs of some of the pioneering investigators in the area. Although somewhat dated, this is still the best available source for the languages of Sumatra. 2) Cense and Uhlenbeck (1958). A useful guide to the older literature on languages of Borneo, containing 323 entries, plus some general reference works. Very dated today, as a fairly large number of publications on the languages of Borneo have appeared since it was compiled. 3) Teeuw (1961). A meticulous piece of bibliographical scholarship, this book contains over 1,050 general entries for Malay and Bahasa Indonesia, together with an appendix of practical manuals, textbooks, school books and the like with roughly another 340 titles. Easily the best source for work published before 1960, but now very dated. 4) Uhlenbeck (1964). A thorough survey of the islands of Java and Madura which contains about 270 entries for Sundanese, 420 for Javanese, 450 for Old Javanese, and 220 for Madurese. Excellent for work done through the early 1960s, but now dated. 5) Pusat (1976). A specialized bibliography devoted to dictionaries. Contains about 320 entries for Malay/Indonesian dictionaries, mostly bilingual dictionaries (Indonesian-Dutch, Indonesian-English, etc.), as well as about 250 entries for dictionaries of other languages of Indonesia, and roughly 240 entries for monolingual dictionaries of special terminology (abbreviations/acronyms, encyclopedias, geological terms, etc.). The definition of 'dictionary' is loose, as some entries refer to short wordlists. Written in Indonesian; unannotated. 6) Collins (1990). A richly annotated survey of works on the Malay dialects of Borneo, with 359 entries. Virtually exhaustive through the late 1980s. 7) Noorduyn (1991). The last volume to appear in the Bibliographical Series of the Koninklijk Instituut voor Taal-, Land- en Volkenkunde. Because it was published more than 30 years after most of its companion volumes, it is far more up to date. Nonetheless, much important work on the languages of Sulawesi has appeared over the past 15 years, making even this bibliography more useful for the older Dutch literature than for work done in the postcolonial period. Of 228 pages, 92 are devoted exclusively to the South Sulawesi languages, reflecting the inordinate amount of attention that has been devoted to both Buginese and Makasarese as compared to most other languages of the island. 8) Collins (1995a). A detailed listing of all known sources for Malay dialects in Sumatra. Updates Voorhoeve (1955) with regard to this one language; written in Indonesian. 9) Collins (1995b). Like the preceding entry, a thorough and richly annotated survey of works on the Malay dialects of Java, Bali, and Sri Lanka, with 687 entries. Virtually exhaustive through the early 1990s; written in Indonesian. 10) Collins (1996). A virtually exhaustive compilation of published sources on Malay dialects in eastern Indonesia up to the middle 1990s; written in Indonesian. 11) Collins (Ms.). Easily the most complete bibliography of peninsular Malay dialects currently available.

New Guinea area: 1) Carrington (1996). Twenty years in the making, this enormous bibliography contains over 14,000 entries on Papuan and AN languages of the New Guinea area (western Pacific and adjoining parts of eastern Indonesia). It casts its net very wide, incorporating many publications that have no obvious connection to the geographical region in question, but which evidently are included because their author has published *other* works that are relevant. In an effort to be maximally inclusive it also lists typescripts by members of the Summer Institute of Linguistics that probably are irretrievable for most interested scholars, anthropological works with miniscule linguistic content, and general books on the Pacific that do little more than mention language names with no linguistic data of any kind. Because it is organized alphabetically by author's last name without reference to the genetic affiliation of the languages, it is difficult without a major investment of time and effort, to determine roughly how many entries relate to AN languages.

Vanuatu: 1) Lynch and Crowley (2001). A survey of the languages of Vanuatu intended to update Tryon (1976), this book also includes a bibliography with about 480 entries covering both the indigenous languages of Vanuatu and Bislama.

Oceanic: 1) Klieneberger (1957). The only bibliography on the Oceanic languages as a whole, this book is now extremely dated. However, it was thorough and useful for its time, with some 2,166 entries covering the full range of Oceanic languages, as well as pidgin and creole languages.

This brief survey exposes the gaps that remain in bibliographic coverage of the AN languages. In bibliographic terms the Philippines as a whole, and Malay dialects are probably the most well-represented areas. For most non-Malay languages of Indonesia, bibliographic coverage is seriously inadequate. The Dutch sources from the 1950s and 1960s, which were excellent for their time, are now far out of date. There is no bibliography of Malagasy linguistics, Chamic linguistics, or of the languages of New Caledonia and the Loyalty Islands. Finally, Klieneberger's bibliography of Oceanic linguistics is now more than half a century old. While a comprehensive bibliography of Oceanic linguistics may seem a daunting task given the enormous amount of work that has been done on these languages over the past thirty years, even more restricted areas in the Pacific have been surprisingly neglected. Although there have been detailed bibliographies of some individual languages, as Tagalog and Malay, for example, there has never been a comprehensive bibliography for the Polynesian (or Central Pacific) languages as a group, or for the languages of Micronesia.

References

Abbreviations used for journals and other periodical publications:

AA   *American Anthropologist*
AP   *Asian Perspectives* (Honolulu)
BIHP   *Bulletin of the Institute of History and Philology, Academia Sinica*. Began in China
              in 1929, transferred to Taiwan with the Nationalist government in 1945. The
              Institute of Linguistics split off as an independent unit in 1997, with a new journal
              Language and Linguistics beginning in 2000, at which time *BIHP* ceased to exist.
BKI   *Bijdragen tot de Taal-, Land- en Volkenkunde*. Began in 1851; until 1949 was
              Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië; earlier
              abbreviation was BTLV.
BRB   *Borneo Research Bulletin*. Published by the Borneo Research Council, Department
              Of Anthropology, College of William and Mary, Williamsburg, Virginia.
BSLP   *Bulletin de la société linguistique de Paris*.
BSOAS   *Bulletin of the School of Oriental and African Studies*, University of London.
CA   *Current Anthropology*.
CAAAL   *Computational Analyses of Asian & African Languages* (Tokyo).
ILDEP   *Indonesian Languages Development Project*. Series of occasional publications on
              languages of Indonesia, University of Leiden.
JAOS   *Journal of the American Oriental Society*.
JASO   *Journal of the Anthropological Society of Oxford*.
JMBRAS   *Journal of the Malayan Branch of the Royal Asiatic Society*.
JPS   *Journal of the Polynesian Society*. Published in New Zealand since 1891.
JRAI   *Journal of the Royal Anthropological Institute of Great Britain and Ireland*.
JRAS   *Journal of the Royal Asiatic Society*.
KITLV   *Koninklijk Instituut voor Taal-, Land- en Volkenkunde*. Royal Institute of
              Linguistics, Geography and Ethnology, Holland (not part of the *Verhandelingen*).
LACITO   *Langues et civilisations a tradition orale*. Paris: Centre National de la Recherche
              Scientifique.
              LL Language and Linguistics. Institute of Linguistics, Academia Sinica, Taipei.
LLMS   *Language and Linguistics Monograph Series*. Institute of Linguistics, Academia
              Sinica, Taipei.
LSP   *Linguistic Society of the Philippines*.
NUSA   *NUSA, Linguistic Studies in Indonesian and Other Languages in Indonesia* Series of
              occasional publications issued through Atma Jaya Catholic University, Jakarta.
OL   *Oceanic Linguistics*. Journal, began at Southern Illinois University, Carbondale in
              1962; shifted to the University of Hawaii the following year, where it has remained
              ever since.
OLSP   *Oceanic Linguistics Special Publications*. Monograph series associated with the
              journal *Oceanic Linguistics*, and issued through the University of Hawaii Press,
              Honolulu.
PJL   *The Philippine Journal of Linguistics*: published by the Linguistic Society of the
              Philippines since 1970 (Manila).

PL   *Pacific Linguistics.* Series of occasional publications issued by Department of
            Linguistics, Research School of Pacific and Asian Studies, Australian National
            University, Canberra. Began as 'Linguistic Circle of Canberra Publications', and the
            institutional unit was earlier called the 'Research School of Pacific Studies'. The
            numbering system changed in 2000.
SELAF   *Société d'études linguistiques et anthropologiques de France* (Paris).
SLCAA   *Study of languages & cultures of Asia and Africa* (Tokyo).
SMJ   *The Sarawak Museum Journal.* Published by the Sarawak Museum, Kuching, since
            1913.
SPLC   *Studies in Philippine Languages and Cultures.* Published by Linguistic Society of
            the Philippines and Summer Institute of Linguistics, Philippines since 1977.
UHWPL   *University of Hawaii Working Papers in Linguistic*s (Honolulu).
VBG   *Verhandelingen van het Bataviaasch Genootschap van kunsten en Wetenschappen.*
            Proceedings of the Batavia Society of Arts and Sciences, published in Indonesia
            during the Dutch colonial period.
VKI   *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde.*
            Monograph series issued through the Royal Institute of Linguistics, Geography and
            Ethnology, Leiden; formerly in The Hague.
VSIS   *Veröffentlichungen des Seminars für Indonesische und Südseesprachen der
            Universität Hamburg.* Monograph series issued through Department of Indonesian
            and Pacific Island Languages, University of Hamburg.
WILC   *Workpapers in Indonesian Languages and Cult*ures. Published by the Summer
            Institute of Linguistics in collaboration with Cenderawasih University in Irian Jaya,
            Hasanuddin University in Sulawesi, and Pattimura University in Maluku, and with
            the cooperation of The Department of Education and Culture, Republic of Indonesia.
ZfE   *Zeitschrift für Ethnologie.* Published in Berlin and then Braunschweig since 1869.
ZfES   *Zeitschrift für Eingeborenen-Sprachen.* Published in Berlin since 1910; began as
            Zeitschrift für Kolonial-Sprachen.

Abinal, R.R. and P.P. Malzac. 1970 [1888]. *Dictionnaire malgache-français*. Paris: Editions
        Maritime et d'Outre-Mer. Marseille.
Abo, T., B.W. Bender, A. Capelle and T. DeBrum. 1976. *Marshallese-English dictionary*.
        PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.
Adam, T. and J.P. Butler. 1948 [1922]. *Grammar of the Malay language*. New York: Hafner
        Publishing Co.
Adelaar, K.A. 1981. Reconstruction of Proto-Batak phonology. In Blust: 1-20.
_____ 1983. Malay consonant-harmony: an internal reconstruction. In Collins: 57-67.
_____ 1989. Malay influence on Malagasy: linguistic and culture-historical implications.
        OL 28: 1-46.
_____ 1991. Some notes on the origin of Sri Lanka Malay. In H. Steinhauer, ed., *Papers in
        Austronesian Linguistics*, No. 1: 23-37. Canberra: Pacific Linguistics.
_____ 1992. *Proto-Malayic: the reconstruction of its morphology and parts of its lexicon
        and morphology*. Canberra: Pacific Linguistics.
_____ 1994a. The classification of the Tamanic languages. In Dutton and Tryon: 1-42.
_____ 1994b. Malay and Javanese loanwords in Malagasy, Tagalog and Siraya (Formosa).
        BKI 150: 50-65.
_____ 1996. Malay in the Cocos (Keeling) islands. In Nothofer: 167-198.
_____ 1997a. Grammar notes on Siraya, an extinct Formosan language. *OL* 36: 362-397.
_____ 1997b. An exploration of directional systems in West Indonesia and Madagascar. In
        Senft: 53-81.

_____ 2004. A la recherche d'affixes perdus en malais. In Zeitoun: 165-176.

_____ 2005a. *Salako or Badameà: sketch grammar, texts and lexicon of a Kanayatn dialect in West Borneo*. Wiesbaden: Harrassowitz.

_____ 2005b. The Austronesian languages of Asia and Madagascar: a historical perspective. In Adelaar and Himmelmann: 1-42.

_____ 2005c. Malayo-Sumbawan. *OL* 44: 357-388.

_____ 2010. The amalgamation of Malagasy. In Bowden, Himmelmann and Ross: 161-178.

_____ 2011. *Siraya: retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Berlin: de Gruyter Mouton.

_____ 2012. Malagasy phonological history and Bantu influence. *OL* 51: 123-159.

Adelaar, K.A. and R. Blust, eds. 2002. *Between worlds: linguistic papers in memory of David John Prentice*. Canberra: Pacific Linguistics.

Adelaar, K.A. and N.P. Himmelmann, eds. 2005. *The Austronesian languages of Asia and Madagascar*. London and New York: Routledge.

Adelaar, K.A. and A.K. Pawley, eds. 2009. *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Canberra: Pacific Linguistics.

Adelung, J.C. 1806-1817. 4 vols. *Mithridates*. Berlin.

Adriani, N. 1893. *Sangireesche spraakkunst*. Leiden: Nederlandsch Bijbelgenootschap.

_____ 1928. *Bare'e-Nederlandsch woordenboek*. Leiden: Brill.

Akamine, J. 2002. The Sinama derived transitive construction. In Wouk and Ross: 355- 366.

Alkire, W.H. 1977. *An introduction to the peoples and cultures of Micronesia* (2nd edition). Menlo Park, California: Cummings Publishing Co.

Ameda, C., G.A. Tigo, V.B. Mesa and L. Ballard. 2011. *Ibaloy dictionary*. Asheville, North Carolina: Biltmore Press.

Anceaux, J.C. 1952. *The Wolio language*. *VKI* 11. The Hague: Nijhoff.

_____ 1961. *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. *VKI* 35. The Hague: Nijhoff.

_____ 1965. Linguistic theories about the Austronesian homeland. *BKI* 121: 417-432.

_____ 1987. *Wolio dictionary (Wolio-English-Indonesian)*. *KITLV*. Dordrecht, Holland: Foris Publications.

Anderbeck, K. 2007. An initial reconstruction of Proto-Lampungic: phonology and basic Vocabulary. *SPLC* 16: 41-165.

Andersen, D. 1999. Moronene numbers. In D. Mead, ed., *Studies in Sulawesi Linguistics*, Part V: 1-72. NUSA 45.

Andersen, E.S. 1978. Lexical universals of body-part terminology. In Greenberg, ed., 3: 335-368.

Anderson, S.R. 1972. On nasalisation in Sundanese. *Linguistic Inquiry* 11.3: 253-268.

Antworth, E.L. 1979. *A grammatical sketch of Botolan Sambal*. Manila: LSP.

Arka, I W. and M. Ross, eds. 2005. *The many faces of Austronesian voice systems*: *some new empirical studies*. Canberra: Pacific Linguistics.

Arms, D.G. 1973. Whence the Fijian transitive endings? *OL* 12: 503-558.

Arndt, P. 1961. *Wörterbuch der Ngadhasprache*. Studia Instituti Anthropos, Vol. 15. Posieux, Fribourg, Switzerland.

Asuncion-Landé, N.C. 1971. *A bibliography of Philippine linguistics*. Athens, Ohio: Center for International Studies, Ohio University.

Atkinson, Q.D. and R.D. Gray. 2005. Are accurate dates an intractable problem for historical linguistics? In C. Lipo, M. O'Brien, S. Shennan and M. Collard, eds., *Mapping our Ancestry: Phylogenetic Methods in Anthropology and Prehistory*: 269-296. Chicago: Aldine.

Atkinson, Q., G. Nicholls, D. Welch and R. Gray. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103.2: 193-219.

Austin, P.K. 2000a. Verbs, voice and valence in Sasak. In Austin: 5-24.

_____ ed. 2000b. *Sasak: Working Papers in Sasak*, vol. 2. Melbourne: Department of Linguistics and Applied Linguistics, The University of Melbourne.

Ayed, S.A., L.B. Underwood, and V.M. van Wynen. 2004. *Tboli-English dictionary*. Manila: Summer Institute of Linguistics.

Aymonier, É. and A. Cabaton. 1906. *Dictionnaire cham-français*. Bulletin de l'école française d'extrême-orient 7. Paris: Leroux.

Baird, Louise. 2002. *A grammar of Kéo: An Austronesian language of East Nusantara*. Ph.D. thesis, Australian National University.

Baldi, P., ed. 1990. *Linguistic change and reconstruction methodology*. Trends in Linguistics Studies and Monographs 45. Berlin: Mouton de Gruyter.

Ball, Douglas. 2007. On ergativity and accusativity in Proto-Polynesian and Proto-Central Pacific. *OL* 46: 128-153.

Bancel, P.J. and A. Matthey de l'Etang. 2002. Tracing the ancestral kinship system: the global etymon KAKA. *Mother Tongue* VII: 209-222.

Barber, C.C. 1979. *Dictionary of Balinese-English*. 2 vols. Aberdeen University Library Occasional Publications, No. 2. University of Aberdeen.

Bareigts, A. 1987. *Notes on Kkef.falan*. Fengpin, Taiwan (privately printed).

Barnes, R.H. 1974. *Kédang: A study of the collective thought of an Eastern Indonesian people*. Oxford Monographs in Social Anthropology. Oxford: Clarendon Press.

_____ 1977. *Mata* in Austronesia. *Oceania* 48.4: 300-319.

_____ 1980. Fingers and numbers. *JASO* 11.3: 197-206.

Bauer, W. 1993. *Maori*. Descriptive Grammars Series. London: Routledge.

Beaujard, P. 1998. *Dictionnaire malgache-français: dialecte Tañala, Sud-Est de Madagascar*. Paris: l'Harmattan.

Beaumont, C.H. 1976. Austronesian languages: New Ireland. In Wurm 2: 387-397.

_____ 1979. *The Tigak language of New Ireland*. Canberra: Pacific Linguistics.

Behrens, D. 2002. *Yakan-English dictionary*. Manila: Linguistic Society of the Philippines.

Bell, F.L.S. 1977. *Tanga-English, English-Tanga dictionary*. Oceania Linguistic Monographs, No. 21. Sydney: Department of Anthropology, University of Sydney.

Bellwood, P. 1979. *Man's conquest of the Pacific: the prehistory of Southeast Asia and Oceania*. New York: Oxford University Press.

_____ 1997 [1985]. *Prehistory of the Indo-Malaysian archipelago*. 2nd, rev. edition. Honolulu: University of Hawaii Press.

Belo, M., J. Bowden, J. Hajek and N. Himmelmann. 2005. *The Waimaha language of East Timor*. Website accessible at http://rspas.anu.edu.au/linguistics/Waimaha/eng/sounds.html.

Bender, B.W. 1968. Marshallese phonology. *OL* 7: 16-35.

_____ 1969a. Vowel dissimilation in Marshallese. *Working Papers in Linguistics* 1.1: 88-96. Honolulu: Department of Linguistics, University of Hawaii.

_____ 1969b. *Spoken Marshallese*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

_____ 1971. Micronesian languages. In Sebeok: 426-465.

_____ W.H. Goodenough, F.H. Jackson, J.C. Marck, K.L. Rehg, H.M. Sohn, S. Trussel and J.W. Wang. 2003a. Proto-Micronesian reconstructions – 1. *OL* 42: 1-110.

_____ W.H. Goodenough, F.H. Jackson, J.C. Marck, K.L. Rehg, H.M. Sohn, S. Trussel and J.W. Wang. 2003b. Proto-Micronesian reconstructions – 2. *OL* 42: 271-358.

_____ ed. 1984. *Studies in Micronesian linguistics*. Canberra: Pacific Linguistics.

Benedict, P.K. 1941. A Cham colony on the island of Hainan. *Harvard Journal of Asiatic Studies* 4: 129-134.

_____ 1942. Thai, Kadai and Indonesian: a new alignment in southeastern Asia. *American Anthropologist* 44: 576-601.

_____ 1972. *Sino-Tibetan: a conspectus*. New York: Cambridge University Press.

_____ 1975. *Austro-Thai: language and culture, with a glossary of roots*. New Haven: Human Relations Area Files.

_____ 1984. Austro-Tai parallel: a tonal Cham colony on Hainan. *CAAAL* 22: 83-86.

_____ 1990. *Japanese/Austro-Tai.* Ann Arbor, Michigan: Karoma Publishers.

Benjamin, G. 1976. Austroasiatic subgroupings and prehistory in the Malay peninsula. In P.N. Jenner, L.C. Thompson and S. Starosta (eds.), *Austroasiatic Studies,* Part 1: 37-128. *OLSP* 13.

Bennardo, G., ed. 2002. *Representing space in Oceania: Culture in language and mind*. Canberra: Pacific Linguistics.

Benton, R.A. 1971. *Pangasinan reference grammar*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Benveniste, E. 1973 [1969]. *Indo-European language and society*. Miami Linguistics Series No. 12. Coral Gables, Florida: University of Miami Press.

Bergsland, K. and H. Vogt. 1962. On the validity of glottochronology. *CA* 3: 115-153 (with comments).

Berlin, B. and P. Kay. 1969. *Basic color terms: their universality and evolution*. Berkeley: University of California Press.

Bermejo, J. 1894. *Arte compendiado de la lengua Cebuana* (2nd edition). Tambobong. Pequena Tipo-litografia del Asilo de Huerfanos.

Besnier, N. 2000. *Tuvaluan: a Polynesian language of the central Pacific*. Descriptive Grammars Series. London and New York: Routledge.

Bhat, D.N.S. 1978. A general study of palatalisation. In Greenberg, 2: 47-92.

Biggs, B. 1965. Direct and indirect inheritance in Rotuman. *Lingua* 14: 383-415.

_____ 1971. The languages of Polynesia. In Sebeok: 466-505.

_____ 1994. New words for a new world. In Pawley and Ross: 21-29.

Blacking, J. 1977. *The anthropology of the body*. Association of Social Anthropologists monograph 15. London: Academic Press.

Blagden, C.O. 1902. A Malayan element in some of the languages of southern Indo- China. *Journal of the Straits Branch of the Royal Asiatic Society* 38: 1-27.

_____ 1916. Preface. In Brandstetter 1916: v-ix.

Blake, F.R. (1906). Expression of case by the verb in Tagalog. *JAOS* 27: 183-189.

_____ 1917. Reduplication in Tagalog. *American Journal of Philology* 38: 425-431.

_____ 1925. *A grammar of the Tagalog language, the chief native idiom of the Philippine islands*. American Oriental Series, Vol. 1. New Haven: American Oriental Society.

_____ 1930. A semantic analysis of case. In James T. Hatfield, *et al*, eds., *Curme volume of linguistics studies*: 34-49. Baltimore: Waverly Press.

Blench, R. and M. Spriggs, eds. 1997. *Archaeology and language I: theoretical and methodological orientations*. London and New York: Routledge.

Blevins, J. 2003. A note on reduplication in Bugotu and Cheke Holo. *OL* 42: 499-505.

_____ 2004a. The mystery of Austronesian final consonant loss. *OL* 43: 208-213.

_____ 2004b. *Evolutionary phonology: the emergence of sound patterns*. Cambridge University Press.

_____ 2005. The role of phonological predictability in sound change: privileged reduction in Oceanic reduplicated substrings. *OL* 44: 517-526.

_____ 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32: 117-166.

_____ 2007. A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands. *OL* 46: 154-198.

_____ 2008. Phonetic explanation without compromise: the evolution of Mussau syncope. *Diachronica* 25: 1-19.

_____ 2009a. Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *OL* 48: 264-273.

_____ 2009b. Low vowel dissimilation outside of Oceanic: the case of Alamblak. *OL* 48: 477-483.

Blevins, J. and A. Garrett. 1998. The origins of consonant-vowel metathesis. *Language* 74: 508-556.

_____ and S.P. Harrison. 1999. Trimoraic feet in Gilbertese. *OL* 38: 203-230.

_____ and D. Kaufman. 2012. Origins of Palauan intrusive velar nasals. *OL* 51: 18-32.

Blood, D.W. 1961. Women's speech characteristics in Cham. *Asian Culture* 3: 139-143.

_____ 1962. Reflexes of Proto-Malayo-Polynesian in Cham. *Anthropological Linguistics* 4.9: 11-20.

Bloomfield, L. 1917. 2 parts. *Tagalog texts with grammatical analysis*. University of Illinois Studies in Language and Literature. Urbana, Illinois: University of Illinois.

_____ 1925. On the sound system of Central Algonquian. *Language* 1: 130-156.

_____ 1933. *Language*. New York: Holt, Rinehart and Winston.

Blust, R. 1969. Some new Proto-Austronesian trisyllables. *OL* 8: 85-104.

_____ 1970a. Proto-Austronesian addenda. *OL* 9: 104-162.

_____ 1970b. *i* and *u* in the Austronesian languages. *Working Papers in Linguistics* 2.6: 113-145. Honolulu: Department of Linguistics, University of Hawaii.

_____ 1971. A Tagalog consonant cluster conspiracy. *PJL* 2: 85-91.

_____ 1972. Note on PAN *qa(R)(CtT)a 'outsiders, alien people'. *OL* 11: 166-171.

_____ 1973a. Additions to "Proto-Austronesian addenda" and "Proto Oceanic addenda with cognates in non-Oceanic Austronesian languages – II". *UHWPL* 5.3: 33-61.

