Asia-Pacific Linguistics

Open Access

College of Asia and the Pacific

The Australian National University

New Advances in Formosan Linguistics

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu (eds.)

A-PL 017 / SAL 003

Asia-Pacific Linguistics

SAL: Studies on Austronesian Languages

Published by Asia-Pacific Linguistics
College of Asia and the Pacific
The Australian National University
Canberra ACT 2601
Australia

First published: 2015

National Library of Australia Cataloguing-in-Publication entry:

Title: New Advances in Formosan Linguistics / edited by Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu.

URL: http://hdl.handle.net/1885/14354

ISBN: 978-1-922185-17-4 (ebook)

Series: Asia-Pacific Linguistics 017 / Studies on Austronesian Languages 003

Subjects: Taiwan languages.
Austronesian languages.
Taiwan languages--Dialects--Phonology.
Taiwan languages--Morphosyntax.

Other Authors/Contributors:
Elizabeth Zeitoun, editor.
Stacy Fang-Ching Teng, editor.
Joy J. Wu, editor.
Australian National University. Asia-Pacific Linguistics.

Dewey Number: 499.2

Cover illustration: Taiwan seen from the Pacific (2014), courtesy of Yi-zhen Lin
Typeset by Elizabeth Zeitoun, with the help of Kanyu Yeh and Wanda China

Contents

Foreword[*]

ELIZABETH ZEITOUN, STACY F. TENG AND JOY J. WU

1 Purpose of the present volume

The present volume is a festschrift in honour of Lillian M. Huang, who, in a very few years, became a leading figure in Formosan linguistics after she obtained her PhD degree in 1987. Over the past twenty-eight years, she has been involved in important groundwork, in both academia and indigenous language policies in Taiwan, as we will show below (sections 3 and 4). She has been engaged in the development of both through her pre-eminent role in projects relating to typological studies on Formosan languages in the early 1990s, and on language teaching materials and proficiency tests since the late 1990s and early 2000s.

Lillian may retire in a few years. Before she does, we thought it would be most appropriate to honour her by putting together papers by a number of scholars and students who have benefitted from or have been in contact with her in one way or another (e.g. through collaborative work, teaching, supervising, advising etc.). The idea of such a volume was conceived by Elizabeth Zeitoun in the autumn of 2009. Further plans were initially worked out with Stacy F. Teng, soon joined by Joy J. Wu. The three editors have been close to Lillian since the early and mid 1990s. Of the three, Zeitoun, who has been working with her on diverse projects for over twenty years, is her closest collaborator on the academic level. Both Wu and Teng were Lillian's MA supervisees. Through her fieldwork courses, she introduced Wu to Amis and Teng to Puyuma, languages on which they are still working.

The title of the present volume, *New advances in Formosan linguistics*, reflects our pursuit of publishing cutting-edge, provocative, and thoughtful papers that explore new directions and perspectives on Formosan languages and linguistics. It is worth noticing that this is the first collected volume on Formosan languages that has not issued from a workshop or a conference—the papers included in this volume are thus varied in terms of topic coverage—and the first that specifically deals with (and covers nearly all) the Formosan languages, a grouping understood in its broader context, that is, including Yami, a Batanic (Philippine) language spoken on Orchid Island under the political jurisdiction of Taiwan.

In what follows, we begin by providing a few chronological notes regarding Lillian (section 2), then outlining her work in the context of the past twenty flourishing years of research on Formosan languages (section 3). In section 4, we provide a bird's eye view of

* The first draft of this foreword was written by Elizabeth Zeitoun, and corrected with the help of Stacy F. Teng and Joy J. Wu. We are grateful to Malcolm Ross and Raleigh Ferrell who kindly polished the present foreword, and to Amy Pei-jung Lee, Hui-chuan Huang and Gujing Lin who provided helpful suggestions on it.

the past twenty years' revitalisation movement on Formosan languages in Taiwan and explain her social engagement. In section 5, we provide a brief account of the papers included in this volume. Section 6 closes this foreword with our acknowledgments. We need to make it clear here that we will only concentrate on Lillian's contribution with regard to the field of Formosan linguistics, and are thus led to exclude her contributions in other fields (most notably English teaching).

2 Chronological notes

Lillian was born 13 November 1953 at Wanhua, Taipei. She is the sixth child of a large family composed of two brothers and six sisters. She was admitted to the Taipei First Girls High School (the best girls' high school in Taiwan). Though she was talented in mathematics, she decided to major in English at National Taiwan Normal University. Upon graduation in 1976, she worked as an English teacher for one year in Taipei Municipal Dali High School and then as a teaching assistant at the Department of English, National Taiwan Normal University for another year. She headed for the United States in 1978 and obtained an MA in linguistics at the University of Florida at Gainesville. She was again a teaching assistant at NTNU from 1980 to 1982 and lecturer from 1982 to 1987, but obtained a leave of absence to pursue higher studies. In 1983, she started her PhD studies at the Department of Linguistics at Rice University, Texas under the supervision of Philip W. Davis. She graduated in May 1987 and published her PhD dissertation (Huang 1988) in Taiwan a year later. To date, her dissertation remains an important reference in Chinese linguistics. On her return to Taiwan in the summer of 1987, Lillian continued to work in the Department of English, NTNU as Associate Professor—for a few years, she was also a part-time professor at the Department of Ethnology, National Chengchi University (NCCU)—and became full professor in August 1995. She was chair of the department from August 1996 to July 1999 and the director of the Office of Academic Development (now the Office of Research and Development). She retired from NTNU in 2006 and started a new career the same year at the Department of Applied Foreign Languages, Shih Chien University. She was chair of the department from August 2006 to July 2007.

Lillian is one of the founders, and was the first secretary, of the Linguistic Society of Taiwan (10 December 1999-31 January 2002). For three years from 2010 to 2012, she was head of the linguistics section of the National Science Council (now Ministry of Science and Technology). She was also a member of the Language Development Committee, Council of Indigenous Peoples from September 2006 to August 2008, and a member of the Indigenous Peoples Commissions of Taipei City (2007–2009) and New Taipei City (2011–2013). She has been a consultant for the Indigenous Languages Research and Development Centre, Council of Indigenous Peoples since June 2014. For many years, she has also been an advisor to the Ministry of Education.

It was Philip W. Davis who advised her to investigate Formosan languages, and on return to Taiwan in 1987, she started her long–life research on Atayal. She first studied Wulai Atayal, where Pastor Jin-chuan You was her major informant. Six publications appeared between 1989 and 1994 on Wulai Atayal (Huang 1989, 1991, 1992, 1993a, 1993b, 1994). In 1988 Stanley Starosta, then professor at the University of Hawai‘i at Mānoa, spent a year as a visiting professor at National Tsing Hua University where

Lillian was able to follow his classes on lexicase. In the spring semester of 1989, her PhD supervisor, Philip W. Davis, was invited to the NTNU Department of English, and she had the opportunity to continue her study of Atayal with him. In 1992–1993, she spent ten months as a visiting scholar at the University of Hawai'i at Mānoa and published a paper in *Oceanic Linguistics* (OL) on Atayal ergativity (Huang 1994)—the first paper devoted to a Formosan language to be published in that journal since Li (1974). In the autumn of 1993, Paul Jen-kuei Li taught a fieldwork course on Mayrinax Atayal at National Tsing Hua University. A number of students (among others, Zeitoun, Saillard and H. Chang) and scholars (among them, Lillian, and Mei) attended his classes. After these classes had finished, Lillian continued to investigate Mayrinax Atayal with her major informant, Chin-fa Tang. These years of research resulted in a number of publications on Mayrinax Atayal (Huang 1995b, 1995c, 1996a, 1996b, 2000a, 2001, 2002). In the early 1990s, Lillian also worked on Amis (Huang 1995a) and Bunun (Huang 1997a). In the mid and late 1990s, she had the opportunity to work on Thao (Huang 2000b), and Puyuma (Huang 2000c). In the early 2000s, she started to work on Plngawan (C'uli') Atayal (Huang 2006), and around 2006, she switched her interest to Jianshi (Squliq) Atayal (Huang 2008, Huang & Tali' Hayung 2008, 2011). In Fall 2000, she spent about four months at University at Buffalo, the State University of New York with Joy J. Wu, who was studying in a PhD program there, and attended Van Valin's classes on Role and Reference Grammar.

3 Lillian's contributions to the field of Formosan linguistics

To retrace Lillian's academic contributions to Formosan linguistics, there follows a short overview of the development of this field.

Although some of the Formosan languages (Siraya, Favorlang) were recorded during the Dutch period in the seventeenth century, only in the late nineteenth century did Japanese scholars start to investigate the cultures and languages of the Formosan ethnic tribes more extensively. However, despite a number of compilations and research studies carried out during the Japanese era—to date Ogawa & Asai's (1935) monograph is still a crucial reference—the field did not flourish as it should have, due perhaps to a fifty-year government policy stigmatising the term "aborigine" by imposing Mandarin Chinese as the only official language.[1] Efforts were made by two renowned linguists, Paul Jen-kuei Li and Shigeru Tsuchida to present exhaustive studies on the phonology of various languages and to offer lists of vocabularies on the basis of which phonological and lexical reconstructions were built up, helping them to advance subgrouping hypotheses. It was not until the late 1980s and early 1990s, however, that the Formosan languages started to attract more linguists and students, and Lillian was instrumental in that growing interest as she offered classes on field method and supervised a number of students.[2] She was the first to train and supervise an aboriginal MA student (Yan 1992) working on his own language, Amis. In the following ten years, she (co-)supervised six MA students working

[1] The term "aborigine" was replaced by "indigenous" in the early 2000s. Thus, the Council of Aboriginal Affairs was renamed Council of Indigenous Peoples on 25 March, 2002.

[2] At about the same time, Paul Jen-kuei Li was doing the same at National Tsing Hua University.

on Formosan linguistics (Wu 1995, Tan 1997, Teng 1997a, Shih 2008, and Tali' Hayung 2008) and one PhD student (Yeh 2003).

It would not be too much of an exaggeration to say that in the late 1980s, Lillian was the first to regenerate interest in this field, and provide novel ideas on the morpho-syntax of Atayal after years of its lying fallow. While her first book on (Wulai) Atayal (Huang 1993a) had only discussed a limited number of syntactic issues, her grammar on Mayrinax Atayal (Huang 1995d) was the first truly extensive work on a Formosan language and in our view the best at the time (the first grammar had been Tung 1964 on Tsou, and the second, describing simple sentences in Tanan Rukai, was Li 1973). Two things need to be noted here. First, her formal publications, although representing only a small portion of her outstanding professional contributions, are consistently reliable in their presentation of data, and her research topics have provided a foundation for the discussion of new issues in the field. Second, most of her publications are impregnated with her profound belief that semantics plays a major role in the grammar of a language.

A turning point in her career, and in Formosan linguistics, was marked by the projects on typology. Zeitoun (then a PhD student) first approached Paul Jen-kuei Li to ask whether he would help her carry out a project on the typology of Formosan languages. While he declined, he advised her to go to Lillian, then visiting scholar at the University of Hawai'i at Mānoa. Though such a project seemed too gigantic to be viable, it actually worked out admirably, for two reasons. First, it was Lillian who gave shape and order to Zeitoun's ideas. Second, she realised that such an ambitious project could not be carried out by just two members. We asked other MA graduates (initially Marie M. Yeh, Joy J. Wu and Anna H. Chang and soon also Stacy F. Teng) to join the team. Most of these team members would have been forced to abandon Formosan linguistics if they had not been called back by their former supervisor to participate in these projects. We all collaborated on the morpho-syntactic typology of the Formosan languages for about six years, until 1999 when the members of the group started to be busy with their own careers or were engaged in higher studies.

The goals of these projects were twofold: Trying to understand variations among languages and dialects while acquiring a better understanding of a particular language—each of the project members was in charge of the investigation of one or two Formosan languages—by reaching cross-linguistic generalisations that might have otherwise been overlooked. A third and perhaps more far-reaching and challenging goal was to determine how the Formosan languages had evolved after their split. This goal was, of course, not achieved in this short time period. In these six years, three major projects were carried out. Between 1994 and 1996 we focussed on simple sentences (the title of the NSC project was "A typological study of grammatical relations in Formosan languages"). In 1996–1998 we concentrated on complex sentences (the project's title was the same as above). In 1998–1999, we dealt with verb classification. During that period, two dozen manuscripts were written, with more than half of them presented in conferences and/or published. The projects also allowed members to write two dozen articles and/or books on specific languages and to pursue PhD studies in Taiwan and abroad. Aside from a short introduction on the typology of Formosan languages (Zeitoun 2004), published papers included the study of case markers (Huang et al. 1998), pronouns (Huang et al. 1999a), negation (Yeh et al. 1998), tense/aspect and modality (Zeitoun et al.

1996, Zeitoun & Huang 1997), existential, possessive, and locative constructions (Zeitoun et al. 1999, Zeitoun 2000b), interrogative constructions (Huang et al. 1999b), verb classification (Zeitoun & Huang 2000, Zeitoun 2000a). Unpublished papers and theses include studies on reciprocals (Zeitoun 2002), serial verb constructions (Huang 1997b), pivotal constructions (Yeh 1997), cognition and uterrance clauses (A. Chang 1997), adverbial clauses (Zeitoun 1997), relative clauses (Wu 1997), and coordination (Teng 1997b).

In 1994 or 1995, the idea (again initiated by Zeitoun) grew that we needed grammatical overviews (covering the phonology, morphology and syntax) of the Formosan languages. Again, it was Lillian who formed the team and for about two years we worked on one or more languages. A series of twelve grammars on most Formosan languages (except Saaroa and Kanakanavu, for which we had no data) was published in 2000. This collaborative work was not without defects and we were all aware of them. Some of us (Lillian, H. Chang, Zeitoun) had to write more than one grammar as there were still very few students/scholars involved in the study of Formosan languages. Others (e.g. A. Chang and Wu) had previously worked only on specific topics. While A. Chang had only focused on causatives in Paiwan, Wu had dealt only with complex sentences in Amis. Last but not least, none of us had many years of research practice and the writing of these grammars, though coordinated, proved more difficult than had been expected and of varying completedness and depth (see A. Chang 2000; H.-h. Chang 2000; H. Chang 2000a, 2000b; Huang 2000a, 2000b, 2000c, Lin 2000, Wu 2000, Yeh 2000, Zeitoun 2000c, 2000d, 2000e).

In 2011, that is eleven years after the publication of this book series, Lillian decided to convene a new team to revise these earlier grammars while working on new dialects/languages. For three years, she coordinated the writing of grammars of fourteen languages (November 2011–October 2014). Three languages were left out because they were not recognised at the time of the project application: Saaroa, Kanakanavu and Pazeh. The governmental recognition of Saaroa and Kanakanavu as the fifteenth and sixteenth indigenous ethnic groups only took place in June 2014 and we did not possess much information on Pazeh—whose last speaker passed away in October 2010—or Kaxabu. The new working team included members that had worked on the first edition of the grammars (Huang, Zeitoun, H. Chang, Yeh, Wu, A. Chang), scholars that started working actively on Formosan languages in the 1990s (Li-May Sung, Victoria D. Rau), scholars who obtained their PhDs in/or after 2000 and who started working on Formosan languages in the 1990s or later (Amy P. Lee, Fu-hui Hsieh, Stacy F. Teng, Hui-chuan Huang), MA and/or PhD students (Tali' Hayung (Hsin-sheng Wu), Yu-han Chao, Wen-chi Shen, Chao-kai Shih, Shih-lang Jian, Lowking Nowbucyang (Wei-cheng Hsu), Hui-huan Ann Chang) and a number of aboriginal assistants. This project is important enough to be mentioned here because it is the first time that so many Formosanists began working together.

We have outlined Lillian's major achievements as a scholar and shown that she has always encouraged younger colleagues and students to do research on various Formosan languages. She has published co-authored papers with them to help them survive in a harried academic atmosphere, and this has proven to be a driving force in Formosan linguistics. We turn to another side of her career in the following section, her contributions to indigenous language policies.

4 Lillian's contributions to indigenous language policies

For over a hundred years, the Formosan languages have been affected by language policies imposed by different governments. Taiwan was politically dominated by Japan for 50 years (1895–1945) and for another forty years or so (1949–1987), governmental policy dictated that Mandarin was the only official language to be used on the island. This has, of course, accelerated the loss of Formosan languages among the younger generation. With the end of martial law in 1987, Taiwan indigenous ethnic groups started to realise the importance of preserving their cultures and languages. In 1990, Atayal began to be officially taught in the Wulai primary and junior high schools. This was the dialect Lillian was working on at the time, and she designed one of the textbooks (Huang 1992).

Revitalisation movements were initiated through governmental initiatives. The Ministry of Education and the Council of Indigenous Peoples have been instrumental in implementing various language policies. We cannot survey all the policies that have been implemented since the early 1990s but will touch on those that are relevant to our honoree.

In 1991 and 1992, the Ministry of Education entrusted Paul Jen-kuei Li with the design of more appropriate orthographic systems for all the Formosan languages (Li 1991). There was no attempt, however, to implement these orthographic systems in official documents (e.g. textbooks), and different orthographic systems were used for the same language or dialect, leading to violent controversies within ethnic groups. Things got worse when the first indigenous language proficiency tests was held by the Council of Indigenous Peoples in 2001. A meeting was called on 16 April 2003 to help revise and develop more unified orthographic systems and Lillian helped coordinate the discussions for all Formosan languages between native speakers and linguists. The revised writing systems for all Formosan languages were promulgated by the Ministry of Education and the Council of Indigenous Peoples on 15 December 2005. Ten years later, it is clear that an even more unified system that would fit all the Formosan languages is necessary and Lillian has set herself the 2015 goal of working out a single orthographic system that will be definitively accepted.

In 1993 and 1994, the Ministry of Education entrusted a small group of linguists (among them Lillian, see Huang 2001) and indigenous people to design the major principles regarding the compilation of textbooks and announced the inclusion of a chapter on indigenous peoples and languages in junior high school curricula. In 2002, after a significant change in policy, the Ministry of Education and the Council of Indigenous Peoples granted the Center for Aboriginal Studies, National Chengchi University (with Lim Siu-theh as its head) a new project to re-design indigenous language textbooks. Nine sets (covering the officially recognised languages and dialects) amounting to thirty-eight volumes were compiled and edited within three years between 2002 and 2005. These textbooks are available online (http://www.alcd.nccu.edu.tw/). Lillian was not involved in the textbook redesign but published her assessment of teaching materials and textbooks as Huang (1995e, 1995f, 1998, 2007b). For many years she was also contracted by the Ministry of Education to evaluate indigenous language teaching in elementary schools in Taipei City and New Taipei City (Huang 2011).

Lillian was also very often present in the training camps that were organised for indigenous language teachers. In 2001, she designed a seventy-two hour program that

included introductory courses on Taiwan indigenous cultures and languages (eighteen hours), Formosan linguistics (twenty hours) and language teaching (32 hours). She later organised more varied training camps for a larger public, following the policies implemented by the Council of Indigenous Peoples (often policies she has proposed). These training camps included courses on language teaching, dictionary compilation, cultural anthology, and computer-assisted programs.

Her major contribution—and perhaps the most challenging—was to design proficiency tests for indigenous languages. These proficiency tests are now divided into four levels and concern all the Formosan languages and dialects (Huang 2003, 2007a) and cover four language skills: listening, speaking, reading, and writing. Lillian helped compile written and audio materials for test preparation, based on the official textbooks designed by NCCU, for the test takers of the beginning and intermediate levels.

She was also instrumental in helping the Council of Indigenous Peoples to plan different long-term programs, the last two of which each span a six-year period (2008–2013, 2014–2019) and specifically concern language revitalisation. She even drafted the first program (2008–2013). One of the goals she was able to have implemented was the compilation of a dictionary in Chinese for each of the sixteen officially recognised indigenous languages. This was a long term project that spanned nine years (2005–2014) and one of the first and most impressive collaborative efforts between linguists and native speakers. Some of these dictionaries (Bunun, Kanakanavu, Saisiyat, Thao, Truku, Tsou and Yami) are available online (http://e-dictionary.apc.gov.tw). Lillian headed the team working on the Atayal dictionary for two years (2010–2012).

In view of the above, it is no overstatement to claim that Lillian has made history with regard to Formosan languages. This aspect of her work is not well known to most Formosanists, because she is very modest and has never accepted any awards for her practical contributions. As outlined above, while she took part as a scholar in projects mostly funded by the Ministry of Education in the mid 1990s, at the turn of the twenty–first century she was the only Formosanist who had enough energy, devotion, vision and charisma to collaborate with governmental agencies on applied issues, bringing her administrative, pedagogical and scholastic experience to bear on devising many policies that have changed the face of Taiwan. As in her academic life, she has been the driving force that has enabled different communities, linguists and native speakers alike, to work together despite major differences. In the process, she was led to sacrifice her academic career, something she does not seem to regret too much, while accepting all the frustrations brought by her work with governmental institutions as policies have not always been carried out as she had advised.

In the next section, we provide a short overview of the content of this volume that we dedicate to Lillian as a small token of our gratitude for her many contributions.

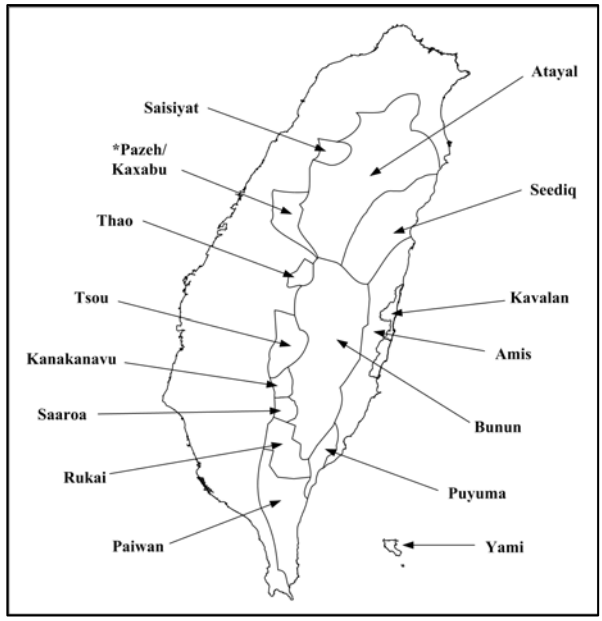
5 A short overview on the content of this volume

This volume covers data from nearly all the extant Formosan languages, Atayal, Amis (five dialects: Nataoran, Sakizaya, Kakacawan, Paheko, and Falangaw), Bunun, Kanakanavu, Kavalan, Rukai (three dialects: Budai, Tona and Mantauran), Paiwan,

Puyuma (three dialects: Nanwang, Katripul and Tamalakaw), Saaroa, Saisiyat, Seediq (three dialects: Toda, Tgdaya and Truku), Thao, Tsou, and Yami.[3]

The volume is arranged as follows. After papers on diachronic and synchronic phonology (chapters 1–4), advances in Formosan morpho-syntax (chapters 5–14) are presented. A feature of Saisiyat discourse is examined in chapter 15, followed by three cross-dialectal studies (chapters 16–18). The four final chapters focus on the lexicon, from different perspectives: semantic extension, language contact and corpus linguistics (chapters 19–22).

Malcolm Ross provides a reassessment of his (1992) PAn reconstructed phoneme paradigm by discussing certain phonemes which are still potentially controversial. Paul Jen-kuei Li's paper examines the occurrence of preglottalised stops in Thao, Bunun and Tsou from a language contact perspective. Hui-chuan Huang discusses permissible syllable types in Bunun, Saisiyat, and Atayal with respect to word-edge within Optimality Theory. Hui-shan Lin re-examines two types of reduplication in Squliq Atayal, C- and CC- reduplication within Optimality Theory and infers that the first reduplication type involves the copying of a bare consonant and the second, the copying of two syllables.



Map 1: Geographical distribution of the Formosan languages

Gujing Lin examines two types of serial verb constructions in Tsou which differ in their voice marking (identical, e.g., AV-AV or different, e.g., PV-AV), this discrepancy being accounted for in terms of tight versus loose structure. Maya Y. Yeh investigates a specific construction headed by *blaq* 'good' followed by a verb marked as UV in Squliq Atayal. Henry Y. Chang investigates nominal aspect in Tsou from a generative

[3] Officially, the Formosan languages are divided into sixteen ethnic groups, adding to those mentioned above Truku and Sakizaya. Linguistically speaking, however, this classification is not without problem since it singles out Truku and Sakizaya as distinct languages while they form dialects of larger languages. Truku is part of Seediq and Sakizaya may be a dialect of Amis (Tsuchida 1982).

perspective, with a special focus on the particle *nia*. Adlay K. Liu re-examines externally-headed and internally-headed relative clauses in Jianshi (Squliq) Atayal within the Lexical-Functional Grammar framework. Wei-tien D. Tsai proposes to account for the (quasi-)similar distribution of temporal adjuncts and subjects in Atayal, Seediq, and Tsou from the perspective of the Minimalist Program. Naomi Tsukida proposes that in Seediq, the Patient of monotransitive verbs, the Theme and the Recipient of ditransitive verbs show secundative alignment (in Malchukov et al.'s 2010 terminology). Dong-yi Lin focuses on the morphosyntactic distribution of interrogative words such as 'what', 'how', 'where', and 'how many' in Amis and Kavalan and provides tests from a generative perspective to recognise those which can be used as verbs and those which cannot. Li-May Sung examines the syntactic, semantic and pragmatic behavior of *why* exclamatives (in comparison to *why* interrogatives) in Budai Rukai. Fuhui Hsieh looks into the function of *sa* in Paiwan, which combines cosubordinate clauses. Chia-jung Pan explores the morphosyntactic distribution of reported evidentials in Saaroa, Kanakanavu and Tsou.

Marie M. Yeh examines the multiple functions of the idiomatic expression *ma' 'isa:a'* in Saisiyat in order to account the relationship that unifies these different functions from a discourse perspective. She shows that this marker operates at three different levels: intra-clausal, inter-clausal and inter-personal.

Joy J. Wu revisits Huang's (1995) claims regarding Amis personal pronouns. The first has to do with case marking and the second with the free status of the genitive pronouns. To do so, she compares the pronominal systems of five Amis sub-dialects, Nataoran, Sakizaya, Kakacawan, Paheko, and Farangaw and concludes that these dialects exhibit various degrees of case neutralisation and that the so-called "free" genitive sets should be analysed as clitics. Stacy F. Teng provides a comparative study of the the forms and functions of the personal pronouns in three Puyuma dialects, Nanwang, Katripul and Tamalakaw and accounts for their discrepancies from a diachronic perspective. Elizabeth Zeitoun discusses the syntactic changes that have occurred in Mantauran (Rukai) by examining in particular the verb glossed as 'say', providing a cross-dialectal comparison and proposing a reconstruction of the verb 'say' at the Proto–Rukai level.

Amy Pei-jung Lee investigates body part nomenclature and categorisation in the three dialects of Seediq, Toda, Tgdaya and Truku from the perspectives of structure and semantic extension. Apay A. Tang compares metaphorical and metonymic expressions of anger and happiness in English and Truku Seediq. She shows that Truku Seediq tends to use more internal and external bodily expressions than English in conceptualizing anger and happiness. Yuehchen Chien outlines the sociohistorical background of Yilan Creole that demonstrates that it is indeed a creole and then goes on to describe the lexical system in Yilan Creole in order to clarify linguistic changes that have taken place in this Atayal dialect.Victoria Rau, Yi-hsin Wu and Meng-chien Yang provide preliminary results from a corpus-based approach identifying and classifying emotion terms in Yami.

6 Acknowledgments

We are grateful to the numerous people and institutions that have, at various stages, supported the editing of this volume. This was a lengthy process. The first call for papers was sent out to a selected list of potential contributors in June 2011 and we started to

receive abstracts and papers a year later. We thank all the colleagues and students who agreed to submit a paper to this volume for their patience and cooperation during the long and at times tedious editing process. All the authors (except for three, who decided to withdraw their submission at that time) first made revisions after receiving remarks by the three editors during a preliminary screening phase. Once their papers had been reviewed, they kindly went through a revision period which was, for some, prolonged, as they had to reply to the reviewers' comments while taking into consideration the editors' remarks.

We thank all the reviewers (Alexander Adelaar, Alexandra Aikhenvald, Edith Aldridge, Wayan Arka, Loren Billings, Ellen Broselow, Lea Brown, Niclas Burenhult, Rik De Busser, Henry Y. Chang, Melody Y. Chang, N. J. Enfield, Bethwyn Evans, Raleigh Ferrell, Catherine Fortin, Lilián Guerrero, William Foley, Arthur Holmer, Feng-fan Hsieh, Fuhui Hsieh, Shu-kai Hsieh, Hui-chuan Huang, Daniel Kaufman, Zoltán Kövecses, Olivia Lam, Randy LaPolla, Paul Jen-kuei Li, Hsiu-chuan Liao, Roger W. Liao, Dorinda Liu, Scott Paauw, Eleana Paul, Matt Pearson, Paul H. Portner, Lawrence Reid, Keren Rice, Malcolm Ross, Peter Sells, Jeff Siegel, Hooi Ling Soh, Li-May Sung, Michael Tanangkingsing, Shigeru Tsuchida, Naomi Tsukida, John Wolff, Iris H. Wu, Foong-Ha Yap, Marie M. Yeh, David Zorc) for their insightful critiques that have helped to shape the papers into their published form. We would also like to thank colleagues who help read certain papers after they had been revised, in particular Dong-yi Lin, Roger W. Liao, and Hui-shan Lin. We are indebted to Paul Jen-kuei Li who went through many of the examples found in this volume providing numerous comments to the authors.

For the formatting of the papers, we are grateful to Kanyu Yeh (葉侃彧) for her initial help. We owe a debt of gratitude to Chih-hsien Lin (林志憲) for re-drawing maps and figures and for putting the final touch to the cover page and to Hsiu-lien Lin (林秀蓮) for her help with the references that were truly inconsistent when the papers were put together. We thank Lori Paximadis for her copy-editing and Wanda China for her help in the pre-final production of this volume. The drawing of the cover—that represents an eye to express the notion of far-sighted, with Taiwan seen from the Pacific as the pupil of the eye—was designed by Yi-zhen Lin (林意真), Elizabeth Zeitoun's daughter.

The manuscript was submitted to Asia-Pacific Linguistics and we are extremely grateful to Paul Sidwell for very enthusiastically accepting our proposal and for his assistance in making this publication possible.

Last but not least, we acknowledge a grant from the Chiang Ching Kuo Foundation (#SP001-D-14) during the latest phase of our book editing and the financial support of the Institute of Linguistics, Academia Sinica for assistantship.

References

Chang, Anna H.-C. 1997. Cognition and utterances constructions in Formosan languages. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Chang, Hsiu-chuan. 2000. *paiwanyu cankao yufa* [*A reference grammar of Paiwan*]. Formosan Languages Series 9. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Chang, Hsun-huei. 2000. *yameiyu cankao yufa* [*A reference grammar of Yami*]. Formosan Languages Series 13. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Chang, Henry Y. 2000a. *saidekeyu cankao yufa* [*A reference grammar of Seediq*]. Formosan Languages Series 6. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Chang, Henry Y. 2000b. *gemalanyu cankao yufa* [*A reference grammar of Kavalan*]. Formosan Languages Series 12. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 1988. *Aspect: A general system and its manifestation in Mandarin Chinese.* Taipei: Student Book Co.

Huang, Lillian M. 1989. The pronominal system in Atayal. *Studies in English Literature and Linguistics* 15:37–50.

Huang, Lillian M. 1992. *taiyayu chuji duben* [*Atayal for beginners*]. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1991. The semantics of *s-* in Atayal. *Studies in English Literature and Linguistics* 17:37–50.

Huang, Lillian M. 1992. Atayal myths: A text with linguistic analysis. *Studies in English Literature and Linguistics* 18:111–141.

Huang, Lillian M. 1993a. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1993b. A bear and a woman: an Atayal text with linguistic analysis. *Studies in English Literature and Linguistics* 19:47–69.

Huang, Lillian M. 1994. Ergativity in Atayal. *Oceanic Linguistics* 33.1:129–143.

Huang, Lillian M. 1995a. The case markers and pronominal system in Amis. *The Journal of National Chengchi University* 70:217–257.

Huang, Lillian M. 1995b. The syntactic structure of Wulai and Mayrinax Atayal: A comparison. *Bulletin of National Taiwan Normal University* 40:261–294.

Huang, Lillian M. 1995c. The voice system in Atayal. *Papers from the First International Symposium on Languages in Taiwan*, ed. by Feng-fu Tsao and Mei-hui Tsai, 73–98. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995d. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995e. taiwan nandao minzu muyu xianyou jiaocai zhi bijiao han wojian [Comparison and review of contemporary aboriginal language textbooks in Taiwan] *taiwan nandao minzu muyu yanjiu lunwenji* [*Papers on Formosan languages*], 77–96 Taipei: Education committee, Ministry of Education. [In Chinese]

Huang, Lillian M. 1995f. taiwan yuanzhumin muyu jiaoxue: jiu xianyou jiaocai han shizhi tanqi [On aboriginal language teaching in Taiwan: Starting from current teaching materials and in-service teachers] *diyijie taiwan bentu wenhua xueshu yantaohui lunwenji* [*Papers from the First Symposium on Taiwan Native Cultures*], 247–262. Taipei: National Taiwan Normal University. [In Chinese]

Huang, Lillian M. 1996a. A syntactic and semantic study of *-ay*, *-aw* and *-anay* in Mayrinax Atayal. *Studies in English Literature and Linguistics* 22:15–36.

Huang, Lillian M. 1996b. Interrogative constructions in Mayrinax Atayal. *Bulletin of National Taiwan Normal University* 41:263–296.

Huang, Lillian M. 1997a. gaoxiongxian de bunongyu [(Isbukun) Bunun in Kaohsiung County]. *gaoxiongxian nandao yuyan* [*Formosan Austronesian languages in Kaohsiung County*], ed. by Paul Jen-kuei Li, 351–410. Kaohsiung: Kaohsiung County Government. [In Chinese]

Huang, Lillian M. 1997b. Serial verb constructions in some Formosan languages: A typological view. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Huang, Lillian M. 1998. taiwan yuanzhumin muyu jiaoxue xiankuang han zhanwang [The current situation of Taiwan aboriginal language teaching and its prospect] *taiwan yuanzhumin wenhua yu jiaoyu zhi fazhan* [*Development of Taiwan indigenous culture and education*], 77–106. Taipei: National Taiwan Normal University. [In Chinese]

Huang, Lillian M. 2000a. *taiyayu cankao yufa* [*A reference grammar of Atayal*]. Formosan Language Series 1. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000b. *shaoyu cankao yufa* [*A reference grammar of Thao*]. Formosan Language Series 4. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000c. *beinanyu cankao yufa* [*A reference grammar of Puyuma*]. Formosan Language Series 9. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000d. taiwan nandao yuyanxue yanjiu: huigu yu zhanwang [Formosan Linguistics: Past and Future]. *hanxue yanjiu* [*Chinese Studies*] 18: 79-110. Taipei: Center of Chinese Studies, National Library. [In Chinese]

Huang, Lillian M. 2000e. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364–390.

Huang, Lillian M. 2001. Focus system of Mayrinax Atayal: a syntactic, semantic and pragmatic perspective. *Journal of Taiwan Normal University: Humanities & Social Sciences* 46.1-2:51–69.

Huang, Lillian M. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3.2:197–225.

Huang, Lillian M. 2003. yuanzhumin zuyu nengli renzheng: huigu yu zhanwang [Indigenous language proficiency tests: Past and future]. *yuanzhumin jiaoyu jikan* [*Aboriginal Education Quaterly*] 29:5–27.

Huang, Lillian M. 2006. Case marking in Plngawan Atayal. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang and Dah-an Ho, 205–238. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

Huang, Lillian M. 2007a. taiwan yuanzhumin zuyu nengli renzheng kaoshi zhi huigu yu zhanwang [Prospect and retrospect on Taiwan indigenous language proficiency tests]. *yuanzhuminzu yuyan fazhan lunchong—lilun yu shiwu* [*The indigenous languages* (*of Taiwan*): *Theory and practice*], 82–100. Taipei: Council of Indigenous Peoples, Executive Yuan. [In Chinese]

Huang, Lillian M. 2007b. taiwan yuanzhumin zuyu jiaocai zhi huigu yu zhanwang [Taiwan indigenous language textbooks: Past and future]. *yuanzhuminzu yuyan fazhan lunchong—lilun yu shiwu* [*The indigenous languages* (*of Taiwan*): *Theory and practice*] 213–224. Taipei: Council of Indigenous Peoples, Executive Yuan. [In Chinese]

Huang, Lillian M. 2008. Grammaticalization in Squliq Atayal. *Concentric: Studies in Linguistics* 34.2:1–46.

Huang, Lillian M. and Tali' Hayung. 2008. Syntax and semantics of *p-* in Squliq Atayal. *Language and Linguistics* 9.3:491–521.

Huang, Lillian M. and Tali' Hayung. 2011. Coordination and comitativity in Squliq Atayal. *Language and Linguistics* 12.1:1–47.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna H. Chang and Joy 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 Publishing Co., Ltd.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna H. Chang and Joy J. Wu. 1999a. A typological study of pronominal systems in some Formosan languages. In *Selected Papers from The Fifth International Conference on Chinese Linguistics*, ed. by H. Samuel Wang, Feng-fu Tsao and Chin-fa Lien, 165–198. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna H. Chang and Joy J. Wu. 1999b. Interrogative constructions in some Formosan languages. In *Chinese languages and linguistics V: Interactions in language*, ed. by Yuen-mei Yin, I-li Yang, Hui-chen Chan, 639–680. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 2. Taipei: Academia Sinica.

Hayung, Tali'. 2008. jianshi taiyayu qianzhui yanjiu [A study of the prefixes in Jianshih Atayal]. MA thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

Li, Paul Jen-kuei. 1973. *Rukai Structure.* Institute of History and Philology Special Publications no. 64. Taipei: Academia Sinica.

Li, Paul Jen-kuei. 1974. Alternations between semi-consonants and fricatives and liquids. *Oceanic Linguistics* 13:164–186.

Li, Paul Jen-kuei. 1991. [*taiwan nandao yuyan de yuyin fuhao xitong*] *Orthographic systems for Formosan languages*. Taipei: Ministry of Education. [In Chinese]

Lin, Ying-chin. 2000. *bazehaiyu* [*Pazeh*]. Formosan Languages Series 3. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Malchukov Andrej, Martin Haspelmath and Bernard Comrie. 2010. Ditransitive constructions: A typological overview. In *Studies in ditransitive constructions: A comparative handbook*, ed. by Andrej Malchukov, Martin Haspelmath and Bernard Comrie, 1–64. Berlin/New York: Mouton de Gruyter.

Ross, Malcolm. 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31.1:23–64.

Shih, Pei-ru Cindy. 2008. Interrogative constructions in Plngawan Atayal. MA thesis. Taipei: National Taiwan Normal University.

Tan, Cindy Ro-lan. 1997. A study of Puyuma simple sentences. MA thesis. Taipei: National Taiwan Normal University.

Teng, Stacy F. 1997a. Complex sentences in Puyuma. MA thesis. Taipei: National Taiwan Normal University.

Teng, Stacy F. 1997b. Coordination in Formosan languages Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Tsuchida, Shigeru. 1982. Subclassification of Amis dialects. Unpublished Manuscript.

Tung, T'ung-ho, with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng and Margaret M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publications 48. Taipei: Academia Sinica. Reprinted in 1991 with notes and errata prepared by Paul Jen-kuei Li.

Wu, Joy J. 1995. Complex sentences in Amis. MA Thesis. Taipei: National Taiwan Normal University.

Wu, Joy J. 1997. Relative clauses in Formosan languages. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Wu, Joy J. 2000. *ameiyu cankao yufa* [*A reference grammar of Amis*]. Formosan Languages Series 11. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Yan, Zhi-kuang. 1992. ameiyu de yufa jiegou: canyuzhe yu shijian de yanjiu [Syntactic structure of Amis: A study of participants and events]. MA Thesis. Taipei: National Chengchi University. [In Chinese]

Yeh, Marie M. 1997. Pivotal constructions in Formosan languages. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Yeh, Marie M. 2000. *saixiayu cankao yufa* [*A reference grammar of Saisiyat*]. Formosan Languages Series 2. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Yeh, Marie M. 2003. A syntactic and semantic study of Saisiyat verbs. PhD dissertation. Taipei: National Taiwan Normal University.

Yeh, Marie M., Lillian M. Huang, Elizabeth Zeitoun, Anna H. Chang and Joy Wu. 1998. A preliminary study on the negative constructions in some Formosan languages. In *Selected Papers from the Second International Symposium on Languages in Taiwan*, 79–110. Taipei: Crane Publishing Co., Ltd.

Zeitoun, Elizabeth. 1997. Temporal, hypothetical and counterfactual clauses in Formosan languages. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Zeitoun, Elizabeth. 2000a. Dynamic vs. Stative verbs in Mantauran (Rukai). *Oceanic Linguistics* 39.2:415–427.

Zeitoun, Elizabeth. 2000b. Notes on a possessive construction in the Formosan languages. In *Grammatical analysis: Morphology, syntax and semantics*, ed. By Videa DeGuzman and Byron Bender (eds), 241–257. Oceanic Linguistics Special Publication 29. Honolulu: University of Hawai‘i Press.

Zeitoun, Elizabeth (Qi, Li-sha). 2000c. *bunongyu cankao yufa* [*A reference grammar of Bunun*]. Formosan languages Series 5. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Zeitoun, Elizabeth (Qi, Li-sha). 2000d. *zouyu cankao yufa* [*A reference grammar of Tsou*]. Formosan Languages Series 7. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Zeitoun, Elizabeth (Qi, Li-sha). 2000e. *lukaiyu cankao yufa* [*A reference grammar of Rukai*]. Formosan Languages Series 8. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Zeitoun, Elizabeth. 2002. Reciprocals in Formosan languages. Paper presented at the Ninth International Conference on Austronesian Linguistics, 8–11 January. Canberra: Australian National University.

Zeitoun, Elizabeth. 2004. Introduction. *Faits de langues: Les langues austronésiennes* 23–24:41–58.

Zeitoun, Elizabeth and Lillian M. Huang. 1997. Toward a typology of tense, aspect and modality in the Formosan languages: a preliminary study. *Chinese languages and linguistics IV: Typological studies of languages in China*, ed. by Tseng Chiu-yu, 595–618. Symposium Series of the Institute of History and Philology, Academia Sinica No. 2.Taipei: Academia Sinica.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391–414.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh and Anna H. Chang. 1999. Existential, possessive, and locative constructions in Formosan languages. *Oceanic Linguistics* 38.1:1–42.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual and modal systems of Formosan languages: a typological perspective. *Oceanic Linguistics* 35.1:21–56.

Lillian M. Huang: Bibliography

Lillian M. Huang's publications related to Formosan languages and linguistics

Huang, Lillian M. 1989. The pronominal system in Atayal. *Studies in English Literature and Linguistics* 15:37–50.

Huang, Lillian M. 1991. The semantics of *s-* in Atayal. *Studies in English Literature and Linguistics* 17:37–50.

Huang, Lillian M. 1992. Atayal myths: A text with linguistic analysis. *Studies in English Literature and Linguistics* 18:111–141.

Huang, Lillian M. 1993a. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1993b. A bear and a woman: an Atayal text with linguistic analysis. *Studies in English Literature and Linguistics* 19:47–69.

Huang, Lillian M. 1994. Ergativity in Atayal. *Oceanic Linguistics* 33.1:129–143.

Huang, Lillian M. 1995a. The case markers and pronominal system in Amis. *The Journal of National Chengchi University* 70:217–257.

Huang, Lillian M. 1995b. The syntactic structure of Wulai and Mayrinax Atayal: A comparison. *Bulletin of National Taiwan Normal University* 40:261–294.

Huang, Lillian M. 1995c. The voice system in Atayal. *Papers from the First International Symposium on Languages in Taiwan*, ed. by Feng-fu Tsao and Mei-hui Tsai, 73–98. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995d. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1996a. A syntactic and semantic study of *-ay*, *-aw* and *-anay* in Mayrinax Atayal. *Studies in English Literature and Linguistics* 22:15–36.

Huang, Lillian M. 1996b. Interrogative constructions in Mayrinax Atayal. *Bulletin of National Taiwan Normal University* 41:263–296.

Huang, Lillian M. 1997a. gaoxiongxian de bunongyu [(Isbukun) Bunun in Kaohsiung County]. *gaoxiongxian nandao yuyan* [*Formosan Austronesian languages in Kaohsiung County*], ed. by Paul Jen-kuei Li, 351–410. Kaohsiung: Kaohsiung County Government. [In Chinese]

Huang, Lillian M. 1997b. Serial verb constructions in some Formosan languages: A typological view. Paper presented at the Eighth International Conference on Austronesian Linguistics, 28–30 December. Taipei: Academia Sinica.

Huang, Lillian M. 2000a. *taiyayu cankao yufa* [*A reference grammar of Atayal*]. Formosan Language Series 1. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000b. *shaoyu cankao yufa* [*A reference grammar of Thao*]. Formosan Language Series 4. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000c. *beinanyu cankao yufa* [*A reference grammar of Puyuma*]. Formosan Language Series 9. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000d. taiwan nandao yuyanxue yanjiu: huigu yu zhanwang [Formosan linguistics: Past and future]. *hanxue yanjiu* [*Chinese Studies*] 18:79–110. Taipei: Center of Chinese Studies, National Library. [In Chinese]

Huang, Lillian M. 2000e. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364–390.

Huang, Lillian M. 2001b. Focus system of Mayrinax Atayal: A syntactic, semantic and pragmatic perspective. *Journal of Taiwan Normal University: Humanities & Social Sciences* 46.1-2:51–69.

Huang, Lillian M. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3.2:197–225.

Huang, Lillian M. 2006. Case marking in Plngawan Atayal. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 205–238. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

Huang, Lillian M. 2008. Grammaticalization in Squliq Atayal. *Concentric: Studies in Linguistics* 34.2:1–46.

Huang, Lillian M., and Tali' Hayung. 2008. Syntax and semantics of *p-* in Squliq Atayal. *Language and Linguistics* 9.3:491–521.

Huang, Lillian M., and Tali' Hayung. 2011. Coordination and comitativity in Squliq Atayal. *Language and Linguistics* 12.1:1–47.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna H. Chang and Joy J. Wu. 1998. A typological overview of nominal case marking of the Formosan languages. In *Selected papers from the Second International Symposium on Languages in Taiwan*, 21–48. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna Chang and Joy Wu. 1999a. A typological study of pronominal systems in some Formosan languages. In *Selected Papers from The Fifth International Conference on Chinese Linguistics*, ed. by H. Samuel Wang, Feng-fu Tsao and Chin-fa Lien, 165–198. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M., Elizabeth Zeitoun, Marie Yeh, Anna H. Chang and Joy J. Wu. 1999b. Interrogative constructions in some Formosan languages. In *Chinese languages and linguistics V: Interactions in language*, ed. by Yuen-mei Yin, I-li Yang, Hui-chen Chan, 639–680. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 2. Taipei: Academia Sinica.

Yeh, Marie M., Lillian M. Huang, Elizabeth Zeitoun, Anna H. Chang and Joy J. Wu. 1998. A preliminary study on the negative constructions in some Formosan languages. In *Selected Papers from the Second International Symposium on Languages in Taiwan*, 79–110. Taipei: Crane Publishing Co., Ltd.

Zeitoun, Elizabeth and Lillian M. Huang. 1997. Toward a typology of tense, aspect and modality in the Formosan languages: a preliminary study. In *Chinese languages and linguistics IV: Typological studies of languages in China*, ed. by Tseng

Chiu-yu, 595–618. Symposium Series of the Institute of History and Philology, Academia Sinica No. 2. Taipei: Academia Sinica.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391–414.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh and Anna H. Chang. 1999. Existential, possessive, and locative constructions in Formosan languages. *Oceanic Linguistics* 38.1:1–42.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual and modal systems of Formosan languages: a typological perspective. *Oceanic Linguistics* 35.1:21–56.

Publications on aboriginal language teaching and proficiency tests

Huang, Lillian M. 1992. *taiyayu chuji duben* [*Atayal for beginners*]. Taipei: Crane Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 1995a. taiwan nandao minzu muyu xianyou jiaocai zhi bijiao han wojian [Comparison and review of contemporary aboriginal language textbooks in Taiwan] *taiwan nandao minzu muyu yanjiu lunwenji* [*Papers on Formosan languages*], 77–96. Taipei: Education committee, Ministry of Education. [In Chinese]

Huang, Lillian M. 1995b. taiwan yuanzhumin muyu jiaoxue: jiu xianyou jiaocai han shizhi tanqi [On aboriginal language teaching in Taiwan: Starting from current teaching materials and in-service teachers] *diyijie taiwan bentu wenhua xueshu yantaohui lunwenji* [*Papers from the First Symposium on Taiwan Native Cultures*], 247–262. Taipei: National Taiwan Normal University. [In Chinese]

Huang, Lillian M. 1998a. taiwan yuanzhumin muyu jiaoxue xiankuang han zhanwang [The current situation of Taiwan aboriginal language teaching and its prospect] *taiwan yuanzhumin wenhua yu jiaoyu zhi fazhan* [*Development of Taiwan indigenous culture and education*], 77–106. Taipei: National Taiwan Normal University. [In Chinese]

Huang, Lillian M. 1998b. taiwan yuanzhumin muyu shizhi yingyou zhi xueyang [Essential scholarship of teaching indigenous teachers in Taiwan] *disanjie taiwan bentu wenhua xueshu yantaohui lunwenji* [*Papers from the Third Symposium on Taiwan Native Cultures*], 189–212. Taipei: National Taiwan Normal University. [In Chinese]

Huang, Lillian M. 2001a. taiwan xiangtu jiaocai di si ce: yuanzhumin pian [*Teaching materials for Taiwan native cultures, Volume 4: Indigenous peoples*]. Taipei: Huatai wenhua. [In Chinese]

Huang, Lillian M. 2003. yuanzhumin zuyu nengli renzheng: huigu yu zhanwang [Indigenous language proficiency tests: Past and future]. *yuanzhumin jiaoyu jikan* [*Aboriginal Education Quaterly*] 29:5–27. [In Chinese]

Huang, Lillian M. 2005. taiwan yuanzhumin zuyu jiaoxue jinxi guan [The past and current situation of Taiwan indigenous language teaching]. *xuexi xing buluo lilun*

yu shiji [*Theory and practice of learning tribes*], 167–186. Taipei: Shida Shuyuan. [In Chinese]

Huang, Lillian M. 2007a. taiwan yuanzhumin zuyu nengli renzheng kaoshi zhi huigu yu zhanwang [Prospect and retrospect on Taiwan indigenous language proficiency tests]. *yuanzhuminzu yuyan fazhan lunchong—lilun yu shiwu* [*The indigenous languages* (*of Taiwan*)*: Theory and practice*], 82–100. Taipei: Council of Indigenous Peoples, Executive Yuan. [In Chinese]

Huang, Lillian M. 2007b. taiwan yuanzhumin zuyu jiaocai zhi huigu yu zhanwang [Taiwan indigenous language textbooks: Past and future]. *yuanzhuminzu yuyan fazhan lunchong—lilun yu shiwu* [*The indigenous languages* (*of Taiwan*): *Theory and practice*] 213–224. Taipei: Council of Indigenous Peoples, Executive Yuan. [In Chinese]

Huang, Lillian M. 2008. di yi zhang yuanzhumin yuyan [Chapter one: Indigenous languages]. *xuxiu taibeixianzhi juansan zhuminzhi—di er pian yuyan (shang)* [*Monograph Series of Taipei County Vol.3, fasc. 2: Language*]. Taipei: Taipei County Government. [In Chinese]

Huang, Lillian M. 2009. taiwan yuanzhumin zuyu jiaoxue zhi huigu yu zhanwang [Indigenous language teaching in Taiwan: Past and future]. *qingyun xuebao* [*Journal of Chingyun University*] 29.4:139–166. [In Chinese]

Huang, Lillian M. 2011. taiwan yuanzhuming zuyu shizi peiyu zhi huigu yu zhanwang [Training of indigenous language teachers in Taiwan: Past and future]. *taiwan yuwen yanjiu* [*Journal of Taiwanese Languages and Literature*] 6.1:69–114. [In Chinese]

Huang, Lillian M., Yu-yang Liu and Tali' Hayung. 2014. cong taiyayu cidian tan taiwan yuanzhuminzu cidian zhi bianzuan [On compiling Taiwan indigenous language dictionaries: An illustration of Atayal dictionaries]. *Taiwan yuanzhuminzu yanjiu jikan* [*Taiwan Journal of Indigenous Studies*] 7.2:73–118. [In Chinese]

1 *Some Proto Austronesian coronals reexamined*[*]

MALCOLM ROSS

1 Introduction

Although the vast majority of Austronesian languages lie outside Taiwan, they all belong to Malayo-Polynesian, a single branch of the family (Blust 1977). The Formosan languages, on the other hand, make up several first-order branches, and this gives them a considerable importance in the reconstruction of Proto Austronesian (PAn).[1] Zorc (1987:758), observing that Philippine evidence does not clarify the reconstruction of the PAn coronal voiced oral obstruents, the topic of section 2 below, writes, '…if there is any justification for a distinction amongst voiced apicals [coronals] at the PAn level, it will come from Formosan languages…'. Earlier he had commented with regard to the reconstruction of PAn *S, re-examined in section 3, that it 'is only justified by various sibilant reflexes among Formosan languages' (Zorc 1982:117).

Ross (1992) reviewed the reconstructed phoneme paradigm of Proto Austronesian (PAn) and discussed the probable phonetic values of its phonemes, based mainly on

* My thanks go to Juliette Blevins and Hui-chuan Huang for answering my questions about sound changes and about the phonetics of Formosan languages, respectively to Stacy F. Teng, who checked various Puyuma data for me, to Elizabeth Zeitoun and Joy J. Wu, who commented on an earlier draft and to Shigeru Tsuchida, who provided me with an electronic copy of unpublished wordlists of Amis dialects he collected in 1980–81. Much of the research on which this paper is based was conducted during periods of research at the Institute of Linguistics of Academia Sinica (ILAS), Taipei, from September to December 2003 and from September 2008 to January 2009. During the earlier period I was a Visiting Research Fellow under the project 'A morpho-syntactic study of Rukai' (Principal Investigator: Elizabeth Zeitoun), supported by National Science Council of Taiwan grant NSC 92-2811-H-001-010. I am grateful to Elizabeth Zeitoun, Dah-An Ho, and Jackson Sun for facilitating these visits, and to the NSC and ILAS for their support.

[1] Wolff (1997:587 n. 9) puts a contrary view when he writes, 'Methodologically it would seem preferable to leave forms which occur exclusively in the languages of Taiwan out of consideration.' The underlying reason for this statement remains unstated, namely that when he wrote his 1997 paper he did not accept the 'Out of Taiwan' hypothesis. (He has since changed his position—see Wolff 2007). He goes on to suggest that contact and borrowing are sufficient to account for the distinctions alleged to reflect *d_1, *d_2 and *d_3, but offers no evidence for this statement.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 1–38
Asia-Pacific Linguistics, 2015

Formosan data. The most problematic part of the paradigm was and is the coronals, and they are revisited here in the light of more recent lexical research. The paradigm presented in the 1992 paper corresponds in most respects with the paradigm presented by Robert Blust (1990, 1999, 2009:547), differing from it mainly in distinguishing three obstruents notated as *d_1, *d_2 and *d_3 where he has just *d. In section 2 of the present paper I reexamine evidence for the three *d's. In section 3 I look again at alleged PAn sibilants and propose, on the basis of sound correspondences provided by Tsuchida (1976), that the segments notated as *S and *s were probably each two PAn phonemes. The two additional phonemes evidently had low functional loads. In section 4.1 I present a version of the paradigm of PAn consonants which takes account of these additions. Ross (1992:39) remarked that it is difficult to reconstruct the sound changes that led to the Paiwan reflexes of PAn coronals, and in section 4.2 I suggest a set of phonetic values for the PAn coronals that permit a better account of their Paiwan reflexes than hitherto.

A standard principle of reconstruction asserts that a distinction reconstructed in PAn should be reflected in at least two first-order branches of the family. This presupposes, of course, that we know what those branches are. This is a source of some controversy. There are several current hypotheses about the subgrouping of Formosan languages (Ho 1998; Blust 1999; Sagart 2004; Ross 2009). Two of these subgroupings (Ho's, with six Formosan groups, and Blust's, with nine) are based on shared phonological innovations, and to use either of them for phonological reconstruction would entail circular reasoning. Sagart's and Ross's subgroupings are based respectively on lexical and morphosyntactic innovations, but each posits a first-order subgroup which includes a number of Formosan languages and Proto Malayo-Polynesian (PMP), and the membership of this subgroup differs significantly under the two hypotheses. Using either of them would threaten the integrity of the reconstruction. I have therefore chosen to treat equally each of the twenty Formosan languages for which data are available. Thus very little in this paper depends on a subgrouping hypothesis, other than my use of low-order reconstructed languages, i.e. Proto Rukai and Proto Atayalic, in order to reduce the citation of data. Proto Rukai has five daughter dialects: Tanan, Budai, Maga, Tona and Mantauran. Proto Atayalic, has two daughter languages, Atayal and Seediq, each with a network of dialects, but no one has disputed the longstanding claim that Atayal and Seediq exclusively share Proto Atayalic as their common ancestor.[2]

Appendix 1 is a tabulation of the Formosan consonant correspondences reflected in my sources, together with the reconstructed phonemes of PAn as reconstructed by Tsuchida (1976) and Blust (2009) and in the current paper. Two versions of the phonetic values of these protophonemes are provided, labelled 'step 1' and 'step 2': these terms are explained in section 4. The row labelled 'Position' refers to the step 1 values. The bottom row of the table provides for reference the symbols used by Wolff (2010). I have deliberately separated them from the symbols at the top of the table, as they represent an analysis which has a separate scholarly history. The phonetic symbols alongside Wolff's symbols are mine. He says that 'for the series *j, *c, and *ɬ the tongue tip was further front than for the series *d, *t, *n, *l and *s' (Wolff 2010:31 n. 50). He describes *c and *t respectively as an apico-dental and an apico-alveolar stop (Wolff 2010:32).

---

[2] See the review in Tsuchida (1976:2-10) and the reconstruction by Li (1981).

2 The voiced oral obstruents

2.1 Previous research and a proposal

The proposal that *d_1, *d_2 and *d_3 were distinct PAn phonemes is based on earlier work by several scholars working on Formosan data. It reflects the set of distinctions made by Dahl (1973, 1976, 1981), with the difference that he claims that the distinctions are also reflected in Tagalog, where I claim that PAn *d_1, *d_2 and *d_3 are merged as Proto Malayo-Polynesian (PMP) *d.[3] The three subscripted phonemes correspond to distinctions made by Tsuchida (1976) in the rather complex way described by Ross (1992). Li (1985) combines *d_2 and *d_3 as his *D, but maintains the distinction between this and *d_1, his *d.

Blust (1999:50–51) argues against *d_1, *d_2, and *d_3 and in favour of simple *d. He notes that Paiwan has the three segments *dj* [ɻ], *z* [z], and *ḍ* [ɖ] (in Ferrell's 1982 Kułałau Paiwan orthography),[4] and that I claim these to reflect PAn *d_1, *d_2, and *d_3 respectively. He also notes Dahl's and my claim that the Puyuma reflexes correlate well with the Paiwan reflexes: *d_1 is reflected as Puyuma *d* [d], whilst *d_2 and *d_3 merge as Puyuma *z* ([ɖ] or [ʐ], varying by dialect). In his opinion, however, the correspondence between Paiwan and Puyuma is not good, and the exceptions are such as to indicate borrowing rather than shared inheritance. He suggests that borrowing was in both directions, but more heavily from Puyuma into Paiwan, and that both languages have borrowed from Rukai. I do not deny that borrowing occurred (and I agree that Paiwan and Puyuma do not form a subgroup), but I shall seek to show here that the correlation between Paiwan and Puyuma reflexes is better than Blust thought in 1999, and that there are other fragments of evidence that suggest that PAn *d_1, *d_2, and *d_3 were distinct and are all three reflected in Paiwan, and only in Paiwan.

Why should my results differ from Blust's? The main reason is that descriptions of the Formosan languages have improved considerably in recent years. The earlier dictionaries of Paiwan (Ferrell 1982), Puyuma (Cauquelin 1991), and Amis (Fey 1986) have been joined by the second edition of Egerod's Atayal dictionary (1999), new dictionaries of Pazih (Li & Tsuchida 2001), Paiwan (Egli 2002), Thao (Blust 2003), and Kavalan (Li & Tsuchida 2006) and Li's (2003) edition of Naoyoshi Ogawa's *English–Favorlang vocabulary*. Li & Toyoshima's (2006) edition of Ogawa's comparative Formosan vocabulary has also made an additional quantity of lexical material accessible.[5]

---

[3] Blust (1976) discusses the improbability that Tagalog reflects these distinctions.

[4] The same orthography is used by Early & Whitehorn (2003). Chang (2006) replaces *ḍ* with *D*. In Egli (1990, 2002) the orthographic symbols are *j*, *z*, and *dr*. Ferrell and Egli show both *s* and *z* as sharing the same point of articulation as retroflex *ḍ*. Chang, however, who provides the only phonetic account, disagrees, showing *s* and *z* as alveolar but *ḍ* as retroflex, and the phoneme paradigms in Ho (1978) reflect this account.

[5] Other data sources include the compilations of Austronesian cognate sets in Dahl (1973), Tsuchida (1976), Wolff (1997), and Blust & Trussel (2010), compilations of cognate sets for groups of Formosan dialects (Ho 1978; Li 1977, 1978, 1981, 1988; Ting 1978; Tsuchida 1969, 1971); Tsuchida's (1981) unpublished wordlists for Amis dialects and his 1982 comparative vocabulary of near-extinct languages; wordlists in Tryon (1995); and my field notes on

Dahl's symbols *d_1, *d_2, and *d_3, which Ross (1992) adopted, raise an analytical (as well as a notational) problem. For Dahl (1976:60), subscripted letters represent distinct protophonemes. For example, his *t_1 and *t_2 represent what are more conventionally represented as *t and *C, but he uses *t_1 and *t_2 because it allows him to write *t where there is no criterial witness to distinguish between them. In my notation, Dahl's *t is *(t,C), and I prefer this because it is more explicit. Tsuchida (1976:123–124), on the other hand, uses subscripts to distinguish correspondence sets which differ from each other 'slightly' and which are provisionally assumed to reflect the same protophoneme. For example, the correspondence sets reflecting *S_1 and *S_2 are provisionally assumed both to reflect *S. However, Tsuchida writes (p. 124), 'It remains a possibility that these subnumbered protophonemes may turn out to be different protophonemes in PAN [PAn] in future study.' There is thus an ambiguity in the literature about the use of subscripts, and, as I am claiming explicitly that *d_1 and *d_2 were distinct phonemes, and that *d_3 was possibly a distinct phoneme too, I prefer for the sake of clarity and consistency to use distinct symbols. In the Dyen/Blust PAn orthographic convention, each protophoneme is represented by a single letter, upper- or lowercase (not a digraph), and IPA symbols other than *ŋ are avoided. I return to this matter below.

Table 1 shows the reflexes of PAn *Z and putative *d_1, *d_2, and *d_3 in all Formosan languages for which adequate data are available to me. The presentation differs in one significant respect from Table 2 of Ross (1992). In that paper I inferred that PAn *d_1 occurred only morpheme-initially and was in complementary distribution with PAn *r, which occurred only morpheme-medially and -finally. This may well be true, but Formosan reflexes of putative *-r- are too few to be sure. There are also two apparent instances of morpheme-final *d_1—PAn *qañud 'drift' in (2a-v), PAn *qapid 'braid (hair)' in (11f)—which would require explanation if reflexes of putative morpheme-final *r in fact reflect morpheme-final *d_1.

Western Oceanic languages. Some Puyuma data, marked ST, were checked with native speakers and supplied by Stacy F. Teng.

Table 1: Reflexes of PAn coronal voiced oral obstruents in Formosan languages and PMP

PAn (Ross 1992)	*Z		*d_1		*d_2		*d_3	
PAn (this paper)	*z		*d		*Z		*D	
Puyuma	d	[d]	d	[d]	ḏ/z	[ḏ/ẓ]	ḏ/z	[ḏ/ẓ]
Proto Rukaic	*d	*[d]	*D	*[ḏ]	*D	*[ḏ]	*D	*[ḏ]
Tsou	c	[ts, ʧ]	c	[ts, ʧ]	c	[ts, ʧ]	c	[tts, ʧ]
Kanakanavu	c	[cç, ʧ]	c	[cç, ʧ]	c	[cç, ʧ]	c	[cç, ʧ]
Saaroa	s	[s]	s	[s]	s	[s]	s	[s]
Paiwan	dj	[ɟ, dʲ]	dj	[ɟ, dʲ]	z	[z]	ḏ	[ḏ]
Bunun	d	[d, ʔd]	d	[d, ʔd]	d	[d, ʔd]	d	[d, ʔd]
Thao	s	[s]	s	[s]	s	[s]	t	[t̪]
Taokas	t		t		…		t	
Favorlang-Babuza	t		t		d, r		r	
Papora	t, d		d		…		…	
Hoanya	dz		d		d		…	
Pazih	d	[d]	d	[d]	d	[d]	d	[d]
Saisiyat	r	[ɾ]	r	[ɾ]	r	[ɾ]	r	[ɾ]
Proto Atayalic	*d	*[d]	*d	*[d]	*d	*[d]	*d	*[d]
Siraya	d, l		s		s		s	
Amis	l	[ɾ]	r, l	[r, ɾ]	r, l	[r, ɾ]	r	[r]
Kavalan	z	[ẓ]	z	[ẓ]	z	[ẓ]	z	[ẓ]
PMP	*Z		*d		*d		*d	

In Table 1 and in examples below, protoforms for Atayalic (Li 1981) and Rukaic (Li 1977) are used as a form of shorthand, as the reconstructions are uncontroversial. PAn *Z (Blust's *z) is included because it is impossible to discuss *d_1 without discussing *Z, as the two usually have the same reflexes in Formosan languages. The table shows the orthographic representation of each phoneme,[6] followed by its default phonetic manifestation, except in the cases of the extinct languages Taokas, Favorlang-Babuza, Papora, Hoanya and Siraya. The Puyuma alternants *ḏ* and *z* occur in different dialects. Papora, Siraya and Amis each appear to have two reflexes of certain protophonemes. In

[6] Ferrell's *d̤* is replaced with *ḏ* in the interests of legibility.

the case of Amis, the 'original' reflex appears to be *r*, as *l* generally occurs in morphemes which contain another *l*, i.e., it is the result of assimilation.

A number of the phonetic representations in Table 1 have an underdot, e.g. [ḍ], marking segments that in the Formosan literature are often termed 'retroflex'. The term is somewhat misleading. The so-called retroflexes of Formosan languages have similar articulations to the so-called 'retroflexes' of Mandarin (Hui-chuan Huang, pers. comm.), and, although these are conventionally included under the rubric 'retroflex' (Hamann 2003:22–23, 45), they do not entail curling of the tongue-tip and are more appropriately described as flat laminal post-alveolars (Ladefoged & Wu 1984; Ladefoged & Maddieson 1996:154).[7] It is reasonable to infer this articulation for PAn.[8] Ladefoged & Maddieson (1996) adopt the convention of printing a dot under a character, e.g., ḍ rather than ɖ, to distinguish a laminal postalveolar from an apical [retroflex] postalveolar. The underdot is used in phonetic transcriptions, but the hooked retroflex symbols for postalveolars are retained in Formosan words and in correspondence tables as they are more readable.

In view of the desirability of notation without subscripts, from this point on, PAn

- *Z becomes *z, following Blust's (1999:34–35, footnote) revised notation
- *d_1 becomes *d
- *d_3 becomes *D, following Li (1985)

But how do we represent *d_2? A reading of Table 1 suggests the initial hypothesis that *d_1 and *d_2 were the voiced equivalents of voiceless *t and *C, i.e., respectively a stop and an affricate (see section 4). If so, then using *d for *d_1 is appropriate, and uppercase *Z is the obvious choice for *d_2, so

- *d_2 becomes *Z

Unfortunately *Z has previously been used by various scholars for what has recently become *z. However, the only semi-practical alternative is to introduce a digraph (*dz) into the system, and this is potentially even more confusing, especially when the investigator is performing searches on electronic files. Under either of the hypotheses about the phonetic values of PAn coronals presented in section 4 (Tables 6 and 8), the existing PAn orthographic convention is inconsistent in the use of uppercase letters, as they represent more than one point of articulation. I have been sorely tempted to revise the orthography so that, e.g., uppercase letters are used for postalveolars, lowercase for dentals and alveolars, but this presupposes that scholars are both certain about and agree upon the phonetics of PAn phonemes, and neither presupposition is true. In any case a new orthography would lead to confusion.[9]

---

[7] The retroflex is a flat post-alveolar as opposed to the English domed post-alveolars [ʃ] and [ʒ].

[8] It is tempting to infer that modern Formosan languages have adapted their articulation to Mandarin, in which their speakers are bilingual, but Hui-chuan Huang (pers. comm.) points out that in Taiwan Mandarin and Taiwanese Southern Min the retroflex obstruents are typically replaced by the corresponding apical dentals, so this is not possible.

[9] Wolff (2010) uses a PAn orthography (see Appendix 1) which more closely reflects putative phonetic values of the protosegments, but it reflects a different analysis of PAn phonology from mine and is also, I suspect, unfamiliar to most Austronesianists.

Table 2 lists languages which witness to one or more of the distinctions among *z, *d, *Z, and *D.[10] Table 3 is largely an extract from Table 1 and is intended as a point of reference for the protophoneme sections below. It shows the reflexes of PAn *z, *d, *Z, and *D in the witness languages. The languages in the lower section are witnesses to the PMP distinction between *z and the other coronal voiced obstruents. The data supporting Tables 1 and 2 are presented in the following sections.

Table 2: Witness languages for one or more of the distinctions among *z, *d, *Z and *D

Distinction	*z vs *d	*d vs *Z	*d vs *D	*Z vs *D
Puyuma	no	yes	yes	no
Proto Rukaic	yes	no	no	no
Kanakanavu	no	no	medially?	medially?
Paiwan	no	yes	yes	yes
Thao	no	no	yes	yes
Favorlang-Babuza	no	yes	yes	no
Hoanya	yes	no	…	…
Siraya	yes	no	no	no
PMP	yes	no	no	no

Table 3: Reflexes of *z, *d, *Z and *D in witness languages, with criterial reflexes in bold

PAn	*z	*d	*Z	*D
Puyuma	**d**	**d**	d̦/z	d̦/z
Proto Rukaic	***d**	*D	*D	*D
Kanakanavu	c-c-	**s-** (?)	c-c-c	c-**s-**
Paiwan	dj	dj	**z**	**d̦**
Thao	s	s	s	**t**
Favorlang-Babuza	**t**	**t**	d, r	d, r
Hoanya	**dz**	d	d	r- (?)
Siraya	**d, l**	s	s	s
Malay	**j**	d	d-d-t	d
Ngaju Dayak	**j**	d- -t	d-d-t	d
Manam	**s, -r-, z**	d, r	d, r	…
Roviana	**s, h, z**	d, r	d, r	…

[10] The distinctions *z vs *Z and *z vs *D are omitted. They are redundant, as every language in Table 2 has a 'no' in one of the first two columns.