_____ 1973b. The origins of Bintulu *ɓ, ɗ*. BSOAS 36: 603-620.

_____ 1974a. A double counter-universal in Kelabit. *Papers in Linguistics* 7: 309-324.

_____ 1974b. *The Proto-North Sarawak vowel deletion hypothesis*. Unpublished Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

_____ 1974c. A Murik vocabulary, with a note on the linguistic position of Murik. *SMJ* 22.43 (new series): 153-189.

_____ 1976a. A third palatal reflex in Polynesian languages. *JPS* 85: 339-358.

_____ 1976b. Austronesian culture history: some linguistic inferences and their relations to the archaeological record. *World Archaeology* 8: 19-43.

_____ 1977a. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9.2: 1-15. Honolulu: Department of Linguistics, University of Hawaii.

_____ 1977b. A rediscovered Austronesian comparative paradigm. *OL* 16: 1-51.

_____ 1977c. *Sketches of the morphology and phonology of Bornean languages 1: Uma Juman (Kayan)*. Canberra: Pacific Linguistics: 7-122.

_____ 1978a. *The Proto Oceanic palatals*. Memoir 43, The Polynesian Society. Wellington.

_____ 1978b. Eastern Malayo-Polynesian: a subgrouping argument. In Wurm and Carrington, Fasicle 2: 181-234.

_____ 1978c. Review of Thomas A. Sebeok, ed., *Current Trends Linguistics, vol. 8: Linguistics in Oceania*. *Language* 54: 467-480.

_____ 1979. Proto-Western Malayo-Polynesian vocatives. *BKI* 135: 205-251.

_____ 1980a. More on the origins of glottalic consonants. *Lingua* 52: 125-156.

_____ 1980b. Austronesian etymologies. *OL* 19: 1-181.

_____ 1980c. Iban antonymy: a case from diachrony? In D.J. van Alkemade et al, eds., *Linguistic studies offered to Berthe Siertsema*: 35-47. Amsterdam: Rodopi.

_____ 1981a. Some remarks on labiovelar consonants in Oceanic languages. In Hollyman and Pawley: 229-253.

_____ 1981b. The reconstruction of Proto-Malayo-Javanic: an appreciation. *BKI* 137: 456-469.

_____ 1981c. Linguistic evidence for some early Austronesian taboos. *AA* 83: 285-319.

_____ 1982a. The Proto-Austronesian word for 'female'. In Carle et al: 17-30.

_____ 1982b. The linguistic value of the Wallace Line. *BKI* 138: 231-250.

_____ 1982c. An overlooked feature of Malay historical phonology. *BSOAS* 45: 284- 299.

_____ 1983/1984a. More on the position of the languages of eastern Indonesia. *OL* 22/23: 1-28.

_____ 1983/1984b. Austronesian etymologies - II. *OL* 22/23: 29-149.

_____ 1984a. Malaita-Micronesian: an Eastern Oceanic subgroup? *JPS* 93: 99-140.

_____ 1984b. On the history of the Rejang vowels and diphthongs. *BKI* 140: 422-450.

_____ 1984c. A Mussau vocabulary, with phonological notes. Papers in New Guinea Linguistics No. 23: 159-208. Canberra: Pacific Linguistics.

_____ 1986. Austronesian etymologies - III. *OL* 25: 1-123.

_____ 1986/1987. Language and culture history: two case studies. *AP* 27: 205-227.

_____ 1987a. Rennell-Bellona /l/ and the "Hiti" substratum. In D.C. Laycock and W. Winter, eds., *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday*: 69-79. Canberra: Pacific Linguistics.

_____ 1987b. Lexical reconstruction and semantic reconstruction: the case of Austronesian "house" words. *Diachronica* 4: 79-106.

_____ 1988a. *Austronesian root theory: an essay on the limits of morphology*. Amsterdam/Philadelphia: John Benjamins.

_____ 1988b. Dempwolff's contributions to Austronesian linguistics. *Afrika und Übersee* 71.2: 90-96.

_____ 1988c. Sketches of the morphology and phonology of Bornean languages, 2: Mukah Melanau. In Steinhauer 1988b: 151-216.

_____ 1989a. The adhesive locative in Austronesian languages. *OL* 28: 197-203.

_____ 1989b. Austronesian etymologies – IV. *OL* 28: 111-180.

_____ 1991a. The Greater Central Philippines hypothesis. *OL* 30: 73-129.

_____ 1991b. On the limits of the "Thunder complex" in Australasia. *Anthropos* 86: 517-528.

_____ 1991c. Sound change and migration distance. In Blust, ed.: 27-42.

_____ 1992. On speech strata in Tiruray. In M.D. Ross, ed., *Papers in Austronesian Linguistics*, No. 1: 1-52. Canberra: Pacific Linguistics.

_____ 1993a. Kelabit-English vocabulary. *SMJ* 44.65 (new series): 141-226.

_____ 1993b. Central and Central-Eastern Malayo-Polynesian. *OL* 32: 241-293

_____ 1993c. *S metathesis and the Formosan/Malayo-Polynesian language boundary. In Øyvind Dahl, ed., *Language: a doorway between human cultures. Tributes to Otto Chr. Dahl on his ninetieth birthday*: 178-183. Oslo: Novus.

_____ 1993d. Austronesian sibling terms and culture history. *BKI* 149: 22-76 (also published in Pawley and Ross 1994: 31-72).

_____ 1994a. The Austronesian settlement of mainland Southeast Asia. In K.L. Adams and T.J. Hudak, eds. *Papers from the Second Annual Meeting of the Southeast Asian*

Linguistics Society: 25-83. Tempe, Arizona: Program forSoutheast Asian Studies, Arizona State University.

_____ 1994b. Obstruent epenthesis and the unity of phonological features. *Lingua* 93: 111-139.

_____ 1995a. Notes on Berawan consonant gemination. *OL* 34: 123-138.

_____ 1995b. Sibilant assimilation in Formosan languages and the Proto-Austronesian word for 'nine': a discourse on method. *OL* 34: 443-453.

_____ 1995c. The prehistory of the Austronesian-speaking peoples: a view from language. *Journal of World Prehistory* 9: 453-510.

_____ 1996a. Low vowel dissimilation in Ere. *OL* 35: 96-112.

_____ 1996b. Low vowel dissimilation in Oceanic languages: an addendum. *OL* 35: 305-309.

_____ 1996c. Notes on the semantics of Proto-Austronesian *-an 'locative'. In M. Klamer, ed., *Voice in Austronesian*: 1-11. NUSA 39.

_____ 1996d. The Neogrammarian hypothesis and pandemic irregularity. In Durie and Ross: 135-156.

_____ 1996e. Some remarks on the linguistic position of Thao. *OL* 35: 272-294.

_____ 1996f. The linguistic position of the Western Islands, Papua New Guinea. In Lynch and Pat: 1-46.

_____ 1997a. Semantic change and the conceptualisation of spatial relationships in Austronesian languages. In Senft: 39-51.

_____ 1997b. Ablaut in northwest Borneo. *Diachronica* 14: 1-30.

_____ 1997c. Nasals and nasalisation in Borneo. *OL* 36: 149-179.

_____ 1997d. Rukai stress revisited. *OL* 36: 398-403.

_____ 1997e. Review of Darrell T. Tryon, ed., *Comparative Austronesian dictionary: An introduction to Austronesian studies*. *OL* 36: 404-419.

_____ 1998a. Seimat vowel nasality: a typological anomaly. *OL* 37: 298-322.

_____ 1998b. Some problems in Thao phonology. In S. Huang, ed., *Selected Papers from the Second International Symposium on Languages in Taiwan (ISOLIT)*: 1- 20. Taipei: Crane.

_____ 1998c. A note on the Thao patient focus perfective. *OL* 37: 346-353.

_____ 1998d. Squib: a note on higher-order subgroups in Oceanic. *OL* 37: 182-188.

_____ 1998e. The position of the languages of Sabah. In M.L.S. Bautista, ed., *Pagtanáw: essays on language in honor of Teodoro A. Llamzon*: 29-52. Manila: LSP.

_____ 1998f. Ca- reduplication and Proto-Austronesian grammar. *OL* 37: 29-64.

_____ 1999a. Notes on Pazeh phonology and morphology. *OL* 38: 321-365.

_____ 1999b. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Zeitoun and Li: 31-94.

_____ 1999c. A note on covert structure: Ca- reduplication in Amis. *OL* 38: 168-174.

_____ 2000a. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. In C. Renfrew, A. McMahon and L. Trask, eds., *Time depth in historical linguistics*, vol. 2: 311-331. Papers in the Prehistory of Languages. Cambridge: The McDonald Institute for Archaeological Research.

_____ 2000b. Rat ears, tree ears, ghost ears and thunder ears in Austronesian languages. *BKI* 156: 687-706.

_____ 2000c. Chamorro historical phonology. *OL* 39: 83-122.

_____ 2000d. Low vowel fronting in northern Sarawak. *OL* 39: 285-319.

_____ 2001a. Some remarks on stress, syncope, and gemination in Mussau. *OL* 40: 143-150.

_____ 2001b. Thao triplication. *OL* 40: 324-335.

_____ 2001c. Reduplicated colour terms in Oceanic languages. In A.K. Pawley, M. Ross and D. Tryon, eds., *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*: 23-49. Canberra: Pacific Linguistics.

_____ 2001d. Historical morphology and the spirit world: the *qali/kali- prefixes in Austronesian languages. In J. Bradshaw and K.L. Rehg, eds., *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*: 15-73. Canberra: Pacific Linguistics.

_____ 2001e. Language, dialect and riotous sound change: the case of Sa'ban. In G.W. Thurgood, ed., *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*: 249-359. Arizona State University Program for Southeast Asian Studies Monograph Series. Tempe: Arizona State University.

_____ 2001f. Malayo-Polynesian: new stones in the wall. *OL* 40: 151-155.

_____ 2002a. Formalism or phoneyism? The history of Kayan final glottal stop. In Adelaar and Blust: 29-37.

_____ 2002b. The history of faunal terms in Austronesian languages. *OL* 41: 89-139.

_____ 2002c. Kiput historical phonology. *OL* 41: 384-438.

_____ 2002d. Notes on the history of 'focus' in Austronesian languages. In Wouk and Ross: 63-78.

_____ 2003a. *Thao dictionary*. *LLMS* A5. Taipei.

_____ 2003b. *A short morphology, phonology and vocabulary of Kiput, Sarawak*. Canberra: Pacific Linguistics.

_____ 2003c. Three notes on Early Austronesian morphology. *OL* 42: 438-478.

_____ 2003d. The phonestheme ŋ- in Austronesian languages. *OL* 42: 187-212.

_____ 2003e. Vowelless words in Selau. In Lynch: 143-152.

_____ 2004a. Austronesian nasal substitution: a survey. *OL* 43.1: 73-148.

_____ 2004b. *t to k: an Austronesian sound change revisited. *OL* 43: 365-410.

_____ 2005a. Review of Lynch, Ross and Crowley, *The Oceanic Languages*. *OL* 44: 536-550.

_____ 2005b. A note on the history of genitive marking in Austronesian languages. *OL* 44: 215-222.

_____ 2005c. The linguistic macrohistory of the Philippines: some speculations. In Liao and Rubino: 31-68.

_____ 2005d. Whence the Malays? In J.T. Collins and A. Sariyan, eds., *Borneo and the homeland of the Malays: four essays*: 64-88. Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ 2005e. Must sound change be linguistically motivated? *Diachronica* 22: 219-269.

_____ 2006a. The origin of the Kelabit voiced aspirates: a historical hypothesis revisited. *OL* 45: 311-338.

_____ 2006b. Anomalous liquid : sibilant correspondences in western Austronesian. *OL* 45: 210-216.

_____ 2006c. Supertemplatic reduplication and beyond. In Chang, Huang and Ho: 439-460.

_____ 2007a. The prenasalized trills of Manus. In J. Siegel, J. Lynch and D. Eades, eds., *Language description, history and development: linguistic indulgence in memory of Terry Crowley*: 297-311. Amsterdam/Philadelphia: John Benjamins.

_____ 2007b . Disyllabic attractors and antiantigemination in Austronesian sound change. *Phonology* 24: 1-36.

_____ 2007c. Òma Lóngh historical phonology. *OL* 46: 1-53.

_____ 2007d . The linguistic position of Sama-Bajaw. *SPLC* 15: 73-114.

_____ 2008a. Is there a Bima-Sumba subgroup? *OL* 47: 46-114.

_____ 2008b. Remote Melanesia: one history or two? An addendum to Donohue and Denham. *OL* 47: 445-459.

_____ 2009a. *The Austronesian languages*. 1st edition. Canberra: Pacific Linguistics.

_____ 2009b. Palauan historical phonology: whence the intrusive velar nasal? *OL* 48: 307-346.

_____ 2009c. The position of the languages of eastern Indonesia: a reply to Donohue and Grimes. *OL* 48: 36-77.

_____ 2009d. In memoriam, Isidore Dyen, 1913-2008. *OL* 48: 488-508.

_____ 2010. The Greater North Borneo hypothesis. *OL* 49: 44-118.

_____ 2011a. The problem of doubletting in Austronesian languages. *OL* 50: 399-457.

_____ 2011b. Dempwolff reinvented: a review of Wolff (2010). *OL* 50: 560-579.

_____ 2012a. The marsupials strike back: a reply to Schapper (2011). *OL* 51: 261-277.

_____ 2012b. One mark per word? Patterns of dissimilation in Austronesian and Australian languages. *Phonology* 29.3: 355-381.

_____ n.d. (a). Fieldnotes on 41 languages of northern and central Sarawak.

_____ n.d. (b). Fieldnotes on 32 languages of the Admiralty islands and adjacent areas.

_____ n.d. (c). Fieldnotes on 6 aboriginal languages of Taiwan.

_____ n.d. (d). Lexicostatistical lists of 231 Austronesian languages, with retention percentages from Proto-Malayo-Polynesian. Ms.

_____ n.d. (e). The eye as center: a semantic universal. Ms.

_____ n.d. (f). Fieldnotes on Jarai.

_____ ed. 1981. *Historical linguistics in Indonesia*, Part I. NUSA 10.

_____ ed. 1991. *Currents in Pacific Linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*. Canberra: Pacific Linguistics.

____, and Jürg Schneider, eds. 2012. *A world of words: Revisiting the work of Renward Brandstetter (1860-1942) on Lucerne and Austronesia*. Wiesbaden: Harrassowitz.

Blust, R. and S. Trussel. Ongoing. *Austronesian comparative dictionary*. Available online at ww.trussel2.com/ACD .

Bolton, R.A. 1989. Nuaulu phonology. *WILC* 7: 89-119.

Bopp, F. 1841. *Über die Verwandtschaft der malaisch-polynesischen Sprachen mit den indo-europäischen. Gelesen in der Akademie der Wissenschaften am 10. Aug. und 10. Dec. 1840*. Berlin: Dümmler.

Boutin, M.E. 1993. Bonggi phonemics. In M.E. Boutin and I. Pekkanen, eds., *Phonological descriptions of Sabah languages*: 107-130. Kota Kinabalu: Sabah Museum.

Bowden, J. 2001. *Taba: description of a South Halmahera language*. Canberra: Pacific Linguistics.

Bowden, J., N.P. Himmelmann, and M. Ross, eds. 2010. *A journey through Austronesian and Papuan linguistic and cultural space: Papers in honour of Andrew Pawley*. Canberra: Pacific Linguistics.

Bradshaw, J. 1979. Obstruent harmony and tonogenesis in Jabêm. *Lingua* 49: 189-205.

_____ 1982. *Word order change in Papua New Guinea Austronesian languages*. Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

Brainard, S. and D. Behrens. 2002. *A grammar of Yakan*. Manila: Linguistic Society of the Philippines.

Brand, D.D. 1971. The sweet potato: an exercise in methodology. In Riley, Kelly, Pennington, and Rands: 343-365.

Brandes, J.L.A. 1884. *Bijdrage tot de vergelijkende klankleer der westerse afdeeling van de Maleisch-Polynesische taalfamilie*. Utrecht: P.W. van de Weijer.

Brandstetter, R. 1906. *Ein Prodromus zu einem vergleichenden Wörterbuch der Malaiopolynesischen Sprachen*. Luzern.

_____ 1916. *An introduction to Indonesian linguistics: being four essays by Renward Brandstetter*. Trans. by C.O. Blagden. Royal Asiatic Society Monographs, Vol. XV. London.

_____ 1937. Die Verwandtschaft des Indonesischen mit dem Indogermanischen. *Wir Menschen der indonesischen Erde, II*. Luzern.

Bril, I. 2000. *Dictionnaire nêlêmwa-nixumwak-français-anglais*. Langues et cultures du Pacifique 14, SELAF 383. Paris: Peeters.

_____ 2002. *Le nêlêmwa (Nouvelle-Calédonie)*. Langues et cultures du Pacifique 16, SELAF 403. Paris: Peeters.

Brockelmann, C. 1908-1913. 2 vols. *Grundriss der vergleichende Grammatik der semitischen Sprachen*. Berlin.

Brosius, J.P. 1988. A separate reality: Comment on Hoffman's The Punan: hunters and gatherers of Borneo. *BRB* 20.2: 81-106.

Brown, C.H. 1983. Where do cardinal direction terms come from? *Anthropological Linguistics* 25: 121-161.

Brown, C.H. and S.R. Witkowski. 1981. Figurative language in a universalist perspective. *American Ethnologist* 8.3: 596-615.

Brownie, J. and M. Brownie. 2007. *Mussau grammar essentials*. Data Papers on Papua New Guinea Languages. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics-Papua New Guinea Academic Publications.

Brugmann, K. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeitschrift fur allgemeine Sprachwissenschaft* 1: 226-256.

Brunelle, M. 2005. A phonetic study of Eastern Cham register. In A. Grant and P. Sidwell, eds., *Chamic and beyond: Studies in mainland Austronesian languages*: 1-35. Canberra: Pacific Linguistics.

_____ 2009. Contact-induced change? Register in three Cham dialects. *Journal of the Southeast Asian Linguistics Society* 2: 1-22.

Budenz, J. 1966 [1873-81]. *A comparative dictionary of the Finno-Ugric elements in the Hungarian vocabulary*. Indiana University Uralic and Altaic Series, vol. 78. Bloomington: Indiana University Press.

Burger, P.A. 1946. Voorlopige Manggaraise Spraakkunst. *BKI* 103.1-2: 15-265.

Burrow, T. and M.B. Emeneau. 1984. *A Dravidian etymological dictionary*. Oxford: Clarendon Press.

Buschmann, J.C.E. 1859. Die Spuren der Aztekischen Sprache im nordlichen Mexiko und höheren amerikanischen Norden. *Abhandlungen der Königlichen Akademie der Wissenschaften*, 1854, Supplement-Band II: 512-576.

Busenitz, R.L. and M.J. Busenitz. 1991. Balantak phonology and morphophonemics. In J.N. Sneddon, ed., *Studies in Sulawesi Linguistics*, Part II: 29-47. NUSA 33.

Cachey, T.J. Jr., ed. 2007. *First voyage around the world, 1519-1522: an account of Magellan's expedition*. Toronto: University of Toronto Press.

Campbell, J. 1892. Remarks on George Patterson, 'Beothick vocabularies'. *Transactions of the Royal Society of Canada*, Sect. II: 19-26.

Campbell, L. 1997. *American Indian languages: the historical linguistics of native North America*. Oxford Studies in Anthropological Linguistics. New York and Oxford: Oxford University Press.

Campbell, W. 1888. *The gospel of St. Matthew in Formosan (Sinkang dialect), edited from Gravius' edition of 1661*. London.

_____ 1903. *Formosa under the Dutch*. London: Kegan Paul.

Capell, A. 1943. *The linguistic position of South-eastern Papua*. Sydney: Australasian Medical Publishing Co, Ltd.

_____ 1956. *A new approach to Australian linguistics*. Handbook of Australian Languages, Part 1. Oceania Linguistic Monographs, No. 1. University of Sydney.

_____ 1962. Oceanic linguistics today. *CA* 3: 371-428 (with comments and reply).

_____ 1968. *A new Fijian dictionary* (3rd edition). Suva, Fiji: The Government Printer.

_____ 1969. *Grammar and vocabulary of the language of Sonsorol-Tobi*. Oceania Linguistic Monographs, No. 12. University of Sydney.

_____ 1971. The Austronesian languages of Australian New Guinea. In Sebeok: 240- 340.

Carle, R., M. Heinschke, P.W. Pink, C. Rost and K. Stadtlander. 1982. *GAVA': Studies in Austronesian languages and cultures dedicated to Hans Kähler*. VSIS 17. Berlin: Reimer.

Carrington, L. 1996. *A linguistic bibliography of the New Guinea area*. Canberra: Pacific Linguistics.

Carroll, V. and T. Soulik. 1973. *Nukuoro lexicon*. PALI Language Texts: Polynesia. Honolulu: The University Press of Hawaii.

Catford, J.C. 1988. Notes on the phonetics of Nias. In McGinn: 151-200.

Cauquelin, J. 2008. *Ritual texts of the last traditional practitioners of Nanwang Puyuma*. LLMS A23. Taipei: Institute of Linguistics, Academia Sinica.

_____ to appear. *Nanwang Puyuma-English dictionary*. *LLMS* W5. Taipei: Institute of Linguistics, Academia Sinica

Cavalli-Sforza, L.L., P. Menozzi and A. Piazza. 1994. *The history and geography of human genes*. Princeton, New Jersey: Princeton University Press.

Cense, A.A. 1979. *Makassaars-Nederlands woordenboek. KITLV*. The Hague: Nijhoff.

Cense, A.A. and E.M. Uhlenbeck. 1958. *Critical survey of studies on the languages of Borneo. KITLV* Bibliographical Series 2. The Hague: Nijhoff.

Chai, C.K. 1967. *Taiwan aborigines: a genetic study of tribal variations*. Cambridge, Mass.: Harvard University Press.

Chang, H.Y. 2006. Rethinking the Tsouic subgroup hypothesis: a morphosyntactic perspective. In Chang, Huang, and Ho: 565-583.

Chang, H.Y., L.M. Huang, and D.A. Ho, eds. 2006. *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*. Language and Linguistics Monograph Series W-5. Taipei: Institute of Linguistics, Academia Sinica.

Chang, K.C. 1969. *Fengpitou, Tapenkeng and the prehistory of Taiwan*. New Haven: Yale University Publications in Anthropology, No. 73.

_____ 1986. *The archaeology of ancient China* (3rd edition). New Haven: Yale University Press.

Chang, M.L. 1998. Thao reduplication. *OL* 37.2: 277-297.

Chang, Y.L. 1997. *Voice, case and agreement in Seediq and Kavalan*. Unpublished Ph.D. dissertation, Tsing-hua University. Hsinchu, Taiwan: Department of Linguistics, Tsing-hua University.

Charles, M. 1974. Problems in the reconstruction of Protophilippine phonology and the subgrouping of the Philippine languages. *OL* 13: 457-509.

Chen, C.L. 1988 [1968]. *Material culture of the Formosan aborigines* (3rd edition). Taipei: Southern Materials Center, Inc.

Chowning, A. 1996. POC *mata: how many words, how many meanings? In Lynch and Pat: 47-60.

Chrétien, C.D. 1965. The statistical structure of the Proto-Austronesian morph. *Lingua* 14: 243-270.

Chung, S. 1978. *Case marking & grammatical relations in Polynesian*. Austin & London: University of Texas Press.

_____ 1998. *The design of agreement: evidence from Chamorro*. Chicago and London: The University of Chicago Press.

Churchward, C.M. 1940. *Rotuman grammar and dictionary*. Sydney: The Australasian Medical Publishing Co.

_____ 1959. *Tongan dictionary*. Tonga: The Government Printing Press.

CIA-The World Factbook (www.cia.gov/cia/publications/factbook/geos/kr.html#People).

Clark, R. 1976. *Aspects of Proto-Polynesian syntax*. Auckland, New Zealand: Linguistic Society of New Zealand.

_____ 1982a. Proto-Polynesian birds. In J. Siikala, ed., *Oceanic Studies: essays in honour of Aarne A. Koskinen*: 121-143. Transactions of the Finnish Anthropological Society, No. 11. Helsinki.

_____ 1982b. 'Necessary' and 'unnecessary' borrowing. In Halim, Carrington, and Wurm 3: 137-143.

_____ 1985. Languages of north and central Vanuatu: groups, chains, clusters and waves. In Pawley and Carrington: 199-236.