2.2 PAn *z

The reconstruction of PAn *z is uncontroversial (Blust 1999:36–37, 2009:571–572), although *d and *z have merged in a majority of Formosan languages. Because of this merger, however, reflexes of *z need to be identified in order to distinguish them from *d (section 2.3). The criterial reflexes of PAn *z (i.e., contrasting with *d, *Z, and *D) are Hoanya *dz*, Proto Rukaic *d, Siraya *d* or *l*,[11] and reflexes in many Malayo-Polynesian languages. Reflexes in those Malayo-Polynesian languages which are represented in examples below are shown in the lower part of Table 3. Blust uses examples from Babuza rather than Hoanya, but Babuza *t* reflects both *z and *d and is therefore not a distinctive witness of *z.

Etyma illustrating these reflexes of PAn *z are listed in (1). The number of examples is few because a reflex is required from at least one Formosan language (in order to guarantee that the etymon is of PAn ancestry) and because two of the Formosan languages with criterial reflexes (Hoanya and Siraya) are extinct and data sources are limited. If one wished to exemplify etyma containing PAn *z with criterial reflexes only in Malayo-Polynesian languages and non-criterial reflexes in Formosan, then the list would be considerably longer.[12]

(1) a. *zalan ‘road’ > **PRuk** *dalane **Hoa**[13] *dzalan* **Sir** *daraŋ* **Mal** *jalan* [also **Puy** *daLan* **Pai** *djalan* **Kav** *razan* (metathesis)]

b. *zaLiH ‘near’ > **Maga Rukai** *me-d-dali* [also **Kan** *ara-cani* **Saa** *ma-saɬi* **Pai** *djaɬi* ‘soon’, **Sed** *dalix* **Ilk** *dan-dani*]

c. *zaRami[an] ‘straw’ > **Puy** *daramian* **Pai** *djamia* ‘rice straw’ **Mal** *jerami* [**Kav** *rami* ‘rice straw’]

d. *zaRum ‘needle’ > **Sir** *lakim* (for †*laxim* **Mal** *jarum* [also **Puy** *daʔum* **Pai** *djaum*]

e. *zawaC ‘walk’ > **PRuk** *davace **Hoa** *dz‹am›awat* [also **Pai** *djavats* **Ami** *lafac*, *rafac* ‘roam’]

f. *zawiL ‘far’ > **Puy** *dawil* **PRuk** *daili [also **Tso** *covhi* **Tao** *tavit* **Kav** *laul* (assimilation)]

g. *zuluʔ ‘earlier, in advance’ > **Pai** *pa-djulu* ‘prepare something in advance’ **Sasak** *juluʔ* ‘do one thing before another’

---

[11] Variation between the Siraya reflexes appears to be free, and one cannot tell whether it is dialectal, idiolectal, or transcriptional.

[12] I have generally stuck to the orthographies of my sources, with three exceptions: (i) Siraya forms are given in the orthography of Adelaar (1999); (ii) *ə* is replaced with *e* in all languages except Tsou (Tsou is the one language where *ə* and *e* represent different phonemes); and (iii) *o* is replaced with *u* in languages where these vowels are not in contrast.

[13] Abbreviations of languages names used in the examples are the first three letters of the language name as listed in Tables 1 and 3, with the addition of **Bab** Babuza, **Bas** Basay, **Fav** Favorlang, **ND** Ngaju Dayak, **PAta** Proto Atayalic, **PRuk** Proto Rukaic, **PRuTs** Proto Rukai-Tsouic, **Tro** Trobiawan. The symbol † is used to mark forms that are expected but do not in fact occur.

h. *quzaL ‘rain’ > **PRuk** *udale **Hoa** *m-udzas* **Sir** *udal* **Mal** *hujan* [also **Puy** *ʔudal* **Kan** *ʔucáne* **Saa** *usałe* **Pai** *qudjał* **Bun** *qudan* **Tha** *qusað* **Paz** *ʔudal* **Sai** *ʔæ-ʔœral* **Ami** *ʔurad* **Kav** *uzan*]

2.3 PAn *d

In Ross (1992:40–45) I claimed that the only criterial reflexes of PAn *d (i.e., contrasting with *Z and *D) were Paiwan *dj* and Puyuma *d*. I did not take account of data from extinct Formosan languages and missed the fact that its Favorlang-Babuza reflex, *t*, is also criterial.

In (2) are listed etyma containing PAn *d. Since my claim is that the criterial reflexes of *d are correlated across the three languages and are not generally the outcome of borrowing or of random change, only etyma reflected in at least two of the languages are listed, Paiwan and Puyuma in (2a), Favorlang-Babuza and Paiwan in (2b), and Favorlang-Babuza and Puyuma in (2c). In order to show that the etyma in (2) contain *d, not *z, reflexes from languages which are criterial for *z (section 2.2) are added in square brackets.

Very few cases of morpheme-medial or -final PAn *d occur in the data, perhaps because of the putative complementary relationship between PAn *d- and PAn *-r-/*-r mentioned in section 2.1.

(2) a. i. *danaw ‘lake’ > **Puy** *danaw* **Pai** *djanaw* [**ND** *danaw*]
ii. *dapaL ‘sole’ > **Puy** *dapal* **Saa** *sapałe* **Pai** *djapał* ‘hind legs’ [**PRuk** *Dapale **Sir** *sapal*]
iii. *daRaq ‘blood’ > **Puy** *daraʔ* **Saa** *caraʔə* **Pai** *djaq* **Bab** *tagga* [**Hoa** *ddoi* **Mal** *darah*]
iv. *(dz)away ‘create’ > **Puy** *d‹in›away* ‘creation’ **Pai** *ma-djavay* ‘fruitful’
v. *qañud ‘drift’ > **Puy** *laʔud* (metathesis < *ʔalud) **Pai** *qałudj* [**PRuk** *aluDu]

b. i. *daqiS ‘forehead’ > **Pai** *djaqis* **Bab** *tees* [**Kav** *zais* **Mal** *dahi*]
ii. *dilaq ‘lick, kiss’ > **Saa** *s‹um›a-silae* **Pai** *dj‹m›ilaq* **Bab** *ta-tsira* (*ts* = affrication before /i/) [**Hoa** *da-lida* (metathesis) **PRuk** *Dila]
iii. *duRi ‘thorn’ > **Pai** *djui* **Bun** *duli* **Bab** *toi* [**Mal** *duri*]

c. *daReq ‘soil, clay’ > **Puy** *dareʔ* **Saa** *sare:* **Fav** *ta* [**PRuk** *daʔe (*d- for **†***D-) **Hoa** *rubrul-daxu* **Roviana** *raro* ‘clay pot’]

Tsuchida (1976:153–154, 181–182) divides PAn *d into his *D_3 and his *d. The criterial reflex of his *d is Saaroa *c*, in Saaroa *caraʔə* < PAn *daRaq ‘blood’ above, whereas his *D_3 is reflected as Saaroa *s*, as in Saaroa s: Saaroa *sapałe* < PAn *dapaL ‘sole’, Saaroa *s‹um›a-silae* < PAn *dilaq ‘lick, kiss’, Saaroa *sare:* < PAn *daReq ‘soil clay’. Another instance of Saaroa *c* reflects PAn *Z: Saaroa *tala-la-laucu* ‘look down’ < PAn *lahuZ ‘west, seawards’. In general, Saaroa reflects PAn *z, *d, *Z and *D indifferently as *s*.

2.4 PAn *Z

Just one criterial reflex reflects *only* PAn *Z: this is Paiwan *z*. However, my claim in Ross (1992) was that this corresponds to Puyuma *ḏ* or *z* (but not *d*). It also corresponds to Favorlang-Babuza *d* or *r* (but not *t*).

The variation between Puyuma *ḏ* and *z* is dialectal (both are retroflex). The variation between *d* and *r* in the Favorlang-Babuza data is free, and we cannot tell whether it is dialectal, idiolectal, or transcriptional.

Etyma reflected in two or more of these languages are listed in (3): Paiwan and Puyuma in (3a), Favorlang-Babuza and Paiwan in (3b), and Favorlang-Babuza and Puyuma in (3c).

(3) a. i. *ZaLum ‘water’ > **PPuy** *ɖanum **Pai** *załum* **Bab** *dalom* **Kav** *zanum* [also **Hoa** *ladum* (metathesis) **Sir** *salum* **ND** *danum*]
ii. *Zaya ‘east inland’ > **Puy** *ḏaya* ‘west’ **Pai** *i-zaya* [also **PRuk** *Daða **Sir** *taxa-seya* **Kav** *zaya* ‘west’ **Mal** *barat-daya* ‘southeast’]
iii. *Zekel ‘stuck in throat’ > **Puy** *ḏekel* **Pai** *zekel*
iv. *ZemZem ‘dark’ > **Puy** *ḏeme-kerem* ‘dawn twilight’ **Pai** *zemzem* [also **Hoa** *ma-dum* **Sir** *ma-simdim* **Kav** *mRi-zemzem* ‘be dim’ **ND** *daræm*]
v. *ZeRuŋ ‘distant thunder’ > **Puy** *ḏeruŋ* **Pai** *zuŋ* (also *deruŋ* apparently borrowed from Puyuma) [**Kav** *zeŋzeŋ* ‘thunder’, cf **PRuk** *Dereder **Sir** *siŋdiŋ* < **PAn** *ZeRZeR ‘thunder’]
vi. *Zuma ‘other’ > **Puy** *ḏuma* **Pai** *zuma* **Fav** *roman* [also **PRuk** *Duma **Kav** *zuma* **Ilokano** *ag-duma* ‘differ’]
vii. *quZaŋ ‘shrimp, lobster’ > **Katripul Puyuma** *qezaŋ* ‘shrimp’ **Pai** *quzaŋ* **Mal** *hudaŋ*
viii. *quZem ‘black’ > **Puy** *uḏe-uḏem* **Pai** *quze-quzem* ‘dull’ **Fav** *ma-udum* [also **Mal** (*h*)*udam* ‘dulled (of lustre); faded (of coloring)’]
ix. *CukuZ ‘cane prop, walking stick’ > **Puy** *sa-rekuḏ-an* (ST) **Pai** *tukuz-an* [also **PRuk** *ukuDu (for **†***cukuDu) **Sir** *hukas* **Ami** *cukur*]
x. *lahuZ ‘west’ > **Puy** *ļauḏ* ‘east’ **Pai** *łauz* [also **PRuk** *LauDu ‘below’ **Sir** *r‹m›aus* **Mal** *laut* ‘sea’]
xi. *layaZ ‘*Sambucus formosana*’ > **Puy** *layaḏ* **Pai** *łayaz* [also **PRuk** *laðaD **Kav** *layas*]
xii. *likuZ ‘back’ > **Puy** *ļikuḏ-an* **Pai** *likuz* [also **Sir** *rikus* **Kav** *ku-rikuz* ‘follow’ **ND** *ba-rikor* ‘on one’s back’]
xiii. *SateZ ‘deliver’ > **Puy** *ateḏ* ‘send, deliver’ **Pai** *satjez* ‘return s.t., send’ [also **Mal** *hantar* ‘deliver’]
xiv. *timiZ ‘chin’ > **Puy** *timiḏ* **Pai** *tjimiz* [also **Ruk** *timiļ* (Ogawa)][14]
b. i. *ZaZaL ‘old (things)’ > **Pai** *zazał-an* ‘last year’s harvest’ **Fav** *raras* **PAta** *dadal* **Kav** *zan*
ii. *quZip ‘alive’ > **Pai** *q‹m›uzi-quzip* **Fav** *orich* [also **PRuk** *m-uDipi]
c. i. *qiZus ‘spoon’ > **Puy** *ʔiḏus* **Fav** *idod*
ii. *qalaZ ‘fence’ > **Puy** *ʔaļaḏ* **Bab** *arrar*

[14] Ogawa transcribes a word-final retroflex stop as a lateral.

2.5 PAn *D

In Ross (1992) I claimed that PAn *D is reflected by Paiwan *ḏ* (which reflects *only* PAn *D) and Puyuma *ḏ* or *z* (which also reflects PAn *Z). It is also reflected by Favorlang-Babuza *d* or *r* (which, like the Puyuma reflex, also reflects PAn *Z). Morpheme-medially it is reflected in (4b) by Kanakanavu *-s-* (not *-c-*, which reflects *z and *Z), but it is possible that this is a case of assimilation, i.e., *ʔusase* from earlier *ʔucase. With the publication of Blust (2003) it appears that Thao may also have a separate reflex of PAn *D, namely *t* in (4a-ii) and (4b), which also reflects *t and *s but does not reflect *z, *d or *D (they are reflected by Thao *s*). This makes initial *sh-* of Thao *shakish* in (4a-iv) and initial *d-* of *duku-duku* in (4a-v) irregular, the former perhaps through assimilation and the latter probably through borrowing from Bunun.[15]

Of the reconstructed protophonemes proposed here, PAn *D is the least well attested, as (4) shows, and the most problematic. It is a problem because it is most widely attested in *DuSa 'two', which displays irregularities in several Formosan reflexes. However, it is also attested in a few other etyma, and I think it possible that it was a distinct PAn phoneme, albeit one with a low functional load.

Etyma reflected in two or more of the criterial languages are listed in (4), Paiwan and Puyuma in (4a) and Favorlang-Babuza and Paiwan in (4b).

(4) a. i. *Deme[lR] 'thick (as of a board)' > **Puy** *ke-zemer* **Pai** *ke-ḏemel* [also **PRuk** *Demele **Kan** *maki-cemere*]

ii. *DuSa 'two' > **Puy** *ḏua* **Pai** *ḏusa* **Tha** *tuʃa* **Bab** *na-doa, na-roa*[16] [also **Hoa** *rusa*[17] **PRuk** *Dusa **Sir** *sa-sua* **Kav** *zusa* **Mal** *dua*]

iii. *paŋuDaL 'pandanus' > **Puy** *paŋuḏal* 'pineapple' (ST)[18] **Pai** *paŋuḏal* 'pineapple' [also **PRuk** *paŋuDale 'pineapple' **Mal** *pandan*]

iv. *Dakes 'camphor laurel, *Cinnamomum camphora*' > **Puy** *ḏakes* **Pai** *ðakes* **Tha** *ʃakiʃ* [also **PRuk** *Dakese **Tso** *cʔosə* **Kan** *cakése* **Bun** *dakus* **Paz** *dakes* **Sai** *rakeʃ* **PAta** *dakus **Ami** *rakes* **Kav** *zaqes*]

v. *Dukuŋ 'bent' > **Pai** *ḏukuŋ* **Tha** *duku-duku* (for †**tuku*; Bunun loan?) [**ND** *roko* 'be in a bowing position' **Cebuano** *lukuŋ* 'make a coil, form a circle from something stiff' **Javanese** *ruŋkuŋ* 'walk with a bent back']

b. *quDaS 'grey hair' > **Kan** *ʔusase* **Pai** *quḏas* **Tha** *qutaʃ* **Fav** *uras*

A further etymon which possibly contained PAn *D is *DaqaLi[an] 'daylight' in (5). The Paiwan reflex with the root *raqałi* is reported by Ferrell (1982) to be from a western dialect of Paiwan. It is not clear in terms of Ho's (1978) comparison of Paiwan dialects which dialect this might be, but if it is the one that he labels simply 'Paiwan', then *r* reflects both *r and *D. However, the Thao reflex is *saqazi*, seemingly reflecting either

---

15 Thao has numerous Bunun loans (Blust 1996).

16 PAn *S is irregularly lost in Favorlang-Babuza, as it is also in closely related Taokas *gua*.

17 This is the only Hoanya reflex of PAn *D. Initial *r-* here may represent the same phoneme as *d* reflecting *d and *Z.

18 Cauquelin (1991) records Puyuma *paŋudal* 'pineapple', but this appears to be an error.

*d or *Z but not *D—unless my inference that PAn *D is reflected by Thao *t* is incorrect.[19]

(5) *DaqaLi[an] (?) 'daylight' > **Tha** *saqaði* **Bab** *dalen-, lalian* (assimilation) **Pai** (western dialect) *r‹m›aqaɬi-aɬi-an* 'mid-morning' [also **Tso** *cohzona* **Paz** *dalian* **Kav** *llan* (assimilation) **Chamorro** *haʔani* **Roviana** *rane*]

2.6 Irregular correspondence sets

There are, of course, other counter-examples besides the irregularities noted in the cognate sets above. The important question is whether these are so many as to support Blust's (1999) borrowing hypothesis or so few as to favour my hypothesis of inherited phonemic distinctions. There are in fact remarkably few of them. They are given below, with comments following each example.

(6) a. *dekeC 'adhere' > **Puy** *ɖekeʈ* 'paste' (ST; for †**dekeʈ*) **Pai** *se-djekets* **Tagalog** *dikit* 'joined'
b. *(tC)uduq 'point, instruct' > **Puy** *tuʐuʔ* (Wolff 1997:591) **Pai** *tsudjuq* 'finger' **Bun** *tuduq* **Paz** *tudu* 'educate, forefinger' **Ami** *turuʔ* **Javanese** *tuduh*

In both examples, Paiwan *dj* corresponds with Puyuma *ɖ/z* where the Paiwan form leads us to expect Puyuma *d*. In (6a) Puyuma initial *ɖ-* perhaps reflects assimilation of point of articulation to final *-ʈ*.

(7) *(dZ)aqu 'soapberry: *Sapindus mukrossi*' > **Puy** *daʔu* (for †**ɖaʔu*) **Ruk** *daw* (Wolff 1997, dialect not identified) **Pai** *zaqu* **Ami** *rauʔ* **Mal** *rau* 'tree sp.: *Dracontomelum edule*'

Paiwan *z-* reflects *Z and Puyuma *d-* reflects *D. Presumably one is borrowed.

(8) PAn *SiZi 'Formosan wild goat' > **Puy** *siri* (for †**iɖi*) **PRuk** *ki-sisi (for †**ki-siDi) **Tso** *sizi* **Pai** *sizi* **Tha** *sisi* (sibilant assimilation, for †*ʃisi*; (Blust 1995) **Sai** *ɟiriʔ* **Ami** *siri* **Bun** *sidiʔ* **Kav** *sizi*

I assume that the Proto Rukaic and Puyuma forms are borrowed (Puyuma from Amis?).

(9) *ŋalay 'spittle, saliva' > **Puy** *ŋaɭay* **PRuk** *so-ŋaLay 'spit' **Tso** *ŋroi* **Kan** *ŋaai* **Saa** *ŋaliʔi* **Pai** *ŋadjay* (for †**ŋalay*) **Kav** *ŋaŋay* (assimilation < earlier **ŋanay*)

This item includes *l, and not *d, *Z and *D, in its reconstructed form. It is included because it is correctly mentioned by Blust (1999:51) as an example of borrowing: when the range of data is considered, it is indeed clear that the Paiwan form is the odd man out and is a borrowing.

(10) *ke(zZ)em 'close the eyes' > **Pai** *tji-kezem* 'have the eyes closed' **Mal** *kejam* 'deliberately close the eye'

Paiwan *-z-* points to *-Z-, Malay *-j-* to *-z-. It is possible that Paiwan *kezem* is a borrowing from Puyuma, but the relevant Puyuma term is not recorded in our sources.

---

[19] Blust (2009:274) believes that Thao *tusha* is an irregular reflex (for expected **shusha*), derived by assimilating the initial to that of 'one' and 'three' when counting: *tata, tusha, turú*.

2.7 Further etyma

There are a number of other PAn etyma in which *d, *Z, and *D can be partially or fully disambiguated because they contain a criterial witness. They are included here for the sake of completeness. They are not included above as they do not have two correlated criterial reflexes.

(11) ***d**

a. *dakep ‘seize, catch’ > **Pai** *djekep* **Tha** *sakup* **Sai** *rakep* **Mal** *dekap* ‘embrace’
b. *dakuC ‘grasp, grab’> **Pai** *djakuts* **Tag** *dakot* ‘a handful, a graspful’ **Chamorro** *hakot* ‘snatch, grab’ (*-k-* for †*-h-*; Blust & Trussel 2010)
c. *demdem ‘thoughts’ **Puy** *demdem* ‘heart, feelings’ **PRuk** *Deme-Deme ‘heart, mind’ **Ami** *demdem* ‘endure (illness, feelings)’ **Mal** *dendam* ‘yearning, longing’
d. *disdis > **Tamalakaw Puyuma** *disdis* ‘cut many times, giving slight wounds’ **Mal** *didis* ‘cut into thin slices’
e. *ludaq ‘spittle’ > **Pai** *ludjaq* ‘betelnut spittle’ **Mal** *ludah*
f. *qapid ‘braid (hair)’ > **Pai** *q‹em›apidj* ‘braid hair’ (Egli 2002:232),[20] **Bontok** *apid* ‘plait, braid (as the hair)’

(12) ***z** or ***d**

*(zd)akiS ‘climb’ > **Fav** *t‹umm›aies* **Paz** *dakis* **Kav** *zaqis*

(13) ***Z**

a. *ZamaR ‘torch, light’ > **PRuk** *Damare ‘moon’ **Pai** *zama-n*[21] **Ami** *lamal* **Kav** *zamaR* ‘fire’ **Mal** *damar* ‘resin’
b. *ZaRiŋ ‘groan, moan’ > **Tso** *troe-creŋi* **Saa** *tara-a-sariŋi* **Pai** *z‹m›aiŋ* **Bun** *da-daliŋ* **Ami** *laliŋ*
c. *Zeles ‘bowstring’ > **Maga Rukai** *ḍrese* **Kan** *cee-* **Saa** *e-selee* **Pai** *zelet*
d. *ZiRi ‘stand’ > **PRuk** *ʔi-diʔi (*d-* for †**D-*) **Tso** *ya-cʔə* (Szakos n.d.) **Pai** *mi-zi* ‘standing’ **Tha** *mi-łiłi* (sibilant assimilation, for †**-siłi*; (Blust 1995) **Sai** *mi-riḻiʔ* **Ami** *sa-lili* ‘housepost’ **Mal** *diri*
e. *ZaŋZaŋ ‘roast’ > **PRuk** *Daŋe **Tso** *t‹m›a-cŋə-cŋə* **Kan** *c‹um›a-caŋe-caŋe* **Saa** *s‹um›a-saasaŋe* **Pai** *zaŋzaŋ* **Tha** *sansan* **Paz** *dadaŋ* **Ami** *raŋraŋ* ‘hold near fire to dry’ **Kav** *zazaŋ* ‘dry by fire’ **Mal** *dandaŋ* ‘cooking pot’
f. *kiZiŋ ‘spoon’ > **PRuk** *kiDiŋi **Saa** *ta-isiŋi* **Pai** *kiziŋ* **Ami** *k-al-iliŋ*
g. *seZu ‘hiccup’ > **Pai** *tezu* **Mal** *sedu, sendu*
h. *tuZuR ‘sleep’ > **Saa** *ma-ta-tusoro* (Ferrell 1969) **Fav** *torro* **Hoa** *ma-tulu* **Bun** *ma-tudul* **Mal** *tidur*
i. *tuZuq ‘leak, drip’ > **PRuk** *tuDu **Kan** *tucuʔu* **Saa** *tusu-a* **Pai** *tjuzuq* **Bun** *tuduq-an* **Tha** *tusuq* **Paz** tuqut (metathesis) **PAta** *tudiq **Ami** *tuluq*
j. *CuZud ‘cane prop’ > **Fav** *ta-tear*
k. *tuqeZ ‘tree stump’ > **Kan** *t‹an›eʔece* **Saa** *t‹ał›ese* **Pai** *tjuqez* **Tag** *tuʔod*

---

[20] Ferrell (1982) has *qapiz* ‘something braided in six strands’, reflecting putative *qapiZ.

[21] Dahl (1973:58) adds **Sir** *ma-dama* ‘tomorrow’ and Wolff (1997:593) **Pai** *djama* ‘morning’ as reflexes, but they both reflect PAn *zaman ‘morning, tomorrow, daylight’ (cf **Pap** *daman-an* **Puy** *an-daman* ‘tomorrow’ **Manam** *zama* ‘tomorrow’ **Bam** *jam* ‘daytime, tomorrow’).

l. *qeluZ ‘main housepost’ > **PRuk** *uluD **Kan** *ʔuucu* **Saa** *ʔulusu* **Pai** *qeluz* **Bun** *hauʔ* **Tha** *krus* **Paz** *urut* **PAta** *qelud

(14) ***Z** or ***D**

a. *(ZD)aRam ‘accustomed to’ > **Fav** *ma-darram* **Paz** *daxam*

b. *dekiŋ ‘to bark, of a deer’ > **Tamalakaw Puyuma** *d‹em›a-dekiŋ* ‘(of deer) to cry “eking, eking”’ **Mal** *deŋkiŋ* ‘sharp, high-pitched bark (of dog or barking-deer); yelp’

c. *a(ZD)ak ‘go up’ > **Tso** *ocʔo* **Saa** *m-a-asake* **Fav** *dadach* (initial *d-* unexplained) [**Kap** *alak* ‘raise price’]

d. *qi(ZD)aS ‘moon’ > **Bun** *idas* **Bab** *idass* **Sai** *iras* (Ogawa) **PAta** *idas [also **Puy** *ʔilas* (borrowed form: *S is lost in inherited items in Puyuma) **Pai** *qiłas* **Paz** *ilas*]

e. *kuku(ZD) ‘leg’ > **Puy** *kukuɻ* **PRuk** *kuku **Kan** *ke:ke* **Ilokano** *kukod* ‘shank, shin of animals’

(15) ***D**

a. *DiSi ‘wipe’ > **PRuk** *Disi-Disi **Tso** *ma-hisi* **Pai** *ɻisi-ɻis* ‘rub’ **Paz** *mu-didis*

b. *puDeR ‘kidneys’ > **PRuk** **peDeʔe* **Tso** *pcərə* **Saa** *pesere* **Pai** *puɻu*

c. *kuDkuD ‘hoe, dig’ > **Pai** *kuɻkuɻ* ‘small hoe for weeding’, **Ami** *kurkur* ‘dig soil’, **ND** *kukur* ‘rasp’

2.8 Borrowing or inheritance?

Blust (1999:50–51) suggests that Paiwan *dj*, *z* and *ɻ* and Puyuma *d* and *ɻ/z* may be accounted for by borrowing of reflexes of PAn *d and *z between the two languages and from Rukai (and perhaps elsewhere). He counts the reflexes listed by Dahl (1973) and Ross (1992) and arrives at the totals in Table 4. He hypothesises that

(16) a. the inherited reflexes of *d are Paiwan *dj* and Puyuma *ɻ/z*

b. Paiwan *z* (and possibly *ɻ*) reflects borrowings with Puyuma *ɻ/z*

c. Puyuma *d* reflects borrowings with Paiwan *dj*

It is easy to see how (16b) and (16c) might be claimed to account for the correspondences between Paiwan *z* and Puyuma *ɻ/z* and between Paiwan *dj* and Puyuma *d* respectively, but hard to see how in (16a) inheritance can be said to account for a barely occurring correspondence, Paiwan *dj* and Puyuma *ɻ/z*—a problem which Blust mentions (1999:51).

A count of my own examples together with their locations in this paper is also shown in Table 4. This count only includes examples with both Paiwan and Puyuma reflexes. It is clear by inspection that the 22 examples supporting the proposed correspondences far outweigh the four ‘irregular’ examples. This observation is confirmed by the Fisher’s exact statistic, which shows that there is a probability of 0.8% that the distribution of examples shown in Table 4 arose by chance. The distribution is thus significant, and the hypothesis that most effectively accounts for it is one based on shared inheritance. Blust’s sample was only half this size, and its distribution, as he surmised, was not significant. The details are set out in Appendix 2.

Table 4: Numbers of reflexes of *d, *Z and *D in Paiwan and Puyuma

PAn	Puyuma	Paiwan	Blust's count	my count	examples
*d	d	dj	2	5	(2a)
*Z	ɖ/z	z	6	14	(3a)
*D	ɖ/z	ɖ	2	3	(4a)
—	ɖ/z	dj	1	2	(6)
—	d	z	2	1	(7)
—	r	z	—	1	(8)
Totals			13	26	

If the distribution of correspondences does indeed reflect shared inheritance, then there are two possible classes of hypothesis. To the first class belong hypotheses to the effect that PAn indeed had just the voiced apical obstruent *d, which underwent an unconditioned split or splits into *d, *Z and *D in a language or languages ancestral to Favorlang-Babuza, Paiwan and Puyuma (this could have been a two-stage event: first a split into *d and *Z/*D, as reflected in Favorlang-Babuza and Puyuma, then a split of the *Z/*D protophoneme into Paiwan *Z and *D). The second class consists of the hypothesis that *d, *Z, and *D are indeed of PAn antiquity, but have merged in most daughter languages.

Hypotheses of the first class would place Favorlang-Babuza, Paiwan, and Puyuma in a subgroup. This would be problematic, as Tsuchida (1982:9–11) shows that Favorlang-Babuza, Taokas, Papora and Hoanya probably formed a subgroup, which Blust (1999) labels 'Central Western Plains' and which, to my knowledge, no investigator has ever questioned. Furthermore, despite my temerity in suggesting otherwise (Ross 1992:43), there is no evidence for anything other than a contact relationship between Paiwan and Puyuma, which are strikingly different in historical phonology, morphology, and lexicon.

The difficulty with the hypothesis that *d and *Z are of PAn antiquity is that it implies that they have merged in most daughter languages, and, unless we wish to claim that all these languages are a subgroup, we must infer that the mergers have occurred several times over. This is in fact not difficult to believe. Proponents of the Malayo-Polynesian hypothesis (Blust 1977, 1999) claim that *t and *C merged several times over, in PMP and in Siraya, Amis, Bunun, Kavalan, and Ketagalan. A similar claim is made with regard to *z and *d (in its conventional embodiment) in Saisiyat, Pazih, Atayalic, Thao, Rukaic, Tsouic, Paiwan, Puyuma, Amis, Bunun, and Kavalan. Against this background, it is reasonable to infer that *d and *Z, the putative voiced equivalents of *t and *C, have merged in Saisiyat, Pazih, Atayalic, Hoanya, Thao, Rukaic, Tsou, Siraya, Amis, Bunun, Kavalan, and PMP.

*D was excluded from the discussion in the previous paragraph, as its status is obviously more doubtful. It is represented by only the three Paiwan-Puyuma examples in (4a) and the Favorlang-Thao-Kanakanavu-Paiwan example in (ab). There are, on the other hand, no cases of Paiwan *ɖ* with a Puyuma correspondent other than *ɖ*. Could these

four examples and the three listed under *D in section 2.7, and indeed all other etyma with *ḍ* listed by Ferrell (1982) and Egli (2002), be due to borrowing? The question is a difficult one, and requires much larger lexical databases in Formosan languages other than Paiwan for its solution. Thus Paiwan *ke-ḍemel* (4a-i) may well be a borrowing from Puyuma *ke-zemer*, the more so as the latter reflects putative PAn *ZemeR, and inherited Paiwan terms reflect *R as zero. Paiwan *paŋuḍal* 'pineapple' in (4a-iii) may be a borrowing of the Puyuma word with the same form, the more so as terms for plants are often borrowed Wolff (1994). On the other hand, there is no evidence that Paiwan numerals are borrowed, and Paiwan *ḍusa* 'two' in (4a-ii) does not reflect a borrowing of Puyuma *ḍua*. For the time being, the safer assumption is that *D did occur in PAn, and that the merger of *Z and *D has apparently occurred everywhere except in Paiwan and perhaps Thao and Kanakanavu. However, I do not think this assumption is safe enough to use in subgrouping hypotheses.

There is, on the other hand, good reason to think that Paiwan items containing reflexes of *d and *Z are generally inherited rather than borrowed. These words display reflexes of PAn protophonemes that are Paiwan-specific. The most salient is the loss of PAn *R, illustrated in (17).

(17) a. *djaq* < **PAn** *daRaq 'blood' (2a-iii)
b. *djui* < **PAn** *duRi 'thorn' (2b-iii)
c. *z‹m›aiŋ* < **PAn** *ZaRiŋ 'groan, moan' (13b)
d. *zama-n* < **PAn** *ZamaR 'torch, light' (13a)
e. *zuŋ* < **PAn** *ZeRuŋ 'distant thunder' (3a-v)
f. *mi-zi* 'standing' < **PAn** *ZiRi 'stand' (13d)

Note that *zuŋ* coexists with another Paiwan term, *deruŋ*, evidently borrowed from Puyuma *ḍeruŋ*.

Conversely, as noted in the case of Paiwan ḍusa 'two' above, Paiwan words containing reflexes of *d, *Z and putative *D regularly reflect PAn *S as Paiwan *s*, whereas Puyuma has lost it: see also (2b-i), (3a-xii), (4b), (8), and (15a).

3 Putative sibilants

Blust (2009:547) and others recognise two PAn phonemes which are written as sibilants: *s and *S. Where they have *s, however, Tsuchida (1976:127–131) follows Dyen (1971) in reconstructing two phonemes, in Tsuchida's notation *s and *θ.[22] Also following Dyen (1965:299–302), Tsuchida (1976:159–162, 234–236, 250) recognises six subscripted variants of *S, as well as a further variant that he labels *X. Two of these, his *S_1 and *S_2, represent significantly different correspondence sets and are candidates for recognition as separate PAn phonemes, a possibility recognised by Li (1985), who reconstructs four PAn sibilants, *t' (Tsuchida's *s), *θ (following Tsuchida), *s

[22] Actually, Dyen and Tsuchida reconstruct a third phoneme—or a variant of *s—to account for the reflexes of the word for 'nine', but Blust (1995) has shown that this is explained if one reconstructs PAn *Siwa but PMP *siwa, an idiosyncratic change for which he provides a motivation.

(Tsuchida's *S_1) and *ʃ (Tsuchida's *S_2).[23] To avoid subscripts, from this point onward *S will be used for *S_1, in line with Blust's usage, and—for want of a better symbol—*x for *S_2. Neither of these putative contrasts has received critical attention since Tsuchida drew attention to the relevant sound correspondences nearly 40 years ago.

Table 5 shows the Formosan reflexes of hypothetical PAn *θ, *s, *S and *x. Criterial reflexes of *θ and *x, i.e. reflexes distinct from those of *s and *S respectively, are shown in bold. Both *θ and *x are reflected in relatively few cognate sets, but Rukaic *θ, the most distinctive criterial reflex of putative PAn *θ, occurs in many more Rukai words than those Tsuchida lists. I return to this in section 3.1.

The crucial question, of course, is whether PAn *θ and *x existed or whether their reflexes can be accounted for in some other way. These are the questions addressed in section 3.1 and section 3.2. The reconstruction of PAn *s and *S stands on a very sure footing and will not be further discussed.

A glance at Table 5 tells us (i) that PAn *S was a sibilant, and so was putative *x, if it occurred, and (ii) that PAn *s may or may not have been a sibilant, but became one at some point in the prehistories of several Formosan languages and PMP; the same is true of *θ, if it occurred. All four columns in Table 5 display outcomes of a sequence of lenitions, namely debuccalisation, [s] > [h], and elision, [h] > Ø, which occurs with monotonous regularity in the world's languages (Ferguson 1990).