_____ 1991. Fingota/fangota: shellfish and fishing in Polynesia. In Pawley: 78-83.

_____ 2009. *Leo Tuai: a comparative lexical study of North and Central Vanuatu languages*. Canberra: Pacific Linguistics

_____ n.d. POLLEX: Comparative Polynesian Lexicon (computer database). Auckland: University of Auckland, Department of Anthropology.

Clayre, B. 1988. The changing face of focus in the languages of Borneo. In Steinhauer 1988a: 51-88.

_____ 1991. Focus in Lundayeh. *SMJ* 42: 413-434.

Clynes, A. 1994. Old Javanese influence in Balinese: Balinese speech styles. In Dutton and Tryon: 141-179.

_____ 1997. On the Proto-Austronesian "diphthongs". *OL* 36: 347-361.

_____ 1999. Rejoinder: Occam and the Proto-Austronesian "diphthongs". *OL* 38: 404- 408.

Codrington, R.H. 1885. *The Melanesian languages*. Oxford: The Clarendon Press.

Codrington, R.H. and J. Palmer. 1896. *A dictionary of the language of Mota, Sugarloaf island, Banks' islands*. London: Society for Promoting Christian Knowledge.

Coedès, G. 1971 [1964]. *The Indianized states of Southeast Asia*, ed. by W.F. Vella, trans. by S.B. Cowing. Honolulu: The University Press of Hawaii.

Cohn, A.C. 1990. Phonetic and phonological rules of nasalisation. *UCLA Working Papers in Phonetics*, 76. Los Angeles: University of California at Los Angeles.

_____ 1993a. The status of nasalized continuants. *Phonetics and phonology* 5: 329-367.

_____ 1993b. Voicing and vowel height in Madurese: a preliminary report. In Edmondson and Gregerson: 107-121.

_____ 1994. A phonetic description of Madurese and its phonological implications. *Working Papers of the Cornell Phonetics Laboratory* 9: 67-92.

Cohn, A.C., and W.H. Ham. 1999. Temporal properties of Madurese consonants: a Preliminary report. In Zeitoun and Li: 227-249.

Cohn, A.C., W.H. Ham and R.J. Podesva. 1999. The phonetic realisation of singleton-geminate contrasts in three languages of Indonesia. *Proceedings of the International Congress of Phonetic Sciences*: 587-590.

Collinder, B. 1955. *Fenno-Ugric vocabulary: an etymological dictionary of the Uralic languages*. Stockholm: Almqvist & Wiksell.

Collins, J.T. 1980. *Ambonese Malay and creolisation theory*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ 1982. Linguistic research in Maluku: a report of recent field work. *OL* 21: 73- 146.

_____ 1983a. *The historical relationships of the languages of Central Maluku, Indonesia.* Canberra: Pacific Linguistics.

_____ 1983b. *Dialek Ulu Terengganu* (The Upper Trengganu dialects). Monograf 8, Fakulti Sains Kemasyarakatan dan Kemanusiaan, Universiti Kebangsaan Malaysia. Bangi, Selangor: National University of Malaysia.

_____ 1990. *Bibliografi dialek Melayu di pulau Borneo* (An annotated bibliography of Malay dialects in Borneo). Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ 1992. Preliminary notes on Berau Malay. In Martin: 297-333.

_____ 1995a. *Bibliografi dialek Melayu di pulau Sumatera* (An annotated bibliography of Malay dialects in Sumatra). Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ 1995b. *Bibliografi dialek Melayu di pulau Jawa, Bali dan Sri Lanka* (An annotated bibliography of Malay dialects in Java, Bali and Sri Lanka). Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ 1996. *Bibliografi dialek Melayu di Indonesia Timur* (An annotated bibliography of Malay dialects in East Indonesia). Kuala Lumpur: Dewan Bahasa dan Pustaka.

_____ n.d. *Bibliografi dialek Melayu di Semenanjung Melayu* (An annotated bibliography of Malay dialects in the Malay Peninsula). Kuala Lumpur: Dewan Bahasa dan Pustaka. Ms.

_____ ed. 1983. *Studies in Malay dialects*. NUSA 16.

Collins, M.A., V.R. Collins and S.A. Hashim. 2001. *Mapun-English dictionary*. Manila: Summer Institute of Linguistics – Philippines, Inc.

Conant, C.E. 1911. The RGH law in Philippine languages. *Journal of the American Oriental Society* 31: 74-85.

_____ 1912. The pepet law in Philippine languages. *Anthropos* 5: 920-947.

Conklin, H.C. 1953. *Hanunóo-English vocabulary*. University of California Publications in Linguistics, vol. 9. Berkeley and Los Angeles: University of California Press.

_____ 1956. Tagalog speech disguise. *Language* 32: 136-139.

_____ 1980. *Ethnographic atlas of Ifugao: a study of environment, culture, and society in northern Luzon*. New Haven and London: Yale University Press.

Constantino, E. 1971. Tagalog and other major languages of the Philippines. In Sebeok: 112-154.

Cook, K.W. 1997. The Samoan transitive suffix as an inverse marker. In M. Verspoor, K.D. Lee and E. Sweetser, eds., *Lexical and syntactic constructions and the construction of meaning*: 347-361. Amsterdam/Philadelphia: John Benjamins.

_____ 2004. The role of metathesis in Hawaiian word creation. In A.S. da Silva, A. Torres and M. Gonçalves, eds., *Linguagem, cultura e cognição: Estudos de linguistica cognitiva*, vol. 1: 207-214. Coimbra: Almedina.

Coolsma, S. 1985 [1904]. *Tata Bahasa Sunda* (Trans. of *Soendaneesche spraakkunst*). ILDEP 21. Jakarta: Djambatan.

Corston, S.H. 1996. *Ergativity in Roviana*. Canberra: Pacific Linguistics.

Corston-Oliver, S. 2002. Roviana. In Lynch, Ross and Crowley: 467-497.

Costello, N. 1966. Affixes in Katu. In D.D. Thomas, N.D. Hoa and D. Blood, eds., *Mon-Khmer Studies II*: 63-86. Saigon: The Linguistic Circle of Saigon and the Summer Institute of Linguistics.

Costenoble, H. 1940. *Die Chamorro Sprache*. The Hague: Nijhoff.

Counts, D.R. 1969. *A grammar of Kaliai-Kove*. OLSP 6. Honolulu: University of Hawaii Press.

Court, C. 1967. Some areal features of Mentu Land Dayak. *OL* 6: 46-50.

Coward, N.E. 1989. A phonological sketch of the Selaru language. *WILC* 7: 1-42.

Croft, W. 2003. *Typology and universals* (2nd edition). Cambridge University Press.

Crowley, T. 1982. *The Paamese language of Vanuatu*. Canberra: Pacific Linguistics.

_____ 1990. *Beach-la-mar to Bislama: the emergence of a national language in Vanuatu*. Oxford: Clarendon Press.

_____ 1991. Parallel development and shared innovation: some developments in central Vanuatu inflectional morphology. *OL* 30: 179-222.

_____ 1992. *A dictionary of Paamese*. Canberra: Pacific Linguistics.

_____ 1995. Melanesian languages: do they have a future? *OL* 34: 327-344.

_____ 1998. *An Erromangan (Sye) grammar. OLSP* 27.

_____ 1999. *Ura: a disappearing language of southern Vanuatu*. Canberra: Pacific Linguistics.

_____ 2003. *A new Bislama dictionary*. 2nd ed. Suva: Institute of Pacific Studies, University of the South Pacific.

_____ 2006a. *The Avava language of central Malakula* (*Vanuatu*). Edited by John Lynch. Canberra: Pacific Linguistics.

_____ 2006b. *Tape: a declining language of Malakula (Vanuatu)*. Edited by John Lynch. Canberra: Pacific Linguistics.

_____ 2006c. *Naman: a vanishing language of Malakula (Vanuatu)*. Edited by John Lynch. Canberra: Pacific Linguistics.

_____ 2006d. *Nese: a diminishing speech variety of Malakula (Vanuatu)*. Edited by John Lynch. Canberra: Pacific Linguistics.

Cumming, S. 1991. *Functional change: the case of Malay constituent order*. Berlin and New York: Mouton de Gruyter.

Dahl, O.C. 1951. *Malgache et Maanjan: une comparaison linguistique*. Studies of the Egede Institute, No. 3. Oslo: Egede-Instituttet.

_____ 1954. Le substrat bantou en malgache. *Norsk tidsskrift for Sprogvidenskap* 17: 325-362.

_____ 1976 [1973]. *Proto-Austronesian*. 2nd, rev. edition. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London: Curzon Press.

_____ 1978. The fourth focus. In Wurm and Carrington: 383-393.

_____ 1981a. *Early phonetic and phonemic changes in Austronesian*. Oslo: Instituttet for Sammenlignende Kulturforskning.

_____ 1981b. Austronesian numerals. In Blust: 46-58.

_____ 1986. Focus in Malagasy and Proto-Austronesian. In Geraghty, Carrington, and Wurm 2: 21-42.

_____ 1991. *Migration from Kalimantan to Madagascar*. Norwegian University Press. Oslo: The Institute for Comparative Research in Human Culture.

Darlington, P.J. 1980. *Zoogeography: the geographical distribution of animals*. Huntington, New York: Robert E. Krieger Publishing Co.

Davidson, J., G. Irwin, F. Leach, A. Pawley, and D. Brown, eds., *Oceanic culture history: essays in honour of Roger Green*. Dunedin, New Zealand: New Zealand Journal of Archaeology Special Publication.

Davis, K. 2003. *A grammar of the Hoava language, western Solomons*. Canberra: Pacific Linguistics.

Deck, N.C. 1933-1934. A grammar of the language spoken by the Kwara'ae people of Mala, British Solomon Islands. *JPS* 42: 33-48, 133-144, 241-256; 43: 1-16, 85-100, 163-170, 246-257.

De Guzman, V.P. 1988. Ergative analysis for Philippine languages: an analysis. In McGinn: 323-345.

Dempwolff, O. 1905. Beiträge zur Kenntnis der Sprachen von Deutsch Neuguinea. *Mitteilungen des Seminars für Orientalische Sprachen* 8: 182-254.

_____ 1920. Die Lautentsprechungen der indonesischen Lippenlaute in einigen anderen austronesischen Südseesprachen. *ZfES*, Supplement 2. Berlin: Reimer.

_____ 1922. Entstehung von Nasalen und Nasalverbindungen im Ngaju (Dajak). *ZfES* 13: 161-205.

_____ 1924-1925. Die l-, r-, und d- Laute in austronesischen Sprachen. *ZfES* 15: 19-50, 116-138, 223-238, 273-319. Berlin: Reimer.

_____ 1927. Das austronesische Sprachgut in den melanesischen Sprachen. *Folia Ethnoglossica* 3: 32-43.

_____ 1934-1938. 3 vols. *Vergleichende Lautlehre des austronesischen Wortschatzes. ZfES*, Supplement 1. *Induktiver Aufbau einer indonesischen Ursprache* (1934), Supplement 2. *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* (1937), Supplement 3. *Austronesisches Wörterverzeichnis* (1938). Berlin: Reimer.

_____ 1939. *Grammatik der Jabêm-Sprache auf Neuguinea*. Abhandlungen aus dem Gebiet des Auslandskunde, vol. 50. Hamburg: Friederichsen, de Gruyter.

Diffloth, G. 1976. Expressives in Semai. In P.N. Jenner, L.C. Thompson and S. Starosta, eds. *Austroasiatic Studies, Part I*: 249-264. *OLSP* 13.

_____ 1994. The lexical evidence for Austric, so far. *OL* 33: 309-321.

Dixon, R.M.W. 1977. Where have all the adjectives gone? *Studies in Language* 1: 19-80.

_____ 1980. *The languages of Australia*. Cambridge Language Surveys. Cambridge University Press.

_____ 1988. *A grammar of Boumaa Fijian*. University of Chicago Press.

_____ 1994. *Ergativity*. Cambridge Studies in Linguistics, No. 69. Cambridge: Cambridge University Press.

Dixon, R.M.W. and A.Y. Aikhenvald, eds. 2000. *Changing valency: case studies in transitivity*. Cambridge and New York: Cambridge University Press.

Dixon, R.B. and A.L. Kroeber. 1912. Relationship of the Indian languages of California. *American Anthropologist*, n.s. 14.4: 691-692.

Djajadiningrat, H. 1934. 2 vols. *Atjèhsch-Nederlandsch woordenboek met Nederlandsch-Atjèhsch register*. Batavia.

Djawanai, S. 1977. A description of the basic phonology of Nga'da and the treatment of borrowings. In I. Suharno, ed., *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part IV: 10-18. NUSA 5.

Donohue, M. 1999. *A grammar of Tukang Besi*. Mouton Grammar Library 20. Berlin: Mouton de Gruyter.

_____ 2002. Tobati. In Lynch, Ross and Crowley: 186-203.

_____ 2004. The pretenders to the Muna-Buton group. In J. Bowden and N.P. Himmelmann, eds., *Papers in Austronesian subgrouping and dialectology*: 21-35. Canberra: Pacific Linguistics.

_____ 2007. The Papuan language of Tambora. *OL* 46: 520-537.

Donohue, M., and C.E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia. *OL* 47: 114-158.

Dreyfuss, J. 1983. The backwards language of Jakarta youth (JYBL), a bird of many language feathers. In Collins: 52-56.

Dubois, C.D. 1976. *Sarangani Manobo: an introductory guide*. Manila: LSP.

Dunnebier, W. 1951. *Bolaang Mongondowsch-Nederlandsch woordenboek. KITLV*. The Hague: Nijhoff.

Durie, M. 1985. *A grammar of Acehnese on the basis of a dialect of North Aceh. VKI* 112. The Hague: Nijhoff.

Durie, M. and M. Ross, eds. 1996. *The comparative method reviewed: regularity and irregularity in language change.* New York and Oxford: Oxford University Press.

Dutton, T.E. 1975. South-eastern Trans-New Guinea Phylum languages. In Wurm 1: 613-664.

_____ 1976. Magori and similar languages of southeast Papua. In Wurm 2: 581-636.

_____ 1986. Police Motu and the Second World War. In Geraghty, Carrington, and Wurm 2: 351-406.

Dutton, T.E. and D.T. Tryon, eds. 1994. *Language contact and change in the Austronesian-speaking world.* Trends in Linguistics Studies and Monographs 77. Berlin: Mouton de Gruyter.

Dyen, I. 1947a. The Malayo-Polynesian word for 'two'. *Language* 23: 50-55.

_____ 1947b. The Tagalog reflexes of Malayo-Polynesian D. *Language* 23: 227-238.

_____ 1949. On the history of the Trukese vowels. *Language* 25: 420-436.

_____ 1951. Proto-Malayo-Polynesian *Z. *Language* 27: 534-540.

_____ 1953a. Dempwolff's *R. *Language* 29: 359-366.

_____ 1953b. *The Proto-Malayo-Polynesian laryngeals.* William Dwight Whitney Linguistic Series. Baltimore: Linguistic Society of America.

_____ 1956a. The Ngaju-Dayak 'Old speech stratum'. *Language* 32: 83-87.

_____ 1956b. Language distribution and migration theory. *Language* 32: 611-626.

_____ 1962. Some new Proto-Malayopolynesian initial phonemes. *JAOS* 82: 214-215.

_____ 1963. The position of the Malayopolynesian languages of Formosa. AP 7: 261-271.

_____ 1965a. *A lexicostatistical classification of the Austronesian languages.* Indiana University Publications in Anthropology and Linguistics, and Memoir 19 of the International Journal of American Linguistics. Baltimore: The Waverly Press.

_____ 1965b. *A sketch of Trukese grammar.* American Oriental Series, Essay 4. New Haven: American Oriental Society.

_____ 1965c. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14: 285-305.

_____ 1971a. The Austronesian languages and Proto-Austronesian. In Sebeok: 5-54.

_____ 1971b. Malagasy. In Sebeok: 211-239.

_____ 1971c. Review of Die Palau-sprache und ihre Stellung zu anderen indonesischen Sprachen, by Klaus Pätzold. *JPS* 80.2: 247-258.

_____ 1971d. The Austronesian languages of Formosa. In Sebeok: 168-199.

Dyen, I. and D.F. Aberle. 1974. *Lexical reconstruction: the case of the Proto-Athapaskan kinship system.* Cambridge University Press.

Dyen, I., A.T. James and J.W.L. Cole. 1967. Language divergence and estimated word retention rate. *Language* 43: 150-171.

Edmondson, J.A. and K.J. Gregerson, eds. 1993. *Tonality in Austronesian languages. OLSP* 24. Honolulu: University of Hawaii Press.

Edmondson, J.A., J.H. Esling, J.G. Harris and T.C. Huang. 2005. A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis – an Austronesian language of Taiwan. *Language and Linguistics* 6: 381-396.

Egerod, S. 1965. Verb inflection in Atayal. *Lingua* 15: 251-282.

_____ 1966. A statement on Atayal phonology. *Artibus Asiae Supplementum* XXIII (Felicitation volume for the 75th birthday of Professor G.H. Luce) 1: 120-130.

Elbert, S.H. 1953. Internal relationships of Polynesian languages and dialects. *Southwestern Journal of Anthropology* 9: 147-173.

_____ 1965. Phonological expansions in Outlier Polynesian. *Lingua* 14: 431-442.

_____ 1972. *Puluwat dictionary*. Canberra: Pacific Linguistics.

_____ 1974. *Puluwat grammar*. Canberra: Pacific Linguistics.

_____ 1975. *Dictionary of the language of Rennell and Bellona. Language and culture of Rennell and Bellona islands*, Vol. III, Part I. Copenhagen: The National Museum of Denmark.

_____ 1988. *Echo of a culture: a grammar of Rennell and Bellona. OLSP* 22. Elbert, S.H. and M.K.Pukui. 1979. *Hawaiian grammar*. Honolulu: The University Press of Hawaii.

Engelbrecht, W.A., and P.J. van Hervarden. 1945. *Ontdekkingsreis van Jacob Le Maire en Willem Cornelisz. Schouten in de jaren 1615-1617: Journalen, documenten en andere bescheiden.* 's-Gravenhage: Nijhoff.

Esser, S.J. 1927. *Klank- en vormleer van het Morisch*, part 1. Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen, No. 67.3. Leiden: Vros.

_____ 1930. Renward Brandstetter—29 Juni—1930. *Tijdschrift voor indische taal-, land-, en volkenkunde* LXX.2-3: 146-156.

_____ 1938. Talen. Sheet 9 of *Atlas van Tropisch Nederland*. Batavia: Topographische Dienst in Nederlandsch-Indië.

Evans, B., ed. 2009. *Discovering history through language: papers in honour of Malcolm Ross*. Canberra: Pacific Linguistics.

Evans, B. and M. Ross. 2001. The history of Proto Oceanic *ma-. *OL* 40: 269-290.

Evans, I.H.N. 1923. *Studies in religion, folk-lore & custom in British North Borneo and the Malay peninsula*. Cambridge University Press.

Ewing, M. and M.A.F. Klamer. 2010. *East Nusantara: typological and areal analyses*. Canberra: Pacific Linguistics.

Ferguson, C.A. 1963. Assumptions about nasals. In J.H. Greenberg, ed., *Universals of language*: 53-60. Cambridge, Mass: The MIT Press.

Ferrell, R. 1968. 'Negrito' ritual and traditions of small people on Taiwan. In N. Matsumoto and T. Mabuchi, eds., *Folk religion and worldview in the Southwestern Pacific*: 63-72. Tokyo: Keio University.

_____ 1969. *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica, Monograph No. 17. Taipei: Academia Sinica.

_____ 1970. The Pazeh-Kahabu language. *Bulletin of the Department of Archaeology and Anthropology, National Taiwan University* 31/32: 73-96.

_____ 1971. Aboriginal peoples of the southwestern Taiwan plain. *Bulletin of the Institute of Ethnology*, Academia Sinica 32: 217-235.

_____ 1982. *Paiwan dictionary*. Canberra: Pacific Linguistics.

Fey, V. 1986. *Amis dictionary*. Taipei: The Bible Society.

Firth, R.C. 1985. *Tikopia-English dictionary*. Auckland: Auckland University Press.

Fischer, J.L., with the assistance of A.M. Fischer. 1970. *The Eastern Carolines*. New Haven: Human Relations Area Files Press.

Florey, M. 2005. Language shift and endangerment. In Adelaar and Himmelmann: 43- 64.

Fokker, A.A. 1895. *Malay phonetics*. Leiden: Brill.

Foley, W.A. 1976. *Comparative syntax in Austronesian.* Unpublished Ph.D. Dissertation. Berkeley, California: Department of Linguistics, University of California.

_____ 1986. *The Papuan languages of New Guinea*. Cambridge Language Surveys. Cambridge University Press.

_____ 2012a. A comparative look at nominalisations in Austronesian. Keynote address presented at 12-ICAL, Denpasar, Bali, July 2, 2012. Ms., 50pp.

_____ 2012b. Review of Claire Moyse-Faurie and Joachim Sabel, eds, Topics in Oceanic morphosyntax. *Language* 88: 910-914.

Fonzaroli, P. 1975. On the common Semitic lexicon and its ecological and cultural background. In J. and T. Bynon, eds., *Hamito-Semitica*: 43-53. The Hague: Mouton.

Forman, M.L. 1971. *Kapampangan dictionary*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Förster, J.R. 1996 [1778]. *Observations made during a voyage round the world*, ed. by N. Thomas, H. Guest and M. Dettelbach, with a linguistic appendix by K.H. Rensch. Honolulu: University of Hawaii Press.

Fortgens, J. 1921. *Bijdrage tot de kennis van het Sobojo (Eiland Taliaboe, Soela-groep)*. The Hague: Martinus Nijhoff.

Foster, M.L. 1998. The transoceanic trail: the Proto-Pelagian language phylum. *Precolumbiana: a journal of long-distance contacts* 1.1-2: 88-113.

Fox, C.E. 1955. *A dictionary of the Nggela language* (*Florida, British Solomon Islands*). Auckland: The Unity Press.

_____ 1970. *Arosi-English dictionary*. Canberra: Pacific Linguistics.

Fox, G.J. 1979. *Big Nambas grammar*. Canberra: Pacific Linguistics.

Fox, J.J. 1971. Sister's child as plant: metaphors in an idiom of consanguinity. In R. Needham, ed., *Rethinking kinship and marriage*: 219-252. London and New York: Tavistock Publications.

_____ 1988. *To speak in pairs: essays on the ritual languages of eastern Indonesia*. Cambridge University Press.

_____ 1993. *Dictionary of Rotinese formal dyadic language*. Unpublished revision of 1972 manuscript, with English to Rotinese glosses. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, Australian National University.

_____ 1995. Origin structures and systems of precedence in the comparative study of Austronesian societies. In Li, Tsang, Huang, Ho, and Tseng: 27-57.

_____ 2005. Ritual languages, special registers and speech decorum in Austronesian languages. In Adelaar and Himmelmann: 87-109.

Fox, J.J. and C.E. Grimes. 1995. Roti. In Tryon 1995, Part 1, Fascicle 1: 611-622.

François, A. 2002. *Araki: a disappearing language of Vanuatu*. Canberra: Pacific Linguistics.

_____ 2003a. Of men, hills, and winds: space directionals in Mwotlap. *OL* 42: 407-437.

_____ 2003b. *La sémantique du prédicat en Mwotlap (Vanuatu)*. Société de Linguistique de Paris LXXXIV. Leuven-Paris: Peeters.

_____ 2004. Reconstructing the geocentric system of Proto Oceanic. *OL* 43: 1-31.

_____ 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *OL* 44: 443-504.

_____ 2010. Phonotactics and the prestopped velar lateral of Hiw. Resolving the ambiguity of a complex segment. *Phonology* 27.3: 393-434.

_____ 2011. Where *R they all? The geography and history of *R-loss in Southern Oceanic languages. *OL* 50: 140-197.

_____ 2012. The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. In P. Unseth & L. Landweer (eds), *Language Use in Melanesia. International Journal of the Sociology of Language* 214: 85–110.

Friberg, T. and B. Friberg. 1991. Notes on Konjo phonology. In J.N. Sneddon, ed., *Studies in Sulawesi Linguistics*, Part. II: 71-115. NUSA 33.

Friederici, G. 1912-1913. 3 vols. *I. Wissenschaftliche Ergebnisse einer Amtlichen Forschungsreise nach dem Bismarck-Archipel im Jahre 1908: II. Beiträge zur Völker- und Sprachenkunde von Deutsch-Neuguinea* (1912)*, III. Untersuchungen über eine melanesische Wanderstrasse* (1913). Berlin: Ernst Siegfried Mittler und Sohn.

Gabelentz, H.C. von der. 1861-1873. *Die melanesischen Sprachen nach ihrem grammatischen Bau und ihrer Verwantschaft unter sich und mit den malaiischpolynesischen Sprachen*. Abhandlungen der philologisch-historischen Classe der Königlich Sächsischen Gesellschaft der Wissenschaften, 3, 7.