Table 5: Reflexes of PAn sibilants in Formosan languages and PMP

PAn (Tsuchida 1976)	*θ	*s	*S_1	*S_2
PAn (Li 1985)	*θ	*t'	*s	*ʃ
PAn (this paper)	*θ	*s	*S	*x
Puyuma	s	s	Ø	Ø
Proto Rukaic	***θ**-**θ**-s	*s	*s	***Ø**
Tsou	s	s	s	**Ø**
Kanakanavu	**s**-**s**-s	Ø-Ø-s/Ø	s-s/Ø-s	Ø
Saaroa	**s**	Ø	s/Ø-Ø-s/Ø	Ø
Paiwan	t	t	s	s
Bunun	c [ts]	c [ts]	s	s-s/Ø-s
Thao	t	t	ʃ (*sh*)	Ø
Taokas	t	t	ʃ (*sh*)	Ø
Favorlang-Babuza	t	t	s	Ø
Papora	t	t	s	Ø
Hoanya	t	t	s-s-t	Ø
Pazih	z-z-t	z-z-t	s	**h**
Saisiyat	h	h	ʃ (*sh*)	**h**
Proto Atayalic	*h	*h	*s	***h**
Siraya	-s-k	-s-k	x-Ø-x/ŋ/Ø	Ø
Amis	c [ts]	c [ts]	s	s
Kavalan	s	s	s	s
Basay	ts-**s**-ts	ts-ts-	s	**ts**-**ts**-
Trobiawan	ts-**s**-ts	ts-ts-	-s-s	-**ts**-
PMP	*s	*s	*h	*h

[23] The variants of *S and *X are also discussed by Dahl (1973:32–35), but he draws no conclusions about them.

3.1 PAn *x

The reconstruction of PAn *S is uncontroversial, and supporting data have been offered by Dahl (1973), Tsuchida (1976) and Wolff (1997). In (18), (19), and (20) cognate sets containing respectively initial, medial and final reflexes of PAn *x are presented. They have previously been reconstructed with *S.

(18) a. *xapuy 'fire' > **Puy** *apuy* **PRuk** *apuy **Tso** *puzu* **Kan** *apulu* **Saa** *apułu* **Pai** *sapuy* **Bun** *sapuð* **Tha** *apuy* **Tao** *yaway* **Bab** *chau* **Pap** *dapu* **Hoa** *dzapu* (Taokas, Favorlang-Babuza and Papora reflect earlier *****zapuy) **Paz** *hapuy* **Sai** *hapuy* **PAta** *p-hepuy-an* **Sir** *apuy* **Mal** *api*

b. *xema 'tongue'[24] > **Puy** *sema* (for †**ema*; borrowed from Paiwan) **Tso** *umo* (Szakos n.d.) **Pai** *sema* **Tha** *zama* (Ferrell 1969) **Paz** *za-hama* **Sai** *kæ-hmaʔ* **PAta** **hemaʔ* **Ami** *sema*

c. *xeθi > 'meat, content' **Puy** *isi* (Ferrell 1969) **Kan** *ʔa-isi* 'exist' **Saa** *maa-is-ana* 'content' **Pai** *seti* **Ami** *heci* **Bun** *cici* (Ferrell 1969) **PAta** *hei **Kav** *siʔ* **Nias** *ösi* 'contents, flesh'

d. *xuni 'chirp' > **Tso** *m-uni* **Kan** *um-a-uni* **Saa** *um-a-ani* **Tha** *ʃ‹m›a-ʃuni* (root *ʃuni* for †*tuni*; borrowed from Bunun?) **Paz** *huni-* **Sai** *ma-huniʔ* **Ami** *huni* **Kav** *suni* **Ilokano** *uni* 'sound, voice, noise, cry (of animals), song (of birds)' **Balinese** *uni* 'sound, make a sound'

e. *xuReLa 'snow' > **Puy** *urla* **Mantauran Rukai** *uʔuɭa* **Tso** *ruho* **Kan** *erena* **Saa** *uruła* **Pai** *sułа* **Tha** *ułða* (Ferrell 1969) **Fav** *oela* **Hoa** *o-hut-ta* **Paz** *hahela* **Sai** *hæ-hælæʔ* (assimilation: medial *-h-* for †*-z-*) **Cʔuliʔ Atayal** *hulaʔi:* **Sir** *uxla* **Ami** *suqla* (Ferrell 1969) **Kav** *suRna*

(19) a. *CuxuR 'thread (needle)' > **Puy** *aʈur* **Kan** *c‹um›a-cuuru* **Pai** *ts‹m›usu* **Bun** *ma-tusul* **Fav** *sa-ssies* **Paz** *siuh-* **Sai** *sœhœ-œn* **PAta** *loho **Kav** *tusuR*

b. *kaxiw 'tree, wood' > **Puy** *kawi* **Tso** *evi* (< earlier *kaivi) **Kan** *kaalu* (< earlier *kazhu) **Saa** *kiuʔu* **Pai** *kasiw* **Bun** *kawiʔ* 'wooden' **Tha** *kawi* **Tao** *yayeb* **Bab** *eeb* **Pap** *he* **Hoa** *hai* **Paz** *kahuy* **Sai** *kæhœy* **PAta** *kahuy **Sir** *kayu* **Ami** *kasui* **Mal** *kayu*

c. *Lixepis 'thin' > **Puy** *aripit* **PRuk** **ma-Du-Lipsi* **Tso** *hipsi* **Kan** *ma-nipi* **Saa** *ma-łipii* **Pai** *łusepit* **Paz** *lipit* **Sai** *lihpih-an* **Ami** *ki-hepic* **Bun** *ma-nisbis* **Kav** *inpis* (metathesis, for †*nipis*) **Ilokano** *iŋpis* 'thin, tenuous' (metathesis) Sundanese *ipis* 'thin, fine (hair')

d. *quxuŋ 'mushroom' > **PRuk** *ʔuŋu **Tso** *uŋo* **Kan** *uuŋu* **Saa** *uʔuŋa* **Bun** *huuŋ* **Tha** *qun* **PAta** *qehuŋ **Ilokano** *ooŋ* 'mushroom; toadstool'

(20) a. *kuxkux 'fingernail, claw' > **Pai** *kuskus* 'scrape' **Bun** *kuskus* **Tha** *kuku* **Sai** *ka-k‹l›okœh* **PAta** *kukuh **Ami** *k‹an›us* **Kav** *q‹n›uqus* **Mal** *kuku*

b. *Cumex 'clothes louse' >**Paz** *sumah* **Sai** *sumæh* **Bun** *tombos* **Kav** *tumes*

Several pieces of circumstantial evidence favour the reconstruction of *x as a separate phoneme. First, the differences between the reflexes of *S and *x shown in Table 5 are sharp enough to favour the reconstruction of a separate protophoneme. Most of the items

[24] Blust & Trussel (2010) records Central Malayo-Polynesian forms alongside Formosan, but none is a certain cognate of the Formosan forms.

containing *x are well attested and the cognate sets are regular. This speaks against the alternative hypothesis that items containing *x are the result of borrowing (if they **are** the result of borrowing, it must have happened very early in Austronesian history). Supposing the four protophonemes in Table 5 did exist in PAn, it is also very clear that *x underwent these changes before the other three (and the fact that its reflex is so often zero may explain why few cognates sets that include it have been recorded). This has a consequence that shows up in (19a). Kanakanavu, Saaroa, Tsou and the Maga and Tona dialects of Rukai reflect a change whereby, in PAn items that also contain *S, initial *C- is reflected as if it were *t- (Tsuchida 1976:148–150). If PAn *CuxuR were in fact hypothetical *CuSuR, then the expected Kanakanavu reflex would not be *c‹um›a-cuuru* but †*t‹um›a-tuuru.* Evidently *x, presumably originally [s], had been deleted before this local *C- > *t- change occurred.

3.2 PAn *θ

The data reflecting putative *θ differ qualitatively from those reflecting *x in several ways. The sound correspondence shown under *θ in Table 5 differs from the correspondence reflecting *s only in its Rukai, Kanakanavu and Saaroa reflexes. According to Tsuchida (1976:127), the Maanyan and Malagasy reflexes of *θ also differ from those of *s, but Dahl (1981:81–83) indicates that this is not the case. One etymon, shown in (21), is interpreted as containing *θ on the basis of its Malagasy reflexes alone. Without these, the relevant correspondence reflects *S, as Blust & Trussel's (2010) citation of Cebuano *kikhi* (from PMP *kihkih) confirms.

(21) *kiSkiS (Tsuchida's *kiθkiθ) 'shave, scrape' > **Kan** *k‹um›a-kisíkisi* **Saa** *k‹um›i-kisikisi* **Tso** *m-ru-ʔsiʔsi* 'shave fur' **Bun** *kiskis-un* 'shave beard' **Ata** *ke-kiskis* 'a file' **Kav** *kiskis* 'shave' **Ami** *kiskis* 'scrape away surface' **Cebuano** *kikhi* [Malagasy *hihís-ana* 'be scraped', *kikís-ana* 'be scraped']

Tsuchida cites two other monosyllabic reduplicates with syllable-final *θ, shown in (22). The reconstruction of *θ rather than *S is justified by reflexes of PMP *s in Ilokano, Tagalog and Bikol.

(22) a. Tsuchida's *ŋuθ-ŋuθ 'snore' > **Kan** *puru-ŋusú-ŋusu* **Saa** *puri-ŋusu-ŋusu* [**Tso** *roŋo-sŋu-sŋu* (metathesis) **Ilokano** *ŋosóŋos* 'growl' **Bikol** *ŋosŋos* 'growl']

b. Tsuchida's *Raθ-Raθ 'grate, scratch' > **Kan** *r‹um›a-rasé-rase* 'grate' **Saa** *r‹um›a-raa-rase* 'grate' [**Ilokano** *rasrás* 'scratch, rub oneself over the clothes' **Tagalog** *gas-gas* 'scratched' **Bikol** *gásgás* 'scratch (not an itch)']

However, the meaning correspondences between the Formosan and Philippine items are only approximate and the frequency of reduplicated monosyllabic roots in Formosan and Philippine languages is so great that the alleged Formosan and Philippine cognates in (22) may be chance lookalikes. In addition, it seems likely that (22a) is historically related in some way to PAn *ŋuSuL 'nosebleed' and PAn *ŋuSuR 'nasal mucus', both with Formosan reflexes (Blust & Trussel 2010) and that the form was actually PAn *ŋuS-ŋuS 'snore'.

Unlike items containing *x, items containing *θ are often not well attested and their cognate sets display irregularities. This qualitative poverty suggests that at least some items that appear to be reconstructable with *θ have loans among their reflexes. A number of Rukai items containing *θ belong to apparent cognate sets that also include Paiwan and, except for (23a), are apparently restricted to languages around Rukai:

(23) a. **Budai Rukai** 'compensate' *θavuŋu* **Saa** *s‹um›a-savuŋu* 'compensate for, pay fine' **Puy** *pia-savuŋ* 'compensation, fine' **Pai** *tavuŋ* 'compensate' **Old Javanese** *sambuŋ* 'contribute' (Tsuchida 1976:130) (< PAn *sa(b,w)uŋ)
b. **Tanan** and **Budai Rukai** *takiθi* 'small knife' **Pai** *tjakit* 'knife for hunting or fighting'
c. **PRuk *** kalaθe 'bracelet' **Pai** *kałat* 'bracelet'
d. **Budai Rukai** *baɭiθi* 'exchange' **Pai** *v‹en›alit* 'exchange' **Ami** *mi-balic* 'exchange'
e. **Budai Rukai** *paθaɭa* 'envy, jealousy' **Pai** *patalaq* 'jealous' (Ho 1978)
f. **Budai Rukai** *ta-si-kaɭa-kaɭaθ-ane* 'wrist' **Pai** *kała-kałat-an* 'wrist' (Ho 1978)

In each item in (23) Rukai *θ* corresponds to Paiwan *t* (which usually reflects PAn *s). It is reasonable to surmise that the Rukai items in (23) are loans from a language which at the time of the loan reflected PAn *s as a dental obstruent, and Paiwan is an obvious candidate for that source. In (23f) there is further evidence of borrowing: Rukai *ɭ* reflects PAn *l, but Paiwan *ɬ* reflects PAn *L.

Li (1977) shows that both *θ and *s are reconstructable for Proto Rukai, with the Rukai dialect reflexes shown in (24):

(24)

Proto Rukaic	*θ	*s
Tanan	θ	s
Budai	θ	s
Maga	θ	s
Tona	θ	s
Mantauran	s	ʔ

However, several of the Rukaic cognate sets listed by Li show an unexpected crossover between the two correspondences, as he is careful to note. For the sake of comparison, both of Li's Proto Rukaic reconstructions are given in each case as (i) and (ii), then the PAn reconstruction and other Formosan cognates as (iii).

(25) a. i. **PRuk** *beraθe 'husked rice' > **Tanan** *beráθe* **Budai** *beraθe*
ii. **PRuk *** beʔase > **Maga** *beése* **Tona** *beʔase* **Mantauran** *veʔao*
iii. **PAn** *beRas > **Puy** *beras* **Kan** *véera* **Saa** *e-verae* **Tso** *fərsə*, **Pai** *vat* **PAta** *buwax **Sir** *pxak* **Ami** *felac* **Kav** *vRas* **Bas** *balatce* **Tro** *vulac* **PMP** *beRas
b. i. **PRuk** *θaviki > **Maga** *θveke* **Tona** *θaviki* **Mantauran** *saviki*[25]
ii. **PRuk** *sabiki 'areca nut' > **Tanan**, **Budai** *sabiki*

[25] The Mantauran form reflects PRuk *-v- as -*v*- rather than expected zero, and may be a loan.

iii. **PAn** *Sawiki > **Kan** *aviki* **Tso** *fiʔi* **Pai** *saviki* **Bun** *saviki?* **Tha** *ʃawiki* **Hoa** *abiki* **Sir** *sabiki*

c. i. **PRuk** *θenay > **Maga** *u-θnée*, **Tona** *θenáy*
ii. **PRuk** *senay 'sing' > **Tanan**, **Budai** *senáy*
iii. **PAn** *Senay > **Pai** *s‹m›enay*

Notice that in (25a) it is the Tanan and Budai reflexes that contain *θ*, whereas in (25b) and (25c) it is the Maga, Tona, and Mantauran reflexes. This underlines the fact that we are not dealing here with an extra correspondence set, but with mixtures of the sets shown in (24), i.e., with borrowing of the items with *θ*. Also diagnostic of borrowing is that in (25b-i) and (25c-i) *θ* does not reflect PAn *s but *S—and in (26a) it reflects PAn *C.

(26) a. **PRuk** *gauθu > **Maga** *góθu* **Tona** *gáuθu* **Mantauran** *hausu*
b. **PRuk** *garucu 'comb' > **Tan** *garúcu* **Bud** *gárucu*
c. **PAn** *ga(rR)uC > **Puy** *garuʈ* **Pai** *garuts*

Other possible loans are shown in (27). The probable PAn source of *-θ-* is based on the assumption that Puyuma *ʈeb-ʈeb* etc. are regular reflexes.

(27) a. **PRuk** *ua-θiabe 'chop' **Puy** *ʈeb-ʈeb* (< PAn *C-)
b. **PRuk** *tebeθe 'belt' **Bas** *tisup* (< PAn *-S)
c. **PRuk** *θiraubu 'dance' **Kan** *mu-surauvu* **Saa** *mu-suarauvu* (< PAn *S-)
d. **PRuk** *θuŋu 'raincape' **Saa** *sauŋa* [**Jav** *sòŋ-sòŋ* 'parasol'] (< PAn *ṣ)

However, even if all the Rukaic items listed in (23), (25), (26), and (27) are attributed to borrowing, there remain six cognate sets, three with initial PRuk *θ shown in (28) and three with medial PRuk *θ shown in (29), which are less readily explained away.

(28) a. *θuθu 'breast' > **PRuk** *θuθu **Saa** *ʔu-susu* [**Puy** *susu* **Pai** *tutu* **Bun** *susuʔ* **Tha** *tutu* **Bab** *tsitu* **Hoa** *tutu* **Paz** *zuzu* **Sai** *hœhœʔ* **Ami** *cucu* **Kav** *sisu* **Bas** *tsjitsu* **Tro** *tsjitsu* **PMP** *susu]
b. *θiLaR 'ray, light, bright' > **Kan** *sinaŋe* **Saa** *siłaŋe* [**Tha** *tiłað* **Ami** *ciʃal* **Bas** *tsenal* 'sun' **Tro** *tsenal* 'sun', **PMP** *sinaR 'ray of light' (Blust & Trussel 2010)]
c. *θepθep 'suck' > **PRuk** *θepeθepe [**Puy** *sipsip* **Pai** *t‹m›eptep* **Bun** *supsup-un* **Bab** *tsiptsiep* **Paz** *zezep* **Ami** *mi-cepcep* **Kav** *sepsep* **Bas** *tsuptsup*]

(29) a. *ŋuθu 'mouth' > **Saa** *ŋusuu* [**Tao** *ya-nu[t]* **Bab** *nut* 'nose' **Pap** *nud* **Hoa** *nut* **Sir** *ŋuŋusu* **Ami** *ŋusuʔ* **Proto Oceanic** *ŋusu- 'mouth']
b. *aθu 'dog' > **PRuk** *aθu [**Pai** *vatu* **Bun** *ʔasuʔ* **Tha** *atu* **Tao** *outu* **Bab** *atu* **Pap** *hatu* **Hoa** *atu* **Paz** *wazu* **Sai** *ʔæhœʔ* **Sir** *asu* **Ami** *wacu* **Kav** *wasu* **Bas** *watsu* **Tro** *watsu* **PMP** *asu (Blust & Trussel 2010)]
c. *xeθi 'meat, content' > **Kan** *ʔa-isi* 'exist' **Saa** *maa-is-ana* 'content' [**Puy** *isi* (Ferrell 1969) **Pai** *seti* **Bun** *cici* (Ferrell 1969) **PAta** *hei **Ami** *heci* **Kav** *ʔsi* **Bas** *tsji* **PMP** *hesi 'flesh, meat' (Blust & Trussel 2010)]

All six are well attested in Formosan languages and, except for (28c), in Malayo-Polynesian languages. All reflexes, apart perhaps from Rukaic, reflect PAn *s. Two items, (28a) and (29a), are body parts, normally considered to be resistant to borrowing. There is reasonable evidence here, then, that PRuk *θ was a regular reflex of

PAn *s, and that its presence allowed some or all of the items listed in (23), (25), (26) and (27) to be borrowed into Rukai in such a way that a dental obstruent was analysed as the pre-existing *θ*.[26] We should note, however, that body parts are perhaps not as resistant to borrowing as received linguistic wisdom asserts. The PRuk term for 'liver' is reconstructed as *aθay, where *-θ- reflects PAn *-C-:

(30) *qaCay 'liver' > **PRuk** *aθay **Saa** *ʔaciʔi* **Puy** *ʔaʈay* **Pai** *qatsay* **Bun** *qatað* **Paz** *asay* **Ami** *qatay* **PMP** *qatay*

There is one more fact that is also not easily explained by a borrowing hypothesis. As Tsuchida (1976:127) recognised, the Kanakanavu and Saaroa phonemes that correspond with PRuk *θ are not the same as those that correspond with PRuk *s. On my analysis of the data, the relevant correspondences, extracted from Table 5 with criterial reflexes of putative PAn *θ in bold, are:

(31)

PAn	*θ	*s	*S	*x
Proto Rukaic	***θ**-**θ**-s	*s	*s	*Ø
Kanakanavu	**s**-**s**-s	Ø-Ø-s/Ø	s-s/Ø-s	Ø
Saaroa	**s**	Ø	s/Ø-Ø-s/Ø	Ø

What is more, the criterial reflexes occur in items that I have sought to explain away above, in (22a), (22b), (23a), (27c), (27d), (28a), (28b), (29a) and (29c) (but not in 26a-iii). Thus if PRuk *θ and *s are the outcome of a phonemic split in PAn *s under some now lost condition (stress would be a possible candidate),[27] then the same split has occurred in Kanakanavu and Saaroa under the same condition, implying the possibility (but not the necessity) that Rukaic, Kanakanavu and Saaroa have an exclusively shared ancestor. But there are no other grounds that I am aware of for thinking that this is the case. This being so, it is more economic to posit a PAn *θ, albeit tentatively, to account for the correspondence under *θ in (31) and for the items in (28) and (29).

A potential consequence of reconstructing PAn *θ is that an Austronesian subgroup could be proposed, in the shared ancestor of which PAn *θ and *s were merged. This subgroup would include all Austronesian languages except Rukaic, Kanakanavu and Saaroa, and perhaps Basay and Trobiawan, which also display a criterial medial reflex of PAn *θ in Table 5.[28] However, I would not posit a subgroup on the basis of a single phonological change unless it was a very unusual one. Since PAn *θ and *s were probably not very different either auditorily or in articulation, their merger could readily have occurred independently in different early languages.

---

[26] Note that this is not an argument for the pre-existence of *θ, as languages often acquire loan phonemes.

[27] I do not believe that PAn stress is reconstructable, and I accept the arguments of Blust (2009:547–551), including those against the stress hypothesis in Ross (1992). I have also looked carefully at Pejros' (1994) argument that the correspondences reflecting PAn *t and *C can be explained as a split conditioned by stress in a language ancestral to Tsou and Kanakanavu and find that the data do not support it at all.

[28] The quality of the data from Basay and Trobiawan, both extinct, is not particularly good, and I am reluctant to place any weight on them.

To sum up, I think the balance of the evidence favours the reconstruction of PAn *θ, but its existence is less certain than that of any other reconstructed PAn phoneme. This being so, it does not lend itself readily to any subgrouping hypothesis.

4 The phonetic values of PAn coronals and the 'Paiwan problem'

It is possible to treat a reconstructed phoneme paradigm simply as a system of contrasts, but the controversy among Indo-Europeanists evoked by the Glottalic Theory (Hopper 1973; Gamkrelidze & Ivanov 1973, 1995) shows that at least some historical linguists do attribute phonetic values to reconstructed phonemes because this allows them to assess the typological plausibility of the reconstructed paradigm and the reconstructed phonological histories of present-day languages.

There are effectively two steps in determining the phonetic values of a paradigm of protophonemes. Step 1 is to infer the phonetic value of each protophoneme from its reflexes in the data. This entails checking the plausibility of the sound changes required to get from each reconstructed phoneme to its attested reflexes. Step 2 is to check the typological plausibility of the reconstructed paradigm as a system and of the sound changes affecting classes of protophonemes (e.g., lenition of intervocalic voiced stops) that are required to get to attested systems.[29]

Two PAn protoparadigms are presented here. The first is based on step 1 and is close to those in Ross (1992) and Blust (2009:547). The second results from an application of step 2 and entails some rather radical claims. The two hypotheses are offered for consideration. I remain uncertain which I prefer.

I will refer to the two hypotheses as the 'step 1 paradigm' and the 'step 2 paradigm'.

4.1 A step 1 protoparadigm

The step 1 consonant protoparadigm is shown in Table 6. It differs from that presented in Ross (1992:31) in reversing the positions of *z and *Z (see below) and in accommodating the additional phonemes *x and *θ, about which more below. The glottal stop has disappeared in view of the arguments of Blust (2009:561–568). PAn *ñ has been added. In 1992 I thought that PMP *ñ reflected PAn *L, but PAn *L and *ñ are both reconstructable, as they are reflected differently in Tsou and Kanakanavu (Blust 2009:579 n. 79).

My reliance on Formosan data means that Table 6 differs a little from Blust (2009:547), whose *c (as distinct from *s) and initial *r- (as distinct from *d-) are reconstructed on the basis of distinctions in western Malayo-Polynesian languages (Blust 2009:557–560, 581), i.e., in a small number of languages in the western Indo-Malaysian archipelago that belong to just one first-order subgroup. For this reason they are excluded from Table 6, but on grounds of probability rather than certainty. If *r- did occur in PAn, then it was probably an alveolar flap, whilst *R was a trill.

---

[29] The Indo-European Glottalic Theory claimed that the paradigm produced by step 1 was typologically implausible and proposed a more plausible paradigm and sound changes, i.e., it arose from an application of step 2.

Table 6: A ‘step 1’ PAn consonant paradigm with phonetic values

	Bilabial	Dental and inter-dental	Alveolar	Laminal post-alveolar	Palatal	Velar	Uvular and pharyngeal
stop vl	p		t		s [c]	k	q
stop vd	b [b, β]		d	D [ḍ]	Z [ɟ]	g	
affricate vl			C [ts]				
affricate vd			z [dz]				
fricative vl		θ [θ]	x [s]	S [ṣ, ʃ]			h [ħ]
fricative vd					j [j, ɣʲ]		
nasal	m		n		ñ [ɲ]	ŋ	
lateral			L [l, ɬ]	l [l, ɾ]			
trill							R [r] or [ʁ]

Blust (2009:547) adds two phonemes to his 1999 paradigm, namely *-D and *h. The latter does not concern us here. *-D is reconstructed only in final position (as distinct from *-d), and is attested by *-r* (as opposed to *-t*) in Iban, Malay, the Batak languages and Balinese (Blust 2009:569–570). Like *c and *r-, *-D is excluded from the reconstructed paradigm and its separate reflex in a few western Malayo-Polynesian languages is attributed to some event(s) in the history of the languages of the western Indo-Malaysian archipelago. Only one item in which Blust & Trussel (2010) reconstruct it has a criterial reflex in a Formosan language: PAn *SateD ‘deliver’, which is reconstructed here as *SateZ (3a-iii).

The phonetic values of *d, *Z, *D, and *z fall out fairly simply from Table 1, and particularly from the criterial reflexes in Table 3. The values of *z and *Z in the 1992 version are here reversed, as similarities between the reflexes of *d and *z suggest that *z was the affricate corresponding to stop *d and the reflexes of *Z suggest that it was postalveolar. The criterial reflexes are important for the obvious reason that two phonemes cannot occupy the same paradigm slot, and these are the reflexes that enable differentiation.

The paradigm in Ross (1992:38–39) had only two potential sibilants, *s and *S, and two hypotheses about their phonetic values, here labelled A and B, were suggested:

(32)

PAn	*s	*S
A	[c]	[s]
B	[s]	[ʃ] or [ṣ]

Hypothesis A accommodated the fact that several Formosan reflexes of *s are stops or affricates (Table 5). By assigning the value of a palatal stop to *s it also accounted for the fact that when PMP homorganic nasal substitution (Blust 2004) was applied to *s-initial roots, the substitute was *ñ. Hypothesis B reflected the fact that other Formosan reflexes of *s are sibilant or, in the case of [h], formerly sibilant, and that three languages reflect *S as [ʃ]. I favoured B, arguing as follows. Sometime during the history of Paiwan the alveolars *t, *d and *L underwent palatalisation, becoming *tj* [c], *dj* [ɟ] and *ɬ* [ʎ], whilst

*s and *j became alveolar stops *t* [t] and *d* [d]. If, under A, *s had been the palatal [c], then we would have *t [t] becoming *tj* [c] and *s [c] becoming *t* [t], a pair of changes that could not have occurred without a merger—and no merger has occurred. Hence, I argued, B has a higher probability of being correct.

Blust (2009:579–580) favours a version of A which avoids merger, suggesting that *s was a palatal fricative ([ç] or [ɕ]?) rather than a stop [c]. For reasons discussed below, it seems more likely to me that *s was indeed a stop. Whether it was a fricative or a stop, it was certainly a palatal, and remained so when PMP homorganic nasal substitution took place,[30] a probability that B disallows.

Hypothesis A also entails the inference that *S was [s],[31] and Blust favours this on the grounds that the [s] > [h] > Ø lenition sequence is attested in certain Formosan languages (Puyuma, Siraya, partially in Saaroa, and incipiently in Kanakanavu) and in PMP. I agree with this reasoning, but the reconstruction of *x changes its application. It is clear from (33), extracted from Table 5, that *x has undergone lenition more widely among Formosan languages that either *s or *S (conversely, *x is lenited in any language in which *S is lenited)[32] and it can therefore be inferred that *x was [s]. The lenition of *s and *S is less straightforward. In the languages in (33a) *s lenition has proceeded further than *S lenition, whilst in the languages in (33b) and (33c) the opposite is true. Rukaic, Tsou, and Kavalan each reflect *s and *S identically.

(33)

		*s	*S	*x
a.	Kanakanavu	Ø-Ø-s/Ø	s-s/Ø-s	Ø
	Saaroa	Ø	s/Ø-Ø-s/Ø	Ø
	Saisiyat	h	ʃ	h
	Proto Atayalic	*h	*s	*h
b.	Puyuma	s	Ø	Ø
	Thao, Taokas	t	ʃ	Ø
	Favorlang, Papora	t	s	Ø
	Hoanya	t	s-s-t	Ø
	Pazih	z-z-t	s	h
	Siraya	-s-k	x-Ø-x/ŋ/Ø	Ø
	PMP	*s	*h	*h
a.	Paiwan	t	s	s
	Bunun	c [ts]	s	s-s/Ø-s
	Amis	c [ts]	s	s
b.	Proto Rukaic	*s	*s	*h
	Tsou	s	s	Ø
	Kavalan	s	s	s

[30] Since *s was [s] in PMP, substitution must have taken place shortly before *s ceased to be a palatal.

[31] Blust (2009:579) attributes this view to me. In fact Ross (1992:31, 39) says that *S was [ʂ] or [ʃ].

[32] The extinct languages Basay and Trobiawan are exceptions, and I am not sure whether this is an artifact of poorly recorded data or a genuine fact.

What this means is that, after lenition of *x, the reflex of *s shifted into the [s] slot in the languages of (33a), but the reflex of *S did so in the languages of (33b). In the languages in (33c) *x did not lenite (or in Bunun did so only incipiently), but *S nonetheless moved to [s], merging with *x. In the languages in (0d) *s and *S both shifted to [s], i.e., they merged. All this suggests strongly that both *S and *s were both in a position readily to move to [s]. The Thao, Taokas, and Saisiyat reflexes in Table 5 suggest that *S was [ʃ] or [ʂ].

Reasons were given above for believing that *s was a palatal. I suggest that it was a palatal stop [c] because it has followed three different trajectories whose common source appears to be [c]. Crucially, in the trajectory whereby *s has become [t], *s [c] has simply undergone depalatalisation (Thao, Taokas, Favorlang, Papora, Hoanya, Paiwan). By a minor detour *s became an affricate [cç], then was depalatalised as [ts] (Bunun, Amis). By a larger detour *s became a fricative [ç] and was depalatalised to [s] (Puyuma, Rukaic, Tsou, Kavalan, PMP) and then lenited in some languages (33a) and voiced in Pazih.

Finally, I take the phonetic value of *θ to have been [θ] as this is the only reflex value that differs from [s]. The cases for the phonetic values of PAn protophonemes not discussed here were made in the 1992 paper.

In regard to the shape of the 1992 protoparadigm as a whole, Blust (1999:35) writes

> Ross considers retroflex consonants, but not palatal consonants to have been present in PAn, since the former are found in a diverse collection of Formosan languages, while the latter are rare, geographically restricted, and in his view secondary.

and goes on to warn the reader that retroflexion may be an areal feature that has spread across Formosan languages since the departure of PMP from Taiwan's shores. He states my case for retroflex consonants admirably, since, as I wrote in section 1, Formosan languages occupy a special position in the reconstruction of PAn. There is no evidence either way that retroflexion has or has not spread since the departure of PMP, and we can only reconstruct on the basis of the evidence we have.[33] The protophonemes in question are in any case laminal postalveolars (section 2.1) and what matters most here is that they contrast with alveolars. Nonetheless, Blust summarises my position too categorically, as I was at pains to argue that *z (or *Z in 1992) was indeed palatal [ɟ] and was in contrast with both a laminal postalveolar and an alveolar (1992:36–37). The protoparadigm in Table 6 now contains as many palatals as laminals, thanks to the reassignment of *s and the readmission of *ñ (*j occupies an uncertain position between a palatal and a palatalised velar). I have even tried to reorganise it so that laminal postalveolars and palatals form a single postalveolar place of articulation, but one cannot do this in a way that is faithful to the data.

[33] Blust appeals for an example to the spread of retroflexion in the Indo-Aryan languages, but this was due to contact with Dravidian. We have no evidence of contact between earlier Formosan languages and a language with laminal postalveolars.

4.2 A step 2 protoparadigm

There is nothing typologically implausible about the step 1 protoparadigm in Table 6, but it does imply some unusual sound changes, such as *d [d] > Tsou and Kanakanavu *c* [ts], *d [d] > Saaroa and Thao *s* [s].

More serious, however, is the 'Paiwan problem', to which I briefly alluded in section 4.1. Table 7 shows the Proto Paiwan consonant paradigm reconstructed by Ho (1978), transcribed into the orthography of Ferrell (1982) and Early & Whitehorn (2003). The reconstruction is straightforward. The phonetic values shown in parentheses are those of the modern dialects, except where mergers have occurred.[34] The dialects that Ho classifies as southeastern have retained the Proto Paiwan consonant paradigm, whilst the northeastern dialects, Paiwan and Stimur, reflect the mergers shown as boxes in (34).

Table 7: Proto Paiwan consonant paradigm with phonetic values
(in the orthography of Ferrell 1982 and Early & Whitehorn 2003)

	Bilabial	Labio-dental	Alveolar	Laminal post-alveolar	Palatal	Velar	Uvular
stop vl	*p*		*t*		*tj* [c, tʲ]	*k*	*q* [q, ʔ]
stop vd	*b*		*d*	*ḍ* [ɖ, ʐ]	*dj* [ɟ, dʲ]	*g*	
affricate vl			*ts* [ʦ, ʈʂ]				
fricative vl		*v*	*s* [s, ɕ/_i]				
fricative vd			*z*				
nasal	*m*		*n*			*ŋ*	
lateral				*l* [ɻ]	*ɬ* [ʎ, ɬ]		
trill			*r* [r, ʁ]				

(34)

PAn	Proto Paiwan	Stimur Paiwan
*t [t]	*tj [c, tʲ]	t
*s [c]	*t [t]	
*C [ʦ]	*ts [ʦ, ʈʂ]	ts
*d [d]	*dj [ɟ, dʲ]	d
*j [ɟ, ɣʲ]	*d [d]	
*D [ɖ]	*ḍ [ɖ, ʐ]	ḍ
*R [r] or [ʁ]	*r [r, ʁ]	r
*l [l̩, ɻ]	*l [ɻ]	l [ɻ]
*L [l, ɬ]	*ɬ [ʎ, ɬ]	[ʎ] [l]

The Proto Paiwan system in Table 7 looks rather similar to the PAn system in Table 6, but diachronically the similarity is deceptive. As (34) shows, crossovers have occurred.

[34] There is considerable disagreement in the literature about how the dialects of Paiwan should be classified and which dialects have what phonetic values (Ho 1978; Ferrell 1982:5-7; Egli 1990:4-5; Chang 2006), but there is agreement about which phonetic values occur and about mergers.

PAn *t [t], *d [d], and *L [l] have undergone palatalisation, becoming Proto Paiwan *tj* [c], *dj* [ɟ], and *ɬ* [ʎ], and PAn *s and *j have undergone depalatalisation, becoming alveolar stops *t* [t] and *d* [d]. A parallel depalatalisation of PAn *l [l̩, ɻ̩] didn't occur, because *l was not a palatal but a laminal postalveolar.

There is a reconstructive problem here. How could these crossovers have occurred without mergers? Two kinds of solution offer themselves. The first kind has two variants.