Gafos, D. 1998. A-templatic reduplication. *Linguistic Inquiry* 29: 515-527.

Galang, R. 1982. The acquisition of Tagalog verb morphology. *PJL* 13: 1-15.

Garvan, J.M. 1963. *The Negritos of the Philippines*, ed. by Hermann Hochegger. Wiener Beiträge zur Kulturgeschichte und Linguistik, vol. XIV. Horn – Vienna: Ferdinand Berger.

Garvey, C.J. 1964. *Malagasy introductory course*. Washington, D.C.: Center for Applied Linguistics.

Garvin, P.L. and S.H. Riesenberg. 1952. Respect behavior on Ponape: an ethnolinguistic study. *American Anthropologist* 54: 201-220.

Geddes, W.R. 1961. *Nine Dayak nights*. London: Oxford University Press.

Geerts, P. 1970. ʾĀreʾāre dictionary. Canberra: Pacific Linguistics.

Geertz, C. 1960. *The religion of Java*. Glencoe, Illinois: The Free Press.

Geraghty, P.A. 1983. *The history of the Fijian languages. OLSP* 19.

_____ 1986. The sound system of Proto-Central-Pacific. In Geraghty, Carrington, and Wurm 2: 289-312.

_____ 1989. The reconstruction of Proto-Southern Oceanic. In Harlow and Hooper I: 141-156.

_____ 1990. Proto-Eastern Oceanic *R and its reflexes. In J.H.C.S. Davidson, ed., *Pacific island languages: essays in honour of G.B. Milner*: 51-93. London: School of Oriental and African Studies, University of London.

_____ 1994. Linguistic evidence for the Tongan Empire. In Dutton and Tryon: 233- 249.

Geraghty, P.A., L. Carrington, and S.A. Wurm, eds. 1986. 2 vols. FO*CAL* I, II: Papers from the Fourth International Conference on Austronesian Linguistics. Canberra: Pacific Linguistics.

Gerdts, D.B. 1988. Antipassives and causatives in Ilokano: evidence for an ergative analysis. In McGinn: 295-321.

Gibson, J.D. and S. Starosta. 1990. Ergativity east and west. In Baldi: 195-210.

Gifford, E.W. 1929. *Tongan society*. Bernice P. Bishop Museum Bulletin 8. Honolulu.

Gifford, E.W. and D. Shutler, Jr. 1956. Archaeological investigations in New Caledonia. Anthropological Records 18: 1-125. Berkeley: University of California Press.

Gil, D. 1996. How to speak backwards in Tagalog. In *Pan-Asiatic Linguistics, Proceedings of the Fourth International Symposium on Language and Linguistics, January 8-10, 1996, Institute of Language and Culture for Rural Development, Mahidol University at Salaya*, vol. 1: 297-306.

_____ 2002. Ludlings in Malayic languages: an introduction. In B.K. Purwo, ed., *PELBBA 15, Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya, Jakarta*: 125-180.

Goddard, I. 1996. The classification of the native languages of North America. In I. Goddard, ed., *Handbook of North American Indians*, Vol. 17, *Languages*: 290- 323. Washington, D.C: Smithsonian Institution.

Golson, J. 2005. Introduction to the chapters on archaeology and ethnology. In Pawley, Attenborough, Golson and Hide: 221-233.

Gonda, J. 1947. The comparative method as applied to Indonesian languages. *Lingua* 1: 86-101.

_____ 1948. The Javanese vocabulary of courtesy. *Lingua* 1: 333-376.

_____ 1950. The functions of word duplication in Indonesian languages. *Lingua* 2: 170-197.

_____ 1952. *Sanskrit in Indonesia*. Nagpur: The International Academy of Indian Culture.

_____ 1975. *Selected studies*. Volume 5: *Indonesian linguistics*. Leiden: Brill.

Gonzalez, A.B. 1973a. 'Classifiers' in Tagalog: a semantic analysis. In Gonzalez, ed: 125-140.

_____ ed., 1973b. *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*. LSP Special Monograph Issue No. 4. Quezon City: Linguistic Society of the Philippines.

Gonzalez, A.B. and L.T. Postrado. 1976. The dissemination of Pilipino. *PJL* 7.1-2: 60- 84.

Goodenough, W.H. 1961. Migrations implied by relationships of New Britain dialects to Central Pacific languages. *JPS* 70: 112-126.

_____ 1962. Comment on Capell, 'Oceanic Linguistics today'. *Current Anthropology* 3: 406-408.

Goodenough, W.H. and H. Sugita. 1980. *Trukese-English dictionary*. Memoirs of the American Philosophical Society, vol. 141. Philadelphia: American Philosophical Society.

Goodenough, W.H. and H. Sugita. 1990. *Trukese-English dictionary*, *Supplementary volume: English-Trukese and index of Trukese word roots*. Memoirs of the American Philosophical Society, vol. 141sg. Philadelphia: American Philosophical Society.

Goudswaard, N. 2005. *The Begak (Ida'an) language of Sabah*. Ph.D. dissertation, Free University of Amsterdam. Utrecht: Landelijke Onderzoekschool Taalwetenschaap (Netherlands Graduate School of Linguistics).

Goulden, R. 1996. The Maleu and Bariai languages of West New Britian. In Ross: 63-144.

Grace, G.W. 1955. Subgrouping of Malayo-Polynesian: a report of tentative findings. *American Anthropologist* 57: 337-339.

_____ 1959. *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family*. Memoir 16, International Journal of American Linguistics.

_____ 1966. Austronesian lexicostatistical classification: a review article. *OL* 5: 13-31.

_____ 1967. Effect of heterogeneity in the lexicostatistical test-list: the case of Rotuman. In G.A. Highland et al, eds., *Polynesian culture history: essays in honor of Kenneth P. Emory*: 289-302. Bernice P. Bishop Special Publication No. 56. Honolulu: Bernice P. Bishop Museum.

_____ 1969. A Proto Oceanic finder list. *Working Papers in Linguistics* 2: 39-84. Honolulu: Department of Linguistics, University of Hawaii.

_____ 1975. *Canala dictionary* (*New Caledonia*). Canberra: Pacific Linguistics.

_____ 1976. *Grand Couli dictionary* (*New Caledonia*). Canberra: Pacific Linguistics.

_____ 1990. The "aberrant" (vs. "exemplary") Melanesian languages. In Baldi: 155- 173.

_____ 1996. Regularity of change in what? In Durie and Ross: 157-179.

Grant, A. and P. Sidwell, eds. 2005. *Chamic and beyond: studies in mainland Austronesian languages*. Canberra: Pacific Linguistics.

Gravius, D. 1661. *Het heylige evangelium Matthei en Johannis overgeset inde Formosaansche tale, voor de inwoonders van Soulang, Mattau, Sinckan, Bacloan, Tavokan, en Tevorang*. Amsterdam (reprinted in Campbell 1888).

Gray, R.D. 2005. Pushing the time barrier in the quest for language roots. *Science* 309, 23 September, 2005.

Gray, R.D., A.J. Drummond, and S.J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323: 479-483.

Green, R.C. 1966. Linguistic subgrouping within Polynesia: the implications for prehistoric settlement. *JPS* 75: 6-38.

_____ 1991. Near and remote Oceania—disestablishing "Melanesia" in culture history. In Pawley: 491-502.

Green, R.C. and A.K. Pawley. 1999. Early Oceanic architectural forms and settlement patterns: linguistic, archaeological and ethnological perspectives. In R. Blench and M. Spriggs, eds., *Archaeology and Language III: Artefacts, languages and texts*: 31-89. London and New York: Routledge.

Greenberg, J.H. 1954. A quantitative approach to the morphological typology of language. In R.F. Spencer, ed., *Method and perspective in anthropology: papers in honor of Wilson D. Wallis*: 192-220. Minneapolis: University of Minnesota.

_____ 1957. *Essays in linguistics*. Chicago: University of Chicago Press.

_____ 1971. The Indo-Pacific hypothesis. In Sebeok: 807-871.

_____ 1978a. Some generalisations concerning initial and final consonant clusters. In Greenberg, ed., 2: 243-279.

_____ 1978b. Generalisations about numeral systems. In Greenberg, ed., 3: 249-295.

_____ ed. 1978. 4 vols. *Universals of Human Language*. Stanford University Press.

Greenhill, S.J., R. Blust, and R.D. Gray. 2008. The Austronesian Basic Vocabulary Database: from bioinformatics to lexomics. *Evolutionary Bioinformatics* 4: 271-283.

Greenhill, S.J. and R. Clark. 2011. Research note: POLLEX-Online: The Polynesian Lexicon project online. *OL* 50: 551-559.

Greenhill, S.J., and R.D. Gray. 2009. Austronesian language phylogenies: myths and Misconceptions about Bayesian computational methods. In A. Adelaar and A. Pawley, eds., Austronesian historical linguistics and culture history: a festschrift for Robert Blust: 375-397. Canberra: Pacific Linguistics.

Grimes, B., ed. 2000. *Ethnologue: languages of the world*. 14th edition. Dallas, Texas: Summer Institute of Linguistics, Inc.

Grimes, C.E. 1991. *The Buru language of eastern Indonesia*. Unpublished Ph.D. dissertation. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

_____ 1997. Compounding and semantic bleaching in languages of eastern Indonesia. In Odé and Stokhof: 277-301.

_____ 2010. Hawu and Dhao in eastern Indonesia: revisiting their relationship. In Michael C. Ewing and Marian Klamer, eds., *East Nusantara: typological and areal analyses*: 251-280. Canberra: Pacific Linguistics.

Grimes, C.E. and K.R. Maryott. 1994. Named speech registers in Austronesian languages. In Dutton and Tryon: 275-319.

Guérin, Valérie. 2011. *A grammar of Mavea*. Honolulu: University of Hawai'i Press.

Guerreiro, A.J. 1998. The Modang men's house in regard to social and cultural values. In R.J. Winzeler, ed., *Indigenous architecture in Borneo: traditional patterns and new developments*:69-87. Borneo Research Council.

Guy, J.B.M. 1974. *A grammar of the northern dialect of Sakao*. Canberra: Pacific Linguistics.

———— 1978. Proto-North New Hebridean reconstructions. In Wurm and Carrington: 781-850.

———— 1983. Glottochronology examined and found wanting. Ms., 42pp. Hage, P. 1998. Was Proto Oceanic society matrilineal? *JPS* 107: 365-379.

Hage, P. 1999. Reconstructing ancestral Oceanic society. *AP* 38: 200-228.

———— 1998. Was Proto Oceanic society matrilineal? *JPS* 107: 365-379.

Hage, P. and F. Harary. 1996. *Island networks: communication, kinship, and classification structures in Oceania*. Cambridge University Press.

Hage, P. and J. Marck. 2003. Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *CA* 44: 121-127.

Hajek, J. 1995. A mystery solved: the forgotten tone languages of New Ireland. *University of Melbourne Working Papers in Linguistics* 14: 9-14.

Hajek, J. and J. Bowden. 2002. Taba and Roma: clusters and geminates in two Austronesian languages. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco*: 1033-1036.

Hale, H. 1846. Notes on the language of Rotuma; Rotuman vocabulary. *United States exploring expedition under the command of Charles Wilkes*, vol. 6: 469-478. Philadelphia: Sherman.

Hale, K. 1968. Review of Hohepa, *A profile-generative grammar of Maori. JPS* 77: 83- 99.

———— 1971. A note on a Walbiri tradition of antonymy. In D.D. Steinberg and L.A. Jakobovits, eds., *Semantics: an interdisciplinary reader in philosophy, linguistics, and psychology*. Cambridge University Press.

Halim, A. 1974. *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Djambatan.

Halim, A., L.Carrington and S.A. Wurm, eds. 1981-1983. 4 vols. *Papers from the Third International Conference on Austronesian Linguistics*. Canberra: Pacific Linguistics.

Hall, K.. 1985. *Maritime trade and state development in early Southeast Asia*. Honolulu: University of Hawaii Press.

Halle, M. 2001. Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In M. Kenstowicz, ed., *Ken Hale: a life in language*: 153-168. Cambridge, MA: MIT Press.

Hamel, P.J. 1994. *A grammar and lexicon of Loniu, Papua New Guinea*. Canberra: Pacific Linguistics.

Happart, G. 1650. *Woordboek der Favorlangsche taal*. Reprinted by W.R. van Hoëvell in *Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen* 18 (1842).

Hardeland, A. 1858. *Einer Grammatik der Dajackschen Sprache*. Amsterdam: Frederik Muller.

———— 1859. *Dajacksch-Deutsches Wörterbuch*. Amsterdam: Frederik Muller.

Harlow, R. and R. Hooper, eds. 1989. 2 Parts. *VICAL 1, Oceanic languages: Papers from the Fifth International Conference on Austronesian Linguistics*. Auckland: Linguistic Society of New Zealand.

Harrison, S.P. 1973. Reduplication in Micronesian languages. *OL* 12: 407-454.

———— 1976. *Mokilese reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Harrison, S.P. and S. Albert. 1977. *Mokilese-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Harrison, S.P. and F.H. Jackson. 1984. Higher numerals in several Micronesian languages. In Bender: 61-79.

Harvey, M. 1982. Subgroups in Austronesian. In Halim, Carrington, and Wurm 2: 47-99.

Hassan, I.U., S.A. Ashley, and M.L. Ashley. 1994. *Tausug-English dictionary: Kabtangan Iban Maana*. Manila: Summer Institute of Linguistics (also published as Sulu Studies 6).

Haudricourt, A.G. 1963. *La langue des Nenemas et des Nigoumak (dialects de Poum et de Koumac, Nouvelle-Calédonie)*. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

_____ 1964. Les consonnes postnasalisées en Nouvelle-Calédonie. *Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass., 1962*: 460-461. The Hague: Mouton.

_____ 1965. Problems of Austronesian comparative philology. *Lingua* 14: 315-329.

_____ 1968. La langue de Gomen et la langue de Touho en Nouvelle-Calédonie. *BSLP* 63: 218-235.

_____ 1971. New Caledonia and the Loyalty islands. In Sebeok: 359-396.

_____ 1984. La tonologie des langues de Hai-nan. *BSLP* 79: 385-394.

Haudricourt, A.G. and F. Ozanne-Rivierre. 1982. *Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie)*. LACITO-documents, Asie- Austronésie, No. 4. Paris: Centre National de la Recherche Scientifique.

Hayes, B. and M. Abad. 1989. Reduplication and syllabification in Ilokano. *Lingua* 77: 331-374.

Hayes, L.H. 1992. On the track of Austric, Part I: Introduction. *Mon-Khmer Studies* 21: 143-177.

_____ 1997. On the track of Austric, Part II: Consonant mutation in early Austroasiatic. *Mon-Khmer Studies* 27: 13-41.

_____ 1999a. On the track of Austric, Part III: Basic vocabulary comparison. *Mon- Khmer Studies* 28: 1-34.

_____ 1999b. The Austric denti-alveolar sibilants. *Mother Tongue* 5: 3-14.

Hazeu, G.A.J. 1907. *Gajōsch-Nederlandsch woordenboek met Nederlandsch-Gajōsch register*. Batavia.

Headland, T.N. 2003. Thirty endangered languages in the Philippines. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 47: 1-12.

Headland, T.N. and J.D. Headland. 1974. *A Dumagat (Casiguran)-English dictionary*. Canberra: Pacific Linguistics.

Healey, P.M. 1960. *An Agta grammar*. Manila: Bureau of Printing.

Held, G.J. 1942. *Grammatica van het Waropensch* (*Nederlandsch Noord Nieuw-Guinea*). VBG 77.1. Bandoeng: A.C. Nix.

Hemley, Robin. 2005. Laurie Reid's importance to the Tasaday controversy. In Liao and Rubino: xxi-xxxii.

Hendon, R. 1964. The reconstruction of *-ew in Proto-Malayopolynesian. *Language* 40: 372-380.

Hendrickson, G.R. and L.E. Newell. 1991. *A bibliography of Philippine language dictionaries and vocabularies*. Manila: LSP.

Hervas y Panduro, L. 1784. *Catalogo delle lingue*. Vol. 17 of Idea dell'Universo, 21 vols. Cesena.

Hewson, J. 1993. *A computer-generated dictionary of Proto-Algonquian*. Seattle: University of Washington Press.

Hicks, D. 1976. *Tetum ghosts and kin: fieldwork in an Indonesian community*. Explorations in World Ethnology. Palo Alto, California: Mayfield Publishing Co.

Himmelmann, N.P. 1991. The Philippine challenge to Universal Grammar. Arbeitspapier Nr. 15 (Neue Folge). Köln: Institut für Sprachwissenschaft, Universität zu Köln.

_____ 2001. *Sourcebook on Tomini-Tolitoli languages: general information and word lists*. Canberra: Pacific Linguistics.

_____ 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. In Adelaar and Himmelmann: 110-181.

Himmelmann, N.P. and J.U. Wolff. 1999. *Toratán* (*Ratahan*). Languages of the World/Materials 130. Munich: Lincom Europa.

Ho, D.A. 1978. A preliminary comparative study of five Paiwan dialects. *BIHP* 49: 565-681.

Hocart, A.M. 1919. Notes on Rotuman grammar. *JRAI* 49: 252-264.

Hockett, C.F. 1958. *A course in modern linguistics*. New York: Macmillan.

Hoffman, C. 1986. *Punan: hunters and gatherers of Borneo*. Ann Arbor, Michigan: University Microfilms International Research Press.

Hohepa, P.W. 1969. The accusative-to-ergative drift in Polynesian languages. *JPS* 78: 295-329.

Hoijer, H. 1956. Lexicostatistics: a critique. *Language* 32: 49-60.

Hollyman, J. and A.K. Pawley, eds. 1981. *Studies in Pacific languages and cultures in honour of Bruce Biggs*. Auckland: Linguistic Society of New Zealand.

Holmer, A.J. 1996. *A parametric grammar of Seediq*. Lund: Travaux de l'institut de linguistique de Lund.

Hooley, B.A. 1971. Austronesian languages of the Morobe District, Papua New Guinea. *OL* 10: 79-151.

Hopper, P.J. and S.A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251-299.

Horne, E.C. 1974. *Javanese-English dictionary*. New Haven: Yale University Press.

Hose, C. and W. McDougall. 1912. 2 vols. *The pagan tribes of Borneo*. London: Macmillan & Co.

Houtman, F. de. 1603. *Spraeck ende woord-boeck, Inde Maleysche ende Madagaskarsche Talen met vele Arabische ende Turcsche woorden*. Amsterdam.

Hovdhaugen, E. and U. Mosel, eds. 1999. *Negation in Oceanic languages: typological studies*. München: Lincom Europa.

Hovdhaugen, E., Å. Næss and I. Hoëm. 2002. *Pileni texts with a Pileni-English vocabulary and an English-Pileni finderlist*. The Kon-Tiki Museum Occasional Papers, Vol. 7. Oslo: The Kon-Tiki Museum.

Howells, W.W. 1973. *The Pacific islanders*. New York: Scribner's; London: Weidenfeld and Nicolson.

Huang, L.M. 1993. *A study of Atayal syntax*. Taipei: Crane.

_____ 2001. Focus system of Mayrinax Atayal: a syntactic, semantic and pragmatic perspective. *Journal of Taiwan Normal University* 46: 51-69.

Huang, L.M., M.M. Yeh, E. Zeitoun, A.H. Chang and J.J. Wu. 1998. A typological overview of nominal case marking systems of Formosan languages. In S. Huang, ed., *Selected Papers from the International Symposium on Languages in Taiwan*: 21-48. Taipei: Crane.

Huang, S. and M. Tanangkingsing. 2005. Reference to motion events in six western Austronesian languages: toward a semantic typology. *OL* 44: 307-340.

Hudson, A.B. 1967. *The Barito isolects of Borneo: a classification based on comparative reconstruction and lexicostatistics*. Data Paper No. 68, Southeast Asia Program, Department of Asian Studies, Cornell University. Ithaca, N.Y.: Cornell University.

Hull, Geoffrey. 2002. *Standard Tetum-English dictionary*. 3rd edition. London: Allen & Unwin.

Humboldt, W. von. 1836-1839. *Über die Kawi-Sprache auf der Insel Java, nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin, part 2. Berlin: Königlichen Akademie der Wissenschaften.

Hymes, D. 1960. Lexicostatistics so far. *CA* 1: 3-44 (with comments).

———— 1962. Comment on Bergsland and Vogt. *CA* 3: 136-141.

Hyslop, C. 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. Canberra: Pacific Linguistics.

Inglis, John. 1882. *A dictionary of the Aneityumese language*. London: Williams and Norgate.

Ingram, D. 1978. Typology and universals of personal pronouns. In Greenberg 3: 213-247.

Institute of Fijian Language and Culture. 2007. *Na iVola Vosa VakaViti* (*The Fijian Monolingual Dictionary*). Suva: Government Printing Office.

Ivens, W.G. 1929. *A dictionary of the language of Sa'a (Mala) and Ulawa, South-East Solomon Islands*. London: Oxford University Press.

Jackson, F.H. 1983. *The internal and external relationships of the Trukic languages of Micronesia*. Ph.D. dissertation, University of Hawaii. Ann Arbor: University Microfilms International.

Jackson, F.H. and J.C. Marck. 1991. *Carolinian-English dictionary*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

Jakobson, R. 1960. Why mama and papa? In B. Kaplan and S. Wapner, eds, *Perspectives in psychological theory dedicated to Heinz Werner*: 124-134. New York: International Universities Press, Inc.

Janhunen, J. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77: 219-274.

Jaspan, M.A. 1984. *Materials for a Rejang-Indonesian-English dictionary*. Canberra: Pacific Linguistics.

Jauncey, Dorothy. 2011. *A grammar of Tamambo, the language of west Malo, Vanuatu*. Canberra: Pacific Linguistics.

Jensen, J.T. 1977a. *Yapese reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

———— 1977b. *Yapese-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Johnson, R. A. 1996. *A bibliography of Philippine linguistics*. Manila: LSP.

Johnson, R., G.O. Tan and C. Goshert. 2003. 4th ed. *Bibliography of the Summer Institute of Linguistics, Philippines 1953-2003. 50th anniversary edition.* Manila: Summer Institute of Linguistics, Philippines, Inc.

Johnston, R.L. 1980. *Language, communication and development in New Britain*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Jones, A.A. 1998. *Towards a lexicogrammar of Mekeo (An Austronesian language of western Central Papua)*. Canberra: Pacific Linguistics.

Jones, R. 1978. *Indonesian etymological project, III: Arabic loan-words in Indonesian*. Published simultaneously by the Indonesian Etymological Project and as Cahier d'Archipel 2, SECMI, Paris. London: School of Oriental and African Studies, University of London.

Jonker, J.C.G. 1906. Over de eind-medeklinkers in het Rottineesch en Timoreesch. *BKI* 59: 263-343.

———— 1908. *Rottineesch-Hollandsch woordenboek*. Leiden: Brill.

_____ 1914. Kan man bij de talen van de Indische Archipel eene westelijke en eene oostelijke afdeeling onderscheiden? *Verslagen en medeelingen der Koninklijke Akademie van Wetenschappen*, 4th series, 12: 314-417.

_____ 1915. *Rottineesche spraakkunst*. Leiden: Brill.

_____ 1932. *Lettineesche taalstudiën*. VBG 69. Bandoeng: Nix.

Josephs, L.S. 1975. *Palauan reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Kadazan Dusun Cultural Association. 1995. *Kadazan Dusun-Malay-English dictionary*. Kota Kinabalu, Sabah.

___ 2004. *Kamus Murut Timugon-Melayu, dengan ikhtisar ethnografi*. Kota Kinabalu, Sabah.

Kähler, H. 1959. *Vergleichendes Wörterverzeichnis der Sichule-Sprache auf der Insel Simalur an der Westküste von Sumatra*. VSIS 1. Berlin: Reimer.

_____ 1965. *Grammatik der Bahasa Indonésia*. 2nd, rev. edition. Wiesbaden: Harrassowitz.

_____ 1987. *Enggano-Deutsches Wörterbuch*. VSIS 14. Berlin: Reimer.

Kawamoto, T. 1977. Toward a comparative Japanese-Austronesian I. *Bulletin of Nara University of Education* 26.1: 23-49.

_____ 1984. Two sets of sound laws between Japanese and Austronesian. *Bulletin of Joetsu University of Education* 3: 31-50.

Kayser, A. 1993 [1936]. *Nauru grammar*. Edited, and with an introduction by K.H. Rensch. Canberra: Embassy of the Federal Republic of Germany.

Keenan, E.L. 1976. Remarkable subjects in Malagasy. In C.N. Li: 247-301.