In the first variant the PAn alveolars *t and *d shifted out of their PAn slots, becoming something else, thus making way for the palatals *s and *j to become the Proto Paiwan alveolars. What might this 'something else' have been? The only possibility I see is that PAn *C and *Z were not in fact alveolar but postalveolar affricates, i.e., [tʂ] and [dʐ],[35] and that *t and *d became [ts] and [dz] before palatalisation. Schematically

(35) PAn *t [t], *d [d] > Post-PAn *[ts], *[dz]
PAn *s [d], *j [j, ɣʲ] > Proto Paiwan *t, *d
Post-PAn *[ts], *[dz], *L [l, ɬ] > Proto Paiwan *tj [c, tʲ], *dj [ɟ, dʲ], *ɬ [ʎ, ɬ]

This is not a particularly plausible scenario, as it entails an intervening stage where the system lacks alveolar stops [t] and [d] and has affricates contrasting at alveolar ([ts], [dz]) and postalveolar ([tʂ], [dʐ]) points of articulation. Typologically this would be a rather odd system.

In the second variant of the first kind of solution it is the palatals that become something else to make way for the palatalising alveolars. Again the obvious solution is that they became (palatal) affricates before they became alveolar stops. And again the transitional system would be typologically odd because it lacked alveolar stops. Both variants also invite us to believe (i) that stops became affricates, then became stops again, and (ii) that PAn *t and *d became Proto Paiwan *tj and *dj, then the change was reversed in the Paiwan and Stimur dialects. Both (i) and (ii) would be diachronically unusual.

The second kind of solution is more radical, but it avoids these difficulties. It infers that Paiwan is more conservative than scholars have assumed, and that the phonetic values of PAn coronals were roughly as they were in Proto Paiwan and that the changes in (35) never occurred. This 'step 2' PAn consonant paradigm is shown in Table 8, and (35) is rewritten as (36), where the only posited change affects *j.[36]

(36) PAn *t [c], *d [ɟ], *L [ʎ] > Proto Paiwan *tj [c, tʲ], *dj [ɟ, dʲ], *ɬ [ʎ, ɬ]
PAn *s [t], *j [z] > Proto Paiwan *t [t], *d [d]

A step 2 paradigm attempts to take account of sound changes affecting classes of protophonemes, in this instance in Paiwan. Inevitably some of the phonetic decisions it entails seem counterintuitive, and it sometimes requires widespread sound changes to account for reflexes in languages other than Paiwan, e.g. PAn *t [c] is depalatalised as [t] in all languages except Paiwan. It also requires some rather complex changes, e.g. PAn *d [ɟ] is depalatalised as [d] in Puyuma, Bunun, Papora, Hoanya, Pazih, Saisiyat,

---

[35] Paiwan *ts* is postalveolar in some dialects.

[36] One could argue by analogy that if *s was [t], then *j was [d], but this does not take account of reflexes of *j such as Rukaic *g, Kanakanavu [l], Saaroa [ɬ], Hoanya *dz*, Pazih [z], Saisiyat [ð, z], and Thao [z], not to mention Siraya, Amis, Basay, Trobiawan and Kavalan [n].

Atayalic, and Amis (and then flapped in Saisiyat and Amis) and as [ɖ] in Rukaic; PAn *d [ɟ] becomes an affricate and is devoiced and depalatalised as [ts] in Tsou and Basay. On the other hand the paradigm in Table 8 simplifies some changes, e.g. PAn *s [t] remains unchanged in Paiwan and the Western Plains languages (Thao, Taokas, Favorlang–Babuza, Papora and Hoanya). The fact that *L is assigned a palatal articulation explains why it merges with *ñ in a number of Formosan languages (Saaroa, Paiwan, Bunun, Thao, Saisiyat, Atayalic, Amis, and Kavalan).[37] But, as an anonymous reviewer points out, neither paradigm is fully explanatory, as both require, for example, that PAn *x [s] become Basay and Trobiawan *ts*, an unlikely strengthening.

Table 8: A hypothetical 'step 2' PAn consonant paradigm with phonetic values

	Bilabial	Dental and inter-dental	Alveolar	Laminal post-alveolar	Palatal	Velar	Uvular and pharyngeal
stop vl	p		s [t]		t [c]	k	q
stop vd	b [b, β]			D [ɖ]	d [ɟ]	g	
affricate vl			C [ts]				
affricate vd			z [dz]		Z [ɟʝ]		
fricative vl		θ [θ]	x [s]	S [ʂ, ʃ]			h [ħ]
fricative vd			j [z]				
nasal	m		n		ñ [ɲ]	ŋ	
lateral				l [l̩, ɾ]	L [ʎ, ɬ]		
trill							R [r] or [ʀ]

The question that has to be answered here is not, Which protoparadigm gives us the least changes? After all, some five thousand years have probably elapsed since the break-up of PAn—time for many changes. If we look at the reconstructed history of Indo-European, we may be surprised at how few changes seem to have occurred in Formosan and Philippine languages. Rather, the relevant question is, Which protoparadigm gives us the more plausible set of changes? I am not sure that I know the answer to this question. I have offered the hypothetical paradigm in Table 8 as an alternative for consideration, and not because I am committed to it. If nothing else, the juxtaposition of the protoparadigms in Tables 6 and 8 is a warning that we cannot subgroup Formosan languages on the basis of phonetic change. Blust (1999) rightly avoids these and subgroups instead on the basis of phonemic mergers, i.e., on losses of distinction that are not dependent on a reconstruction of the sounds of the protolanguage.

5 Conclusion

Dyen and Tsuchida were inclined by their method of work to posit a protophoneme or a variant protophoneme for every correspondence set that they found, even when it

[37] Relevant data are not available for Rukai, Taokas, Favorlang–Babuza, Papora, Hoanya, Basay, and Trobiawan.

differed by only one reflex from another set. This led to a certain ambiguity with regard to subscripted symbols: Did a subscripted symbol represent a variant of an already recognised protophoneme? Or did it represent a new protophoneme? The tendency has been for other scholars to assume that all subscripted variants of a the same symbol are variants of a single phoneme, and sometimes to apply the same assumption to Dahl's work, even when he explicitly uses subscripts to symbolise different phonemes. I have shown here that the Formosan data do support the distinctions between Dahl's *d_1 and *d_2 (my *d and *Z) and between Tsuchida's *S_1 and *S_2 (my *S and *x). There is also a case for a distinction between *d_3 (my *D) on the one hand and *d_1 and *d_2 on the other, but it is less strong. The case for a PAn distinction between Tsuchida's *s and *θ is perhaps weaker still, but the reflexes are difficult to account for otherwise.

Finally, my efforts to ascribe phonetic values to the PAn coronals suggest that it is possible to reconstruct phonetic values with a fair degree of plausibility, but not with anywhere near the level of certainty that would be required to use them as a baseline for determining shared phonetic innovations to identify subgroups.

Appendix I: Formosan sound correspondences

	Stops and affricates												Glides	
Position	Bilabial		Alveolar				Post-alveolar		Palatal	Velar		Uvular	Bilabial	Palatal
PAn (Tsuchida)	*p	*b	*t	*C	*d, *D_3	*D_2	*$D_{1,4}$	*Z	*s	*k	*g	*q	*w	*y
PAn (Blust)	*p	*b	*t	*C	*d	*d	*d	*Z, then *z	*s	*k	*g	*q	*w	*y
PAn (Ross)	*p	*b	*t	*C	*d- (*d_{-1})	*Z^1 (*d_2)	*D (*d_3)	*z^1	*s	*k	*g	*q	*w	*y
‘step 1’[2]	*[p]	*[b, β]	*[t]	*[ts]	*[d]	*[ʣ]	*[ḍ]	*[ʤ, ḍʐ]	*[c]	*[k]	*[g]	*[q]	*[w]	*[j]
‘step 2’	*[p]	*[b, β]	*[c]	*[ʦ]	*[ɟ]	*[ʣ]	*[ḍ]	*[ɟ̬]	*[t]	*[k]	*[g]	*[q]	*[w]	*[j]
Puyuma	p	b	t	ţ [ʈ]	d	ḍ [ḍ/ʐ]	ḍ [ḍ/ʐ]	d	s	k	g	ʔ	w	y
P Rukai	p	b	t	c [ts]	D [ḍ]	D [ḍ]	D [ḍ]	d	s	k	g	Ø	v [β]	-ð-y
Tsou	p	f	t	c [ts, ʧ]	c [ts]	c [ts]	c [ts]	c [ts]	s [s, ʃ]	ʔ, k	k	Ø	v	-z-y
Kanavu	p	v [β]	t	c [cç ʧ]	s-	c [ts]	c-s-	c [ts]	Ø-Ø-s/Ø	k	k	ʔ	Ø	-l-i
Saaroa	p	v	t	c [ts, ʧ]	s	s	s	s	Ø	k	k	ʔ	Ø	-ɬ-i
P Paiwan	p	v	tj [c]	ts	dj [ɟ]	z	ḍ [ḍ]	dj [ɟ]	t	k	g	q	v, Ø	-y
P Bunun	p	b	t	t	d	d	d	d	c [ʧ]	k	k	q	v	-ð
Thao	p	f	t	θ (*c*)	s	s	t	s	t	k	…	q	w; Ø/_i	-y
Taokas	p, Ø	b	t	s	t	…	g	t	t	k/Ø	…	h	…	-y
Fav-Babuza	p, Ø	b	t	ʧ (*ch*)	t	d, r	r	t	t	Ø	…	Ø	…	-Ø
Papora	p, Ø	b	t	s	t, d	…	…	d	t	Ø	…	Ø	…	-Ø
Hoanya	p	b	t	s	d	d	…	dz	t	k	…	Ø	…	-Ø
Pazih	p	b	t	s	d	d	d	d	z-z-t	k	…	Ø	Ø	-y
P Saisiyat	p	b	t	s [θ/ʂ]	r [ɾ]	r [ɾ]	r [ɾ]	r [ɾ]	h	k	…	ʔ	w; Ø/_i	-y
P Atayal	p	b [b/β]	t	c [ʧ]	d	d	d	d	h [ɦ]	k	…	q	w	-y
Siraya	p	v	t	t	s	s	s	d, l	s	k	…	Ø	…	-y
P Amis	p	b [b/β]	t [ʈ]	t [ʈ]	r,l [ɾ]	r,l [ɾ]	r	l [ɾ]	c [ts]	k	…	ʔ	w, b [b/β]	…
Basay	p	b	t	t	ts-	-l-	l-	ts-l-	ts-s-	k	…	Ø	…	…
Trobiawan	p	v	t	t	z			ts	ts-s-	k	…	Ø	…	…
Kavalan	p	v [β]	t	t	z	z	z	z	s	q	…	Ø	…	…
PMP	*p	*b	*t	*t	*d	*d	*d	*Z	*s	*k	*g	*q	*w	*y
PAn (Wolff)	*p	*b	*t	*t	*d	*d	*d	*j [ḍ]	*c [ʈ]	*k	—	*q	*w	*y

Notes: In phonetic representations dialect alternants are separated by a slash (/), other alternants by a comma.

1. Blust and I now write earlier *Z as *z, but in my orthography it contrasts with *Z, a phoneme not recognised in the conventional orthography.

2. For an explanation of ‘step 1’ and ‘step 2’, see the text, section 4.

	Fricatives					Nasals				Liquids		
Position	Interdental	Alveolar	Postalv.	Pal./Vel.	Glottal	Bilabial	Alveolar	Palatal	Velar	Alveolar	Postalv	Uvular
PAn	*θ	*S_2	*S_1	*j	*H	*m	*n	*ñ	*ŋ	*L, *N	*l	*r, R
(Tsuchida)												
PAn (Blust)	*s	*S	*S	*j	*h	*m	*n	*ñ	*ŋ	*N	*l	*r, R
PAn (Ross)	*θ	*x	*S	*-j-j	*h	*m	*n	*ñ	*ŋ	*L	*l	*r, R
‘step 1’	*[θ]	*[s]	*[ʃ, ʂ]	*[j, ɣʲ]	[ħ]	*[m]	*[n]	*[ɲ]	*[ŋ]	*[l, ɬ]	*[l, ɾ]	*[r, ʁ]
‘step 2’	*[θ]	*[s]	*[ʃ, ʂ]	*[z]	[ħ]	*[m]	*[n]	*[ɲ]	*[ŋ]	*[ʎ, ɬ]	*[l, ɾ]	*[r, ʁ]
Puyuma	s	Ø	Ø	d	Ø	m	n	l	ŋ	l	l̪ [l̪]	r
P Rukai	θ-θ-s	s	S	g, Ø	Ø	m	n	…	ŋ	l	L [l̪]	r, ʔ
Tsou	s [s, ʃ]	Ø	s	Ø	Ø	m	n	n	ŋ	h; k/_f,s,h	r [ɾ/j]	r [ɾ]
Kanavu	s [ç, ʃ]	Ø	s-s/Ø-s	l [ɾ], Ø	Ø	m	n	ŋ	ŋ	n	l [ɾ]	r
Saaroa	s [s, ʃ]	Ø	s/Ø-Ø-s/Ø	l [ɬ], Ø	Ø-Ø-ʔ	m	n	l [ɬ]	ŋ	ɬ [ɬ]	l [ɾ]	r
P Paiwan	t	s	s	d	-Ø	m	n	ɬ [ʎ]	ŋ	ɬ [ʎ]	l [l̪]	Ø, r
P Bunun	c [ʧ]	s-s, Ø-s	s	Ø	h [χ]	m	n	n	ŋ	n	Ø	l
Thao	t	Ø	ʃ (*sh*)	ð (z)	-ʔ	m	n	ð (z)	n	ð (z)	r	ɬ (*lh*)
Taokas	t	Ø	ʃ	t, Ø		m	n	…	n	t	l, r	h
Fav-Babuza	t	Ø	s	d		m	n	…	n	s, l	r	Ø
Papora	t	Ø	s	d, Ø		m	n	…	n	l	r	l (?)
Hoanya	t	Ø	s	dz		m	n	…	ŋ	s	l	h
Pazih	z-z-t	h [ħ]	s	z, -t	h [ħ]	m	n	…	ŋ	l	r [ɾ~ɻ]	x
P Saisiyat	h	h	ʃ	z [ð]	h	m	n	l	ŋ	l	l̪ [l̪]	l̪ [l̪]
P Atayal	h [ħ]	h [ħ]	s [ʃ]	j [j̪]	h [ħ]	m	n	l [ɬ]	ŋ	l	r [r~ɾ]	g [g~ɣ]
Siraya	s	Ø	x-Ø-x,Ø,ŋ	n [n̪]		m	n	l	ŋ	l	r	g
P Amis	c [ts]	s	s	n	h-Ø [ħ]	m	n [n̪]	d [d,ð,ʟ, ɬ]	ŋ	d [d,ð,ʟ, ɬ]	l [ɾ]	l [ɾ]
Basay	ts	ts	s	n		m	n	…	ŋ	n	ts	l, Ø
Trobiawan	ts	ts	s	n		m	n	…	ŋ	n	ts	l
Kavalan	s	s	s	n		m	n	n	ŋ	n	r [ɾ]	r [ɾ], R [ʁ]
PMP	*s	*h	*h	*j		*m	*n	*ñ	*ŋ	*l-n-n	*l	*r, R
PAn (Wolff)	—	—	*s	*g	*h [ħ]	*m	*n	*ɬ [l̪]	*ŋ	*ɬ [l̪]	*ɬ	*ɣ [ʁ]

Appendix II: Fisher's exact test

The appropriate test of significance for a two-dimensional data matrix in which some cell values are less than 5 is Fisher's exact test. The frequencies of my examples in Table 4 are set out in the matrix below.

	Pai dj	**Pai** z	**Pai** ɖ	Totals
Puy d	5	1	0	6
Puy ɖ/z	2	14	3	19
Puy r	0	1	0	1
Totals	7	16	3	26

The test yields $p = 0.008$, i.e. there is a probability of 0.8% that the distribution of examples shown in Table 4 arose by chance. The distribution is significant.

Blust's example frequencies from Table 4 are set out in the matrix below.

	Pai dj	**Pai** z	**Pai** ɖ	Totals
Puy d	2	2	0	4
Puy ɖ/z	1	6	1	8
Totals	3	8	1	12

The test yields $p = 0.661$, i.e., there is a probability of 66.1% that the distribution of examples shown in Table 4 arose by chance. The distribution is there not significant.

References

Adelaar, K. Alexander. 1999. Retrieving Siraya phonology: A new spelling for a dead language. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 313–354. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

Blust, Robert. 1976. Review of Dahl (1973). *Language* 52.1:221–237.

Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *University of Hawaii Working Papers in Linguistics* 9.2:1–15.

Blust, Robert. 1990. Patterns of sound change in the Austronesian languages. In *Linguistic change and reconstruction methodology*, ed. by Philip Baldi, 231–267. Trends in linguistics: Studies and monographs 45. Berlin and New York: Mouton de Gruyter.

Blust, Robert. 1995. Sibilant assimilation in formosan languages and the proto-Austronesian word for 'nine': A discourse on method. *Oceanic Linguistics* 34.2:443–453.

Blust, Robert. 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35.2:272–294.

Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

Blust, Robert. 2003. *Thao dictionary*. Language and Linguistics Monograph Series A5. Taipei: Academia Sinica.

Blust, Robert. 2004. Austronesian nasal substitution: A survey. *Oceanic Linguistics* 43.1:73–148.

Blust, Robert. 2009. *The Austronesian languages*. Pacific Linguistics 602. Canberra: The Australian National University. [2013 revised edition on-line at: http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf]

Blust, Robert and Stephen Trussel. 2010. *Austronesian comparative dictionary*. http://www.trussel2.com/acd/.

Cauquelin, Josiane, 1991. *Dictionnaire puyuma-français*. Textes et documents Nousantariens 9. Paris: École Française d'Extrême-Orient.

Chang, Anna Hsiou-chuan. 2006. A reference grammar of Paiwan. PhD dissertation. Canberra: The Australian National University.

Dahl, Otto Christian. 1973. *Proto-Austronesian*. Scandinavian Institute of Asian Studies monograph series No. 15. Oslo: Studentlitteratur Curzon Press.

Dahl, Otto Christian. 1976. *Proto-Austronesian*. Scandinavian Institute of Asian Studies monograph series No. 15. Oslo: Studentlitteratur Curzon Press. Second revised edition.

Dahl, Otto Christian. 1981. *Early phonetic and phonemic changes in Austronesian*. Oslo: The Institute for Comparative Research in Human Culture.

Dyen, Isidore. 1965. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285–305.

Dyen, Isidore. 1971. The Austronesian languages and Proto-Austronesian. In *Current trends in linguistics, vol.8: Linguistics in Oceania*, ed. by Thomas E. Sebeok, 5–54. The Hague: Mouton.

Early, Robert and John Whitehorn. 2003. *One hundred Paiwan texts*. Pacific Linguistics 542. Canberra: The Australian National University.

Egerod, Søren. 1999. *Atayal–English dictionary*. 2nd ed. by Jens Østergaard Petersen. Historisk–filosofiske Skrifter 20. Copenhagen: The Royal Danish Academy of Sciences and Letters.

Egli, Hans. 1990. *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.

Egli, Hans. 2002. *Paiwan Wörterbuch: Paiwan-Deutsch, Deutsch-Paiwan*. Wiesbaden: Otto Harrassowitz.

Ferguson, Charles A. 1990. From esses to aitches: Identifying pathways of diachronic change. In *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, ed. by William A. Croft, Keith Denning and Suzanne Kemmer, 59–78. Typological Studies in Language 20. Amsterdam: John Benjamins.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica Monograph No. 17. Taipei: Academia Sinica.

Ferrell, Raleigh. 1982. *Paiwan dictionary*. Pacific Linguistics Series C-73. Canberra: The Australian National University.

Fey, Virginia. 1986. *Amis dictionary*. Taipei: Bible Society in the Republic of China.

Gamkrelidze, Thomas V. and Vjačeslav V. Ivanov. 1973. Sprachtypologie und die Rekonstruktion der gemeinindogermanischen Verschlüsse. *Vorläufiger Bericht* 27.3:150–156.

Gamkrelidze, Thomas V. and Vjačeslav V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a Proto-language and a Proto-culture*. Translated by Johanna Nichols. Trends in Linguistics: Studies and Monograph 80. Berlin and New York: Mouton de Gruyter.

Hamann, Silke Renate. 2003. The phonetics and phonology of retroflexes. PhD dissertation. Utrecht: University of Utrecht.

Ho, Dah-an, 1978. wuzhong paiwan fangyan de chubu bijiao [A preliminary comparative study of five Paiwan dialects]. *zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* [*Bulletin of the Institute of History and Philology, Academia Sinica*] 49.4:565–681. [In Chinese]

Ho, Dah-an. 1998. Genetic relationships among the Formosan languages. *hanxue yanjiu* [*Chinese Studies*] 16.2:141-167. Taipei: Center of Chinese Studies, National Library. [In Chinese]

Hopper, Paul J. 1973. Glottalized and murmured occlusives in Indo-European. *Glossa* 7.2:141–166.

Ladefoged, Peter and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, Peter and Zongji Wu. 1984. Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12.3:267–278.

Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* 48.1:1–92.

Li, Paul Jen-kuei. 1978. The case-marking systems of the four less-known Formosan languages. In *Second International Conference on Austronesian Linguistics: Proceedings*, ed. by S.A. Wurm and Lois Carrington, 569–615. Pacific Linguistics C-61. Canberra: The Australian National University.

Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52.2:235–301.

Li, Paul Jen-kuei. 1985. The position of Atayal in the Austronesian family. In *Austronesian linguistics at the 15th Pacific Science Congress*, ed. by Andrew Pawley and Lois Carrington, 257–280. Pacific Linguistics C-88. Canberra: The Australian National University.

Li, Paul Jen-kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59.2:479–508.

Li, Paul Jen-kuei. *English–Favorlang vocabulary by Naoyoshi Ogawa, with an introduction by Paul Jen-kuei Li*. Asia–Africa Lexicon Series 49. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Language and Linguistics Monograph Series A-2. Taipei: Academia Sinica.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2006. *Kavalan dictionary*. Language and Linguistics Monograph Series A-19. Taipei: Academia Sinica.

Ogawa, Naoyoshi. 2006. *A comparative vocabulary of Formosan languages and dialects by Naoyoshi Ogawa*, ed. by Paul Jen-kuei Li and Masayuki Toyoshima. Asian and African Lexicon Series, No.49. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Pejros, Ilia. 1994. Some problems of Austronesian accent and *T~*C (notes of an outsider). *Oceanic Linguistics* 33.1:105–127.

Ross, Malcolm. 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31.1:23–64.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert*

Blust, ed. by Alexander Adelaar and Andrew Pawley, 295–326. Pacific Linguistics 601. Canberra: The Australian National University.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43.2:411–444.

Szakos, József. n.d. Tsou dictionary. Unpublished computer file, Providence University, Taitung.

Ting, Pang-Hsin. 1978. gu beinanyu de nice. [Reconstruction of Proto-Puyuma phonology] *zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* [*Bulletin of the Institute of History and Philology, Academia Sinica*] 49.3:321–392. [In Chinese]

Tryon, Darrell T. 1995. *Comparative Austronesian dictionary: An introduction to Austronesian studies*. Trends in Linguistics: Documentation 10. Berlin and New York: Mouton de Gruyter.

Tsuchida, Shigeru. 1969. Wordlist of three Bunun dialects. Unpublished mimeograph.

Tsuchida, Shigeru. 1971. List of words in Formosan languages. Unpublished mimeograph.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No. 5. Tokyo: University of Tokyo.

Tsuchida, Shigeru. 1981. Wordlists of Amis dialects collected in 1980–1981. Unpublished mimeograph.

Tsuchida, Shigeru. 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. Tokyo: University of Tokyo.

Wolff, John U. 1994. The place of plant names in reconstructing Proto Austronesian. In *Austronesian terminologies: Continuity and change*, ed. by A.K. Pawley and M.D. Ross, 511–540. Pacific Linguistics C-127. Canberra: The Australian National University.

Wolff, John U. 1997. The Proto-Austronesian voiced apical and palatal stops *d and *j. In *Proceedings of the Seventh International Conference on Austronesian Linguistics: Leiden 22-27 August 1994*, ed. by Cecilia Odé and Wim Stokhof, 581–601. Amsterdam: Rodopi B.V.

Wolff, John U. 2007. Do the Malayo-Polynesian languages constitute a subgroup of the Austronesian language family? 10-ICAL Historical Comparative Papers. *Studies in Philippine Languages and Cultures* 15:13–24.

Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. Ithaca, NY: Cornell University, Southeast Asia Program Publications.

Zorc, R. David. 1982. Where, O where have the laryngeals gone? Austronesian laryngeals re-examined. In *Papers from the Third International Conference on Austronesian Linguistics, Vol 2: Tracking the travellers*, ed. by Amran Halim, Lois Carrington and S.A. Wurm, 111–144. Pacific Linguistics C-75. Canberra: The Australian National University.

Zorc, R. David. 1987. Austronesian apicals (*dDzZ) and the Philippine non-evidence. In *A world of language: Papers presented to Professor S.A. Wurm on his 65th birthday*, ed. by Donald C. Laycock and Werner Winter, 751–761. Pacific Linguistics C-100. Canberra: The Australian National University.

2 *The preglottalised stops in three Formosan languages**

PAUL JEN-KUEI LI

1 Introduction

This paper is to illustrate a case of contact-induced phonological feature across languages. Preglottalised stops are not commonly found in Austronesian languages and no Austronesian scholar has suggested reconstructing this type of stop for Proto-Austronesian (PAn).

Two preglottalised stops /b, d/[1] appear in three Formosan languages, Tsou (Nevskij 1935:26, 29–30; Tung et al. 1964:12), Bunun (Ogawa & Asai 1935:586) and Thao (F.K. Li, Chen & Tang 1956:23), all spoken in central Taiwan. They are not found in any other Formosan language. There are two alternative hypotheses to account for this areal feature: (1) they originated in one language and then diffused to the other two, or (2) there were parallel developments in each of the three languages. I shall present some evidence to argue for the first hypothesis, although the possibility of parallel development in Tsou and Bunun cannot be excluded. Moreover, I shall try to demonstrate how it was diffused from one language to another.

Languages with series of preglottalised stops are found in continental Southeast Asia, Hainan, and southwestern China.[2] They are all geographically far away from Taiwan, so

* This paper was presented at the conference of Language Contact in Asia and the Pacific, University of Macao, 6–7 September 2012. I would like to thank Shigeru Tsuchida, Elizabeth Zeitoun, and two anonymous reviewers for their valuable comments on the earlier drafts of this paper and Atul, a native speaker of the Takibakha dialect of Bunun, for going through Tung's (1964:443–547) Tsou glossary and my own wordlists of Tsou to find out if there are similar forms in Bunun.

1 Following the general tradition in Formosan linguistics, the preglottalised voiced stops, which are phonetic rather phonemic in the said languages, will be written as plain /b/ and /d/. All Tsou forms are based on the Duhtu dialect, while Bunun forms are mostly based on general Bunun unless specified otherwise, as in (3) below.

2 Series of preglottalised stops have been reported for some languages in the Tai language family, such as Wuming, and the hypothesis of series of preglottalised stops have been posited for Proto Tai in F.K. Li's (1943) reconstructions.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, *39–46*
Asia-Pacific Linguistics, 2015

it is extremely unlikely that languages in such remote areas would have any close contact with Formosan languages to exert any influence on these languages.

2 The relationships of the three Formosan languages and their /b, d/

Based on Blust's (1999:45) subgrouping, Tsou, Bunun and Thao[3] belong to three different major subgroups. If there were exclusively shared phonological innovations, they would be closer to each other. Their reflexes[4] of PAn phonemes are as given in Table 1 below. Sound changes in special phonetic environments may not be specified.

Table 1: Tsou, Bunun and Thao reflexes of PAn

PAn	*p	*b	*t	*C	*d	*D	*Z	*k	*j	*R	*q	*m	*ŋ
Tsou	p	f	t	c	c	c	c	ʔ, Ø[5]	Ø	r	Ø	m	ŋ
Bunun	p	b	t	t	d	d	d	k	Ø	l	q	m	ŋ
Thao	p	f	t	θ	s	s	s	k	ð	ɬ	q	m	n

PAn	*n	*N	*ñ	*l	*s	*S[6]	*H	*w	*y	*a	*i	*u	*ə
Tsou	n	h, k	n,h	r	s	s, Ø	Ø	v	z	o	i	u	ə
Bunun	n	n	n	Ø	s	s	h[7]	v	ð	a	i	u	u
Thao	n	ð	ð	r	t	ʃ, Ø	Ø	w	y	a	i	u	i

It is clear that the three languages under discussion are phonologically quite divergent from each other, with very few shared phonological innovations, such as the merger of *d, *D, and *Z. Let it be noted that only /b/ and /d/ in Bunun are derived from PAn *b and

---

[3] Language abbreviations include: Ami, Amis; Bun, Bunun; Kav, Kavalan; RukLa, Labuan Rukai; Pai, Paiwan; PAn, Proto Austronesian; Sai, Saisiyat; Tha, Thao; Tso, Tsou. The symbol < A stands for assimilation, < D for dissimilation, and < M for metathesis; the abbreviations are AF, actor-focus; PF, patient-focus.

[4] Historical derivations of Tsou phonology are given by Tsuchida (1976:305–311), those of Bunun phonology by P. Li (1988) and Wolff (2010:167–180), and those of Thao phonology by Blust (2003:73–79) and Wolff (2010:97–110). See these references for more details.

[5] In Tsou *k is lost word-initially, unless it is followed by *m, in which case it becomes *b*, and it becomes *ʔ* word-medially.

[6] Dahl (1981:28ff) reconstructs two types of *S, *S_1 and *S_2, for PAn. Tsou and Thao reflexes of *S_1 are *s* and *ʃ* respectively, and their reflexes of *S_2 are zero, while Bunun reflexes of both *S_1 and *S_2 are *s*, though occasionally lost, e.g., *paliS_1i > Tso *prisi-a*, Tha *parʃi* 'to taboo'; *CaqiS_1 > Tso *t‹m›eʔsi* (*t*- irregular), Bun *ma-taqis*, Tha *ʃ‹m›aqiʃ* (< A *s*- > *ʃ*-) 'to sew'; *S_2apuy > Tso *puzu*, Bun *sapuð*, Tha *apuy* 'fire'; *ka S_2iw > Tso *evi*, Tha *kawi* (< M) 'tree, wood'.

[7] Dahl (1981:42) reconstructs two types of *H, *H_1 and *H_2, for PAn. The northern and central dialects of Bunun reflect /h/ for *H_1 in initial and medial positions, e.g., *H_1ules > *huluc* (< A) 'clothes', *buH$_1$et > *puhut* (< A) 'squirrel', but in the final position only in Takituduh, e.g.,*teluH$_1$ > *tauh* 'three'. Bunun /h/ also reflects PAn *l in the initial position, e.g., *lima > *hima* 'five' (Wolff 2010:171, 178). See Wolff (2010:170–171) for a detailed and reasonable account for the status of /h/ in Bunun.

*d/*D/*Z, respectively, and that there is no /b/ or /d/ in Tsou or Thao. Yet /b/ and /d/ do occur in the phonemic inventory of Tsou and Thao.

The preglottalised stops occur word- or syllable-initially and medially before all the main vowels /i, u, a/ in all the three languages, and also before the other vowels /e, o, ə/ in Tsou. No form ends with a coda in Tsou. Word-final /b/ and /d/ are attested only in the Isbukun dialect of Bunun (P. Li 1987), but they are not attested in Thao. In short, the preglottalised stops are not distributionally defective except in the word-final position in these languages.

3 The origins of the preglottalised stops in Tsou

PAn *b has become /f/ in Tsou, e.g., *batu > *fatu* 'stone', *qabuH > *fuu* 'ashes' while PAn *d, *D or *Z has become /c/ in Tsou, e.g., *daŋdaŋ > *cŋəcŋə* 'dry by fire', *DemDem > *cməcmə* 'cloud', *ZaReq > *cro-a* 'earth'.

The preglottalised stops /b, d/ in Tsou are historically derived from a *k immediately followed by a nasal (Tsuchida 1972), as illustrated in (1) and (2) below.

(1)	*k‹um›aən	>	*kumaənə	>	*kmaənə >	*bonə* 'eat'
	*k‹um›a-kita	>	*kmakita	>		*baito* 'see'
	*k‹um›aRaC	>	*kmaracə	>		*borcə* 'bite'

Note the following sound changes that have taken place in Tsou: (1) It has developed an echo vowel after the final consonant, and (2) there is a vowel deletion rule predictable from the position of the penultimate stress. The Tsou forms are cited from the Duhtu dialect, in which /r/ was retained at least until early 1980s.

There is synchronic evidence for the change *km > *b* in Tsou, as shown in the alternation of *b ~ ʔ* in the verb *bir-biŋi* (< M, < D, AF), *rimʔ-a* (< M, PF) 'to seek' < *kilim; see Tsuchida (1976:290) for a detailed account.

There is only one example for the origin of /d/ in Tsou (Tsuchida 1972):

(2) *k‹aN›uSkuS[8] 'nail' > *kaNu-kaNuSkuS > *du-duku > *dudku* 'finger'

Dahl's (1981:35) reconstruction for the above cognate form is *kuS_2kuS_2, and the regular reflex of *S_2 is zero in Tsou. However, there is a semantic discrepancy between 'nail' and 'finger', as noted by Tsuchida.

It is not clear when the sound change of preglottalisation took place in Tsou. It should have been at least a few hundred years ago; see section 4 below. The earliest fieldwork on Tsou was done in 1928 by Nevskij (1935), a Russian scholar. The preglottalised stops /ʔb, ʔd/ already existed at that time, as transcribed by Nevskij (1935:26, 29–30).

4 The preglottalised stops in Bunun

As mentioned above, the preglottalised stops /b/ and /d/ in Bunun are historically derived from PAn *b and *d/*D/*Z, respectively, e.g., *qabuH > *qabu* 'ash',

[8] The fossilised infix *‹aN› is attested in several Formosan languages, e.g., RukLa *k‹al›oko-a*, Pai *k‹alj›uskus-an*, Sai *ka-k‹l›okœh*, Kav *q‹n›uqus*, Ami *k‹an›uʔus* 'fingernail'.

*bi(n)tuq-en > *bintuqan* ‘star’, *daqiS > *daqis* ‘face’, *DemDem > *ma-dumdum* ‘dark’, *quZaN > *qudan* ‘rain’.

Preglottalisation of the voiced stops in Bunun is found in all the five major Bunun dialects of Bunun spoken in Nantou (P. Li 1987). Nevertheless, it is not found in some other varieties of Bunun dialects spoken in some other regions. Such a feature is less common as we go down towards the south and part of the east coast. For instance, /b/ and /d/ in the Isbukun dialect spoken in Kaohsiung, southern Taiwan, are not preglottalised at all (Huang 1997:352; Motoyasu Nojima, pers. comm.), nor in the Takbanuaz dialect spoken in Hualian in the east coast (Amy P. Lee, pers. comm.). It is most likely that preglottalisation was lost in the south and part of the east coast after the Isbukun moved southward and the Takbanuaz eastward.[9] Note that /b/ and /d/ in the Taivatan dialect spoken in Hualian are preglottalised. An alternative explanation is much less likely that only some speech communities (or dialects) of Bunun, such as the ones in Nantou, have had close contact with Tsou in history, while the others have not. If that had been the case, Thao borrowing of the feature of preglottalisation from Bunun would have taken place much too recently.

It is quite possible that the preglottalisation of the voiced stops in Bunun, which is restricted to the central area in Nantou, has been diffused from Tsou.

As indicated in the Dutch written documents in the 17th century, Tsou used to be a dominant language, occupying a major part of southern Taiwan, whereas Bunun was only a minor tribe restricted to a small mountain area in Nantou, central Taiwan, about 360 years ago (Mabuchi 1953–1954).

Tsou and Bunun may have been in close contact in the past. For instance, Takopulan, a Bunun speech community which was an enclave in the Tsou area isolated from the other Bunun, was completely assimilated into Tsou by the mid-20th century (Mabuchi 1953–1954).

If Bunun and Tsou have been in close contact for many years, then the mutual influence should be not only phonetic, but also lexical. Unfortunately I have not found much lexical evidence for Tsou influence on Bunun. I have collected about 800 lexical items for Tsou and all Bunun dialects. After a quick check of the data, I found only a few forms suspected to be loanwords either from Tsou to Bunun or from Bunun to Tsou, as given in (3) below.

(3)	Tsou	Bunun	Gloss
	sapiri	*sapil*	‘shoes’
	raŋui	*laŋui*	‘woman’s name’
	ruru	*lulu*	‘shin bone’
	puktu	*puktu* (S)[10]	‘squirrel’
	paŋka	*paŋka*	‘table, chair’
	bahu	*bahu*	‘square basket or box’
	kakatu	*kakatu*	‘spider’
	anika	*anika*	‘female name’

[9] I am indebted to an anonymous reviewer for the suggestion. With regard to preglottalisation, we do not have enough information for all Bunun varieties in the east coast.

[10] The letter S stands for the southern dialect of Bunun, N for the northern dialects, and C for the central.

tumpu	*tumpu*	'place name'
acipa	*qacipa*	'turtle'
udu	*qudu*	'plant sp., *Amaranthus spinosus* L.'
patikewi	*patikau*	'bird sp., *Spizixos semitorques*'
puutu	*puut*	'Taiwanese, Hakka'
mamespiŋi	*maluspiŋ-að* (S)	'woman'
kutʔi	*kuti*	'vagina'
fainə	*vain* (N)	'thigh'
piʔo	*piqa*	'lame'
kakusuŋu	*kakusuŋ* (S)	'shrimp'
sbukunu	*Isbukun* (S)	'Bunun'

Borrowing must have taken place if Tsou and Bunun have identical lexical forms, such as *paŋka* 'table, chair', *bahu* 'square basket or box', *kakatu* 'spider', and *anika* 'female name'. The problem is how we decide which is the donor language and which is the target language. Perhaps most Tsou words with /a/ are loanwords from Bunun relatively recently; *a from an earlier stage most generally became /o/ in Tsou, e.g., *ina > *ino* 'mother', *maCa > *mcoo* 'eye', cf. *ama > *amo* 'father', where the initial *a did not become *o*.[11] As noted in the Austronesian Comparative Dictionary online by Blust, both Tsou and Thao form *kakatu* 'spider', which appears to be a loanword from Bunun *kakatu* 'spider' < *kakaCu.

Let it be noted that there is a contrast between /d/[12] and /r/ in the Duhtu dialect of Tsou. Since there is no /r/ in Bunun, /r/ in Tsou is all realised as /l/ in Bunun, as in the first three examples above. Borrowing was probably from the Duhtu dialect of Tsou (spoken in Nantou) to Bunun.

However, a few items seem to indicate that borrowing took place the other way around, e.g., Bunun *qacipa*, Tsou *acipa* 'turtle', Bunun *qudu*, Tsou *udu* 'plant sp.' There is no /q/ in Tsou. Where would Bunun get its /q/ in these forms if borrowing was from Tsou? Moreover, Tsou /c/ does not regularly correspond to Bunun /c/. It is, therefore, more likely that Tsou borrowed these two items from Bunun.

The cognate form for 'squirrel' is *buhut* (< PAn *buhet) or its variant forms in many Formosan languages. It is *puhut* in all Bunun dialects except Isbukun, which has the form *puktu*, identical with the Tsou form. The latter is thus believed to be a loanword from Tsou.[13]

If the other items are Bunun loanwords from Tsou, then the borrowing must have taken place before (1) Tsou developed an echo vowel in the word-final position, as in the words for 'shoes', 'thigh' and 'shrimp', and (2) devoicing of /v/ to /f/ in Tsou, as in the word for 'thigh'. Alternatively Tsou *fainə* is a loan from Bunun *vain* 'thigh', because there is no /v/ in Tsou.

---

[11] Tsuchida (1976:279) provides an explanation for why in certain phonetic environments *a does not become *o*.

[12] Tsou /d/, which is phonetically preglottalised, is also written as /l/ in the orthography of Tung (1964) and a few others. Wright (1999) shows that *l* is an allophone of the preglottalised *d* before /a/.

[13] An anonymous reviewer said that the form *puktu* developed independently in Isbukun and Tsou, but unfortunately he did not show how it did.

The main problem with these loanwords is that most of them do not contain the preglottalised stops /b/ and /d/ to account for the diffusion of the feature.

5 The preglottalised stops in Thao

Thao has numerous lexical forms containing the two preglottalised stops, but none of them can be demonstrated to be historically derived by regular sound changes. Many of them can, in fact, be shown to be loans from Bunun, as their lexical forms are identical or have close resemblance in these two languages (P. Li 2013). Thao and Bunun have been in close contact in the past few hundred years, and intermarriage between the peoples is very common, mostly Bunun women married into Thao families living in Thao villages (P. Li 2011:4). A few examples of Thao loanwords from Bunun are given in (4) below:

(4)

Bunun	Thao	Glosses
biqi	*biqi*	'goiter'
bataqan	*bataqan*	'a wooden frame to carry things'
baluku	*baruku*	'bowl'
binanauʔað	*binanawʔað*	'woman'
bantac (N), *bantas*	*bantaθ*	'leg'
tumbus	*tumbuʃ*	'body louse'
ma-daidað	*ma-daydað*	'love'
ma-dumdum	*dumdum*	'dark'
daudauk	*mun-daudauk*	'slow'
qalidaŋ	*qaridan*	'pigeon pea, bean sp.'
ledaðuʔan	*layʔaðuan*	'a Bunun village name'

In fact, Thao has extensive lexical and phonological borrowing from Bunun. There are hundreds of Thao loanwords from Bunun. The consonants /b, d, l, ʔ, h/ have acquired their phonemic status in Thao, due to the great number of loans from Bunun. See P. Li (2013) for more details.

On the other hand, Bunun may also have borrowed a few words from Thao, e.g., Thao *qutaʃ* (< PAn *qutaS), Bunun *qudas* 'grey hair'. While the Thao form is a regular derivation from PAn, *d* in the Bunun form is irregular.

6 Conclusion

With facts such as given above, we can get a reasonably clear picture how the phonetic feature of preglottalised stops originated in one language, Tsou in this case, and then diffused to the other two, one after another. However, there is not sufficient lexical evidence for Tsou influence on Bunun. Hence we cannot exclude the possibility that there might be a parallel development in Bunun.

References

Aikhenvald, Alexandra Y. and R.M.W. Dixon, eds. 2006. *Grammars in contact: A cross-linguistic typology*. Explorations in Linguistic Typology 4. Oxford: Oxford University Press.

Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

Blust, Robert. 2003. *Thao dictionary*. Language and Linguistics Monograph Series A-5. Taipei: Academia Sinica.

Dahl, Otto Christian. 1981. *Early phonetic and phonemic changes in Austronesian*. Oslo: The Institute for Comparative Research in Human Culture.

Huang, Lillian M. 1997. gaoxiongxian de bunongyu [(Isbukun) Bunun in Kaohsiung County]. *gaoxiongxian nandao yuyan* [*Formosan Austronesian Languages in Kaohsiung County*], ed. by Paul Jen-kuei Li, 351–410. Kaohsiung: Kaohsiung County Government. [In Chinese]

Li, Fang-Kuei. 1943. The hypothesis of a pre-glottalized series of consonants in primitive Tai. *Bulletin of the Institute of History and Philology, Academia Sinica* 11:177–188.

Li, Fang-kuei, Chi-lu Chen and Mei-chün Tang. 1956. shaoyu jilue [Notes on the Thao language.] *Bulletin of the Department of Archaeology and Anthropology, National Taiwan University* 7:23–51

Li, Paul Jen-kuei. 1987. The preglottalised stops in Bunun. In *A world of language: Papers presented to professor S. A. Wurm on his 65th birthday*, ed. by Donald C. Laycock and Werner Winter, 381–387. Pacific Linguistics C-100. Canberra: The Australian National University.

Li, Paul Jen-kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59.2:479–508.

Li, Paul Jen-kuei. 2011. *Thao texts and songs*. Language and Linguistics Monograph Series 44. Taipei: Academia Sinica.

Li, Paul Jen-kuei. 2013. Thao loans from Bunun. *Bulletin of Chinese Linguistics* 7:225–241.

Mabuchi, Tōichi. 1953–1954. takasagozoku no idou oyobi bunpu [Migration and distribution of the Formosan aborigines]. *Minzokugaku Kenkyu* [*Japanese Journal of Ethnology*] 18.1/2:123–154, 18.4:23–72.

Nevskij, N. A. 1935. *Materialy po Govoram Jazyka Cou* [*Dialect material of the Tsou language*]. Moskva: Trudy Instituta Vostokovedenija 11.

Ogawa, Naoyoshi and Erin Asai. 1935. *Gengo ni yoru Taiwan takasago-zoku densetsu-shu* [*The myths and traditions of the Formosan native tribes* (*texts and notes*)]. Taihoku [Taipei]: Taihoku Imperial University. [In Japanese]

Tsuchida, Shigeru. 1972. The origins of the Tsou phonemes /b/ and /d/. *Gengo Kenkyu* 62:24–35.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No. 5. Tokyo: University of Tokyo.

Tung, T'ung-ho, with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng, and Margaret M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publications 48. Taipei: Academia Sinica. Reprinted in 1991 with notes and errata prepared by Paul Jen-kuei Li.

Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. Ithaca, NY: Cornell University Southeast Asia Program Publications.

Wright, Richard. 1999. Tsou consonant clusters and auditory cue preservation. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds. *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 277–312. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

3 *Syllable types in Bunun, Saisiyat, and Atayal*[*]

HUI-CHUAN J. HUANG

1 Introduction

Under the assumption that a prosodic word consists of a string of syllables, word edges have been known to exhibit exceptional syllable types. The non-isomorphemism between internal and marginal syllables has been identified in many languages. For example, Borowsky (1989) shows that in English, medial syllables permit VV and VC (ignoring syllable onsets for the moment), but in general not VVC or VCC. The majority of counterexamples (e.g., *ch**am**ber, **an**gel, **an**cient, sh**oul**der, **ant**ler, p**ump**kin, **arc**tic*) contain homorganic nasal+obstruent or sonorant+obstruent sequences. True counterexamples such as *d**eic**tic, d**eix**is, s**eis**mic, abs**orp**tion/abs**orp**tive,* and *sc**ulp**ture/sc**ulp**tor* are rare. However, in English final syllables, VVC and VCC syllables are abundant. In the literature of generative phonology, relevant data are often analyzed by invoking concepts such as extrasyllabicity (e.g., Steriade 1982; Clements & Keyser 1983, among many others). The additional consonants may be analyzed as outside of the final syllable in various representational approaches, in order not to conflict with the canonical syllable shapes in the language.

Although how to analyze exceptional phonotactics at right word edges has received much attention in phonological theorisation (Rubach & Booij 1992; Piggott 1999; Broselow 2003; Rice 2003; Côté 2011; among others), most previous descriptions of the phonology of Formosan languages have remained inexplicit regarding the distribution of coda consonants permitted in nonfinal versus final positions. Relevant observations are

[*] It is my great pleasure to dedicate this paper to Professor Lillian M. Huang. An earlier version of the paper was presented at the workshop on "A Typological Study of Austronesian Languages in Taiwan and Their Revitalisation", which was held at Academia Sinica on 1–2 December 2012. I thank the audience there for their comments and questions. Special thanks go to Marie M. Yeh for discussing the Saisiyat data with me and for her help in many ways. I would also like to express my sincere gratitude to the anonymous reviewers, to the editors for carefully reading the manuscript and offering many helpful suggestions, and to Professor Paul Li for his valuable comments. The usual disclaimer applies. The research reported here has been funded partially by National Science Council of Taiwan (NSC 99-2410-H-001-113, NSC 100-2420-H-001-022-MY3).

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 47-74
Asia-Pacific Linguistics, 2015

often stated in the forms of possible strings of consonants and vowels in a morpheme or a word, or in the forms of syllable templates without mentioning potential word edge effects. For example, the canonical forms of Squliq Atayal are given in Li (1980:356–357) as CV, CVC, CCVC, CCCVC, CCCCVC, CVCV, CVCVC, CCVCVC, CCCVCVC, CVCCVC, C(C)VCVCVC, and CVCCVCVC, while Paiwan syllable structure is described as (C)V(C) in Ho (1994:310). More detailed discussions regarding the phonotactics of Formosan languages are therefore called for. This paper aims to clarify possible syllable types in medial and final positions in three Formosan languages—Bunun, Saisiyat, and Atayal—with the hope that future typological studies will yield a more extensive survey.[1]

Descriptions of syllable types in different positions offer the information regarding the phonotactics of a language, but phonotactics can be alternatively described by simply stating possible segmental sequences that are permitted in medial and marginal positions. For example, Blust (2009:207–213) points out that consonant clusters in most Austronesian languages, if allowed, occur only in medial position. Blust (ibid.) classifies the patterns of two-member medial consonant clusters from a typological perspective, showing that languages may differ in terms of whether they allow medial clusters in reduplicated monosyllables, prenasalised obstruents, geminates, or heterorganic clusters in non-reduplicated bases. Blust (ibid.) further discusses the proportion of permissible final consonants to total consonants in six languages, showing various degrees of final contrast reduction.

Whether consonantal phonotactics should be accounted for in a syllable-based or string-based approach is a controversial issue. To capture the fact that the permitted syllable types in internal and marginal positions are not necessarily the same in a given language, followers of the syllable-based model often attribute the possible asymmetry to some mechanisms or constraints (e.g., Broselow 2003; Wiltshire 2003; Flack 2009; Byrd 2010). Proponents for the string-based model (e.g., Blevins 2003; Steriade 1999, 2001), however, have argued that consonantal phonotactics are determined by the robustness of perceptual cues for a feature in its context, which suggests that using the syllable alone to state phonotactics is inadequate. In a diachronic account of phonotactic development in Latin, Sen (2011) shows that linear segmental sequences determine the patterns of consonantal assimilation whereas syllable structures play only an indirect role in the choice of allophones, which implies that both linear and syllabic information are required to certain extents in characterizing the phonotactics of a language.

Broselow (2003) argues that the string-based approach is clearly problematic for languages such as Balantak (Austronesian) and proposes that interactions of syllabic constraints and position-specific faithfulness constraints, which protect edges or interiors of morphological constituents from being changed, give rise to the attested language typology. In the case of Balantak, phonetic contexts alone cannot account for the wider

---

[1] The three particular languages are chosen for the present study to illustrate different patterns because the author has direct fieldwork experience with these languages. Among the other Formosan languages, Paiwan seems to represent another distinct type of pattern by allowing only nasals and glides as codas of medial syllables but a variety of consonants in final position (S. Yeh 2011). Moreover, Kavalan behaves differently by containing intervocalic geminate consonants, as can be seen in the dictionary data (Li & Tsuchida 2006).

range of contrasts permitted in the first member of a CC sequence across root-suffix boundaries because the more restricted set of morpheme-internal consonants occur in exactly the same contexts. In the present study, the consonantal phonotactics of the three Formosan languages will be stated in terms of syllable types and the concepts of symmetrical/asymmetrical patterns as stated in Broselow (ibid.). In the languages that symmetrically allow codas in both internal and final positions, whether there is asymmetry in the range of possible feature contrasts in the two positions will be examined. Since cue-based studies have observed that the perceptibility of a feature contrast is tied to the environment where it occurs, the present study will also note if there are restrictions on medial consonant clusters, similar to the approach in Blust (2009).

In Broselow (2003), a set of constraints within the framework of Optimality Theory (OT; Prince & Smolensky 1993/2004; McCarthy & Prince 1995) is proposed to account for a sample of languages showing various symmetrical or asymmetrical patterns. The OT constraints predict that there should be languages in which closed syllables are banned in medial position but are required word-finally. However, Broselow (ibid.) has not found such languages in her survey and comments that this is a potentially problematic gap for the predicted typology. It will be shown that the conservative form of Squliq Atayal in Jianshi discussed in section 6 potentially fills the typological gap in Broselow's study and therefore strengthens the syllable-based analysis.

The organisation of the paper is as follows. In section 2 the classification of symmetrical and asymmetrical patterns as discussed in Broselow (2003) will be summarised, followed by a discussion of the Bunun data in section 3, Saisiyat in section 4, Mayrinax Atayal in section 5, and two Squliq Atayal varieties in section 6. In the Saisiyat section, the patterns of early Saisiyat as described in Li (1978) are compared with the Saisiyat data collected for the present study, showing the influence of the loss of flap consonants on syllable types. In the section on Squliq Atayal, two varieties of Squliq are included, highlighting the subdialectal differences among Squliq which have not been revealed in previous studies. Section 7 discusses related issues and concludes the paper.

2 Symmetrical vs asymmetrical patterns

Since it is known from the phonology literature that the possible segmental sequences in word-initial and word-final positions may not be identical to those allowed in, respectively, word-internal syllable onset and coda positions, it is helpful to first of all identify whether a language exhibits the fit between interior/edge patterns. Focusing on the right edge, Broselow (2003) examines languages that allow coda consonants only at edges (CV.CV**C**) or internally (CV**C**.CV), as well as the symmetrical languages (CV**C**.CV**C**) to see whether a wider range of place and/or manner contrasts are allowed in one position than in another.

In symmetrical languages, structures found at word-final and word-medial positions are the same. For example, Hawaiian syllables are uniformly open, while Manam uniformly allows both open and closed syllables in interior and right edge positions, so both languages are classified as symmetrical. In terms of OT constraints, NoCoda ranks higher than faithfulness constraints in Hawaiian while in Manam, NoCoda is lower than faithfulness constraints:

(1) Symmetrical patterns: (adapted from Broselow 2003:162)

Inside	Edge	Language	Syllable description
V.	V#	Hawaiian	Uniformly open
V.	V#	Manam	Uniformly faithful (open or closed)
VC.	VC#		

Under the assumption that words simply consist of possible syllables, symmetry is expected of word-final and preconsonantal positions, but in some languages, structures found at the right edges are not the same as those in preconsonantal positions. Some of these 'asymmetrical' languages allow an extra consonant at the right word edge (e.g., Kamaiurá, Yapese), while others permit additional elements in internal position (e.g., Diyari).

(2) Asymmetrical patterns: (adapted from Broselow 2003:163–165)

Inside	Edge	Language	Syllable description
Right edge more marked			
V.	V#	Kamaiurá	Edge: Coda possible (open or closed)
	VC#		Inside: Coda not possible (open)
V.	VC#	Yapese	Edge: Coda required (closed)
VC.			Inside: Coda possible (open or closed)
Right edge less marked			
V.	V#	Diyari	Edge: Coda not possible (open)
VC.			Inside: Coda possible (open or closed)

Broselow (ibid.) proposes that the following three OT constraints are responsible for the asymmetrical patterns, all of which are position-specific in nature:

(3) Position-specific constraints:
   a. Anchor-R(GrW): A segment at the right edge of the grammatical word in the output has a correspondent at the right edge of the grammatical word in the input (McCarthy & Prince 1995)
   b. I-O Contiguity: The correspondents of a contiguous input string must be contiguous in the output, and vice versa.
   c. Final-C: Prosodic words must end in consonants (McCarthy 1993; McCarthy & Prince 1993).

It is argued in Broselow (ibid.) that while coda markedness constraints give rise to symmetrical patterns, some position-specific constraints in (3) are high-ranking in the asymmetrical types of languages and lead to the occurrence of structures at edges not found in interior position, or vice versa. The rankings of Anchor-R >> NoCoda >> Faithfulness in Kamaiurá account for the possibility of having final consonants but no CC sequences anywhere, and the rankings of Contiguity >> NoCoda >> Faithfulness in Diyari are responsible for marked closed syllables in medial position. In Yapese, the obligatory word-final consonants come from the rankings of Final-C >> Faithfulness >> NoCoda.[2]

[2] There are some discrepancies in the literature regarding whether Yapese contains word-internal codas. Jensen (1977:47) generalises that all major morphemes in Yapese consist

Broselow (ibid.) discusses a problematic gap in the typology predicted by the constraints in (3), and states that the gap would necessitate a revision of the theory if it is proven to be an impossible pattern. The gap corresponds to a type of language predicted by the rankings Final-C >> NoCoda >> Faithfulness, which generally forbid closed syllables but require all words to end with consonants. In section 6.1 the conservative forms of Squliq Atayal in Jianshi Township, Hsinchu County (as well as Fuxing District, Taoyuan City), which display such a pattern, will be discussed.

Broselow (ibid.) further examines languages which symmetrically allow codas in both word-internal and word-final positions to see if they exhibit asymmetrical patterns in terms of possible feature contrasts. In languages such as Eastern Ojibwa, internal CC is limited to nasal-stop or fricative-stop sequences, although final consonants are unrestricted (Piggott 1999). Conversely, in languages such as Lardil (Hale 1973), word-final nasals are limited to a single place or articulation, but internal nasals show more variety because they are homorganic with the following obstruents.

(4) Asymmetrical patterns in terms of feature contrasts (adapted from Broselow 2003:167)

Inside	Edge	Language	Description
V.CV VN.TV[3] VS.TV	V# VC#	Eastern Ojibwa	Edge: more contrasts
V.CV VC.CV	Vn#	Lardil	Edge: fewer contrasts

In languages such as Eastern Ojibwa, a high-ranking Syllable Contact constraint (SyllContact), which requires that the last segment in a syllable must be higher in sonority than the first segment of the onset of the following syllable, reduces the possible contrasts in internal but not final codas. In languages such as Lardil, an Agree-Place constraint dictates that nasals must share place with a following obstruent, and the rankings of AgreePlace >> CodaMarkedness >> Faithfulness lead to the wider varieties of nasal consonants in internal position. There are other types of constraints whose high-ranking status leads to asymmetrical feature contrasts in edge/interior positions; readers are referred to Broselow (ibid.) for the details of the other cases.

The idea of classifying languages as either symmetrical or asymmetrical patterns would clarify the similarities and differences in the phonotactics of Formosan languages.

---

of groups of open or closed syllables and that the last syllable is always closed. Piggott (1999), citing Jensen (ibid.), states that nonfinal syllables of Yapese are basically open but syllables closed by liquids might be possible (p. 145). Broselow (2003), citing both Jensen (ibid.) and Piggott (ibid.), classifies Yapese as the type of language which permits nonfinal codas. Côté (2011:853–857), citing Piggott's and Broselow's work, states that Yapese has no internal codas. It will be pointed out in section 4 that early Saisiyat displays an asymmetrical pattern just as the illustrating case of Yapese discussed in Broselow (ibid.); the classification is based on the assumption that Yapese tolerates internal codas because Saisiyat clearly permits internal closed syllables.

[3] The symbol N.T stands for nasal-stop sequences and S.T below means fricative-stop sequences.

In the three Formosan languages to be examined below, it will be shown that Isbukun Bunun resembles Manam in symmetrically allowing open and closed syllables. While Saisiyat at an earlier stage resembles Yapese in requiring word-final coda consonants, modern Saisiyat has become a symmetrical type of language through the historical loss of flap consonants. Mayrinax Atayal allows many internal coda consonants due to a historical syncope process, showing more of a symmetrical pattern, but the two varieties of Squliq dialects in Hsinchu County exhibit asymmetrical patterns. The data of these languages are discussed in detail in the following sections.

3 Isbukun Bunun

Isbukun Bunun has three vowels /i u a/ and fourteen consonants /p t k ʔ b d s h v ð l m n ŋ/ in its phonemic inventory (He et al. 1986; H. Lin 1996; Li 1997; L. Huang 1997; Zeitoun 2000). The syllables of Isbukun have been characterised as CVX (H. Lin 1996:32) or CV(C) (Li 1997:306; Zeitoun 2000:46), or maximally as four slots CVVC (He et al. 1986:7) or CGVC/CVGC in which the surface glides come from underlying vowels (H. Huang 2008:3).[4] Postconsonantal onglides (e.g., in CGVC) are in the nucleus rather than in the onset based on evidence from stress assignment (H. Huang 2005). Complex syllable margins (i.e., CC) are not permitted. Syllables in Isbukun allow up to four segments, which can be found in both nonfinal and final positions.

It has been mentioned in the literature that all Isbukun consonants can appear in initial, medial, and final positions (Li 1997:302; Zeitoun 2000:42). However, as far as the medial position is concerned, the illustrating examples in previous studies usually include either onset or coda positions, so it is unclear whether all consonants can appear as codas of nonfinal syllables. Blust (2009:207) includes Bunun in his typological survey of medial consonant clusters in Austronesian languages and states that Bunun allows clusters in reduplicated monosyllables and hetero-organic clusters in non-reduplicated bases but prohibits prenasalised obstruents and geminates.

Blust's (ibid.) statements have implied that Bunun is a language that tolerates many different kinds of nonfinal coda consonants. The data below show that all consonants can appear in the codas of word-final syllables and that almost all the consonants can be internal codas.[5]

---

[4] While the symbol C stands for a consonant, including true consonants and glides, the symbol G is used to represent a surface glide throughout the paper whenever there is a need to exclude the possibility of a true consonant in statements.

[5] While H. Lin (1996), Li (1997), and Zeitoun (2000) describe the Nantou variety of Isbukun Bunun, the data recorded by He et al. (1986) and L. Huang (1997) are based on the Isbukun variety spoken in Kaohsiung (spelt as gaoxiong in pinyin). The data below are based on the author's field work on Kaohsiung Isbukun, or words from the literature with confirmation of native speakers. The transcriptions here are broad in the sense that surface glides are all given in the forms of underlying vowels and that allophonic processes such as palatalisation are not shown; for example, (5m) *hansiap* in narrow transcriptions would be [han.ʃjap].

(5) Isbukun Bunun word-internal and word-final codas[6]

		Word-internal codas		Word-final codas	
a.	/p/	tau**p.**las	'radish'[7]	ʔai**p**	'today'
b.	/t/	ma.pi**t.**ʔia	'cook (AV)'	ka.la**t**	'bite'
c.	/k/	ta**k.**na	'yesterday'	tul.ku**k**	'chicken'
d.	/ʔ/	(ma**ʔ.**maʔ)	'tongue'	(maʔ.ma**ʔ**)	'tongue'
e.	/b/	ma.ŋa**b.**ðan	'wide'	kit.ŋa**b**	'begin'
f.	/d/	su**d**.nut	'flood'	hu**d**	'drink'
g.	/s/	i**s.**mut	'grass'	ŋu.lu**s**	'mouth'
h.	/h/	ma.sau**h.**ðaŋ	'hungry'	su.pa**h**	'many'
i.	/v/	(lu**v.**luv)	'wind'	saŋla**v**	'vegetables'
j.	/ð/	ma**ð.**ʔav	'shy'	ʔu.va**ð**	'child'
k.	/l/	du**l.**pin	'sparrow'	ma.pu.tu**l**	'short'
l.	/m/	tu**m**.bus	'louse'	da.nu**m**	'water'
m.	/n/	ha**n.**siap	'understand'	mas.ʔa**n**	'ten'
n.	/ŋ/	ha**ŋ.**vaŋ	'cow'	has.bi**ŋ**	'sneeze'

Some data are placed in parentheses because they are not considered satisfactory examples illustrating legitimate codas in the phonology of the language. Based on an examination of the author's own field notes and the data in previous studies, including He et al. (1986); Li (1988); H. Lin (1996); T. Lin et al. (2001), it appears that internal coda consonants are somewhat more restricted than word-final consonants in Isbukun. The glottal stop /ʔ/ and labio-dental fricative /v/ seem to show up as internal codas only if they precede a morpheme boundary or end the first half of a word that has the shape of fossilised reduplication. In the data in (5), words of fossilised reduplication (e.g., (5d) *maʔ.maʔ* 'tongue' and (5i) *lu**v.**luv* 'wind') are forced to be chosen to illustrate internal /ʔ v/ codas because words in which medial /ʔ v/ followed by another consonant without a morpheme boundary in between have not been found.

It is well known that many phonological processes, including syllabification, are sensitive to morphological structures (Laeufer 1995; Noske 1999; Kaisse 2005; Byrd 2010; among others). Since morphology may interfere with syllabification, exceptional syllable types are often found at morpheme boundaries. It is also known that reduplication may exhibit unusual phonology (Raimy 2011). The exceptional patterns in reduplication are often attributed to some identity constraints which hold between the base and the reduplicant (Wilbur 1973; McCarthy & Prince 1995). Fossilised reduplication is said to reflect earlier productive reduplication patterns (Blust 1988), so it would not be surprising to find unexpected phonotactics in words of reduplicated shapes, which is why reduplicated and non-reduplicated bases often need to be distinguished when stating permissible consonant clusters in a word (e.g., in Blust 2009). Therefore, in

---

[6] The abbreviations used in the paper follow the Leipzig Glossing Rules. Those which are not found there are as follows. AV: actor voice; BV: beneficiary voice; NAV: non-actor voice.

[7] In consonant sequences such as *pl* in (5a) and *sm* in (5g), syllabification is justified by the phonological distribution of the consonants: syllable boundaries fall between the two consonants because the same sequences do not appear in word-initial positions.

forming generalisations regarding internal codas, the present study will exclude fossilised reduplication and those forms in which the internal codas are immediately followed by a morpheme boundary, such as in compounds (*mað.ʔa**v.**dai.ŋað* 'very shy') or in affixed/cliticised forms, in order to factor out the effects of morphological structures. The phonologically exceptional syllable types or syllabification induced by morphology can be attributed to some high-ranking alignment constraints on the edges of morphological and prosodic constituents (e.g., Noske 1999). To summarise, the paper would like to reveal the purely phonological generalisations regarding the shapes of syllables; influences from morphology, which might obscure the phonological generalisations, can be separately dealt with by invoking other higher-ranked constraints and therefore will be left aside in the present study.

One thing worth mentioning in the Isbukun data is that although glottal stops in internal codas might be restricted, glottal stops in final positions are in fact limited, too. Word-final glottal stops have been transcribed for Isbukun in the literature, for example *vauʔ* 'eight' in Li (1988), which focuses on the Nantou variety of Isbukun, but the corresponding words in the Isbukun spoken in Kaohsiung seem to be vowel-ending, as suggested by the lack of glottal stops in both unsuffixed and suffixed forms (e.g., *vau* 'eight' and *p-in-vau-un* [pin.vaw.wun]). In He et al. (1986) and T. Lin et al. (2001), both of which focus on the Kaohsiung variety of Isbukun, almost all words ending with a glottal stop have the shape of duplicated monosyllables, such as *luʔ.luʔ* 'wound' (He et al. 1986:114) and *suʔ.suʔ* 'nipple' (T. Lin et al. 2001:39).[8] The presence of word-final glottal stops in this type of word is presumably due to the tendency to keep corresponding consonants in reduplicated forms; in words that do not have the shape of fossilised reduplication, final glottal stops are usually dropped. If glottal stops tend to get lost word-finally, the absence of glottal stops in internal preconsonantal position in fact does not induce asymmetry.

Notice that internal codas would not be the same as the following onsets in Isbukun, as suggested by the lack of geminates stated in Blust (ibid.).[9] The avoidance of identical consonant sequences can also be observed in the loss of consonants in the concatenation of morphemes, such as *ʔis-suað* becoming *ʔisuað* 'plant' (Li 1997:305). The data suggest that there is a general tendency to avoid identical consonant sequences both within and across morphemes in Isbukun; this tendency is responsible both for the absence of internal codas followed by identical onsets, and for the segment deletion rule as stated in Li (ibid.).[10]

---

[8] There are two words in T. Lin et al. (2001), which end with a glottal stop but do not have the shape of duplicated monosyllables: *taŋiaʔ* 'ear' (p.10) and *iahluaʔ* 'ant' (p.13). (The lateral consonant is consistently written as [ɬ] in He et al. (1986) and T. Lin et al. (2001) because it is voiceless for some speakers; the symbol has been simplified to a regular *l* here.)

[9] The other languages to be discussed below, including Saisiyat, Mayrinax Atayal and two varieties of Squliq, all lack geminate consonants.

[10] The segment deletion rule in Li (1997:305–306) applies to both consonants and vowels. In the case of identical vowel sequences in the input, H. Huang (2008) has argued that vowel coalescence rather than vowel deletion has taken place, because they count as bimoraic in stress assignment. As to the avoidance of identical consonant sequences, this issue awaits more research to justify whether deletion or coalescence takes places in examples such as *ʔis-suað* becoming *ʔisuað*. Perhaps an acoustic measurement of the duration of deleted/coalesced consonants, in contrast with regular single consonants, would provide a clue.

Open syllables in Isbukun are found in nonfinal positions as shown in (5), and they appear in final positions, too, as implied in the above discussion regarding the dropping of final glottal stops in the language. Although in terms of permissible features internal coda consonants are somewhat more restricted than word-final consonants,[11] Isbukun Bunun in general exhibits the symmetrical pattern because both open and closed syllables appear in nonfinal and final positions and syllables across positions uniformly obey the upper limit of four segments. The OT constraints Faithfulness >> NoCoda account for the Isbukun data as in the case of Manam, and the three position-specific constraints leading to the asymmetrical types are all ranked low.

4 Saisiyat

Saisiyat has two dialects, the northern Taai (Daai in pinyin) dialect in Hsinchu County and the southern Tungho (i.e., Donghe) dialect in Miaoli (Tsuchida 1964; Li 1978). The inventory of the language contains sixteen consonants /p t k ʔ b[β] z s ʃ h m n ŋ r l j w/ and six vowels /i œ ə o æ a/ (Chao 1954; Zeitoun et al. 2011).[12] Tsuchida (1964:43) states that the main phonetic difference between the two dialects is the loss in Tungho of the voiced dental/uvular fricative, which corresponds to the description of the flap consonant in other studies such as Li (1978). The flap consonant was still documented in Taai in Li (1978) although it was already lost in the Tungho dialect at that time. M. Yeh (1991, 2000, 2003) include the flap consonant /ɭ/ in the inventory for the Taai dialect and comments that it is retained in some words by a couple of speakers (M. Yeh 2000:47). At the present time it seems that Taai speakers who consistently pronounce the flap consonants are very difficult to come by. In addition to the loss of the flap consonant, the phonetic differences between the two dialects that are frequently discussed in the literature include the articulation of coronal fricatives /s z/ in Taai as interdentals /θ ð/ in the Tungho dialect (Tsuchida 1964; Li 1978; M. Yeh 2000; Kaybaybaw 2009; Zeitoun et al. 2011).[13]

As far as syllables are concerned, M. Yeh (2000) states that Saisiyat allows both CV and CVC syllables. Zeitoun & Wu (2005) point out that there are three syllable types: CV, CVC, and CVV (e.g., *tal.kaa* 'table'). The so-called CVV syllable type results from the loss of the flap consonant in the coda of CVC accompanied by lengthening, so it is conceptually equivalent to a syllable with long vowels (Deng 2007; Zeitoun et al. 2011).[14] For example, the word *tal.kaa* 'table' was originally pronounced as *tal.kaL* as documented by Li (1978), in which the symbol *L* stands for the flap consonant in the coda,

---

[11] The data also show that the distribution of the coda consonant /m/ in nonfinal syllables seems to be affected by the following consonants: it is mostly followed by homorganic /p b/ (excluding compounds and fossilised reduplication), in contrast with /n ŋ/, which can appear before heterorganic consonants. The distribution of nasal codas awaits more data collection in the future.