Keesing, F.M. 1962. *The ethnohistory of northern Luzon*. Stanford, California: Stanford University Press.

Keesing, R.M. 1988. *Melanesian pidgin and the Oceanic substrate*. Stanford: Stanford University Press.

Kelly, R.L., J-F. Rabedimy and L. Poyer. 1999. The Mikea of Madagascar. In R.B. Lee and R. Daly, eds., *The Cambridge Encyclopedia of hunters and gatherers*: 215- 219. Cambridge University Press.

Kempler Cohen, E.M. 1999. *Fundamentals of Austronesian roots and etymology*. Canberra: Pacific Linguistics.

Kennedy, R. 2003. *Confluence in phonology: evidence from Micronesian reduplication*. Ph.D. dissertation. Tucson: Department of Linguistics, University of Arizona.

Keraf, G. 1978. *Morfologi dialek Lamalera*. Ph.D. dissertation, University of Indonesia. Ende, Flores: Arnoldus.

Kern, H. 1886. De Fidji-taal vergeleken met hare verwanten in Indonesië en Polynesië. *Verhandelingen van het Koninklijke Nederlandsche Akademie van Wetenschappen, afdeeling Letterkunde* 16: 1-242. Reprinted in *Verspreide Geschriften* 4: 243-343, 5: 1-141 (1917).

_____ 1889. Taalkundige gegevens ter bepaling van het stamland der Maleisch-Polynesische volken. *Verslagen en Mededeelingen der Koninklijke Akademie van Wetenschappen, afdeeling Letterkunde*, 3: 270-287. Reprinted in *Verspreide Geschriften* 6: 105-120 (1917).

_____ 1906a. Taalvergelijkende verhandeling over het Aneityumsch, met een aanhangsel over het klankstelsel van het Eromanga. *Verhandelingen van het Koninklijke Akademie van Wetenschappen, afdeling Letterkunde* 8.1: 1-146. Reprinted in *Verspreide Geschriften* 5: 149-285 (1917).

_____ 1906b. J.L.A. Brandes. Levensberichten der afgestorven medeleden van de Maatschappij der Nederlandsche Letterkunde. Reprinted in *Verspreide Geschriften* 15: 299-310 (1917).

_____ 1908. Austronesisch en Austroasiatisch (review of Schmidt 1906). *BKI* 60: 166- 172. Reprinted in *Verspreide Geschriften* 14: 317-325 (1917).

_____ 1913-1928. *Verspreide Geschriften* (Collected Works). 15 vols. The Hague: Nijhoff.

Kern, R.A. 1948. The vocabularies of Iacob Le Maire. *Acta Orientalia XX*: 216-237.

Key, M.R. 1984. *Polynesian and American linguistic connections*. Edward Sapir Monograph Series in Language, Culture, and Cognition 12. Supplement to *Forum Linguisticum* 8.3.

_____ 1998. Linguistic similarities between Austronesian and South American Indian languages. *Pre-Columbiana: A journal of long-distance contacts* 1: 59-71.

Kikusawa, R. 2002. *Proto Central Pacific ergativity: its reconstruction and development in the Fijian, Rotuman and Polynesian languages*. Canberra: Pacific Linguistics.

King, J.K. and J.W. King, eds. 1984. *Languages of Sabah: a survey report*. Canberra: Pacific Linguistics.

Kirch, P.V. 1997. *The Lapita peoples: ancestors of the Oceanic world*. The Peoples of South-East Asia and the Pacific. Oxford: Blackwell.

_____ 2000. *On the road of the winds: an archaeological history of the Pacific islands before European contact*. Berkeley: University of California Press.

Kirch, P.V. and R.C. Green. 2001. *Hawaiki, ancestral Polynesia: an essay in historical anthropology*. Cambridge University Press.

Klamer, M. 1998. *A grammar of Kambera.* Berlin: Mouton de Gruyter.

_____ 2002. Typical features of Austronesian languages in central/eastern Indonesia. *OL* 41: 363-383.

Klieneberger, H.R. 1957. *Bibliography of Oceanic linguistics*. London and New York: Oxford University Press.

Klinkert, H.C. 1918. *Nieuw Maleisch-Nederlandsch Zakwoordenboek*. Leiden.

Krauss, M.E. 1964. Proto-Athapaskan-Eyak and the problem of Na-Dene I: The phonology. *International Journal of American Linguistics* 30: 118-131.

_____ 1979. Na-Dene and Eskimo-Aleut. In L. Campbell and M. Mithun, eds., *The languages of native America: historical and comparative assessment*: 803-901. Austin and London: University of Texas Press.

Krauss, M.E. and V. Golla. 1981. Northern Athapaskan languages. In W.C. Sturtevant, general editor, *Handbook of North American Indians*, vol. 6: J. Helm, volume editor, Subarctic: 67-85. Washington, D.C.: Smithsonian Institution.

Krishnamurti, B. 1961. *Telugu verbal bases: a comparative and descriptive study*. University of California Publications in Linguistics 24. Berkeley and Los Angeles: University of California Press.

_____ 2003. *The Dravidian languages*. Cambridge Language Surveys. Cambridge University Press.

Kroeger, P.R. 1988. Verbal focus in Kimaragang. In Steinhauer 1988b: 217-240.

_____ 1990. Asu vs. tasu: on the origins of Dusunic moveable t-. In J.T. Collins, ed., *Language and oral traditions in Borneo*. Borneo Research Council Proceedings Series, vol. 2: 93-114. Williamsburg, Virginia: Department of Anthropology, The College of William and Mary in Virginia.

_____ 1992. Vowel harmony systems in three Sabahan languages. In Martin: 279-296.

_____ 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, California: Center for the Study of Language and Information, Stanford University.

Kuiper, F.B.J. 1948. Munda and Indonesian. *Orientalia Nederlandica. A volume of Oriental Studies*: 372-401. Leiden: Sijthoff.

Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Ladefoged, P. 1971. *Preliminaries to linguistic phonetics*. Chicago: The University of Chicago Press.

Ladefoged, P. and I. Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lafeber, A. 1922. *Vergelijkende klankleer van het Niassisch*. Ph.D. dissertation, University of Leiden. The Hague.

Laidig, W.D. 1993. Insights from Larike possessive constructions. *OL* 32: 311-351.

Laidig, W.D. and S.D. Maingak. 1999. Barang-Barang phonology: a preliminary description. In W.D. Laidig, ed., *Studies in Sulawesi linguistics*, Part VI: 46-83. NUSA 46.

Lakoff, G. and M. Johnson. 1980. *Metaphors we live by*. Chicago and London: The University of Chicago Press.

Larish, M.D. 1999. *The position of Moken and Moklen within the Austronesian language family*. Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

Latta, F.C. 1977. Nasalisation in Sundanese: a closer look. Ms, 14pp. Laves, G. 1935. Review of Dempwolff 1934. *Language* 11: 264-267.

Laycock, D.C. 1975. Observations on number systems and semantics. In Wurm 1: 219- 233.

_____ 1978. A little Mor. In Wurm and Carrington 1: 285-316.

_____ 1981. Metathesis in Ririo and other cases. In Halim, Carrington, and Wurm 1: 269-281. Canberra: Pacific Linguistics.

Lebar, F.M., ed. 1972. *Ethnic groups of insular Southeast Asia, vol. 1: Indonesia, Andaman islands, and Madagascar*. New Haven: Human Relations Area Files Press.

_____ ed. 1975. *Ethnic groups of insular Southeast Asia, vol. 2: Philippines and Formosa*. New Haven: Human Relations Area Files Press.

Lebar, F.M., G.C. Hickey and J.K. Musgrave, eds. 1964. *Ethnic groups of mainland Southeast Asia*. New Haven: Human Relations Area Files Press.

Lee, E.W. 1966. *Proto-Chamic phonologic word and vocabulary*. Ph.D. dissertation, Indiana University. Ann Arbor, Michigan: University Microfilms International.

Lee, K.D. 1975. *Kusaiean reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

_____ 1976. *Kusaiean-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Lee, K.D. and J.W. Wang. 1984. Kosraean reflexes of Proto Oceanic phonemes. In Bender: 403-442.

Leenhardt, M. 1946. *Langues et dialects de l'Austro-Mélanésie*. Travaux et mémoires de l'institute d'ethnologie. Paris: Institute d'ethnologie, Musée de l'homme.

Lees, R.B. 1953. The basis of glottochronology. *Language* 29: 113-127.

Lehmann, W.P. 1992. *Historical linguistics: an introduction* (3rd edition). London and NewYork: Routledge.

Lemaréchal, A. 1982. Sémantisme des parties du discourse et sémantisme des relations. *BSLP* 77: 1-39.

Lemle, M. 1971. Internal classification of the Tupi-Guarani linguistic family. In D. Bendor-Samuel, ed., *Tupi Studies I*: 107-129. Norman, Oklahoma: Summer Institute of Linguistics.

Lepsius, C.R. 1863. *Standard alphabet*. London and Berlin.

Lewis, M. Paul, ed. 2009. *Ethnologue: languages of the world*. 16th edition. Dallas, Texas: Summer Institute of Linguistics, Inc.

Li, C.N. ed. 1976. *Subject and topic*. New York: Academic Press, Inc.

Li, P.J.K. 1972. On comparative Tsou. *BIHP* 49: 311-348.

_____ 1973. *Rukai structure*. Institute of History and Philology Special Publications No. 64. Nankang, Taipei: Institute of History and Philology, Academia Sinica.

_____ 1977a. Morphophonemic alternations in Formosan languages. *BIHP* 48: 375- 413.

_____ 1977b. The internal relationships of Rukai. *BIHP* 48: 1-92.

_____ 1980a. The phonological rules of Atayal dialects. *BIHP* 51: 349-405.

_____ 1980b. Men's and women's speech in Mayrinax. In P.J.K. Li et al, eds., *Papers in honor of Professor Lin Yü-k'eng on her seventieth birthday*: 9-17. Taipei: Wen Shin Publishing Co.

_____ 1981. Reconstruction of Proto-Atayalic phonology. *BIHP* 52: 235-301.

_____ 1982a. Male and female forms of speech in the Atayalic group. *BIHP* 53: 265- 304.

_____ 1982b. Atayalic final voiced stops. In Halim, Carrington, and Wurm 2: 171-185. Canberra: Pacific Linguistics.

_____ 1982c. Kavalan phonology: synchronic and diachronic. In Carle et al: 479-495.

_____ 1983. Types of lexical derivation of men's speech in Mayrinax. *BIHP* 54: 1-18.

_____ 1985. The position of Atayal in the Austronesian family. In Pawley and Carrington: 257-280.

_____ 1988. A comparative study of Bunun dialects. *BIHP* 59: 479-508.

_____ 1992. *The internal and external relationships of the Formosan languages*. Taipei: National Museum of Prehistory Planning Bureau.

_____ 1995a. Formosan vs, Non-Formosan features in some Austronesian languages in Taiwan. In Li et al.: 651-681.

_____ 1995b. Is Chinese genetically related to Austronesian? In Wang: 93-112.

_____ 2004a. Origins of the East Formosans: Basay, Kavalan, Amis, and Siraya. In Li 2004b: 927-940.

_____ 2004b. *Selected papers on Formosan languages*. 2 vols. Language and Linguistics Monograph Series C3. Taipei: Institute of Linguistics, Academia Sinica.

_____ 2011. *Thao texts and songs*. *LLMS* 44. Taipei: Institute of Linguistics, Academia Sinica.

Li, P.J.K., C.H. Tsang, Y.K. Huang, D.A. Ho, and C.Y. Tseng. 1995. *Austronesian studies relating to Taiwan*. Symposium Series of the Institute of History and Philology, Academia Sinica, No. 3. Taipei: Academia Sinica.

Li, P.J.K. and S. Tsuchida. 2001. *Pazih dictionary*. *LLMS* A2. Taipei: Institute of Linguistics, Academia Sinica.

_____ 2002. *Pazih texts and songs*. *LLMS* A2-2. Taipei: Institute of Linguistics, Academia Sinica.

_____ 2006. *Kavalan dictionary*. *LLMS* A-19. Taipei: Institute of Linguistics, Academia Sinica.

Liao, H.C. 2004. *Transitivity and ergativity in Formosan and Philippine languages*. Unpublished Ph.D. dissertation. 582pp. Honolulu: Department of Linguistics, University of Hawaii.

_____ 2008. A typology of first person dual pronouns and their reconstructibility in Philippine languages. *OL* 47: 1-29.

Liao, H.C. and C.R.G. Rubino, eds. 2005. *Current issues in Philippine linguistics and anthropology parangal kay Lawrence A. Reid*. Manila: The Linguistic Society of the Philippines and SIL, Philippines.

Lichtenberk, F. 1983. *A grammar of Manam*. Honolulu: *OLSP* 18.

_____ 1984. To'aba'ita language of Malaita, Solomon Islands. *Working Papers in Anthropology, Archaeology, Linguistics, Maori Studies*. Auckland: Department of Anthropology, University of Auckland.

_____ 1985. Possessive constructions in Oceanic languages and in Proto Oceanic. In Pawley and Carrington: 93-140.

_____ 1986. Leadership in Proto Oceanic society: linguistic evidence. *JPS* 95: 341- 356.

_____ 1988. The Cristobal-Malaitan subgroup of Southeast Solomonic.*OL* 27: 24-62.

_____ 1994. Reconstructing heterogeneity. *OL* 33: 1-36.

_____ 2008a. *A grammar of Toqabaqita*. Berlin: Mouton de Gruyter.

_____ 2008b. *A dictionary of Toqabaqita (Solomon Islands)*. Canberra: Pacific Linguistics.

_____ 2009. Oceanic possessive classifiers. *OL* 48: 379-402.

Lieber, M.D. and K.H. Dikepa. 1974. *Kapingamarangi lexicon*. PALI Language Texts: Polynesia. Honolulu: The University Press of Hawaii.

Lincoln, P.C. 1978. Reef-Santa Cruz as Austronesian. In Wurm and Carrington 2: 929- 967.

Lindstrom, L. 1986. *Kwamera dictionary*. Canberra: Pacific Linguistics.

Lister-Turner, R. and J.B. Clark. 1930. *A grammar of the Motu language of Papua* (2nd edition). Sydney: Government Printer.

Lobel, J.W. 2004. Old Bikol *–um-* vs. *mag-* and the loss of a morphological paradigm. *OL* 43: 469-497.

_____ 2005. The angry register of the Bikol languages of the Philippines. In Liao and Rubino: 149-166.

_____ 2010. Manide: an undescribed Philippine language. *OL* 49: 478-510.

_____ 2013. *Philippine and North Bornean languages: issues in description, subgrouping, and reconstruction*. Unpublished Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawai'i.

_____ n.d. Central Philippine verbal focus-aspect-mood morphology and its implications. Ms.

Lobel, J.W. and W.C. Hall. 2010. Southern Subanen aspiration. *OL* 49: 319-338.

Lobel, J.W. and L.H.S. Riwarung. 2009. Maranao revisited: an overlooked consonant contrast and its implications for lexicography and grammar. *OL* 48: 403-438.

_____ 2011. Maranao: a preliminary phonological sketch with supporting audio. Language Documentation and Conservation 5: 31-59. Available online athttp://hdl.handle.net/10125/4487/.

Lopez, C. 1939. *Studies on Dempwolff's 'Vergleichende Lautlehre des austronesischen Wortschatzes'*. Summer Institute of Linguistics, Philippines. Mimeo.

_____ 1965. The Spanish overlay in Tagalog. *Lingua* 14: 467-504.

_____ 1967. Classifiers in Philippine languages. *Philippine Journal of Science* 96: 1-8.

Luomala, K. 1951. *The menehune of Polynesia and other mythical little people of Oceania*. Bernice P. Bishop Museum Bulletin 230. Honolulu.

Lynch J. 1977a. *Lenakel dictionary*. Canberra: Pacific Linguistics.

_____ 1977b. *Notes on Maisin – an Austronesian language of the Northern Province of Papua New Guinea*. Mimeo. Port Moresby: University of Papua New Guinea.

_____ 1978a. *A grammar of Lenakel*. Canberra: Pacific Linguistics.

_____ 1978b. Proto-South Hebridean and Proto Oceanic. In Wurm and Carrington: 717-779.

_____ 1981. Melanesian diversity and Polynesian homogeneity: the other side of the coin. *OL* 20: 95-129.

_____ 1996. Proto Oceanic possessive-marking. In Lynch and Pat: 93-110.

_____ 1998. *Pacific languages: an introduction*. Honolulu: University of Hawaii Press.

_____ 2000. *A grammar of Anejom̃*. Canberra: Pacific Linguistics.

_____ 2001. *The linguistic history of southern Vanuatu*. Canberra: Pacific Linguistics.

_____ 2002. The Proto Oceanic labiovelars: some new observations. *OL* 41: 310-362.

_____ 2003a. Low vowel dissimilation in Vanuatu languages. *OL* 42: 359-406.

_____ 2005. Final consonants in remote Oceanic. *OL* 44: 90-112.

_____ 2009a. Irregular sound change in some Malakula languages. In Alexander Adelaar and Andrew Pawley, eds., *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*: 57-72. Canberra: Pacific Linguistics.

_____ 2009b. At sixes and sevens: the development of numeral systems in Vanuatu and New Caledonia. In Evans: 391-411.

_____ 2010. Vowel loss in Tirax and the history of the apicolabial shift. *OL* 49: 369-388.

_____ ed. 2003b. *Issues in Austronesian historical phonology*. Canberra: Pacific Linguistics.

Lynch J. and T. Crowley. 2001. *Languages of Vanuatu: A new survey and bibliography*. Canberra: Pacific Linguistics.

Lynch J. and D.T. Tryon. 1985. Central-Eastern Oceanic: a subgrouping hypothesis. In Pawley and Carrington: 31-52.

Lynch J. and F. Pat, eds., 1996. *Oceanic studies: Proceedings of the First International Conference on Oceanic Linguistics*. Canberra: Pacific Linguistics.

Lynch J., M. Ross and T. Crowley, eds. 2002. *The Oceanic languages*. Curzon Language Family Series. Richmond, Surrey: Curzon Press.

Maan, G. 1951. *Proeve van een Bulische spraakkunst. VKI* 10. The Hague: Nijhoff.

Mabuchi, T. 1953. Takasagozoku no bunrui: gakushi no kaiko [Retrospective on the classification of the Formosan aborigines]. *Japanese Journal of Ethnology* 18: 1- 11.

Macdonald, D. 1907. *The Oceanic languages. Their grammatical structure, vocabulary, and origin*. London: Henry Frowde.

Macdonald, R.R. and S. Darjowidjojo. 1967. *A student's reference grammar of modern formal Indonesian*. Washington, D.C.: Georgetown University Press.

Maddieson, I. 1984. *Patterns of sounds*. Cambridge Studies in Speech Science and Communication. Cambridge University Press.

_____ 1989a. Aerodynamic constraints on sound change: the case of bilabial trills. *UCLA Working Papers in Phonetics* 72: 91-115.

_____ 1989b. Linguo-labials. In Harlow and Hooper 2: 349-375.

Maddieson, I., and K. F. Pang. 1993. Tone in Utsat. In Edmondson and Gregerson: 75-89.

Madulid, D.A. 2001. *A dictionary of Philippine plant names*. 2 vols. Makati City, Philippines: The Bookmark, Inc.

Mahdi, Waruno. 1988. *Morphophonologische Besonderheiten und historische phonologie des Malagasy*. VSIS, vol. 20. Berlin: Reimer.

_____ 1994a. Some Austronesian maverick protoforms with culture-historical implications – I. *OL* 33: 167-229.

_____ 1994b. Some Austronesian maverick protoforms with culture-historical implications – II. *OL* 33: 431-490.

_____ 1996. Another look at Proto-Austronesian *d and *D. In Nothofer: 1-14.

_____ 1999a. The dispersal of Austronesian boat forms in the Indian Ocean. . In R. Blench and M. Spriggs, eds., *Archaeology and language III: Artefacts, languages and texts*: 144-179. London & New York: Routledge.

_____ 1999b. Linguistic and philological data towards dating Austronesian activity in India and Sri Lanka. In R. Blench and M. Spriggs, eds., *Archaeology and language IV: Language change and cultural transformation*: 160-242. London & New York: Routledge.

_____ 2005. Old Malay. In Adelaar and Himmelmann: 182-201.

_____ 2012. Review of John U. Wolff (2010, *Proto-Austronesian phonology with glossary*, vols. I-II). *Archipel* 83: 214-222.

Makarenko, V.A. 1981. *Preliminary annotated bibliography of Pilipino linguistics (1604-1976)*, ed. by A. Gonzalez and C. Nemenzo Sacris. Manila: De La Salle University Libraries and LSP.

Mallory, J.P. and D.Q. Adams, eds. 1997. *Encyclopedia of Indo-European culture*. London and Chicago: Fitzroy Dearborn.

Marck, J. 1986. Micronesian dialects and the overnight voyage. *JPS* 95: 253-258.

_____ 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.

Maree, J.Y.M., and O.R. Tomas. 2012. *Ibatan to English dictionary*. Manila: Summer Institute of Linguistics.

Marrison, G.E. 1975. The early Cham language and its relationship to Malay. *JMBRAS* 48.2: 52-60.

Marschall, W. 1968. *Metallurgie und frühe Besiedlungsgeschichte indonesiens*. Ethnologica, Band 4. Köln: Brill.

Marsden, W. 1783. *The history of Sumatra*. London.

_____ 1834. On the Polynesian, or East-insular languages. In *Miscellaneous Works of William Marsden*: 1-116. London

_____ 1984 [1812]. *A dictionary and grammar of the Malayan language, with an introduction by R. Jones*. Singapore and New York: Oxford University Press.

Marshall, M. 1984. Structural patterns of sibling classification in Oceania. *CA* 25: 597- 637.

Marten, L. 2006. Bantu classification, Bantu trees and phylogenetic methods. In P. Forster and C. Renfrew, eds., *Phylogenetic methods and the prehistory of languages*: 43-55. Cambridge: McDonald Institute for Archaeological Research, University of Cambridge.

Martens, M. 1995. Uma. In Tryon 1995, Part 1, Fascicle 1: 539-547.

Martin, P.W., ed. 1992. *Shifting patterns of language use in Borneo: Papers from the Second Bi-Ennial International Conference, Kota Kinabalu, Sabah, Malaysia*. Borneo Research Council Proceedings Series, vol. 3. Williamsburg, Virginia: Department of Anthropology, The College of William and Mary in Virginia.

Matisoff, J. A. 1975. Rhinoglottophilia: the mysterious connection between nasality and glottality. In C.A. Ferguson, L.M. Hyman and J.J. Ohala, eds., *Nasálfest: Papers from a symposium on nasals and nasalisation*: 265-287. Stanford, California: Language Universals Project, Department of Linguistics, Stanford University.

_____ 1978. *Variational semantics in Tibeto-Burman*. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, vol. 6. Philadelphia: Institute for the Study of Human Issues.

_____ 1990. On megalocomparison. *Language* 66:106-120.

_____ 1995. Sino-Tibetan numerals and the play of prefixes. *Bulletin of the National Museum of Ethnology* (Osaka) 20.1: 105-252.

_____ 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.

Matteson, E. 1972. Proto Arawakan. In E. Matteson, et al, *Comparative studies in Amerindian languages*: 160-242. Janua Linguarum Series Practica, 127. The Hague: Mouton.

Matthes, B.F. 1858. *Makassaarsche Spraakkunst*. Amsterdam: Frederik Muller.

_____ 1859. *Makassaarsch-Hollandsch woordenboek, met Hollandsch-Makassaarsch woordenlijst*. Amsterdam: Frederik Muller.

_____ 1874. *Boegineesch-Hollandsch woordenboek, met Hollandsch- Boegineesche woordenlijst*. The Hague: M. Nijhoff.

_____ 1875. *Boegineesche Spraakkunst*. The Hague: M. Nijhoff.

Matthews, Peter. 1997. *The concise Oxford dictionary of linguistics*. Oxford and New York: Oxford University Press.

Maxwell, W.E. 1907 [1882]. 8th ed. *A manual of the Malay language, with an introductory sketch of the Sanskrit element in Malay*. London: Kegan Paul, Trench , Trübner & Co.

Mayer, J. 2001. *Code-switching in Samoan: t-style and k-style*. Unpublished Ph.D. dissertation, Department of Linguistics, University of Hawaii.

McCarthy, J.J. and A. Prince. 1990. Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language and Linguistic Theory* 8: 209-283.

_____ 1994. The emergence of the unmarked: optimality in prosodic morphology. In M. Gonzalez, ed., *Proceedings of the North East Linguistic Society* 24: 333-379. Amherst, Mass, Graduate Linguistic Student Association.

McElhanon, K.A. and C.L. Voorhoeve. 1970. *The Trans-New Guinea phylum: explorations in deep-level genetic relationships*. Canberra: Pacific Linguistics.

McFarland, C.D. 1976. *A provisional classification of Tagalog verbs*. *SLCAA* Monograph Series, No. 8. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

_____ 1977. *Northern Philippine linguistic geography*. *SLCAA* Monograph Series, No. 9. Tokyo.

_____ 1980. *A linguistic atlas of the Philippines*. *SLCAA* Monograph Series No. 15. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

McGinn, R. 1989. *The Rejang language: texts and grammatical analysis, based on the Musi dialect*. Ms., 138pp.

_____ 2005. What the Rawas dialect reveals about the linguistic history of Rejang. *OL* 44: 12-64.

_____ ed. 1988. *Studies in Austronesian linguistics*. Ohio University Monographs in International Studies, Southeast Asia Series, No. 76. Athens, Ohio: Ohio University Center for International Studies, Center for Southeast Asian Studies.

McGuckin, C. 2002. Gapapaiwa. In Lynch, Ross and Crowley: 297-321.

McKaughan, H.P. 1958. The inflection and syntax of Maranao verbs. *Publications of the Institute of National Language*. Manila: Bureau of Printing.

_____ 1970. Topicalisation in Maranao – an addendum. In S.A. Wurm and D.C. Laycock, eds, *Pacific linguistic studies in honour of Arthur Capell*: 291-300. Canberra: Pacific Linguistics.

McManus, E.G. and L.S. Josephs. 1977. *Palauan-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Mead, D. 1996. The evidence for final consonants in Proto-Bungku-Tolaki. *OL* 35: 180- 194.

_____ 1998. *Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax*. Ph.D. dissertation, Department of Linguistics, Rice University. Ann Arbor, Michigan: University Microfilms International.

_____ 1999. *The Bungku-Tolaki languages of South-Eastern Sulawesi, Indonesia*. Canberra: Pacific Linguistics.

_____ 2001. The numeral confix *i- -(e)n. *OL* 40: 167-176.

_____ 2003a. The Saluan-Banggai microgroup of eastern Sulawesi. In Lynch: 65-86.

_____ 2003b. Evidence for a Celebic supergroup. In Lynch, ed: 115-141.

Meinhof, C. 1899. *Grundriss einer Lautlehre der Bantusprachen*. Leipzig: F.A. Brockhaus.

Merrill, E.D. 1954. *Plant life of the Pacific world*. New York: Macmillan.

Mettler, T. and H. Mettler. 1990. Yamdena phonology. *WILC 8*: 29-79.

Milke, W. 1958. Ozeanische Verwandtschaftsnamen. *ZfE* 83: 226-229.

_____ 1961. Beiträge zur ozeanischen Linguistik. *ZfE* 86: 162-182.

_____ 1968. Proto Oceanic addenda. *OL* 7: 147-171.

Miller, W.R. 1967. Uto-Aztecan cognate sets. *University of California Publications in Linguistics*, vol. 48. Berkeley and Los Angeles: University of California Press.

Mills, R.F. 1975. *Proto South Sulawesi and Proto Austronesian phonology*. 2 vols. Ph.D. dissertation, Department of Linguistics, The University of Michigan. Ann Arbor, Michigan: University Microfilms International.

Mills, R.F. and J. Grima. 1980. Historical developments in Lettinese. In Naylor: 273- 284.

Milner, G.B. 1958. Aspiration in two Polynesian languages. BSOAS 21: 368-375.

_____ 1961. The Samoan vocabulary of respect. *JRAI* 91.2: 296-317.

_____ 1966. *Samoan dictionary*. London: Oxford University Press.

_____ 1967. *Fijian grammar* (2nd edition). Suva, Fiji: Government Printing Department.

Mintz, M.W. 1971. *Bikol grammar notes*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

_____ 1991. Anger and verse: two vocabulary subsets in Bikol. In R. Harlow, ed., *VICAL 2: Western Austronesian and contact languages. Papers from the Fifth International Conference on Austronesian Linguistics, Parts One and Two*: 231- 244. Te Reo Special Publication. Auckland: Linguistic Society of New Zealand.

_____ 1998. *A course in conversational Indonesian*. Raffles Editions. 2nd rev. edition. Singapore: SNP Publishing Co.

Mintz, M.W. and J. del Rosario Britanico. 1985. *Bikol-English dictionary*. Quezon City, Philippines: New Day Publishers.

Mithun, M. 1994. The implications of ergativity for a Philippine voice system. In B. Fox and P.J. Hopper, eds., *Voice: form and function*: 247-277. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Mithun, M. and H. Basri. 1986. The phonology of Selayarese. *OL* 25: 210-254.

Moeliono, A.M., et al. 1989. 2nd printing. *Kamus besar Bahasa Indonesia*. Jakarta: Balai Pustaka.

Moeliono, A.M., and C.E. Grimes. 1995. Indonesian (Malay). In Tryon: 443-457.

Molony, C.H. and D. Tuan. 1976. Further studies on the Tasaday language: texts and vocabulary. In D.E. Yen and J. Nance, eds., *Further studies on the Tasaday*: 13- 96. Panamin Foundation Research Series, No. 2. Makati, Rizal, Philippines: Panamin Foundation.

Moriguchi, T. 1991. Asai's Basai vocabulary. In Tsuchida, Yamada and Moriguchi: 195-257. Morris, C. 1984. *Tetun-English dictionary*. Canberra: Pacific Linguistics.

Mosel, U. 1980. *Tolai and Tok Pisin: the influence of the substratum on the development of New Guinea Pidgin*. Canberra: Pacific Linguistics.

_____ 1999. Towards a typology of negation in Oceanic languages. In Hovdhaugen and Mosel: 1-19.

Mosel, U. and E. Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.

Mosel, U. and R.S. Spriggs. 1999. Negation in Teop. In Hovdhaugen and Mosel: 45-56.

Motus, C.L. 1971. *Hiligaynon dictionary*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Moyse-Faurie, C. 1993. *Dictionnaire futunien-français*. Langues et cultures du Pacifique 8, SELAF 340. Paris: Peeters.

_____ 1995. *Le Xârâcùù, Langue de Thio-Canala (Nouvelle-Calédonie)*. Langues et cultures du Pacifique 10, SELAF 355. Paris: Peeters.

Moyse-Faurie, C. and M-A. Néchérö-Jorédié. 1986. *Dictionnaire Xârâcùù-Français (Nouvelle-Calédonie)*. Paris: EDIPOP, Les editions populaires.

Mühlhäusler, P. 1979. *Growth and structure of the lexicon of New Guinea Pidgin*. Canberra: Pacific Linguistics.

Murdock, G.P. 1949. *Social structure*. New York: Macmillan.

_____ 1959. *Africa: its peoples and their culture history*. New York: McGraw-Hill.

_____ 1964. Genetic classification of the Austronesian languages: a key to Oceanic culture history. *Ethnology* 3: 117-126.

_____ 1968. Patterns of sibling terminology. *Ethnology* 7: 1-24.

Nababan, P.W.J. 1981. *A grammar of Toba Batak*. Canberra: Pacific Linguistics.

Nathan, G.S. 1973. Nauruan in the Austronesian language family. *OL* 12: 479-501.

Naylor, P.B. 1986. On the pragmatics of focus. In Geraghty, Carrington, and Wurm, 1: 43-57.

_____ ed. 1980. *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*. Michigan Papers on South and Southeast Asia, No. 15. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.

Næss, Åshild. 2006. Bound nominal elements in Äiwoo (Reefs): a reappraisal of the "Multiple noun class systems". *OL* 45: 269-296.

Needham, R., ed. 1973. *Right and left: essays on dual symbolic classification*. Chicago and London: The University of Chicago Press.

Newell, L.E. 1993. *Batad Ifugao dictionary, with ethnographic notes*. Manila: LSP publication 33.

_____ 2006. *Romblomanon dictionary*. Manila: LSP publication 52.

Niemann, G.K. 1891. Bijdrage tot de kennis der verhouding van het Tjam tot de talen van Indonesië. *BKI* 40: 27-44.

Nivens, R. 1993. Reduplication in four dialects of West Tarangan. *OL* 32: 353-388.

_____ 1998. Borrowing vs. code-switching: Malay insertions in the conversations of West Tarangan speakers of the Aru islands of Maluku, eastern Indonesia. Unpublished Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

Noble, G.K. 1965. *Proto-Arawakan and its descendants*. Indiana University Research Center in Anthropology, Folklore, and Linguistics, and Part II of the International Journal of American Linguistics, vol, 31, No. 3. The Hague: Mouton.

Noorduyn, J. 1991. *A critical survey of studies on the languages of Sulawesi. KITLV* Bibliographical Series 1. Leiden: Koninklijik Instituut voor Taal-, Land- en Volkenkunde.

Nothofer, B. 1975. *The reconstruction of Proto-Malayo-Javanic. VKI* 73. The Hague: Nijhoff.

_____ 1980. *Dialektgeographische Untersuchungen in West-Java und im westlichen Zentral-Java*. Wiesbaden: Harrassowitz,

_____ 1981. *Dialektatlas von Zentral-Java*. Wiesbaden: Harrassowitz.

_____ 1984. Further evidence for the reconstruction of *-əy and *-əw. *BKI* 140: 451- 458.

_____ 1986. The Barrier island languages in the Austronesian language family. In Geraghty, Carrington, and Wurm 2: 87-109.

_____ 1990. Review of Blust 1988a. *OL* 22: 132-152.

_____ 2000. A preliminary analysis of the history of Sasak language levels. In Austin: 57-84.

_____ ed. 1996. *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*. Abera Network, Asia-Pacific: 3. Hamburg: Abera Verlag Meyer & Co.

O'Connor, S, M. Spriggs, and P. Veth, eds. 2007. *The archaeology of the Aru islands, eastern Indonesia*. Terra Australis 22. Canberra: The Australian National University Press.

Odé, C. 1997. On the perception of prominence in Indonesian: an experiment. In Odé and Stokhof: 151-166.

Odé, C. and W. Stokhof, eds. 1997. *Proceedings of the Seventh International Conference on Austronesian Linguistics*. Amsterdam and Atlanta: Rodopi.

Ogawa, N. 2003. *English-Favorlang vocabulary*. Edited and with an introduction by Paul Li. Asian and African Lexicon Series No. 43. Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa.

Ogawa, N. and E. Asai. 1935. *The myths and traditions of the Formosan native tribes*. Taipei: Taihoku Imperial University [in Japanese].

O'Grady, G.N. 1998. Toward a Proto-Pama-Nyungan stem list, Part I: Sets J1-J25. *OL* 37: 209-233.

Onvlee, L. 1984. *Kamberaas (Oost-Soembaas)-Nederlands woordenboek. KITLV*. Dordrecht, Holland: Foris Publications.

Ostapirat, Weera. 2000. *Proto-Kra*. Linguistics of the Tibeto-Burman Area 23.1.

_____ 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In Sagart, Blench, and Sanchez-Mazas: 107-131.

Osumi, M. 1995. *Tinrin grammar. OLSP* 25. Honolulu: University of Hawaii Press.

Otsuka, Yuko. 2006. Niuean and Eastern Polynesian: a view from syntax. OL 45: 429-456.

_____ 2011. Neither accusative nor ergative: an alternative analysis of case in Eastern Polynesian. In Claire Moyse-Faurie and Joachim Sabel (eds.), *Topics in Oceanic morphosyntax*: 289- 318. Berlin: Mouton de Gruyter.

Ouyang, J., and Y. Zheng. 1983. The Huihui speech (Tsat) of the Hui nationality in Yaxian, Hainan. *Minzu Yuwen*: 30-40 (in Chinese).

Ozanne-Rivierre, F. 1975. Phonologie du nemi (Nouvelle-Calédonie) et notes sur les consonnes postnasalisées. *BSLP* 70: 345-356.

_____ 1984. *Dictionnaire iaai-français (Ouvéa, Nouvelle-Calédonie)*. Langues et Cultures du Pacifique 6. Paris: SELAF.

_____ 1986. Redoublement expressif et dédoublement des séries consonantiques dan les langues de Îles Loyauté (Nouvelle-Calédonie). *Te Reo* 29: 25-53.

_____ 1992. The Proto Oceanic consonantal system and the languages of New Caledonia. *OL* 31: 191-207.

_____ 1995. Structural changes in the languages of northern New Caledonia. *OL* 34: 45-72.

_____ 1998. *Le Nyelâyu de Balade* (Nouvelle-Calédonie). Langues et cultures du Pacifique 12, SELAF 367. Paris: Peeters.

Ozanne-Rivierre, F. and J.-C. Rivierre. 1989. Nasalisation/oralisation: nasal vowel development and consonant shifts in New Caledonian languages. In Harlow and Hooper 2: 413-432.

Pakir, A. 1986. A linguistic investigation of Baba Malay. Unpublished Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

Pallesen, A.K. 1985. *Culture contact and language convergence*. Manila: LSP publication 24.

Palmer, B. 2009. *Kokota grammar. OLSP* 35.

Palmer, B., and D. Brown. 2007. Heads in Oceanic indirect possession. *OL* 46:1 199-209.

Pampus, K.H. 1999. *Koda Kiwã: Dreisprachiges Wörterbuch des Lamaholot (Dialekt von Lewolema)*. Abhandlungen für die Kunde des Morgenlandes, vol. LII,4. Stuttgart: Franz Steiner.

Panganiban, J.V. 1966. *Talahuluganang Pilipino-Ingles*. Manila.

Parker, G.J. 1970. *Southeast Ambrym dictionary*. Canberra: Pacific Linguistics.

Paton, W.F. 1971. *Ambrym (Lonwolwol) grammar*. Canberra: Pacific Linguistics.

_____ 1973. *Ambrym (Lonwolwol) dictionary*. Canberra: Pacific Linguistics.

Pawley, A.K. 1966. Polynesian languages: a subgrouping based upon shared innovations in morphology. *JPS* 75: 39-64.

———— 1967. The relationships of Polynesian Outlier languages. *JPS* 76: 259-296.

———— 1972. On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds., *Studies in Oceanic culture history*, vol. 3: 1-142. Pacific Anthropological Records, No. 13. Honolulu: Department of Anthropology, Bernice Pauahi Bishop Museum.

———— 1973. Some problems in Proto Oceanic grammar. *OL* 12: 103-188.

———— 1975. The relationships of the Austronesian languages of Central Papua: a preliminary study. In T.E. Dutton, ed., *Studies in languages of Central and South-East Papua*: 3-105. Canberra: Pacific Linguistics.

———— 1981. Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In Hollyman and Pawley: 269-309.

———— 1982. Rubbish-man commoner, big man chief? Linguistic evidence for hereditary chieftainship in Proto Oceanic society. In J. Siikala, ed., *Oceanic Studies: essays in honour of Aarne A. Koskinen*: 33-52. Helsinki: The Finnish Anthropological Society.

———— 1985. Proto-Oceanic terms for 'person': a problem in semantic reconstruction. In V. Acson and R.L. Leed, eds., *For Gordon H. Fairbanks*: 92-104. Oceanic Linguistics Special Publication No. 20. Honolulu: University of Hawaii Press.

———— 1996a. On the position of Rotuman. In Nothofer: 85-119.

———— 1996b. On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In Davidson, Irwin, Leach, Pawley, and Brown: 387-410.

———— 2005. The chequered career of the Trans New Guinea hypothesis: recent research and its implicaions. In Pawley, Attenborough, Golson and Hide: 67-107.

———— 2006. Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *JPS* 115: 215-257.

———— 2009a. Greenberg's Indo-Pacific hypothesis: an assessment. In Evans: 153-180.

———— 2009b. The role of the Solomon Islands in the first settlement of Remote Oceania: bringing linguistic evidence to an archaeological debate. In Adelaar and Pawley: 515-540.

———— 2011. Stability and change in Oceanic fish names. In Ross, Pawley and Osmond: 137-160.

———— ed. 1991. *Man and a half: essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*. Auckland: The Polynesian Society.

Pawley, A.K. and K. Green. 1971. Lexical evidence for the Proto-Polynesian homeland. *Te Reo* 14: 1-36.

Pawley, A.K. and R.C. Green. 1973. Dating the dispersal of the Oceanic languages. *OL* 12: 1-67.

———— 1984. The Proto Oceanic language community. *Journal of Pacific History* 19: 123-146.

Pawley, A.K., and M. Pawley. 1994. Early Austronesian terms for canoe parts and seafaring. In Pawley and Ross: 329-361.

Pawley, A.K. and M.D. Ross. 1993. Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22: 425-459.

Pawley, A.K. and T. Sayaba. 1971. Fijian dialect divisions: Eastern and Western. *JPS* 80: 405-436.

———— 2003. *Words of Waya: a dictionary of the Wayan dialect of the Western Fijian language*. 2 vols. Privately circulated.

Pawley, A.K., R. Attenborough, J. Golson and R. Hide, eds. 2005. *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.

Pawley, A.K. and L. Carrington, eds. 1985. *Austronesian linguistics at the 15th Pacific Science Congress*. Canberra: Pacific Linguistics.

Pawley, A.K. and M.D. Ross, eds. 1994. *Austronesian terminologies: continuity and change*. Canberra: Pacific Linguistics.

Payne, D.L. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In D.C. Derbyshire and G.K. Pullum, eds., *Handbook of Amazonian languages*, vol. 3: 355-499. Berlin: Mouton de Gruyter.

Pejros, I. 1994. Some problems of Austronesian accent and *t ~ *C (Notes of an outsider). *OL* 33: 105-127.

Percival, W.K. 1981. *A grammar of the urbanized Toba-Batak of Medan*. Canberra: Pacific Linguistics.

Peyros, I.I. and S.A. Starostin. 1984. Sino-Tibetan and Austro-Tai. *CAAAL* 22: 123-127.

Philips, S. 1991. Tongan speech levels: practice and talk about practice in the cultural construction of social hierarchy. In Blust: 369-382.

Pinnow, H.J. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Harrassowitz.

Pittman, R. 1959. Jarai as a member of the Malayo-Polynesian family of languages. *Asian Culture* 1.4: 59-67.

Poedjosoedarmo, G.R. 2002. Changes in word order and noun phrase marking from old to modern Javanese: implications for understanding developments in western Austronesian 'focus' systems. In Wouk and Ross: 311-330.

Pokorny, J. 1959. *Indogermanisches Etymologisches Wörterbuch*. 2 vols. Bern and Munich: Francke.

Porter, D. 1977. *A Tboli grammar*. Philippine Journal of Linguistics Special Monograph No. 7. Manila: LSP.

Post, Ursula. 1966. The phonology of Tiruray. *The Philippine Journal of Science* 95: 563-575.

Potet, J-P G. 1995. Tagalog monosyllabic roots. *OL* 34: 345-374.

Prathama, R. and H. Chambert-Loir. 1990. *Kamus Bahasa Prokem* (2nd edition). Jakarta.

Prentice, D.J. 1971. *The Murut languages of Sabah*. Canberra: Pacific Linguistics.

———— 1974. Yet another PAN phoneme? Papers of the First International Conference on Comparative Austronesian Linguistics, 1974 – Proto-Austronesian and Western Austronesian. *OL* 13: 33-75.

Pukui, M.K. and S.H. Elbert. 1971. *Hawaiian dictionary*. Honolulu: University of Hawaii Press.

Pusat. 1976. *Bibliografi perkamusan Indonesia* (Bibliography of Indonesian dictionary writing). Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Quick, P.A. 2007. *A grammar of the Pendau language of central Sulawesi, Indonesia*. Canberra: Pacific Linguistics.

———— n.d. Interpretation of nasal-stop clusters in Pendau. Ms., 11pp.

Ramos, T.V. 1971. *Tagalog structures*. PALI Language Texts: Philippines. Honolulu: The University Press of Hawaii.

Räsänen, M. 1949. *Materialien zur Lautgeschichte der türkischen Sprachen*. Studia Orientalia 15. Helsinki: Finnish Oriental Society.

Rau, D.H.V. 1992. *A grammar of Atayal*. Taipei: Crane.

Rau, D. Victoria and Maa-Neu Dong. 2006. *Yami texts with reference grammar and dictionary*. LLMS A-10. Taipei: Institute of Linguistics, Academia Sinica.

Ray, S.H. 1911. Comparative notes on Maisin and other languages of eastern Papua. *JRAI* 41: 397-405.

_____ 1913. The languages of Borneo. *SMJ* 1.4: 1-196.

_____ 1926. *A comparative study of the Melanesian island languages*. Cambridge: Cambridge University Press in association with Melbourne University Press.

_____ 1938. The languages of the Eastern and South-Eastern divisions of Papua. *JRAI* 68: 153-208.

Rehg, K.L. 1981. *Ponapean reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

_____ 1991. Final vowel lenition in Micronesian languages: an exploration of the dynamics of drift. In Blust: 383-401.

_____ 1993. Proto-Micronesian prosody. In Edmondson and Gregerson: 25-46.

_____ 2004. Linguists, literacy, and the law of unintended consequences. *OL* 43: 498- 518.

Rehg, K.L. and D.G. Sohl. 1979. *Ponapean-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Reid, L.A. 1973a. Diachronic typology of Philippine vowel systems. In T.A. Sebeok, ed., *Current trends in linguistics*, vol. 11: *Diachronic, areal, and typological linguistics:* 485-505. The Hague: Mouton.

_____ 1973b. Kankanay and the problem *R and *l reflexes. In Gonzalez: 51-63.

_____ 1976. *Bontok–English dictionary*. Canberra: Pacific Linguistics.

_____ 1978. Problems in the reconstruction of Proto-Philippine construction markers. In Wurm and Carrington: 33-66.

_____ 1982. The demise of Proto-Philippines. In Halim, Carrington, and Wurm 2: 201-216. Canberra: Pacific Linguistics.

_____ 1987. The early switch hypothesis: linguistic evidence for contact between Negritos and Austronesians. *Man and Culture in Oceania* 3: 41-59.

_____ 1992. On the development of the aspect system in some Philippine languages. *OL* 31: 65-91.

_____ 1994a. Possible non-Austronesian lexical elements in Philippine Negrito languages. *OL* 33: 37-72.

_____ 1994b. Morphological evidence for Austric. *OL* 33: 323-344.

_____ 2002. Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *OL* 41: 295-309.

_____ 2005. The current status of Austric. In Sagart, Blench, and Sanchez-Mazas: 132-160.

_____ 2006. On the origin of the Philippine vowel grades. *OL* 45: 457-73.

_____ 2009. The reconstruction of a dual pronoun to Proto Malayo Polynesian. In Evans: 461-477.

_____ 2010. Palauan velar nasals and the diachronic development of PMP noun phrases: a response to Blust. *OL* 49: 436-477.

_____ ed. 1971. *Philippine minor languages: word lists and phonologies. OLSP* 8.

Reland, H. 1708. *Dissertationum miscellanearum partes tres*. 3 vols. Trajecti ad Rhenum.

Remijsen, B. 2001. *Word-prosodic systems of Raja Ampat languages*. Ph.D. dissertation. University of Leiden.

Renfrew, C. 1987. *Archaeology and language: the puzzle of Indo-European origins*. London: Penguin.

Renfrew, C., A. McMahon and L. Trask, eds. 2000. *Time depth in historical linguistics*. 2 vols. Cambridge: The McDonald Institute for Archaeological Research.

Rensch, C.R., C.M. Rensch, J. Noeb and R.S. Ridu. 2006. *The Bidayuh language: yesterday, today, and tomorrow*. Kuching, Sarawak: Dayak Bidayuh National Association.

Rensch, K.H. 1993. Father Alois Kayser and the recent history of the Nauruan language. In A. Kayser, *Nauru grammar*, ed. by K.H. Rensch: 1-13. Yarralumla, Australia: Embassy of the Federal Republic of Germany.

Revel-Macdonald, N. 1979. *Le Palawan (Philippines): phonologie, catégories, morphologie.* Langues et civilisations de l'Asie du sud-est et du monde insulindien, No. 4. Paris: SELAF.

       1982. Synchronical description at the phonetic and syllabic level of Modang (Kalimantan Timur) in contrast to Kenyah, Kayan, and Palawan (Philippines). In Halim, Carrington and Wurm 2: 321-331. Canberra: Pacific Linguistics.

Richards, A. 1981. *An Iban-English dictionary*. Oxford: Clarendon.

Richardson, J. 1885. *A new Malagasy-English dictionary*. Antananarivo: The London Missionary Society.

Riley, C.L., J.C. Kelly, C.W. Pennington, and R.L. Rands, eds. 1971. *Man across the sea: problems of pre-Columbian contacts*. Austin: University of Texas Press.

Rivers, W.H.R. 1968 [1914]. *The history of Melanesian society*. 2 vols. Oosterhout, Netherlands: Anthropological Publications.