[12] The two cited sources focus on the Tungho dialect of Saisiyat.

[13] For other minor phonetic differences between the two Saisiyat dialects, please see Li (1978:135–138).

[14] Based on acoustic studies on the Tungho dialect, Chang (2009) has found that the loss of the flap consonant has induced diphthongisation of the preceding vowel, which leads to the impression of the so-called 'long' vowels.

but it has changed to [tal.kaː] (e.g., Zeitoun et al. 2011:288).[15] Because consonant clusters are not permitted in syllable margins and long vowels do not precede coda consonants for historical reasons, syllables in Saisiyat are maximally three segments long in both medial and final positions.

Although many previous studies on Saisiyat have given lists of words illustrating the consonants in initial, medial, and final positions, for medial positions they often lump together examples of internal onsets and codas (as in the literature of Isbukun Bunun); either a consonant in intervocalic CC or an intervocalic single (onset) consonant is used to illustrate the consonant in medial position. Therefore, it has been unclear whether there is edge/interior asymmetry in the distribution of syllable types in Saisiyat. An examination of the data shows that modern Saisiyat permits both open and closed syllables in nonfinal and final positions. The following data from Taai Saisiyat illustrate the symmetrical distribution of closed syllables.

(6) (Modern) Taai Saisiyat word-internal and word-final codas[16]

		Word-internal codas		Word-final codas	
a.	/p/	ʔæ**p**.hæl	‘vapor’ (Li 1978)	ʔa.lə**p**	‘wipe (IMP.AV)’
b.	/t/	ta**t**.poʔ	‘hat’	ʃə.βə**t**	‘beat’
c.	/k/	ka**k.**sis	‘itchy’	sa.zə**k**	‘smell’
d.	/ʔ/	kæ**ʔ**.sæʔ	‘husk of rice’(Li 1978)	ʔæ.hæʔ	‘one’
e.	/β/	ka**β**.hœt	‘squirrel’	(h)a.ʃa**β**	‘saliva’
f.	/z/	(ʔæ**z**.ʔəz)	‘bitter’[17]	laŋ.pə**z**	‘ten’
g.	/s/	βi**s**.kol	‘stomach’	ŋa.βa**s**	‘mouth’
h.	/ʃ/	ka**ʃ**.koj	‘sponge gourd’ (Li 1978)	ti.iʃ	‘wipe’
i.	/h/	kæ**h**.maʔ	‘tongue’	sa.βœ**h**	‘all, together’
j.	/m/	ri**m**.ʔan	‘tomorrow’	ki.i**m**	‘seek’
k.	/n/	ʔa**n**.hiʔ	‘bamboo shoots’ (Li 1978)	wa.li.ʃa**n**	‘wild pig’
l.	/ŋ/	si**ŋ.**sis	‘burned rice that sticks to the pan’	kor.ko.ri**ŋ**	‘child’
m.	/r/	pə**r**.ʔœʔ	‘vegetables’[18]	ʔæw.pi**r**	‘sweet potato’
n.	/l/	ka**l**.hiβ	‘cave under a rock’ (Li 1978)	raj.hi**l**	‘money’
o.	/j/	ka**j**.zæh	‘good’	pa.za**j**	‘cooked rice’
p.	/w/	ta**w**.moʔ	‘banana’	ʔæ.la**w**	‘fish’

The data in (6) show that even when the focus is restricted to internal codas (i.e., the first consonant in intervocalic CC), almost all consonants in Saisiyat can still be found in this position. Recall that only those words without a morpheme boundary between the

---

[15] Zeitoun et al. (2011) distinguishes sequences of CVː from CVV (e.g. [tal**kaː**] ‘table’, [ko**maa**t] ‘write (AV)’), which is not recognised in Zeitoun & Wu (2005).

[16] The data are based on the author’s fieldwork on Taai Saisiyat unless noted otherwise.

[17] Words with the internal coda /z/ seem to be rare. It is possible that [ʔæ**z**.ʔəz] ‘bitter’ is composed of reduplication of a /z/-ending syllable plus prefixation historically, as suggested by the (broad) transcriptions [ʔæLəzʔiz] (Taai) and [ʔææzʔəz] (Tungho) in Li (1978:185), and [ʔæəðʔəð] in Zeitoun et al. (2011:296). If so, the appearance of /z/ as an internal coda needs to be justified by better examples in future research.

[18] The word is sometimes pronounced as [pœ**r**.ʔœʔ].

intervocalic CC are used as examples in order to avoid syllable types that might be found only in morphologically complex forms.

As far as open syllables are concerned, they appear in nonfinal positions, as shown in (6), and they are found word-finally, too. Word-final open syllables in Saisiyat are mostly long vowels which come from the historical loss of the flap consonant in final position, as discussed above. The symbol *L* is used in the following data as in the study by Li (1978):

(7) (Modern) Taai Saisiyat word-final open syllables with long vowels

	In the present study	Li (1978)	Gloss
a.	*talkaː*	*talkaL*	'table'
b.	*ʔinaroː*	*ʔinaroL*	'water pipe'
c.	*ʃajβoʃiː*	*ʃajβoʃiL*	'six'
d.	*kərpəː*	*kərpəL*	'fat'
e.	*kaβkaβæhæː*	*kaβkaβæhæL*	'bird'
f.	*r‹om›aʔœː* [19]	*r‹om›œʔœL*	'drink (AV)'

Syllables with long vowels as the result of flap consonant loss are quite abundant in word-final position. However, internal syllables with historical loss of coda flaps may not be long for some speakers (Hsieh 2007), or the vowel length may vary, so *βiLwaʔ* 'thunder' in Li (1978:145), for example, is produced as either [βiːwaʔ] or [βiwaʔ] in the Taai dialect at the present day.[20]

CV open syllables are common in nonfinal position as expected, but word-final CV syllables are in fact unusual in content words and are restricted to function words such as *jami* 'we (exclusive)'and *ʃoʔo* 'you'.[21] Leaving aside the facts that word-final CV syllables are uncommon and that internal CVV syllables may exist only for certain speakers, we can conclude that modern Saisiyat is a symmetrical language which permits both open and closed syllables in medial and final positions.

The patterns in the earlier stage of Saisiyat deserve further discussion. Li (1978) has shown that while Taai Saisiyat preserved the flap consonant, the flap was already lost in Tungho at that time. The Taai dialect in Li's study was therefore conservative and closer to the proto language. An examination of the early Taai Saisiyat data in Li (1978) suggests that word-final open syllables were extremely rare.[22] The avoidance of final open syllables in early Saisiyat, as reflected by the Taai data recorded by Li (1978), suggests that Saisiyat was once an asymmetrical language which required final closed

---

[19] Angle brackets '‹ ›' are used to include surface infixes, following the Leipzig Glossing Rules. Without the infix, the word 'drink' is pronounced variably as either [raʔœː] or [ræʔœː] in the present study.

[20] The transcriptions in Zeitoun et al. (2011) suggest that there is no such variation in Tungho Saisiyat.

[21] It is observed that even the final long vowels are sometimes unstable and tend to be mixed up with a vowel of regular length. See a detailed discussion in Zeitoun et al. (2015). The transcriptions in the literature also show inconsistencies; e.g. *hæhilaL* 'sun' in northern Saisiyat (Li 1978:147) corresponds to *ha-hilaː* in M. Yeh & Kaybaybaw (2010) but *halila* 'day' in Chiang & Chiang (2005:409).

[22] In the Taai Saisiyat data given by Li (1978), the only content word with final CV syllable is *Lima* [ɬima] 'arm and hand', which was *imaʔ* in the Tungho dialect (p.160). Given that the word is [imaʔ] in present-day Saisiyat, the transcription of *Lima* seems to be a typo.

syllables. This is similar to the pattern of Yapese discussed in section 2. The rankings of Final-C >> Faithfulness >> NoCoda are responsible for the asymmetrical pattern in early Saisiyat as in the case of Yapese. The later loss of the flap consonant leads to final open syllables, so modern Saisiyat (Taai and Tungho) has become a symmetrical type of language, suggesting that Final-C is demoted to a lower-ranking status.

5 Mayrinax Atayal (Mtu'uwal)[23]

Mayrinax Atayal has been analyzed in quite a few studies from various perspectives, such as Li (1982a, 1982b, 1983, 1995), L. Huang (1995, 2000a, 2000b), and Lu (2005). Mayrinax has eighteen phonemic consonants /p t k q ʔ b[β] g[ɣ] s c[ʦ] x h m n ŋ r l j w/; the vowel inventory contains either five vowels /i e a o u/ (L. Huang 2000a; C. Lin 2012) or three vowels /i a u/ (Lu 2005). Syllables are said to be maximally (C(C))V(C) (L. Huang 2000a:48) or CGVC/CGVG/CVGC (Lu 2005:24). Nonfinal closed syllables are permitted in Mayrinax (Lu 2005:25), which is different from Squliq, in which nonfinal syllables must be open (Rau 1992:25).

Different from the previous studies in stating that syllables contain at most four segments, the paper argues that, in general, syllables permit maximally three segments (CVC) in both nonfinal and word-final syllables in Mayrinax. Both the onset and coda C in CVC can be a glide in Mayrinax; the use of CVC here entails CVG/GVC/GVG syllable types. Moreover, it is proposed here that in Mayrinax, syllable margins do not allow complex true consonant clusters (CC). For example, words of the shape CCVCCVC in previous studies (e.g., *thawnak* 'sit (male)', L. Huang 2000a:49) are analyzed as three syllables C**V**.CVC.CVC ([tə̯.haw.nak]). The reason against the complex onset analysis here lies primarily in the phonological distribution of such sounds rather than phonetics: If CC is a legitimate type of onset, we expect onset CC (e.g., VC.CCV) in internal position, which is however lacking in Maryinax. In this kind of example with consecutive voiceless consonants at the left word edge, the intervening weak vowel is optionally pronounced as a voiceless schwa due to the voicelessness of the surrounding consonants, which presumably leads to the impression that the words begin with clusters.[24]

---

[23] Many native speakers prefer to adopt Mtu'uwal as the name of their language while Mayrinax is the term commonly used in the literature.

[24] See H. Huang (2006b) for similar arguments for simple onsets in Squliq Atayal. The devoicing of the weak vowel between sequences of voiceless consonants, such as *th* in *thawnak*, is similar to the commonly observed devoicing of the first vowel in the English word *potato*. In Mayrinax, devoicing of the weak vowel becomes less stable if it is followed by a voiced C, and devoicing does not occur if the preceding C is voiced. The devoiced vowels, just as other weak vowels, are omitted in Mayrinax orthography, which leads to apparent CC clusters in the broad transcriptions of many previous studies. In an alternative analysis assuming real consonant clusters, the above-mentioned voiceless vowel would have to be interpreted as the release of the preceding consonant, and the presence of #CC and the absence of *VC.CCV would be presumably attributed to the asymmetrical distribution of consonants in initial and medial positions. That is, medial onset allows only one consonant, but word-initial onset permits an extra one. This alternative analysis is not unlikely, but it

Four-segment syllables seem to appear sometimes in fast speech, although they are normally pronounced as separate syllables CV.GVC~CV.VC,[25] which obeys the upper limit of three segments. The data below illustrate how the three-segment syllables in the present study would be interpreted in a study (e.g., Lu 2005) where Mayrinax syllables are said to be maximally CGVC/CVGC. It is speculated here that the pronunciation of four-segment syllables, if there are any, reflects a possible change by less conservative speakers.[26]

(8) Mayrinax maximal syllables

a.	ra.hu.wal	'big'	CGVC	ra.**hwal**
b.	qu.waw	'wine'	CGVG	**qwaw**
c.	ʦa.wum(~ʦa.um)	'wipe'	CVGC	**ʦawm**
d.	βa.la.jiq(~βa.la.iq)	'good'	CVGC	βa.**lajq**

Because four-segment syllables occur only optionally in Mayrinax in the present study, the following discussion will be based on the analysis that syllables are maximally CVC in the language.

Closed syllables are freely found in interior and edge positions in Mayrinax, as illustrated by the data in (9) below. Notice that although the data contain several affixed forms, these examples are fine because they do not render exceptional syllable types that only occur in these affixed forms, and the morpheme boundaries do not coincide with the right edge of the internal syllables.

In the data in (9), words ending with /ʦ/ have not been found so far although the sound can appear in the codas of nonfinal syllables. There seem to be some variations across speakers about whether the alveolar affricate can occur in word-final position.[27]

---

appears to be more complicated in the need to assume an extrasyllabic consonant at the left word edge.

[25] The point of the discussion here is that the so-called CGVC/CVGC syllables actually contain two vowels. It awaits more research to determine whether there is an intervocalic glide between the two vowels phonologically.

[26] The Mayrinax data are based on the author's own fieldwork in Jinshui Village, Taian Township, Miaoli County. If CGVC/CVGC syllables in (8) are legitimate syllable types in Mayrinax, the right word edge permits an additional segment because syllables allow up to three segments in nonfinal position in this dialect. This is in fact the case found in the variety of Taoshan Squliq to be discussed later. See more discussion in section 6.2 regarding how the Mayrinax data correspond to the four-segment syllables in Taoshan Squliq.

[27] The absence of word-final /ʦ/ is due to historical sound changes (Li 1981:274). In word-final position, /t/ might be pronounced as an affricate [ʦ] by younger speakers (Li 1980:383). Conservative Mayrinax speakers do not show the variation freely; for example, the related forms of 'chop (AV/NAV)' are *k‹um›ut/kut-an/kut-i* in the present study, but they are *k‹um›ut~k‹um›uʦ/kut-an/kut-i* in Lu (2005).

(9) Mayrinax word-internal and word-final codas:

		Word-internal codas		Word-final codas	
a.	/p/	la**p**.ɣ-aw	‘count’	qu.ri**p**	‘ginger (fem.)’
b.	/t/	ma-**t**.ŋiʔ	‘full’	βə.hu**t**	‘squirrel’
c.	/k/	wa**k**.liʔ	‘leopard (fem.)’[28]	ha.wi.nu**k**	‘waist’
d.	/q/	ca**q**.ruɣ	‘stand (fem.)’	ka.hu.ni**q**	‘tree’ (male)
e.	/ʔ/	ɣi**ʔ.**nux	‘tooth’	wa.ɣi**ʔ**	‘sun’
f.	/β/	ka**β**.hul	‘hundred’	h‹u.m›a**β**[29]	‘stab (AV)’
g.	/ɣ/	ha**ɣ**.β-i	‘do magic (IMP.NAV)’	haj.la**ɣ**	‘fast’
h.	/s/	ki**s**.liq	‘heart’	tu.ra.ki**s**	‘millet’
i.	/ʦ/	ʦa**ʦ**.liʔ	‘kick’	—	
j.	/x/	ma-**x.**rix.riq	‘loose’	pa.ju**x**	‘many’
k.	/h/	li**h**.piq	‘thin’	pa.ŋi**h**	‘wound’
l.	/m/	ma.-ha**m**.t-aj	‘act indiscriminately’	h‹u.m›i.ku**m**	‘waste (AV)’
m.	/n/	ri**n**.ʔu	‘dew’	ra.a**n**	‘road (fem.)
n.	/ŋ/	ra**ŋ**.ʔ-an	‘dry’	ha.na.a**ŋ**	‘sound’
o.	/r/	nə.-qa**r**.ʔ-un	‘do completely (NAV)’	k‹u.m›u.wa**r**	‘scrape (AV)’
p.	/l/	pa-**l**.qiŋ	‘hide’	ka.a**l**	‘sky’
q.	/j/	wa**j**.luŋ	‘chicken’	mi.ra**j**	‘turn’
r.	/w/	pa**w**.min	‘eyebrow’	hi.la**w**	‘cover’

The existence of abundant nonfinal codas in Mayrinax is drastically different from the case of the Squliq dialect in the literature, as suggested by Rau’s (1992) statement that nonfinal syllables are all open. While the details of Squliq will be discussed in the next section, a related question concerning Mayrinax is why there seem to be no restrictions on the types of nonfinal codas in this dialect. Based on the historical reconstruction of Proto Atayalic by Li (1981), it is observed that many of the nonfinal codas actually come from the loss of the schwa vowel in Mayrinax, as illustrated below:

(10) Historical loss of schwas (Li 1981)

	Proto Atayalic	Mayrinax	Gloss
a.	*ma-səpat	ma-spat	‘eight’ (p.246)
b.	*ma-hənuk	ma-hnuk	‘soft’ (p.247)
c.	*kitə-hur	kit-huw	‘fat’ (p.248)
d.	*ma-təŋiʔ	ma-tŋiʔ	‘full’ (pp.245-255)
e.	*rakəlid	ʔakliʔ	‘leopard’ (p.254)

The above data show that the historical loss of schwas leads to nonfinal coda consonants such as /s h t k/ in (10). Given that nonfinal codas arise from the loss of word-internal schwas in Mayrinax, the fact that /ʦ/ can appear in internal coda position,

---

[28] Mayrinax Atayal shows certain differences between male and female forms of speech (Li 1982b); *fem.* stands for female forms.

[29] Although L. Huang (2000a:45) mentions that /ʦ β r/ do not appear in word-final positions, the present study has identified words ending with /β r/, similar to Lu (2005).

though not in word-final position, is not surprising since it presumably came from an onset consonant at an earlier stage.

Content words ending with open syllables are limited in Mayrinax literature. In Li (1981), *raŋa* 'trap' is the only vowel-ending content word recorded for Mayrinax; it results from the diachronic loss of final /r/ (Proto Atayalic *daŋar; Li 1981:253). A number of content words with final vowels have been identified in the present study, such as *kihma* 'thick', *taka* 'frog', *pcigha* [pətɕiɣha] 'chest (male)', *mapŋu* 'blunt', *lihu* 'forehead', and *wahi* 'trap (for animals such as wild pigs and bears)'. For words with final vowels [i] or [u], they may alternatively be analyzed as ending with homorganic glide consonants (*-ij* and *-uw*, respectively) for historical reasons, retaining the final consonantal glides even in the synchronic grammar. However, words with final low [a] unambiguously show that content words may end with open syllables in the synchronic phonology through diachronic loss of word-final consonants. Despite the fact that vowel-ending content words may be limited in numbers, Mayrinax in general belongs to the symmetrical type of languages since both open and final closed syllables, maximally with three segments, are allowed in both nonfinal and final positions. The rankings of Faithfulness >> NoCoda in Manam characterise the generally symmetrical pattern in the case of Mayrinax.

6 Squliq Atayal

There have been quite a few studies on Squliq Atayal, especially on its syntax (see e.g., L. Huang 1989, 1993, 2008; L. Huang & Hayung 2008, 2011; Liao 2005; Liu 2005; Tsai 2007; M.Y. Yeh & S. Huang 2013, among others). The complexity of the phonology of the Squliq dialect is manifested primarily in the analysis by Egerod (1965) and Li (1980) and is well recognised in the Austronesian literature, e.g., Himmelmann (2005:125). Extensive data relevant to Squliq can also be found in Li (1981) and the dictionary by Egerod (1980, 1999). Rau (1992) provides a general description, and H. Huang (2006a, 2006b) offers analyses on specific topics.

While the works by Li (ibid.) and Egerod (ibid.) are based on the Fuxing variety in Taoyuan City, Rau (1992) is based on Wulai (New Taipei City) and H. Huang (2006a, 2006b) on Jianshi (Hsinchu County) Squliq. Two varieties of Squliq Atayal, spoken respectively in Jianshi and Wufeng Township, Hsinchu County, will be examined below because they exhibit different patterns in terms of syllable types and distributions.

6.1 Squliq Atayal in Jianshi Township

Jianshi Squliq is in many ways similar to the variety in Fuxing Township investigated by Li (1980, 1981) and Egerod (1965, 1966, 1980, 1999). According to Li (1980), the phonemic inventory of Squliq includes nineteen consonants /p t k q ʔ b[β] z g[ɣ] s c[ʦ] x h m n ŋ r l j w/ and five vowels /i e a o u/. The same inventory is adopted in the theses by M.Y. Yeh (2002) and Hayung (2008) on Jianshi Squliq.

Nonfinal syllables in Jianshi Squliq are all open, similar to the descriptions in Rau (1992) based on the variety in Wulai. Morphologically infixed forms such as /in, quriq,

an/ [q‹i**n**.›ri.q-an] ‘realis/perfective, steal, Locative voice’ tolerate nonfinal closed syllables. However, since there is always a morpheme boundary between intervocalic CC in the forms with internal codas (*q‹i**n**.›ri.qan*), such examples simply show that the exceptional syllable types are induced by morphological structures. In an Optimality-theoretic account, the exceptional medial closed syllables can be attributed to a high-ranking Anchor-R constraint which dictates that the right edge of the infix be aligned with the right edge of a syllable. Once the influence of the morphology is factored out, the phonological generalisation that medial syllables are all open emerges. That is, the data are consistent with the observation that the phonology of the dialect does not allow nonfinal closed syllables. The ban against internal codas holds even for forms of reduplicated monosyllables, such as [pə.h**ə**.pah] ‘flower’.[30] Although all of the four Formosan languages (Paiwan, Puyuma, Bunun, Amis) in Blust’s (2009) extensive survey of Austronesian medial consonant clusters are found to tolerate CC in reduplicated monosyllables, Jianshi Squliq is different in avoiding internal codas by requiring an epenthetic vowel even between the reduplicating CVC bases. The pattern of Jianshi Squliq is unusual among Formosan languages and resembles many of the languages in eastern Indonesia and the Pacific in avoiding medial CC (cf. Blust 2009:207–209).

In the Squliq data given by Li (1980, 1981) and in Jianshi Squliq (Hayung 2008), most words end with a true consonant or a glide.[31] All consonants except /b[β] z g[ɣ] r/ can occur in word-final position (Egerod 1966, Li 1980). Li (1980) argues that the gaps are due to the phonological rules that convert /b[β] z g[ɣ] r/, respectively, to *p j w j* in word-final positions. Words with final homorganic vowels and glide codas, such as *kgij* [kəɣij] ‘hemp’, involve a closed syllable at the right word edge because of the consonantal nature of the final glides.[32] The transcription of a final consonant in these words is justified by its alternation with a true consonant in affixed forms, such as *kgij* [kəɣij] ‘hemp’ and *kgir-i* [kəɣiri] ‘peel hemp (IMP.NAV)’. Li (ibid.) argues that the morphophonemic alternation between /r/ and /j/ and the gap of word-final /r/ together support the rule changing /r/ to *j* in word-final position.

As discussed in section 2, the OT constraints in Broselow (2003) predict that there should exist languages which forbid closed syllables in nonfinal position but require words to end with consonants through the possible rankings of Final-C >> NoCoda >> Faithfulness. Broselow (ibid.) has not found such languages in her survey and comments that this is a potentially problematic gap for the predicted typology. The data of Fuxing

---

[30] The illustrating word *phpah* [pə.h**ə**.pah] is recorded as *hpah* ‘flower’ in Li (1981:263), in which the initial /p/ sound has dropped. Notice that the fossilised reduplication of *pah* in [pə.h**ə**.pah] exhibits pre-tonic vowel reduction to [ə], which is characteristic of Atayal, so it is the repetition of the consonants that suggests its reduplication nature. In words such as [pə.h**ə**.pah], where the (epenthetic) schwa vowel is surrounded by voiceless consonants, the weak vowel is often pronounced as voiceless.

[31] Vowel-ending words in Li (1981) are either forms with the suffix *-i* ‘IMP.NAV’ or pronouns such as *simu* ‘you, PL’.

[32] Word-final sequences of homorganic vowels followed by glide codas in Li’s works are interpreted as *ii* and *uu* in Egerod (1965, 1966, 1980, 1999). The consonantal nature of the final segments is justified by phonological evidence in Li’s work. Li’s descriptions, however, imply that word-final homorganic vowel-glide sequences such as *ij* and *uw* are realized as long vowels at a certain phonetic level (Li 1980:355). Perhaps future phonetic studies could help clarify whether these word-final sequences are VC or vocalic in nature.

Squliq and Jianshi Squliq, as described in Li (1980, 1981) and Hayung (2008), respectively, appear to fill the typological gap in Broselow's study in that they contain only open syllables in medial position while strongly preferring closed syllables in word-final position. The author's own fieldwork on Jianshi Squliq also confirms the existence of such speech forms among older or more conservative speakers.

For younger speakers of Jianshi Squliq, however, words with homorganic vowels and glide codas in final syllables are perhaps vowel-ending. In this less conservative form of speech, a regular CV open syllable instead of a homorganic vowel-glide sequence is produced word-finally, so the alternations manifested in the related words [kəɣi] and [kəɣiri] would involve a rule dropping /r/ in the final position of the unsuffixed form, leading to the existence of a final open syllable.[33] Despite the differences in the articulation of word-final syllables for speakers of different ages, the absence of nonfinal closed syllables and the possible/obligatory presence of word-final closed syllables suggest that the distribution of syllable types is asymmetrical in Jianshi Squliq.

The speech forms of younger Jianshi Squliq speakers are similar to the case of Kamaiurá discussed in Section 2 and can be similarly captured by the ranking of Anchor-R >> NoCoda >> Faithfulness.[34] In contrast, word-final consonants are required in the more conservative speech forms of older Jianshi Squliq speakers; the constraints Final-C >> NoCoda >> Faithfulness account for the data, which fill the important typological gap predicted by Broselow (2003).

6.2 Squliq Atayal in Wufeng Township (Taoshan Village)

The Squliq dialect in Wufeng Township, Hsinchu County, is spoken in Taoshan Village, and is referred to in the literature more often as Taoshan Squliq. Compared with other Squliq varieties, Taoshan Squliq has received much less attention. Although Squliq dialects are more uniform than the other dialect group C'uli'(Li 1980:350), Taoshan Squliq is fairly different from the other Squliq varieties described in previous studies in terms of the distribution of syllable types.

According to Cheng (2001) and Su (2004), Taoshan Squliq has the same nineteen consonants /p t k q ʔ b[β] z g[ɣ] s c[ʦ] x h m n ŋ r l j w/ as other varieties, but they recognise six vowels /i e a o u ə/, with an additional schwa in the inventory. Cheng (2001) has pointed out some differences between Taoshan Squliq and the Squliq in Fuxing

---

[33] The contrast that conservative speakers of Jianshi Squliq retain the final homorganic glides while younger speakers may not is based on unsuffixed forms: /kɣir/ [kəɣij] vs [kəɣi] 'hemp'. Suffixed forms such as *kgir-i* [kəɣiri] are always vowel-ending. Blust's (2009:629) generalisation regarding final glottal stop epenthesis seems to be based on unsuffixed forms, too: 'Many of the Formosan languages, including Atayal, Saisiyat, Pazeh, Rukai, Bunun, Kavalan, Paiwan, Puyuma and Amis, show a largely predictable final glottal stop.' The fact that only the prosodic words that are right-aligned with roots require a final consonant, as in early Saisiyat (see section 4) and the conservative form of Jianshi Squliq discussed here, calls for more research into the role of morphological structures in phonological generalisations.

[34] Notice the dropping of final /r/ as in the illustrating word /kɣir/ [kəɣi] 'hemp' incurs a violation of the Anchor-R constraint, which suggests that Anchor-R is dominated by some constraint avoiding final /r/ in Jianshi Squliq. The rankings Anchor-R >> NoCoda >> Faithfulness are responsible for the asymmetrical patterns in Jianshi Squliq, but the relatively higher ranking status of the constraint does not necessarily mean that it is undominated.

Township (Taoyuan County). First, Taoshan Squliq is more conservative in retaining historical *aj* and *aw*, which have become the monothongs *e* and *o* in Fuxing Squliq. Second, the high vowels /i u/ are pronounced lower in Fuxing than in Taoshan Squliq. Third, unstressed syllables are less likely to be dropped in Taoshan than in Fuxing Squliq. Moreover, according to Cheng (2001:31), while /b[β] z g[ɣ]/ do not occur in word-final position, which is similar to Egerod's and Li's observations of Fuxing Squliq, /r/ can appear word-finally in Taoshan Squliq.[35] In the present study, it is also found that /r/ can appear in word-final position, such as *h‹əm›aw**r*** 'scoop up (AV)', *ʔ‹əm›apa**r*** 'invade (AV)', and *rapa**r*** 'sole of the foot'.[36]

Another important difference not mentioned in Cheng (2001) is that while the Squliq in Fuxing (as well as in Wulai and Jianshi) does not permit nonfinal closed syllables, Taoshan Squliq allows many types of consonants as internal codas. The data below illustrate that most consonants can be internal and word-final codas in Taoshan Squliq.

(11) Taoshan Squliq word-internal and word-final codas

		Word-internal codas		Word-final codas	
a.	/p/	sɪ-**p**.na.jaŋ	'reclaim land'	q‹ə.m›a.ri**p**	'cut (AV)'
b.	/t/			h‹ə.m›a.ku**t**	'move (AV)'
c.	/k/	pə-**k**.sjuɣ-i	'borrow'	ʔu.ju**k**	'baby'
d.	/q/	pə-**q**.hə.l-un	'force'	twa.hi**q**	'far'
e.	/ʔ/	tə**ʔ**.nux-an	'pillow'	qə.βa**ʔ**	'hand'
f.	/β/	kə**β**.hul	'hundred'	—	
g.	/ɣ/	kə-**ɣ**.hə.jaq	'cold (weather)'	—	
h.	/s/	mə-**s**.hə.juʔ	'long'	lu.ku**s**	'clothes'
i.	/z/[37]			—	
j.	/ʦ/			(βju.ri**ʦ**)	'Miaoli (place name)'
k.	/x/			mə.-qwa.la**x**	'rain (AV)'
l.	/h/	β**h**.ja.g-i	'chase'	ja.qi**h**	'bad'
m.	/m/			ju.hu**m**	'gall'
n.	/n/	βi**n**.kis	'old'	qə.lja**n**	'daytime'
o.	/ŋ/	qə**ŋ**.jat	'serious, diligent'	ta.pa**ŋ**	'blanket'
p.	/r/	wə**r**.ŋ-ani	'wrap'	h‹ə.m›aw**r**	'scoop up (AV)'
q.	/l/	kə.-tə**l**.ʔə.tuʔ	'cold'	q‹ə.m›i.hu**l**	'force (AV)'
r.	/j/	ta**j**.luq	'raw'	mə.-ta.ku**j**	'fall down (AV)'
s.	/w/	βa**w**.naw	'peanut'	ʔu.ra**w**	'earth, dirt'

[35] However, Cheng (2001:31–34) does not give specific examples for word-final /r/. In the description of the phoneme /r/, Cheng (ibid.) states that word-final /r/ may become [l], such as *hrhar* [hərəhal] 'phlegm'.

[36] The word 'sole of the foot' is *rapal* in Egerod (1980).

[37] Although the phoneme /z/ has been posited for Squliq in the literature because of its presence in prevocalic position, there is some evidence suggesting that there is no /z/ in some varieties of Squliq (H. Huang 2015).

In word-final position, /β z ɣ/ are not found; the gaps are already identified in the literature for other Squliq varieties as discussed in previous subsections. As far as /ʦ/ is concerned, although it appears in loanwords such as *βjuriʦ*, word-final /ʦ/ has not been found in native words in the present study. A search of the words in the dictionary by Egerod (1980, 1999) and the data in Li (1980) also shows that no words end with /ʦ/.[38] As far as nonfinal position is concerned, many different types of consonants are legitimate internal codas as shown by (11) above.[39] Some of the internal codas in the illustrating examples result from vowel syncope in the synchronic grammar, such as (11p) /war**i**ŋ, ani/ [wər.ŋa.ni] 'wrap (IMP.BV)' The tolerance of internal codas is an important feature of Taoshan Squliq that makes it phonologically distinct from other Squliq varieties discussed in previous studies.

The controversy of whether word-final open syllables exist in Jianshi Squliq, which is discussed in section 6.1, applies similarly to the case of Taoshan Squliq. Most words end with a consonant or a glide, but word-final homorganic vowel-glide sequences may be vowel-ending for less conservative speakers.

Although Taoshan Squliq is similar to Isbukun Bunun, Saisiyat, and Mayrinax in permitting both nonfinal and final closed syllables, it in fact exhibits the same asymmetrical patterns as Jianshi Squliq (and as English) in allowing an additional segment at the right word edge. The following table highlights the point that both the two Squliq varieties are asymmetrical in allowing an extra segment at the right edge despite the fact that they differ somewhat in permissible syllable types.[40] The controversial word-final open syllables (as well as exceptional syllable types induced by infixation) are excluded from the presentation here.

---

[38] However, the Squliq data from Yi-Lan County in Li (1996) contain several /ʦ/-ending words, such as *ramaʦ* 'side food' (Li 1996:207).

[39] The gaps of the internal codas await more field research to see if they are accidental.

[40] Most of the discussion in Broselow (2003) focuses on asymmetrical distribution of syllable types rather than numbers of segments within syllables. It is mentioned briefly in Broselow (ibid.) that cases in which the number of consonants at edges exceeds the number of consonants allowed in internal syllable margins (e.g. English) might support the extraprosodicity approach to edge/interior asymmetry.

(12) Jianshi Squliq and Taoshan Squliq syllable types:

	Nonfinal			Word-Final		
	Syllable types	Jianshi	Taoshan	Syllable types	Jianshi	Taoshan
a.	CV/GV	✓	✓	CV**C**/GV**C**	✓	✓
b.	CGV	✓	✓	CGV**C**	✓	✓
c.	CVG/GVG	—	✓	CVG**C**/GVG**C**	—	✓
d.	CVC/GVC	—	✓			

As manifested in (12a,b,c), the two Squliq varieties are similar in allowing up to three segments in internal syllables but four segments in word-final syllables.[41] The difference in (12c) is due to the monophthongisation of *aj/aw* to *e/o* in Jianshi Squliq. Taoshan Squliq allows nonfinal closed syllables ending with true consonants, as shown by (12d),[42] which could have led to the illusion of a symmetrical language; it in fact belongs to the asymmetrical type as suggested by the additional segments in word-final positions.

The four-segment word-final syllables in Taoshan Squliq presumably came from the contraction of the final two syllables historically, as revealed by a comparison of corresponding words in Mayrinax Atayal.

(13) Taoshan Squliq maximal syllables and corresponding Mayrinax words[43]

	Syllable types	Taoshan Squliq	Mayrinax	Gloss
a.	CGVC	mə.rə.**kjas**	ma.ra.**ki.jas**	‘grow (AV)’
		qə.**ljan**	qa.**li.jan**	‘daytime’
		nə.**qwaq**	ŋa.**qu.waq**	‘mouth’
b.	CVGC	rə.**majʔ**	ra.**ma.jiʔ**	‘horse’
		mə.ɣə.**nawʔ**	ma.ɣa.**na.wuʔ**	‘play (AV)’
c.	CGVG	mə.**ɣjaj**	ma.**ɣi.jaj**	‘escape (AV)’
		qwaw	**qu.waw**	‘wine’

The proposed analysis of the Mayrinax and Squliq data suggests that Proto Atayal (Li 1981) is similar to the conservative form of Jianshi Squliq in requiring final consonants but forbidding internal codas. Mayrinax has developed nonfinal closed syllables through syncope and has obtained final open syllables via the loss of some final consonants,

[41] Cheng (2001) gives examples showing that syllables are maximally CGGVC ([**žjwaw**] ‘thing’, [pəsəhə**žjwaj**] ‘keep straight’) or CGVGC ([təmə**žjajʔ**] ‘propose marriage’) in Taoshan Squliq (p.51); the examples used to illustrate the five-segment syllables all involve [žj]. In the present analysis, the [žj] sequences in Cheng (2001) are interpreted as a single segment [ʑ], so the five-segment syllables are all considered as four segments long. The [j] glide in onset alternates with [ʑ] variably across and within speakers. The Squliq data given here do not show the fricative variant in order to simplify the presentation.