Rivet, P. 1925. Les mélano-polynésiens et les australiens en Amérique. *Anthropos* 20: 51-54.

       1926. Les malayo-Polynésiens en Amérique. *Journal de la Société des Américanistes de Paris* 18: 141-278.

Rivierre, J.-C. 1973. Phonologie comparée des dialects de l'extrême-sud de la Nouvelle Calédonie. LACITO, No. 5. Paris: Centre National de la Recherche Scientifique.

       1983. *Dictionnaire Paicî-Francais (Nouvelle- Calédonie)*. Langues et cultures du Pacifique 4. Paris: SELAF.

       1993. Tonogenesis in New Caledonia. In Edmondson and Gregerson: 155-173.

       1994. *Dictionnaire Cèmuhî-Français*. Langues et cultures du Pacifique 9. Paris: Peeters.

Robins, R.H. 1957. Vowel nasality in Sundanese: a phonological and grammatical study. *Studies in Linguistics* (Special volume of the Philological Society): 87-103. Oxford: Blackwell.

       1959. Nominal and verbal derivation in Sundanese. *Lingua* 8: 337-369.

Robinson, L.C. 2011. *Dupaningan Agta: grammar, vocabulary and texts*. Canberra: Pacific Linguistics.

Robson, S. 2002. *Javanese grammar for students*. 2nd, rev. edition. Victoria, Australia: Monash University Press.

Rodrigues, A.D. 1985. Evidence for Tupi-Carib relationships. In H.E. Manelis Klein and L.R. Stark, eds., *South American Indian languages: retrospect and prospect*: 371-439. Austin: University of Texas Press.

Ross, M.D. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.

       1989. Proto-Oceanic consonant grade and Milke's *nj. In Harlow and Hooper: 433-495.

       1991. How conservative are sedentary languages? Evidence from Western Melanesia. In Blust: 433-451.

       1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *OL* 31: 23-64.

       1993. Tonogenesis in the North Huon Gulf Chain. In Edmondson and Gregerson: 133-153.

       1995a. Reconstructing Proto Austronesian verbal morphology: evidence from Taiwan. In Li et al: 727-791.

       1995b. Proto Oceanic terms for meteorological phenomena. *OL* 34: 261-304.

_____ 1995c. Some current issues in Austronesian linguistics. In Tryon: 45-120.

_____ 1996a. Squib: on the origin of the term "Malayo-Polynesian". *OL* 35: 143-145.

_____ 1996b. Contact-induced change and the comparative method: cases from Papua New Guinea. In Durie and Ross: 180-217.

_____ 1996c. Is Yapese Oceanic? In Nothofer: 121-166.

_____ 1996d. Pottery terms in Proto Oceanic. In Davidson, Irwin, Leach, Pawley, and Brown: 67-82.

_____ 1997. Social networks and kinds of speech-community event. In Blench and Spriggs: 209-261.

_____ 1998a. Proto Oceanic adjectival categories and their morphosyntax. *OL* 37: 85- 119.

_____ 1998b. Possessive-like attributive constructions in the Oceanic languages of northwest Melanesia. *OL* 37: 234-276.

_____ 2002a. The history and transitivity of western Austronesian voice and voicemarking. In Wouk and Ross: 17-62.

_____ 2002b. Jabêm. In Lynch, Ross and Crowley: 270-296.

_____ 2002c. Proto Oceanic. In Lynch, Ross and Crowley: 54-91.

_____ 2003. Talking about space: terms of location and direction. In Ross, Pawley and Osmond: 221-283.

_____ 2005a. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley, Attenborough, Golson and Hide: 15-65.

_____ 2005b. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1.2: 1-24.

_____ 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Chang, Huang and Ho: 521-563.

_____ 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *OL* 46: 456-498.

_____ 2009. Proto Austronesian verbal morphology: a reappraisal. In Adelaar and Pawley: 295-326.

_____ 2010. Lexical history in the Northwest Solomonic languages: evidence for two waves of settlement in the northwest Solomons. In Bowden, Himmelmann, and Ross: 245-270.

_____ 2011. Proto Oceanic *k^w. OL *50*: 25-50.

_____ 2012. In defense of Nuclear Austronesian (and against Tsouic). *LL* 13: 1253-1330.

_____ ed. 1996d. *Studies in languages of New Britain and New Ireland, 1: Austronesian languages of the North New Guinea Cluster in northwestern New Britain.* Canberra: Pacific Linguistics.

Ross, M.D. and Å. Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *OL* 46: 456-498.

Ross, M.D., A.K. Pawley and M. Osmond. 1998. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, Vol. 1: Material culture.* Canberra: Pacific Linguistics.

Ross, M.D., A.K. Pawley and M. Osmond. 2003. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, Vol. 2: The physical environment.* Canberra: Pacific Linguistics.

Ross, M.D., A.K. Pawley and M. Osmond. 2008. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, Vol. 3: Plants* Canberra: Pacific Linguistics.

Ross, M.D., A.K. Pawley and M. Osmond. 2011. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, Vol. 4: Animals* Canberra: Pacific Linguistics.

Ross, M.D., A.K. Pawley and M. Osmond. 2013. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society, Vol. 5: Body and mind*. Berlin: Mouton de Gruyter.

Ross, M.D. and S. Teng. 2005. Formosan languages and linguistic typology. *LL* 6: 739- 781.

Rubino, C.R.G. 2000. *Ilocano dictionary and grammar*. PALI Language Texts. Honolulu: University of Hawaii Press.

Ruhlen, M. 1987. *A guide to the world's languages, vol. 1: Classification*. Stanford, California: Stanford University Press.

_____ 1994. *The origin of language. Tracing the evolution of the Mother Tongue*. New York: John Wiley.

Sabatier, E. 1971. *Gilbertese-English dictionary*. Tarawa, Kiribati: The Catholic Mission.

Safford, W.E. 1909. *The Chamorro language of Guam*. Washington, D.C.: W.H. Lowdermilk and Co.

Sagart, L. 1990. Chinese and Austronesian are genetically related. Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, Oct. 3-7, 1990.

_____ 1993. Chinese and Austronesian: evidence for a genetic relationship. *Journal of Chinese Linguistics* 21: 1-62.

_____ 1994. Old Chinese and Proto-Austronesian. *Oceanic Linguistics* 33: 271-308.

_____ 1995. Some remarks on the ancestry of Chinese. In Wang: 195-223.

_____ 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *OL* 43: 411-444.

_____ 2005. Sino-Tibetan-Austronesian: an updated and improved argument. In Sagart, Blench, and Sanchez-Mazas: 161-176.

_____ 2010. Is Puyuma a primary branch of Austronesian? *OL* 49: 194-204.

Sagart, L., R. Blench and A. Sanchez-Mazas, eds. 2005. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London and New York: RoutledgeCurzon.

Salas Reyes, V, N.L. Prado and R.D.P. Zorc. 1969. *A study of the Aklanon dialect, volume two: dictionary*. Kalibo, Aklan (Philippines).

Salmond, A. 1975. A Luangiua (Ontong Java) word list. *Working Papers in Anthropology, Archaeology, Linguistics, Maori Studies,* No. 41. Auckland: Department of Anthropology, University of Auckland.

Sandin, B. 1967. *The Sea Dayaks of Borneo before White Rajah rule*. East Lansing: Michigan State University Press.

Sanvitores, D.L. de. 1954 [1668]. *Lingua Mariana*. Micro-Bibliotheca Anthropos, vol. 14. Freiburg, Switzerland: Anthropos-Institut.

Sapir, E. 1913-1914. Southern Paiute and Nahuatl, a study in Uto-Aztecan. *Journal de la société des Américanistes de Paris* (n.s.) 10: 379-425 (I), 11: 443-488 (II).

_____ 1915. The Na-Dene languages: a preliminary report. *American Anthropologist*, n.s. 17.4: 534-558.

_____ 1921. *Language: an introduction to the study of speech*. New York: Harcourt, Brace and Company.

_____ 1931. The concept of phonetic law as tested in primitive languages by Leonard Bloomfield. In S.A. Rice, ed., *Methods in social science: a case book*: 297-306. Chicago: University of Chicago Press.

Sato, Hiroko. 2009. Possessive nominalisation in Kove. *OL* 48: 346-363.

Saussure, F. de. 1959 [1915]. *Course in general linguistics*, edited by C. Bally and A. Sechehaye in collaboration with A. Riedlinger. Translated from the French by W. Baskin. New York: McGraw-Hill.

Savage, S. 1980 [1962]. *A dictionary of the Maori language of Rarotonga*. Wellington, New Zealand: Department of Island Territories.

Scaglion, Richard. 2005. *Kumara* in the Ecuadorian Gulf of Guayaquil? In C. Ballard, P. Brown, R.M. Bourke and T. Harwood, eds., *The sweet potato in Oceania, a reappraisal*: 35-41. Pittsburgh and Sydney: Ethnology Monographs 19, and Oceania Monograph 56.

Scebold, R.A. 2003. *Central Tagbanwa: a Philippine language on the brink of extinction*. Manila: LSP publication 48.

Schachter, P. 1976. The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In C.N. Li: 491-518.

Schachter, P. and F.T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Schadeberg, T.C. 2002. Progress in Bantu lexical reconstruction. *Journal of African Languages and Linguistics* 23: 183-195.

Schapper, Antoinette. 2011. Phalanger facts: notes on Blust's marsupial reconstructions. *OL* 50: 258-272.

Schärer, H. 1963. *Ngaju religion: the conception of God among a South Borneo people*. Trans. by R. Needham. The Hague: Nijhoff.

Schefold, R. 1979-1980. *Speelgoed voor de zielen: kunst en cultuur van de Mentawaieilanden*. Delft: Volkenkundig Museum Nusantara, Zürich: Museum Rietberg.

Schleicher, A. 1861-1862. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen: kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache*. Weimar: Böhlau.

Schmidt, H. 2003. Temathesis in Rotuman. In Lynch, ed: 175-207.

Schmidt, J. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.

Schmidt, W. 1899. *Über das Verhälltniss der melanesischen Sprachen zu den polynesischen und untereinander*. Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, philosophish-historisch Classe, vol. CXL. Vienna.

_____ 1900-1901. Die sprachlichen Verhältnisse von Deutsch Neuguinea. *Zeitschrift fur Afrikanische, Ozeanische und Ostasiatische Sprachen* 5: 354-384, 6: 1-97.

_____ 1906. *Die Mon-Khmer Völker: ein bindeglied zwischen Völkern Zentralasiens und Austronesiens*. Braunschweig: Friedrich Vieweg und Sohn.

Schurz, W.L. 1959. *The Manila galleon*. New York: E.P. Dutton & Co.

Schütz, A.J. 1968. A pattern of morphophonemic alternation in Nguna, New Hebrides. In A. Capell et al, eds., *Papers in the linguistics of Melanesia* 1: 41-52. Canberra: Pacific Linguistics.

_____ 1969a. *Nguna texts. OLSP* 4.

_____ 1969b. *Nguna grammar. OLSP* 5.

_____ 1972. *The languages of Fiji*. Oxford: The Clarendon Press.

_____ 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

_____ 1994. *The voices of Eden: a history of Hawaiian language studies*. Honolulu: University of Hawaii Press.

Scott, N.C. 1956. *A dictionary of Sea Dayak*. London: School of Oriental and African Studies, University of London.

Sebeok, T.A., ed. 1971. *Current trends in linguistics*: vol. 8: *Linguistics in Oceania*. 2 parts. The Hague: Mouton.

Sellato, B.J.L. 1981. Three-gender personal pronouns in some languages of central Borneo. *BRB* 13: 48-49.

_____ 1988. The nomads of Borneo: Hoffman and "devolution." *BRB* 20: 106-120.

Senft, G. 1986. *Kilivila: the language of the Trobriand islanders*. Mouton Grammar Library 3. Berlin: Mouton de Gruyter.

_____ ed. 1997. *Referring to space: studies in Austronesian and Papuan languages*. Oxford Studies in Anthropological Linguistics. Oxford: Clarendon Press.

Shen, Y., and D. Gil. 2007. Sweet fragrances from Indonesia: a universal principle governing directionality in synaesthetic metaphors. In W. van Peer, and J. Auracher, eds., *New beginning for the study of literature*: 1-17. Singapore: Cambridge Scholars Press.

Shibatani, M. 1988. Voice in Philippine languages. In M. Shibatani, ed., *Passive and voice*: 85-142. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Shinoda, E.B. 1990. *Annotated chronological bibliography (ACB) of publications and manuscripts in Philippine languages made by Japanese scholars (1902-89)*, ed. by E. Constantino. Quezon City: Cecilio Lopez Archives of Philippine Languages, and the Philippine Linguistic Circle.

Shorto, H. 1976. In defense of Austric. *CAAAL* 6: 95-104.

Sidwell, P. 2005. Acehnese and the Aceh-Chamic language family. In Grant and Sidwell: 211-246.

Simons, G.F. 1982. Word taboo and comparative Austronesian linguistics. In Halim, Carrington, and Wurm 3: 157-226. Canberra: Pacific Linguistics.

Sirk, Ü. 1983. *The Buginese language*. Moscow: Nauka.

_____ 1988. Towards the historical grammar of the South Sulawesi languages: possessive enclitics in the postvocalic position. In Steinhauer 1988b:283-302.

Skeat, W.W. and C.O. Blagden. 1906. *Pagan races of the Malay peninsula*. London: Macmillan.

Skinner, H.D. 1923. *The Morioris of Chatham islands*. *Memoirs of the Bernice P. Bishop Museum*, vol. IX, No. 1. Bernard Dominick Expedition, Publication Number 4. Honolulu: Bernice P. Bishop Museum.

Smythe, W.E. n.d. *Seimat grammar and vocabulary*. Unpublished ms.

Sneddon, J.N. 1970. The languages of Minahasa, north Celebes. *OL* 9: 11-36.

_____ 1975. *Tondano phonology and grammar*. Canberra: Pacific Linguistics.

_____ 1978. *Proto-Minahasan: phonology, morphology and wordlist*. Canberra: Pacific Linguistics.

_____ 1984. *Proto-Sangiric and the Sangiric languages*. Canberra: Pacific Linguistics.

_____ 1993. The drift towards open final syllables in Sulawesi languages. *OL* 32: 1-44.

Sneddon, J.N., K.A. Adelaar, D.N. Djenar, and M. Ewing. 2010. *Indonesian: a comprehensive grammar*. London and New York: Routledge.

Sneddon, J.N. and H.T. Usup. 1986. Shared sound changes in the Gorontalic language group: implications for subgrouping. *BKI* 142: 407-426.

Snouck Hurgronje, C. 1900. Atjèhsche taalstudiën. *Tijdschrift van het Bataviaasch Genootschap* 42: 144-262.

Sohn, H.M. 1975. *Woleaian reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Sohn, H.M. and B.W. Bender. 1973. *A Ulithian grammar*. Canberra: Pacific Linguistics.

Sohn, H.M. and A.F. Tawerilmang. 1976. *Woleaian-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Soriente, Antonia, ed. 2006. *Mencalèny & Usung Marang: a collection of Kenyah stories in the Òma Lóngh and Lebu' Kulit languages*. Jakarta: Atma Jaya University Press.

Southwell, C.H. 1980. *Kayan-English dictionary, with appendices*. Marudi, Baram, Sarawak: Privately printed.

Spaelti, P. 1997. *Dimensions of variation in multi-pattern reduplication*. Ph.D. dissertation. Department of Linguistics, University of California at Santa Cruz.

Sperlich, W., ed. 1997. *Tohi vagahau Niue = Niue language dictionary*. Honolulu: Government of Niue in association with the Department of Linguistics, University of Hawaii.

Spriggs, M.T. 1993. Island Melanesia: the last 10,000 years. In M.A. Smith, M. Spriggs and B. Frankhauser, eds., *Sahul in review: Pleistocene archaeology in Australia, New Guinea and Island Melanesia*: 187-205. Occasional Papers in Prehistory, No. 24. Canberra: Department of Prehistory, Australian National University.

_____ 2011. Archaeology and the Austronesian expansion: where are we now? *Antiquity* 85: 210-228.

Starosta, S. 1986. Focus as recentralisation. In Geraghty, Carrington, and Wurm 1: 73-95.

_____ 1995. A grammatical subgrouping of Formosan languages. In Li, et al: 683-726.

_____ 2002. Austronesian 'focus' as derivation: evidence from nominalisation. *Language and Linguistics* 3: 427-479.

Starosta, S., A.K. Pawley and L.A. Reid. 1982. The evolution of focus in Austronesian. In Halim, Carrington, and Wurm 2: 145-170. Canberra: Pacific Linguistics.

Steinhauer, H. 1993. Notes on verbs in Dawanese (Timor). In G.P. Reesink, ed., *Semaian 11: Topics in descriptive Austronesian linguistics*: 130-158. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

_____ 1996a. Morphemic metathesis in Dawanese (Timor). In Steinhauer, ed.: 217-232.

_____ 2002. More (on) Kerinci sound-changes. In Adelaar and Blust: 149-176.

_____ ed. 1988a. *Papers in Western Austronesian Linguistics*, No. 3. Canberra: Pacific Linguistics.

_____ ed. 1988b. *Papers in Western Austronesian Linguistics*, No. 4. Canberra: Pacific Linguistics.

_____ ed. 1996b. *Papers in Austronesian Linguistics*, No 3. Canberra: Pacific Linguistics.

Sterner, J.K. 1975. Sobei phonology. *OL* 14: 146-167.

Stevens, A.M. 1968. *Madurese phonology and morphology*. American Oriental Series, vol. 52. New Haven: American Oriental Society.

_____ 1994. Truncation phenomena in contemporary Indonesian. In Odé and Stokhof: 167-181.

Stewart, J.M. 2002. The potential of Proto-Potou-Akanic-Bantu as a pilot Proto-Niger-Congo, and the reconstructions updated. *Journal of African Languages and Linguistics* 23: 197-224.

Stresemann, E. 1927. *Die Lauterscheinungen in den ambonischen Sprachen. ZfES*, Supplement 10. Berlin: Reimer.

Strong, W.M. 1911. The Maisin language. *JRAI* 41: 381-396.

Stubbs, Brian. 2008. *A Uto-Aztecan comparative vocabulary*. 4th preliminary edition. Privately circulated.

Summerhayes, G.R. 2007. Island Melanesian pasts: a view from archeology. In J.S. Friedlander, ed., *Genes, language, & culture history in the Southwest Pacific*: 10-35. Oxford University Press.

Sutlive, V.H. 1978. *The Iban of Sarawak*. Arlington Heights, Illinois: AHM Publishing Corporation.

Suzuki, Keiichiro. 1998. *A typological investigation of dissimilation*. Unpublished doctoral dissertation. Tucson: Department of Linguistics, University of Arizona.

Svelmoe, G. and T. Svelmoe. 1974. *Notes on Mansaka grammar*. Language Data, Asian-Pacific Series No. 6. Huntington Beach, California: Summer Institute of Linguistics.

_____ 1990. *Mansaka dictionary*. Language Data, Asia-Pacific Series, No. 16. Dallas: Summer Institute of Linguistics.

Syahdan. 2000. Code-switching in the speech of elite Sasaks. In Austin: 99-109.

Szakos, József. 1994. *Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan*. PhD dissertation, University of Bonn.

Taber, M. 1993. Toward a better understanding of the indigenous languages of Southwestern Maluku. *OL* 32: 389-441.

Tadmor, U. 1995. *Language contact and systemic restructuring: the Malay dialect of Nonthaburi, central Thailand*. Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

_____ 2003. Final /a/ mutation: a borrowed areal feature in western Austronesian languages. In Lynch: 15-35.

Tang, C.C. 2004. Two types of classifier languages: A typological study of classification markers in Paiwan noun phrases. *LL* 5: 377-407.

Teeuw, A. 1961. *Critical survey of studies on Malay and Bahasa Indonesia. KITLV* Bibliographical Series 5. The Hague: Nijhoff.

_____ 1965. Old Balinese and comparative Indonesian linguistics. *Lingua* 14: 271-284.

Teng, S.F.C. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*, Canberra: Pacific linguistics.

Teng, S.F. and M. Ross. 2010. Is Puyuma a primary branch of Austronesian? A reply to Sagart. *OL* 49: 543-558.

Tharp, J.A. and M.C. Natividad. 1976. *Itawis-English wordlist with English-Itawis Finderlist*. New Haven: Human Relations Area Files Press.

Thieberger, N. 2006. *A grammar of South Efate: an Oceanic language of Vanuatu. OLSP* 33.

Thomas, D.M. 1963. Proto-Malayo-Polynesian reflexes in Rade, Jarai, and Chru. *Studies in Linguistics* 17: 59-75.

Thompson, L. 1945. *The native culture of the Marianas islands*. Bulletin 185. Honolulu: Bernice P. Bishop Museum.

Thurgood, E. 1997. Bontok reduplication and prosodic templates. *OL* 36: 135-148.

_____ 1998. *A description of nineteenth century Baba Malay: a Malay variety influenced by language shift*. Unpublished Ph.D. dissertation. Honolulu: Department of Linguistics, University of Hawaii.

Thurgood, G. 1993a. Geminates: a cross-linguistic examination. In J.A. Nevis, G. McMenamin and G. Thurgood, eds., *Papers in honor of Frederick H. Brengelman on the occasion of the twenty-fifth anniversary of the Department of Linguistics, CSU, Fresno*: 129-139. Fresno, California: Department of Linguistics, California State University at Fresno.

_____ 1993b. Phan Rang Cham and Utsat: tonogenetic themes and variants. In Edmonson and Gregerson: 91-106.

_____ 1994. Tai-Kadai and Austronesian: the nature of the historical relationship. *OL* 33: 345-368.

_____ 1999. *From ancient Cham to modern dialects: two thousand years of language contact and change. OLSP* 28.

Thurston, W.R. 1987. *Processes of change in the languages of North-Western New Britain*. Canberra: Pacific Linguistics.

Ting, P.H. 1976. *A study of the Laʔalua language, Formosa – Grammar*. Ms. (in Chinese).

_____ 1978. Reconstruction of Proto-Puyuma phonology. *BIHP* 49: 321-392.

Todd, E. 1978. Roviana syntax. In Wurm and Carrington 2: 1035-1042.

Topping, D.M. 1969. *Spoken Chamorro*. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

_____ 1973. *Chamorro reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Topping, D.M., P.M. Ogo and B.C. Dungca. 1975. *Chamorro-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Tryon, D.T. 1967a. *Nengone grammar*. Linguistic Circle of Canberra Publications B-6.

_____ 1967b. *Dehu-English dictionary*. Canberra: Pacific Linguistics.

_____ 1968a. *Dehu grammar*. Canberra: Pacific Linguistics.

_____ 1968b. *Iai grammar*. Canberra: Pacific Linguistics.

_____ 1976. *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.

_____ ed. 1995. *Comparative Austronesian dictionary, An introduction to Austronesian studies*. 5 vols. Berlin: Mouton de Gruyter.

Tryon, D.T. and M.J. Dubois. 1969. 2 parts. *Nengone dictionary*. Canberra: Pacific Linguistics.

Tryon, D.T. and B.D. Hackman. 1983. *Solomon islands languages: an internal classification*. Canberra: Pacific Linguistics.

Tsang, C.H. 1992. *Archaeology of the P'eng-hu islands*. Institute of History and Philology, Academia Sinica, Special Publication 95. Taipei: Institute of History and Philology, Academia Sinica.

_____ 2005. Recent discoveries at a Tapenkeng culture site in Taiwan: implications for the problem of Austronesian origins. In Sagart, Blench, and Sanchez-Mazas: 63-73.

Tsuchida, S. 1976. *Reconstruction of Proto-Tsouic phonology*. *SLCAA* Monograph Series 5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

_____ 1982. *A comparative vocabulary of Austronesian languages of sinicized ethnic groups in Taiwan, Part I: West Taiwan*. Memoirs of the Faculty of Letters, University of Tokyo, No. 7.

Tsuchida, S. and Y. Yamada. 1991. Ogawa's Siraya/Makatao/Taivoan (comparative vocabulary). In Tsuchida, Yamada and Moriguchi: 1-194.

Tsuchida, S., Y. Yamada and T. Moriguchi, eds. 1991. *Linguistic materials of the Formosan sinicized populations I: Siraya and Basai*. Tokyo: Department of Linguistics, The University of Tokyo

Tung, T.H. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publication No. 48. Taipei.

Turner, R.L. 1966. *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.

Uhlenbeck, E.M. 1955/1956. Review of Isidore Dyen, 'The Proto-Malayo-Polynesian laryngeals'. *Lingua* 5: 308-318.

_____ 1960. *The Javanese pronominal system*. *VKI* 30 (reprinted in Uhlenbeck 1978b: 210-277).

_____ 1964. *Critical survey of studies on the languages of Java and Madura. KITLV* Bibliographical Series 7. The Hague: Nijhoff.

_____ 1971. Indonesia and Malaysia. In Sebeok 1971: 55-111.

_____ 1978a. The Krama-Ngoko opposition: Its place in the Javanese language system. In Uhlenbeck 1978b: 278-299.

_____ 1978b. *Studies in Javanese morphology. KITLV* Translation Series 19. The Hague: Nijhoff.

Urry, J. and M. Walsh. 1981. The lost 'Macassar language' of northern Australia. *Aboriginal History* 5: 91-108.

Vaihinger, H. 1911. *Die Philosophie des Als Ob. System der theoretischen, praktischen und religiösen Fiktionen der Menschheit auf Grund eines idealischen Positivismus.* Berlin.

van den Berg, R. 1989. *A grammar of the Muna language.* Ph.D. dissertation, University of Leiden.

_____ 1991a. Muna dialects and Munic languages: towards a reconstruction. In R. Harlow, ed., *VICAL 2: Western Austronesian and contact languages. Papers from the Fifth International Conference on Austronesian Linguistics*: 21-51. Auckland: Linguistic Society of New Zealand.

_____ 1991b. Muna historical phonology. In J.N. Sneddon, ed., *Studies in Sulawesi linguistics*, Part 2: 1-28. NUSA 33.

_____ 1996a. *Muna-English dictionary.* Leiden: *KITLV* Press.

_____ 1996b. The demise of focus and the spread of conjugated verbs in Sulawesi. In Steinhauer: 89-114.

_____ 2003. The place of Tukang Besi and the Muna-Buton languages. In Lynch: 87-113.

van den Bergh, J.D. 1953. *Spraakkunst van het Banggais. KITLV.* The Hague: Nijhoff.

van der Leeden, A.C. 1997. A tonal morpheme in Ma'ya. In Odé and Stokhof: 327-350.

van der Tuuk, H.N. 1865. Note on the relation of the Kawi to the Javanese. *JRAS* 1: 419-442.

_____ 1971 [1864-1867]. *A grammar of Toba Batak.* Trans. by J. Scott-Kemball. The Hague: Nijhoff.

_____ 1872. 't Lampongsch en zijne tongvallen. *Tijdschrift voor Indische Taal-, Land-, en Volkenkunde* 18: 118-156.

_____ 1897-1912. *Kawi-Balineesch-Nederlandsch woordenboek.* 4 vols. Batavia.

van Engelenhoven, A. 1996. Metathesis and the quest for definiteness in the Leti of Tutukei (East-Indonesia). In Steinhauer: 207-215.

_____ 1997. Indexing the evidence: metathesis and subordination in Leti (Eastern Indonesia). In Odé and Stokhof: 257-275.

_____ 2004. *Leti, a language of Southwest Maluku. VKI* 211. Leiden: *KITLV* Press.

Errington, J.J. 1988. *Structure and style in Javanese: a semiotic view of linguistic etiquette.* Philadelphia: University of Pennsylvania Press.

van Hinloopen Labberton, D. 1924. Preliminary results of researches into the original relationship between the Nipponese and the Malay-Polynesian languages. *JPS* 33: 244-280.

Van Klinken, C.L. 1999. *A grammar of the Fehan dialect of Tetun.* Canberra: Pacific Linguistics.

Verheijen, J.A.J. 1967-1970. *Kamus Manggarai.* 2 vols. 1. *Manggarai-Indonesia*, 2. *Indonesia-Manggarai. KITLV.* The Hague: Nijhoff.

_____ 1977. The lack of formative IN affixes in the Manggarai language. In Ignatius Soeharto, ed., *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part IV: 35-37. *NUSA* 5.

_____ 1984. *Plant names in Austronesian linguistics.* NUSA 20.

Verheijen, J.A.J. and C.E. Grimes. 1995. Manggarai. In Tryon, Part 1, Fascicle 1: 585- 592.

Vérin, P., C.P. Kottak, and P. Gorlin. 1969. The glottochronology of Malagasy speech communities. *OL* 8: 26-83.

Verner, K. 1875. Eine Ausnahme der ersten Lautvershiebung. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der indogermanischen Sprachen* 23.2: 97-130.

Viray, F.B. 1973 [1939]. The infixes *la*, *li*, *lo* and *al* in Philippine languages. In A.B. Gonzalez, T. Llamzon and F. Otanes, eds., *Readings in Philippine linguistics*. Manila: LSP. (Reprinted from *Institute of National Language, Bulletin 3*).

Voorhoeve, P. 1955. *Critical survey of studies on the languages of Sumatra. KITLV* Bibliographical Series 1. The Hague: Nijhoff.

Vovin, A. 1994. Is Japanese related to Austronesian? *OL* 33: 369-390.

Walker, A.T. 1982. *A grammar of Sawu*. NUSA 13.

Walker, A. and R.D. Zorc. 1981. Austronesian loanwords in Yolngu-Matha of northeast Arnhem Land. *Aboriginal History* 5.2: 109-134.

Walker, D.F. 1976. *A grammar of the Lampung language: the Pesisir dialect of Way Lima*. NUSA 2.

Wallace, A.R. 1962 [1869]. *The Malay Archipelago*. New York: Dover Publications.

Walsh, D.S. and B. Biggs. 1966. *Proto-Polynesian word list I*. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

Walworth, Mary. to appear. Eastern Polynesian: The linguistic evidence revisited. OL 52.

Wang, W.S.Y. 1969. Competing changes as a cause of residue. *Language* 45: 9-25.

_____ ed. 1995. *The ancestry of the Chinese language*. Journal of Chinese Linguistics, Monograph Series No. 8.

Ward, J.H. 1971. *A bibliography of Philippine linguistics and minor languages (with annotations and indices bases on works in the library of Cornell University)*. Data Paper No. 83, Southeast Asia Program. Ithaca, New York: Cornell University, Department of Asian Studies.

Warneck, J. 1977 [1906]. *Toba-Batak-Deutsches Wörterbuch. KITLV*. The Hague: Nijhoff.

Waterhouse, J.H.L. 1949. *A Roviana and English dictionary* (revised and enlarged by L.M. Jones). Sydney: Epworth Printing and Publishing House.

Waterson, R. 1990. *The living house: an anthropology of architecture in South-East Asia*. Singapore: Oxford University Press.

Whaley, Lindsay J. 1997. *Introduction to typology: the unity and diversity of language*. Thousand Oaks California, London, New Delhi: Sage Publications.

Whisler, Ronald, and Jacqui Whisler. 1995. Sawai. In Tryon, part 1, fascicle 1: 659-665.

White, W.G. 1922. *The sea gypsies of Malaya*. London: Seeley and Service.

White, G.M., F. Kukhonigita and H. Pulomana. 1988. *Cheke Holo (Maringe/Hograno) dictionary*. Canberra: Pacific Linguistics.

Wikipedia. http://en.wikipedia.org/wiki/Taglish.

Wilkinson, R.J. 1959. *A Malay-English dictionary (Romanised)*. 2 vols. London: Macmillan.

Williams, H.W. 1971 [1844]. 7th ed. *A dictionary of the Maori language*. Wellington, New Zealand: A.R. Shearer, Government Printer.

Williams-van Klinken, C.L. 2011. *Tetun-English interactive dictionary*. Dili: Dili Institute of Technology.

Wilmshurst, J.M., T.L. Hunt, C.P. Lipo, and A.J. Anderson. 2011. High-precision radiocarbon dating shows recent and rapid initial colonisation of Eastern Polynesia. *Proceedings of the National Academy of Sciences* 108.5: 1815-1820.

Wilson, W.H. 1982. *Proto-Polynesian possessive marking*. Canberra: Pacific Linguistics.

_____ 1985. Evidence for an Outlier source for the Proto Eastern Polynesian pronominal system. *OL* 24: 85-133.

_____ 2012. Whence the East Polynesians? Further linguistic evidence for a Northern Outlier source. *OL* 51: 289-359.

Wiltens, C. and S. Danckaerts. 1623. *Vocabularium ofte Woort-boeck naer ordre vanden Alphabet in 't Duytsch-Maleysch ende Maleysch-Duytsch*. The Hague.

Winstedt, R.O. 1927. *Malay grammar* (2nd edition). Oxford: Clarendon Press.

Wise, Claude M., and Wesley D. Hervey. 1952. The evolution of the Hawaiian orthography. *Quarterly Journal of Speech* 38: 311-325.

Wolfenden, E. 2001. *A Masbatenyo-English dictionary*. Manila: LSP publication 38.

Wolff, J.U. 1966. *Beginning Cebuano, Part I*. New Haven and London: Yale University Press.

_____ 1972. *A dictionary of Cebuano Visayan*. Philippine Journal of Linguistics Special Monograph Issue No. 4. Manila: LSP publication 4.

_____ 1973. Verbal inflection in Proto-Austronesian. In Gonzalez: 71-91.

_____ 1974. Proto-Austronesian *r and *d. *OL* 13: 77-121.

_____ 1976. Malay borrowings in Tagalog. In C.D. Cowan and O.W. Wolters, eds., *Southeast Asian history and historiography: essays presented to D.G.E. Hall*: 345-367. Ithaca: Cornell University Press.

_____ 1982. Proto-Austronesian *c, *z, *g, and *T. In Halim, Carrington, and Wurm 2: 1-30. Canberra: Pacific Linguistics.

_____ 1988. The PAN consonant system. In McGinn: 125-147.

_____ 1991. The Proto Austronesian phoneme *t and the grouping of the Austronesian languages. In Blust: 535-549.

_____ 1993. The PAN phonemes *ñ and *N. *OL* 32: 45-61.

_____ 1997. The Proto-Austronesian voiced apical and palatal stops *d and *j. In Odé and Stokhof: 581-602.

_____ 1999. The monosyllabic roots of Proto-Austronesian. In Zeitoun and Li: 139- 194.

_____ 2003. The sounds of Proto Austronesian. In Lynch: 1-14.

_____ 2010. *Proto-Austronesian phonology with glossary*. 2 vols. Ithaca, New York: Southeast Asia Program, Cornell University.

Woollams, G. 1996. *A grammar of Karo Batak, Sumatra*. Canberra: Pacific Linguistics.

Wouk, F. and M. Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

_____ 2010. Proto Austronesian phonology with glossary. 2 vols. Southeast Asia Program Publications. Ithaca, New York: Cornell University.

Wozna, B. and T. Wilson. 2005. *Seimat grammar essentials*. Data Papers on Papua New Guinea Languages. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Wright, R. 1999. Tsou consonant clusters and auditory cue preservation. In Zeitoun and Li: 277- 312.

Wulff, K. 1942. *Über das Verhältniss des Malay-Polynesischen zum Indochinesischen*. Det Künglige Danske Videnskabernes Selskab, Historisk-filologiske Meddelelser 27, ii.

Wurm, S.A. 1978. Reefs-Santa Cruz: Austronesian, but!. In Wurm and Carrington 2: 969-1010.

_____ ed. 1975-1976. *New Guinea area languages and language study*. 2 vols. *Vol. 1: Papuan languages and the New Guinea linguistic scene, Vol 2: Austronesian languages*. Canberra: Pacific Linguistics.

Wurm, S.A. and L. Carrington, eds. 1978. 2 Fascicles. *Second International Conference on Austronesian Linguistics: Proceedings*. Canberra: Pacific Linguistics.

Wurm, S.A. and S. Hattori, eds. 1981. *Language atlas of the Pacific area*. Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy.

Yamada, Y. 1976. *A preliminary dictionary of Itbayaten*. Typescript, 404pp.

_____ 1997. *A bibliography of the Bashiic languages and cultures*, edited by C.N. Zayas. Manila: College of Social Sciences and Philosophy, University of the Philippines Diliman.

_____ 2002. *Itbayat-English dictionary*. Endangered Languages of the Pacific Rim. Kyoto.

Yamada, Y. and S. Tsuchida. 1983. *Philippine languages*. Asian and African Grammatical Manual No. 15b. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Yang, H.F. 1976. The phonological structure of the Paran dialect of Sediq. *BIHP* 47: 611-706.

Yeh, M.M., L.M. Huang, E. Zeitoun, A.H. Chang, and J.J. Wu. 1998. A preliminary study on negative constructions in some Formosan languages. In S. Huang, ed., *Selected Papers from the Second International Symposium on Languages in Taiwan*: 79-110. Taipei: Crane.

Yen, D.E. 1971. Construction of the hypothesis for distribution of the sweet potato. In Riley, Kelly, Pennington, and Rands: 328-342.

_____ 1974. *The sweet potato and Oceania: an essay in ethnobotany*. Bernice P. Bishop Museum Bulletin 236. Honolulu: Bernice P. Bishop Museum.

Zeitoun, E. 2002. Reciprocals in the Formosan languages: a preliminary study. Paper presented at 9-ICAL, January 8-11, Canberra, Australian National University.

_____ 2005. Tsou. In Adelaar and Himmelmann: 259-290.

_____ 2007. *A grammar of Mantauran (Rukai)*. Language and Linguistics Monograph Series A4-2. Taipei: Institute of Linguistics, Academia Sinica.

_____ to appear. Zeitoun, Elizabeth, Tai-hwa Chu, and Lalo a Tahesh Kaybaybaw. *A Study of Saisiyat Morphology*. LLMS. Taipei: Academia Sinica.

_____ ed. 2002. Nominalisation in Formosan languages. *LL* 3.2 (Special Issue). Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

_____ ed. 2004. *Faits de langues: les langues austronésiennes*. Paris: OPHRYS.

Zeitoun, E., T.H. Chu, and L.T. Kaybaybaw. Forthcoming. *A Study of Saisiyat Morphology*. Language and Linguistics Monograph Series. Taipei: Academia Sinica.

Zeitoun, E. and L.M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *OL* 39: 415-427.

Zeitoun, E., L.M. Huang, M. Yeh, A. Chang, and J. Wu. 1996. The temporal, aspectual, and modal systems of some Formosan languages: a typological perspective. *OL* 35: 21-56.

Zeitoun, E., L.M. Huang, M. Yeh, A. Chang, and J. Wu. 1998. A typological overview of nominal case marking of the Formosan languages. *Selected Papers from the Second International Symposium on Languages in Taiwan*: 21-48. Taipei: Crane.

Zeitoun, E., L.M. Huang, M. Yeh, A. Chang, and J. Wu. 1999. Interrogative constructions in some Formosan languages. In Y.M. Yin, I.L. Yang, and H,C. Chan, eds., *Chinese Languages and Linguistics V: Interactions in Language*: 639-680. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 2. Taipei: Academia Sinica.

Zeitoun, E., L.M. Huang, M. Yeh, J. Wu, and A. Chang. 1999. A typological overview of pronominal systems of some Formosan languages. In S.H. Wang, F.F. Tsao, and C.F. Lien, eds., *Selected Papers from the Fifth International Conference on Chinese Linguistics*: 165-198. Taipei: Crane.

Zeitoun, E. and P.J.K. Li, eds. 1999. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 1. Taipei: Academia Sinica.

Zeitoun, E., S. F. Teng, and R. Ferrell. 2010. Reconstruction of '2' in PAN and related issues. *Language and Linguistics* 11: 853-884.

Zeitoun, E., and C.H. Wu. 2006. An overview of reduplication in Formosan languages. In Chang, Huang and Ho: 97-142. *LLMS* W-5. Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, E., M. Yeh, L.M. Huang, A. Chang, and J. Wu. 1998. A preliminary study on the negative constructions in some Formosan languages. *Selected Papers from the Second International Symposium on Languages in Taiwan*: 79-110. Taipei: Crane.

Zeitoun, E., C.H. Yu, and C.X. Weng. 2003. The Formosan Language Archive. *OL* 42: 218-232.

Zewen, F.X.N. 1977. *The Marshallese language: a study of its phonology, morphology, and syntax*. VSIS 10. Berlin: Reimer.

Zoetmulder, P.J. 1982. *Old Javanese-English dictionary*. 2 vols. *KITLV*. The Hague: Nijhoff.

Zorc, R.D. 1971. *Proto-Philippine finder list*. Ms., 122pp.

_____ 1972. Current and Proto-Tagalic stress. *PJL* 3.1: 43-57.

_____ 1977. *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

_____ 1978. Proto-Philippine word accent: innovation or Proto-Hesperonesian retention? In Wurm and Carrington: 67-119.

_____ 1979. On the development of contrastive word accent: Pangasinan, a case in point. In N.D. Liem, ed., *South-East Asian Linguistic Studies*, vol. 3: 241-258. Canberra: Pacific Linguistics.

_____ 1982. Where, o where, have the laryngeals gone? Austronesian laryngeals reexamined. In Halim, Carrington, and Wurm 2: 111-144. Canberra: Pacific Linguistics.

_____ 1983. Proto Austronesian accent revisited. *PJL* 14.1: 1-24.

_____ 1986. The genetic relationships of Philippine languages. In Geraghty, Carrington, and Wurm 2: 147-173.

_____ 1987. Austronesian apicals (*dDzZ) and the Philippine non-evidence. In D.C. Laycock and W. Winter, eds., *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday*: 751-761. Canberra: Pacific Linguistics.

_____ 1990. The Austronesian monosyllabic root, radical or phonestheme. In Baldi: 175-194.

_____ 1996. The reconstruction and status of Austronesian glottal stop – chimera or chameleon. In Nothofer: 41-72.

General Index

828

—C—

—D—

—L—

—M—

—N—

—S—

Index of Names

—A—

Adelaar, K. Alexander, xvi, xix, 32, 40, 50, 64, 69, 72, 152, 156, 186, 230, 237, 247, 252, 261, 267, 311, 515, 516, 517, 563, 568, 569, 591, 596, 606, 625, 630, 650, 655, 734, 735, 736, 744, 769, 770, 776, 785, 786, 789, 792, 799, 805, 809, 812, 813, 819

Anceaux, J.C., 23, 80, 96, 194, 297, 318, 320, 468, 731, 770

Asai, Erin, 50, 229, 513, 551, 552, 553, 559, 577, 802, 804

Austin, Peter K., 320, 771, 780, 795, 803, 808, 814

—B—

Barnes, Robert H., 282, 313, 323, 771

Bellwood, Peter, 25, 28, 724, 771

Bender, Byron W., xix, 114, 115, 141, 209, 210, 268, 308, 332, 517, 596, 599, 611, 618, 620, 666, 725, 769, 771, 776, 790, 796, 812

Benedict, Paul K., 73, 513, 514, 657, 697, 704, 705, 707, 708, 709, 710, 711, 712, 714, 763, 772

Berg, René van den, 224, 308, 384, 452

Biggs, Bruce, 46, 117, 121, 159, 160, 161, 334, 347, 514, 596, 598, 706, 724, 727, 772, 792, 817

Blagden, C.O., 22, 522, 544, 697, 772, 778, 812

Blake, Frank R., 262, 274, 406, 437, 772

Blevins, Juliette, xviii, 199, 257, 262, 272, 273, 410, 517, 620, 622, 627, 628, 634, 639, 646, 713, 714, 772, 773

Bloomfield, Leonard, 57, 332, 333, 335, 336, 543, 760, 773, 810

Bopp, Franz, 22, 512, 688, 695, 696, 714, 777

Bowden, John, x, xviii, 87, 94, 195, 196, 219, 220, 296, 318, 361, 363, 427, 770, 771, 777, 783, 790, 809

Bradshaw, Joel, 98, 261, 659, 776, 777

Brandes, J.L.A., 88, 455, 513, 522, 523, 524, 526, 527, 528, 588, 777, 794

Brandstetter, Renward, x, xiv, 23, 365, 367, 370, 513, 522, 523, 524, 525, 526, 527, 528, 529, 531, 541, 542, 543, 552, 696, 697, 761, 772, 777, 778, 785

Bril, Isabelle, 110, 778

Brunelle, Marc, 73, 188, 657, 778

—C—

Capell, Arthur, 23, 45, 96, 118, 210, 435, 470, 513, 543, 608, 659, 726, 727, 761, 779, 788, 801, 811

Chowning, Ann, 323, 779

Chrétien, C. Douglas, 213, 214, 233, 244, 261, 513, 587, 683, 779

Chung, Sandra, 114, 457, 514, 752, 780

Churchward, C. Maxwell, 118, 272, 780

Clark, Ross, 95, 98, 155, 334, 335, 388, 435, 457, 470, 514, 517, 598, 726, 780, 789, 798

Clynes, Adrian, 129, 130, 657, 780

Codrington, Robert H., 22, 102, 106, 513, 529, 780

Cohn, Abigail, xviii, 184, 192, 230, 241, 257, 652, 780

Collins, James T., xix, 57, 74, 90, 165, 166, 189, 230, 490, 515, 600, 645, 732, 733, 766, 769, 776, 780, 781, 783, 795

Conant, Carlos Everett, 522, 526, 531, 589, 781

Conklin, Harold C., 15, 143, 144, 145, 146, 147, 389, 781

Crowley, Terry, xvi, 106, 124, 164, 201, 204, 206, 211, 245, 283, 304, 414, 420, 428, 429, 432, 438, 469, 472, 479, 497, 498, 502, 509, 516, 672, 693, 694, 725, 727, 729, 730, 734, 767, 776, 781, 782, 783, 799, 801, 809

—D—

Dahl, Otto Chr., xvi, 18, 64, 68, 69, 186, 191, 225, 282, 444, 446, 513, 514, 515, 516, 529, 533, 541, 546, 553, 557, 558, 559, 563, 567, 575, 577, 588, 589, 590, 676, 721, 736, 741, 743, 744, 745, 774, 782

Dempwolff, Otto, xiv, xvi, 10, 22, 23, 31, 95, 155, 158, 191, 213, 214, 225, 233, 242, 261, 315, 436, 491, 513, 514, 522, 528, 529, 530, 531, 532, 533, 534, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 558, 563, 564, 567, 574, 575, 576, 577, 578, 579, 581, 585, 587, 589, 590, 591, 592, 593, 594, 595, 597, 598, 624, 625, 629, 634, 641, 677, 680, 681, 683, 700, 701, 707, 708, 717, 721, 730, 733, 762, 774, 777, 783, 784, 796, 798

Dixon, R.M.W., xvi, xxvii, 107, 118, 458, 461, 494, 706, 761, 783

Donohue, Mark, 3, 80, 282, 295, 316, 471, 734, 735, 777, 783

Dutton, Tom, 44, 692, 693, 694, 769, 776, 780, 784, 787, 789, 805, 811

Dyen, Isidore, 8, 23, 34, 114, 158, 159, 161, 186, 187, 225, 226, 279, 341, 351, 396, 513, 514, 522, 525, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 558, 559, 564, 577, 578, 585, 588, 589, 590, 654, 717, 718, 721, 727, 731, 749, 750, 777, 784, 803, 815

—E—

Edmondson, Jerry, 172, 558, 780, 784, 799, 807, 808

Egerod, Søren, 249, 400, 784

Elbert, Samuel H., 114, 118, 122, 153, 204, 212, 228, 231, 252, 279, 304, 359, 424, 428, 430, 433, 434, 488, 495, 496, 497, 722, 784, 785, 806

Engelenhoven, Aone van, 147, 231, 272, 816

Esser, S.J., 23, 32, 524, 526, 731, 732, 733, 785

Verheijen, J.A.J., 85, 88, 196, 225, 360, 515, 816

—W—

Wallace, Alfred Russel, 4, 6, 76, 82, 154, 515, 719, 774, 817
Wang, William S-Y., 210, 517, 680, 713, 771, 796, 797, 810, 817, 819
Wilson, William H., 488, 496, 505, 509, 660, 722, 724, 765, 789, 817, 818
Wolff, John U., 57, 151, 338, 367, 374, 385, 388, 393, 395, 396, 398, 399, 438, 439, 476, 491, 499, 514, 518, 558, 560, 561, 562, 563, 564, 565, 579, 581, 582, 583, 584, 585, 587, 591, 598, 777, 792, 799, 818
Wurm, S.A., 2, 40, 204, 694, 771, 773, 774, 780, 782, 784, 787, 790, 791, 796, 797, 798, 801, 803, 807, 808, 812, 813, 815, 818, 820

—Y—

Yamada, Yukihiro, 54, 229, 302, 420, 495, 765, 802, 815, 818, 819

—Z—

Zeitoun, Elizabeth, xix, 50, 51, 288, 379, 398, 413, 447, 476, 509, 516, 517, 698, 770, 775, 780, 792, 818, 819, 820
Zorc, R. David Paul, 57, 61, 62, 82, 153, 154, 155, 180, 367, 388, 393, 438, 514, 515, 554, 555, 556, 557, 560, 567, 568, 569, 571, 572, 575, 596, 606, 660, 740, 743, 765, 810, 817, 820