[42] Although nonfinal CVC syllables are allowed in Taoshan Squliq due to the syncope process, final CVCC syllables are not legitimate because complex true consonant clusters are banned in syllable margins in Atayal dialects.

[43] Recall that in the discussion of the Mayrinax data in section 5, whether Mayrinax allows up to three- or four-segment syllables is debatable; sometimes four-segment syllables are possible for some speakers. The inconsistency is presumably related to the contraction phenomenon discussed here. Less conservative speakers of Mayrinax have developed the contraction phenomenon that appears similar to the Squliq data.

becoming more of a symmetrical type of language. Taoshan Squliq has also developed internal codas via syncope, but has remained asymmetrical due to the contraction of the final two syllables which leads to the extra segment at the right word edges. Although Mayrinax is generally considered a conservative dialect of Atayal (Li 1981, 1982b, 1983), it deviates from the proto language more than Squliq in terms of the distribution of syllable types.

7. Conclusion and discussion

This paper has tried to show that syllable-based descriptions allow us to classify the three Formosan languages in terms of symmetrical or asymmetrical patterns, and to capture the motivations of these patterns via the adopted OT analyses. The chart below summarises the languages/dialects discussed in the previous sections. Syllable types (including whether codas are allowed or required) and the maximal number of segments allowed within a syllable are considered for both internal and word-final positions. None of these languages/dialects require internal closed syllables, so the relevant column is omitted for nonfinal syllables.[44]

Table 1: A summary of the patterns of syllable types and their distribution

Languages / dialects	Nonfinal syllables		Word-final syllables			Similar language (type) rankings
	Allow codas	Max #	Allow codas	Need codas	Max #	
Isbukun Bunun	✓	4	✓		4	Manam (symmetrical)
modern Saisiyat	✓	3	✓		3	Manam (symmetrical)
early Saisiyat (Taai in Li 1978)	✓	3	✓	✓	3	Yapese (asymmetrical) Final-C >> Faithfulness
Mayrinax Atayal	✓	3	✓		3	Manam (symmetrical)
Jianshi Squliq (younger)		3	✓		4	Kamaiurá (asym.) Anchor-R >> NoCoda
Jianshi Squliq (conservative)		3	✓	✓	4	(*Fill the gap*) (asym.) Final-C >> NoCoda
Taoshan Squliq	✓	3	✓	(✓)	4	(asymmetrical)[45] Anchor-R/Final-C >> Faithfulness

Isbukun Bunun and modern Saisiyat are typical cases of symmetrical languages despite some possible gaps of feature contrasts in codas. The early Taai dialect of Saisiyat as described in Li (1978), which preserves the flap consonant, exhibits the asymmetrical pattern by requiring word-final codas. Among the Atayal dialects, Mayrinax is relatively more symmetrical, allowing both open and closed syllables and the same maximal syllables in both internal and final positions. Jianshi Squliq is a distinct

[44] Broselow (2003) mentions that languages which require internal closed syllables are expected to be absent since no constraints require all syllables to be closed in all positions (p. 165).

[45] Taoshan Squliq is similar to Kamaiurá and Jianshi Squliq in allowing an extra segment at the right edge, but Taoshan Squliq is different in permitting word-internal closed syllables.

case of asymmetrical languages in that it not only permits an extra segment in final position but also allows final closed syllables that are banned internally. Notice that the asymmetrical patterns of early Saisiyat, Jianshi Squliq, and Taoshan Squliq are based on either distribution of syllable types or maximal syllables, or both. Early Saisiyat manifests asymmetrical patterns in terms of possible syllable types (open or closed), Taoshan Squliq in terms of the number of segments within a syllable, and Jianshi Squliq in terms of both syllable types and sizes. This suggests that both syllable types and maximal syllables are indispensable information in characterizing a language.

The asymmetrical patterns shown in Table 1 are attributed to the high-ranking status of either Final-C or Anchor-R in the OT analysis proposed in Broselow (2003). Final-C could lead to exceptional distribution of a syllable type word-finally, as in early Saisiyat (and in conservative Jianshi Squliq), or it may result in an extra segment at the right edge as in the case of conservative Jianshi Squliq. Anchor-R allows asymmetries at the right edge in the input to surface. In the case of Jianshi Squliq, where infixation leads to unexpected closed syllables within words, the OT framework directly accounts for the apparent exceptions by rendering it possible to rank other independently justified Anchor-R constraints specific to the affixes higher than Final-C or Anchor-R. That is, by attributing the exceptional internal codas in the infixed forms to other high-ranking constraints, the OT framework allows the generalisation that Jianshi Squliq forbids internal closed syllables but permits/requires final closed syllables to be expressed succinctly (Anchor-R >> NoCoda >> Faithfulness; Final-C >> NoCoda >> Faithfulness).

It is observed in the present study that the more conservative forms of Jianshi Squliq (and Proto Atayal) fill an important typological gap in the OT analysis in Broselow (2003) and therefore strengthen the syllable-based approach. Alternatively, one might analyze the same sets of data by stating whether adjacent consonant clusters (CC) are permitted: Isbukun, Saisiyat, Mayrinax, and Taoshan Squliq allow intervocalic CC while Jianshi Squliq bans CC. Although the string-based descriptions are capable of directly capturing the similarities and differences across languages, too, it cannot readily relate the types of Formosan data to the discussion of other languages of more complex structures in the general phonology literature. For example, given that English allows VVC and VCC word-finally but maximally only VV and VC in internal position, as mentioned in the introduction section, the linear *CC approach for the Formosan data here would need to be modified according to the number of intervocalic consonants that is tolerated (*CCC). In contrast, the present approach captures the fact that both Squliq and English exhibit asymmetrical patterns by allowing an extra segment at the right edge and the patterns are presumably accounted for by similar sets of theoretical mechanisms such as Anchor-R.

Identifying potential restrictions on the combinations of intervocalic CC as in Blust (2009) is helpful in understanding whether phonetic contexts play a role in shaping the phonotactics of a language. The languages under investigation in the present study, except Jianshi Squliq, happen to all allow heterorganic clusters in non-reduplicated bases, suggesting that the combinations of intervocalic CC are relatively free in these languages. The paper describes medial and final closed syllables, so possible gaps of (or restrictions on) word-internal codas are more clearly revealed. Although these gaps might appear insignificant when the focus of investigation is only on a few languages, it is believed that large-scale studies of this type would reveal the nature of some of these apparently accidental gaps.

References

Blevins, Juliette. 2003. The independent nature of phonotactic constraints: An alternative to syllable-based approaches. *The syllable in Optimality Theory*, ed. by Caroline Féry, and Ruben van de Vijver, 375–403. Cambridge, UK and New York: Cambridge University Press.

Blust, Robert. 1988. *Austronesian root theory: An essay on the limits of morphology*. Studies in Language Companion Series 19. Amsterdam and Philadelphia: John Benjamins.

Blust, Robert. 2009. *The Austronesian languages*. Pacific Linguistics 602. Canberra: The Australian National University. [2013 revised edition on-line at: http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf]

Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory* 7.2:145–166.

Broselow, Ellen. 2003. Marginal phonology: Phonotactics on the edge. *The Linguistic Review* 20.2-4:159–193.

Byrd, Andrew Miles. 2010. Reconstructing Indo-European syllabification. PhD dissertation. Los Angeles: UCLA.

Chang, Yueh-chin. 2009. *taiyayu han saixiayu de yuyin yanjiu*. [*A phonetic study on Atayal and Saisiyat*]. National Science Council Report (NSC 96-2411-H-007-028-), Taiwan. [In Chinese]

Chao, Jung-lang. 1954. miaolixian nanzhuangxiang donghecun saixiazu yuyanxue diaocha jianbao [Preliminary report of linguistic investigation in the Nan-Chuang Saisiat, Miaoli Prefecture]. *Bulletin of the Department of Anthropology, National Taiwan University* 4:25–27. [In Chinese]

Cheng, Ju Chang. 2001. cong yinyunxue de jiaodu kan taiyayu de biaoyin xitong: yi taoshan fangyan weili [A phonological perspective of phonetic symbolic system of Atayal: The case of Tausan]. MA thesis. Hsinchu: National Hsinchu Teachers' College. [In Chinese]

Chiang, Wen-yu and Fang-mei Chiang. 2005. Saisiyat as a pitch accent language: Evidence from acoustic study of words. *Oceanic Linguistics* 44.2:404–426.

Clements, G. N. and Samuel Jay Keyser. 1983. *CV phonology: A generative theory of the syllable*. Linguistic Inquiry Monograph 9. Cambridge, MA: MIT Press.

Côté, Marie-Hélène. 2011. Final Consonants. *The Blackwell companion to phonology, Vol. 2: Suprasegmental and prosodic phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 848–872. Malden, MA: Wiley-Blackwell.

Deng, Wan-zhen. 2007. Toward a unified analysis of consonant loss and its related phenomena in Saisiyat. MA thesis. Hsinchu: National Tsing Hua University.

Egerod, Søren. 1965. Verb inflexion in Atayal. *Lingua* 15:251–282.

Egerod, Søren. 1966. A statement on Atayal phonology. *Artibus Asiae Supplementum* XXIII (Essays offered to G. H. Luce by his colleagues and friends in honour of his 75th birthday) 1:120–130.

Egerod, Søren. 1980. *Atayal-English dictionary*. Scandinavian Institute of Asian Studies Monograph Series No. 35. London and Malmö: Curzon Press.

Egerod, Søren. 1999. *Atayal–English dictionary*. 2nd ed. by Jens Østergaard Petersen. Historisk–filosofiske Skrifter 20. Copenhagen: The Royal Danish Academy of Sciences and Letters.

Flack, Kathryn. 2009. Constraints on onsets and codas of words and phrases. *Phonology* 26.2:269–302.

Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. *Linguistics in Oceania: Current trends in linguistics Vol. 11 – Diachronic, areal, and typological linguistics*, ed. by Thomas A. Sebeok, 401–458. The Hague: Mouton.

Hayung, Tali'. 2008. jianshi taiyayu qianzhui yanjiu [A study of the prefixes in Jianshih Atayal]. MA thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

He, Rufen, Siqi Zeng, Wensu Li and Qingchun Lin. 1986. *gaoshanzu yuyan jianzhi (bunenyu)* [*An outline of the Formosan native languages* (*Bunun*)]. Beijing: Min-Zu. [In Chinese]

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar, and Nikolaus P. Himmelmann, 110–181. Routledge Language Family Series. London and New York: Routledge.

Ho, Dah-an. 1994. Paiwan. *Comparative Austronesian dictionary: An introduction to Austronesian studies*, Part 1: Fascicle 1, ed. by Darrell T. Tryon, 307–314. Trends in Linguistics: Documentation 10. Berlin and New York: Mouton de Gruyter.

Hsieh, Fang-yin Freda. 2007. Saisiyat phonology: With special reference to vowel variations. MA thesis. Hsinchu: National Tsing Hua University.

Huang, Hui-chuan J. 2005. On the status of onglides in Isbukun Bunun. *Concentric: Studies in Linguistics* 31.1:1–20.

Huang, Hui-chuan J. 2006a. Resolving vowel clusters: A comparison of Isbukun Bunun and Squliq Atayal. *Language and Linguistics* 7.1:1–26.

Huang, Hui-chuan J. 2006b. Squliq Atayal syllable onset: Simple or complex? *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 489–505. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

Huang, Hui-chuan J. 2008. Competition between syllabic and metrical constraints in two Bunun dialects. *Linguistics* 46.1:1–32.

Huang, Hui-chuan J. 2015. The phonemic status of /z/ in Squliq Atayal revisited. *Capturing phonological shades within and across languages*, ed. by Yuchau E.

Hsiao and Lian-Hee Wee, 243-265. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Huang, Lillian M. 1989. The pronominal system in Atayal. *Studies in English Literature and Linguistics* 15:115–133.

Huang, Lillian M. 1993. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1997. gaoxiongxian de bunongyu [(Isbukun) Bunun in Kaohsiung County]. *gaoxiongxian nandao yuyan* [*Formosan Austronesian languages in Kaohsiung County*], ed. by Paul Jen-kuei Li, 351–410. Kaohsiung: Kaohsiung County Government. [In Chinese]

Huang, Lillian M. 2000a. *taiyayu cankao yufa* [*A reference grammar of Atayal*]. Formosan Language Series 1. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Huang, Lillian M. 2000b. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364–390.

Huang, Lillian M. 2008. Grammaticalization in Squliq Atayal. *Concentric: Studies in Linguistics* 34.2:1–46.

Huang, Lillian M. and Tali' Hayung. 2008. Syntax and semantics of *p-* in Squliq Atayal. *Language and Linguistics* 9.3:491–521.

Huang, Lillian M. and Tali' Hayung. 2011. Coordination and comitativity in Squliq Atayal. *Language and Linguistics* 12.1:1–47.

Jensen, John Thayer. 1977. *Yapese reference grammar*. Honolulu: University of Hawai'i Press.

Kaisse, Ellen M. 2005. Word-formation and phonology. *Handbook of word-formation*, ed. by Pavol Štekauer and Rochelle Lieber, 25–47. Studies in Natural Language and Linguistics Theory Vol. 64. Dordrecht: Springer.

Kaybaybaw, Ya'aw Kalahae'. 2009. saixiayu cihui de jiegou yu yuyi yanjiu [A morphological and semantic study on word formation in SaySiyat]. MA thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

Laeufer, Christiane. 1995. Morphology and syllabification domains. *Lingua* 97.2-3:101–121.

Li, Paul Jen-kuei. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 49.2:133–199.

Li, Paul Jen-kuei. 1980. The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 51.2:349–405.

Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52.2:235–301.

Li, Paul Jen-kuei. 1982a. Linguistic variations of different age groups in the Atayalic dialects. *Tsing Hua Journal of Chinese Studies—Presented to Dr. Fang-kuei Li on his eightieth birthday: Studies in linguistics*, ed. by Kun Chang, 14.1-2:167–191.

Li, Paul Jen-kuei. 1982b. Male and female forms of speech in the Atayalic group. *Bulletin of the Institute of History and Philology, Academia Sinica* 53.2:265–304.

Li, Paul Jen-kuei. 1983. Types of lexical derivation of men's speech in Mayrinax. *Bulletin of the Institute of History and Philology, Academia Sinica* 54.3:1–18.

Li, Paul Jen-kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59.2:479–508.

Li, Paul Jen-kuei. 1995. The case-marking system in Mayrinax Atayal. *Bulletin of the Institute of History and Philology, Academia Sinica* 66.1:23–52.

Li, Paul Jen-kuei. 1996. *yilanxian nandao minzu yu yuyan* [*The Formosan tribes and languages in Yilan*]. Monograph Series of Yi-Lan History, Linguistics 1. Yilan: Yilan County Government. [In Chinese]

Li, Paul Jen-kuei. 1997. nantouxian xinyixiang de bunong nanbu fangyan [(Isbukun) Bunun in Xinyi Township, Nantou County]. *Gaoxiongxian nandao yuyan* [*Formosan Austronesian languages in Kaohsiung County*], ed. by Paul Jen-kuei Li, 300–350. Kaohsiung: Kaohsiung County Government. [In Chinese]

Li, Paul Jen-kuei and Shigeru Tsuchida. 2006. *Kavalan dictionary*. Language and Linguistics Monograph Series A-19. Taipei: Academia Sinica.

Liao, Hsiu-chuan. 2005. Another look at the order of clitic pronouns in Wulai Atayal. *Concentric: Studies in Linguistics* 31.1:47–63.

Lin, Cheng-Yuan. 2012. wenshui taiyayu de cihui jiegou [A study of morphological structure in Mayrinax Atayal] MA Thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

Lin, Hsiu-hsu. 1996. Isbukun phonology: A study of its segments, syllable structure and phonological processes. MA thesis. Hsinchu: National Tsing Hua University.

Lin, Tai, Siqi Zeng, Wensu Li and Bukun Ismahasan Islituan. 2001. *Isbukun bunongyu goucifa yanjiu* [*A study on the morphology of Isbukun Bunun*]. Taipei: Duce wenhua. [In Chinese]

Liu, Adlay Kun-long. 2005. The structure of relative clauses in Jianshi Squliq Atayal. *Concentric: Studies in Linguistics* 31.2:89–110.

Lu, Anne Yu-an. 2005. Mayrinax phonology: With special reference to UM affixation. MA Thesis. Hsinchu: National Tsing Hua University.

McCarthy, John. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics* [Special issue on *Constraint-based theories in multilinear phonology*] 38.2:169–195.

McCarthy, John and Alan Prince. 1993. Generalized alignment. *Yearbook of morphology* 1993, ed. by Geert Booij and Jaap van Marle, 79–153. Dordrecht: Kluwer Academic Publishers.

McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. by Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249–384. University of Massachusetts Occasional Papers in Linguistics 18. Amherst, MA: Graduate Linguistics Student Association.

Noske, Manuela. 1999. Deriving cyclicity: syllabification and final devoicing in German. *The Linguistic Review* 16.3:227–252.

Piggott, Glyne L. 1999. At the right edge of words. *The Linguistic Review* 16.2:143–185.

Prince, Alan and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. [Technical report, Rutgers University Center for Cognitive Science and Computer Science Department, University of Colorado at Boulder, 1993].

Raimy, Eric. 2011. Reduplication. *The Blackwell companion to phonology, Vol.4: Phonological interfaces*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2383–2413. Malden, MA: Wiley-Blackwell.

Rau, Der-Hwa Victoria. 1992. *A grammar of Atayal*. Taipei: Crane Publishing Co., Ltd.

Rice, Keren. 2003. On the syllabification of right-edge consonants—evidence from Ahtna (Athapaskan). *Living on the edge: 28 papers in honour of Johnathan Kaye*, ed. by Stefan Ploch, 427–448. Studies in Generative Grammar 62. Berlin and New York: Mouton de Gruyter.

Rubach, Jerzy and Geert E. Booij. 1992. Resolutions of extrasyllabicity in Slovak. *Linguistics* 30:699–729.

Sen, Ranjan. 2011. Diachronic phonotactic development in Latin: the work of syllable structure or linear sequence? *Handbook of the syllable*, ed. by Charles E. Cairns and Eric Raimy, 417–441. Brill's Handbooks in Linguistics 1. Leiden, Boston: Brill.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. PhD dissertation. Cambridge, Mass.: MIT.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. *Proceedings of LP '98: Item order in language and speech, Vol. 1*, ed. by Osamu Fujimura, Brian D. Joseph, and Bohumil Palek, 205–245. Prague: Charles University Karolinum Press.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. *The role of speech perception in phonology*, ed. by Elizabeth Hume, and Keith Johnson, 219–250. San Diego: Academic Press.

Su, Mei-chuan. 2004. taiyayu taoshan fangyan foudingci yanjiu [Negation in Taoshan Atayal]. MA thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

Tsai, Wei-tien Dylan. 2007. Conjunctive reduction and its origin: A comparative study of Tsou, Amis, and Squliq Atayal. *Oceanic Linguistics* 46.2:585–602.

Tsuchida, Shigeru. 1964. Preliminary reports on Saisiyat: Phonology. *Gengo Kenkyu* 46:42–52.

Wilbur, Ronnie. 1973. The phonology of reduplication. PhD dissertation. Urbana-Champaign: University of Illinois at Urbana-Champaign.

Wiltshire, Caroline R. 2003. Beyond codas: word and phrase-final alignment. *The syllable in Optimality Theory*, ed. by Caroline Féry, and Ruben van de Vijver, 254–268. Cambridge and New York: Cambridge University Press.

Yeh, Marie M. 1991. Saisiyat structure. MA thesis. Hsinchu: National Tsing Hua University.

Yeh, Marie M. 2000. *saixiayu cankao yufa* [*A reference grammar of Saisiyat*]. Formosan Language Series 2. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Yeh, Marie M. 2003. A syntactic and semantic study of Saisiyat verbs. PhD dissertation. Taipei: National Taiwan Normal University.

Yeh, Marie M. and Ya'aw Kaybaybaw. 2010. *Saisiyat dictionary*. Technical Report, Indigenous language dictionary compiling project I—Saisiyat. Taipei: Council of Indigenous Peoples, Executive Yuan. (http://e-dictionary.apc.gov.tw/) [In Chinese]

Yeh, Maya Yuting. 2002. Emotion concepts in Squliq Atayal. MA thesis. Taipei: National Taiwan University.

Yeh, Maya Yuting and Shuanfan Huang. 2013. *Hya'* and stance marking in Atayal. *Language and Linguistics* 14.1:135–165.

Yeh, Shih-chi Stella. 2011. Issues in Paiwan phonology. PhD dissertation. Hsinchu: National Tsing Hua University.

Zeitoun, Elizabeth (Qi, Li-sha). 2000. *bunongyu cankao yufa* [*A reference grammar of Bunun*]. Formosan Language Series 5. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Zeitoun, Elizabeth and Chen-huei Wu. 2005. Saisiyat reduplication revisited. *Concentric: Studies in Linguistics* 31.2:31–56.

Zeitoun, Elizabeth (Qi, Li-sha), Tai-hwa Chu and Lalo a Tahesh Kaybaybaw. 2011. chonglun saixiayu shuxie xitong [A discussion on the orthography of Saisiyat]. *yuanzhuminzu yuyan fazhan: lilun yu shiwu luncong* [*The development of indigenous languages: Theory and practice*], ed. by Shih-hui Lin, 276–302. Taipei: Council of Indigenous Peoples, Executive Yuan.

Zeitoun, Elizabeth, Tai-hwa Chu and Lalo a Tahesh Kaybaybaw. 2015. *A study of Saisiyat morphology*. Oceanic Linguistics Special Publication No. 40. Honolulu: University of Hawai'i Press.

4 *Squliq Atayal reduplication: Bare consonant or full syllable copying?*[*]

HUI-SHAN LIN

1 Introduction

Squliq Atayal is a Formosan language spoken in the northern mountain area in Taiwan. Just as in other Formosan languages, Atayal incorporates reduplication as a means of word formation. This paper is interested in the so called *bare consonant* reduplication reported to exist in the Squliq dialect of Wulai spoken in Wulai District, New Taipei City (earlier Wulai Township, Taipei County) (Rau 1992)[1] and of Mrqwang spoken in Jianshi Township, Hsinchu County (W. Lin 2004). Bare consonant reduplication refers to the kind of reduplication in which only the consonant is copied from the base. Bare consonant reduplication in Squliq Atayal is of two types: the C-type that exists in both Wulai and Mrqwang (e.g., *qaniq* 'to eat' > *q̲~qaniq* 'way of eating' (Rau 1992:116),[2] where the reduplicant is underlined and separated from the base by ~), and the CC-type that exists

[*] This paper is dedicated to Prof. Lillian M. Huang, who is a leading expert in Formosan linguistics. The paper has benefited a lot from comments and remarks from anonymous reviewers and the editors of this volume (especially Elizabeth Zeitoun), to whom I would like to express my deepest gratitude. I would also like to thank my Wulai Atayal consultant, Hong Long Lin (林紅隆), and my Mrqwang Atayal consultant, Tingting Payan (頂定・巴顏), for their help with the language data. This study was supported by National Science Council (now Ministry of Science and Technology) with the project number NSC 98-2410-H-003-055-.

[1] Rau (1992) refers to the dialect as Gogan. However, since data in Rau are collected in the speech community of Wulai and Gogan dialect is described as covering a much wider area, including Taipei, Taoyuan, and Yilan (Rau 1992:3), the present paper prefers to follow L.M. Huang (1995) and refers to the Squliq dialect spoken in Wulai as Wulai Atayal.

[2] The reduplication pattern is referred to as *bare consonant* reduplication in W. Lin (2004) and as partial reduplication in Rau (1992). Rau doesn't use the term *bare consonant reduplication*. But her description of the partial reduplication pattern in Wulai Atayal (i.e., "the first segment at the beginning of the full word may be reduplicated" (Rau 1992:116)) matches the definition of bare consonant reduplication. Since Rau doesn't consider reduplication on vowel initial words (Rau 1992:116 n. 1), the reduplicant (according to her description of the partial reduplication pattern) can only be the copy of the word initial consonant.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, *75–100*
Asia-Pacific Linguistics, 2015

only in Mrqwang (e.g., *lihun* 'door' > *lh*~*lihun* 'a lot of doors' (speakers are very sure that many doors are there.) (W. Lin 2004:32). The two types of reduplication patterns are special because bare consonant reduplication is not only rare in the world's languages but is also unreported in other Formosan languages. Due to the fact that Squliq Atayal has a rule of vowel reduction that affects pre-penultimate vowels (Egerod 1965, 1980; Li 1980, 1981; H. Huang 2006b; Rau 1992), whether the reduplication patterns truly involve the duplication of bare consonant(s) is dubious, since the so-called bare consonant reduplication may well be derived from full syllable copying (which refers to the kind of reduplication in which an entire syllable is copied) which later undergoes vowel reduction (e.g., *qaniq* 'to eat' > *qa*~*qaniq* > *q*~*qaniq* 'way of eating'). The purpose of this paper is thus to investigate whether the C-type and the CC-type of reduplication involve bare consonant or full syllable copying.

Based on language internal and typological evidence, this paper argues that while the C-type of reduplication truly involves bare consonant copying, the CC-type of reduplication actually involves reduplication of two syllables. Framed in an OT framework, this paper shows how the differences between the two patterns are properly captured by the slight re-ranking of a set of universal constraints.

The remainder of this paper is organised as follows: Section 2 provides a brief introduction to Squliq Atayal phonology and previous studies of Squliq Atayal reduplication. Section 3 re-examines the C-type and the CC-type of reduplication to determine whether bare consonant reduplication or full syllable copying is involved. Section 4 provides analyses to the observed generalisations based on Optimality Theory (McCarthy & Prince 1993; Prince & Smolensky 1993[2004]). Section 5 concludes the paper.

2 Squliq Atayal phonology and literature on Squliq Atayal reduplication

2.1 Squliq Atayal phonology

The Atayal language has two subgroups: C'uli' and Squliq (Li 1980, 1981, 1985). Unlike the C'uli' dialects, Squliq dialects are fairly uniform phonologically (Li 1980, 1981). The phonological system below is based on the Squliq dialect of Wulai and Mrqwang. Squliq Atayal has nineteen consonants /p, t, k, q, ʔ, c, b[β], s, z, x, g[ɣ], h, m, n, ŋ, r, l, j, w/,[3] and five vowels /i, e, a, o, u/ (Li 1980; H. Huang 2006a, 2006b).[4] Stress generally falls on the last syllable of a word. Squliq Atayal has two important phonological features. First, there is a rule of vowel reduction that changes all pre-penultimate vowels to *ə* (Egerod 1965, 1980; Li 1980, 1981; H. Huang 2006b; Rau

[3] In the Atayal literature, the two phonetic sounds *β* and *ɣ* are usually represented as *b* and *g*, respectively. I will follow this convention here.

[4] According to Rau (1992:24), Wulai Atayal has an additional vowel /y/. The high front unrounded vowel is excluded in the present paper because it has very restricted distribution (occurring only in the final syllable of a couple of words) and might be borrowed from Chinese (cf. Van Coetsem 1987; Rau 1989).

1992). Examples in (1) show that a root vowel will undergo reduction when shifted to pre-penultimate position due to suffixation.[5]

(1) Vowel reduction in pre-penultimate position

a.	/huziq, an/ >	*hziqan*	[həžiqan]	'wet (LF)'	(H. Huang 2006b:491)
b.	/betaq, an/ >	*btaqan*	[bətaqan]	'stab (LF)'	(H. Huang 2006b:492)

Another important feature of Squliq Atayal is that no complex syllable margins (i.e., syllable margins containing consonant clusters) are allowed on the surface (Egerod 1966:121, 1980; Li 1980:369; Rau 1992; H. Huang 2006a, 2006b). Underlying consonant clusters at syllable margins are separated by *ə* (as illustrated in (2)), which is conventionally omitted in the transcription due to its predictable nature.

(2) Consonant clusters at syllable margins are separated by a weak vowel

a.	*blaq*	[bəlaq]	'good'
b.	*mpuw*	[məpuw]	'count (AF)'
c.	*nbuw*	[nə**b**uw]	'drink (AF)'

The two important features of Squliq Atayal show that schwas in the language have two possible sources. For schwas that correspond to full vowels in morphologically related words (as in (1)), the schwas are the result of pre-penultimate vowel reduction; on the other hand, for schwas that lack full vowel correspondents in morphologically related forms (such as those in (2)), the schwas might come from epenthesis to repair consonant clusters (W. Lin 2004, cf. also H. Huang 2006b:492).

2.2 Literature on Squliq Atayal reduplication

This section reviews previous studies on Squliq Atayal reduplication.

Rau (1992:116) examines reduplication of the Squliq dialect of Wulai and points out that Wulai Atayal has a partial reduplication pattern that reduplicates the initial consonant of a word to emphasise "plurality" or "regularity",[6] as illustrated in (3).

(3) Reduplication in Wulai Atayal (Rau 1992:116)

a.	*btunux*	'stone(s)'	>	*b*~*btunux*	'stones'
b.	*wajaj*	'thread(s)'	>	*w*~*wajaj*	'threads'
c.	*hilaw*	'to cover'	>	*h*~*hilaw*	'covering (on the bed)'
d.	*qaniq*	'to eat'	>	*q*~*qaniq*	'way of eating'

Notice that though the reduplicant is considered as a single consonant and is transcribed as such in Rau, in actual pronunciation there is a schwa between the reduplicant and the base because Squliq Atayal does not allow complex syllable margins.

---

[5] Symbols and abbreviations used in the paper follow those of the Leipzig Glossing Rules, except for: AF, Agent Focus; LF, Locative Focus; NEU, neutral; PF, Patient Focus; PART, particle; RED, reduplicant. The glosses in the cited examples may be modified to be consistent with the list of symbols/abbreviations above.

[6] The function denoted by the reduplicated forms in (3c) and (3d) is often referred to as *instrument nominalisation*, in which a verb turns into a noun.

For ease of discussion, inter-consonantal schwa, which is conventionally omitted in the literature for its predictability, will be specified henceforth. Thus, (3) is modified as (4).

(4) Reduplication in Wulai Atayal (revised)

a.	*btunux*	'stone(s)'	>	*bə~bətunux*	'stones'
b.	*wajaj*	'thread(s)'	>	*wə~wajaj*	'threads'
c.	*hilaw*	'to cover'	>	*hə~hilaw*	'covering (on the bed)'
d.	*qaniq*	'to eat'	>	*qə~qaniq*	'way of eating'

The same reduplication pattern is also found in the Squliq dialect of Mrqwang spoken in Jianshi Township, as illustrated by the examples cited from W. Lin (2004) in (5). W. Lin refers to this kind of reduplication as C-reduplication and considers it to denote meanings such as "plurality" in nouns and "irrealis" (interpreted as "future" in the examples), "continuation" or "repetition" in verbs. Notice that since her C-reduplication is considered as involving bare consonant reduplication, the schwas before the bases in (5) are epenthesised vowels rather than the copied vowels undergone reduction in pre-penultimate position.

(5) C-reduplication in Mrqwang Atayal (W. Lin 2004:31–34)

a.	*ŋasal*	'house'	>	*ŋə~ŋasal*	'houses'
b.	*hozil*	'dog'	>	*hə~hozil*	'dogs'
c.	*kaciŋ*	'cow'	>	*kə~kaciŋ*	'cows'
d.	*lukus*	'wear'	>	*lə~lukus*	'will wear (clothes)'
e.	*maniq*	'eat'	>	*mə~maniq*	'will eat'
f.	*rəmaw*	'help'	>	*rə~rəmaw*	'help continuously'

In addition to C-reduplication, W. Lin (2004) observes another type of bare consonant reduplication, which she refers to as CC-reduplication, which reduplicates not only the first but also the second consonant of the base.[7] Examples of her CC-reduplication are

---

[7] In addition to bare consonant reduplication, W. Lin (2004) mentions that Mrqwang Atayal also incorporates total reduplication that copies the whole word, including the word-final coda, to express "intensity", "distributivity", "continuation" or "repetition", as illustrated below:

(i) W. Lin (2004:29–30)

a.	*cipoʔ*	'small'	>	*cipoʔ~cipoʔ*	'very small'
b.	*hopaʔ*	'big'	>	*hopaʔ~hopaʔ*	'very big'
c.	*həbaw*	'light'	>	*həbaw~həbaw*	'very light'
d.	*bətunux*	'stone'	>	*qutux~qutux bətunux*	'every stone'
e.	*laqiʔ*	'child'	>	*qutux~qutux laqiʔ*	'every child'

Notice that the examples above involve duplication of a whole word (e.g., (a)–(c)) as well as part of a phrase (e.g., (d) and (e)). Nonetheless, whether (d) and (e) should be considered as reduplication is questionable since they involve the copying of a phrase. Rau (1992:117) actually refers to the total duplication of word or noun phrase relating to time or measurement to convey the meaning of 'each & every so-and-so' as *doubling* rather than reduplication (e.g., *sasan sasan* 'every morning' < *sasan* 'morning', *qutux kawas qutux kawas* 'year after year' < *qutux kawas* 'one year'). Since the reduplicated forms in (d) and (e) (but not (a)–(c)) also convey the meaning of 'each & every so-and-so' they are considered as doubling rather than reduplication in this paper. For examples (a)–(c), though they involve the complete copying of the word, which is less common in Formosan languages that tend to drop the final coda in total

given in (6). Just as in her C-reduplication, the schwas before the bases in (6) are epenthesised vowels rather than the reduced form of the copied vowels since W. Lin also considers her CC-reduplication to involve bare consonant reduplication.

(6) CC-Reduplication in Mrqwang Atayal (W. Lin 2004:32–33)

a.	*lihun*	'door'	>	*<u>ləhə</u>~lihun*	'a lot of doors' (speakers are very sure that many doors are there.)
b.	*qulih*	'fish'	>	*<u>qələ</u>~qulih*	'a lot of fish' (speakers are very sure that a lot of fish are there.)
c.	*ŋasal*	'house'	>	*<u>ŋəsə</u>~ŋasal*	'a lot of houses' (speakers are very sure that a lot of houses are there.)
d.	*guqoh*	'banana'	>	*<u>gəqə</u>~guqoh*	'a lot of bananas' (speakers are very sure that a lot of bananas are there.)
e.	*ŋarux*	'bear'	>	*<u>ŋərə</u>~ŋarux*	'a lot of bears' (speakers are very sure that a lot of bears are there.)

For the function of her CC-reduplication, W. Lin also considers it to denote the meaning of "plurality" (W. Lin 2004:32). Nonetheless, a comparison of the glosses given by W. Lin for her C-reduplication and CC-reduplication (e.g., *ŋasal* 'house' > *<u>ŋə</u>~ŋasal* 'houses' vs *ŋasal* 'house' > *<u>ŋəsə</u>~ŋasal* 'a lot of houses' (speakers are very sure that a lot of houses are there)) shows that the meanings denoted by the two reduplication patterns are not exactly the same. The following data collected firsthand by the author also support that the CC-type and the C-type of reduplication are used to convey different meanings in Mrqwang Atayal. Thus, the two types of reduplication patterns are more properly considered as functionally distinct and their reduplicants different morphemes.

(7)

	root	I. C-type of reduplication	II. CC-type of reduplication
a.	*laqiʔ* 'child'	*<u>lə</u>~laqiʔ* 'children'	*<u>ləqə</u>~laqiʔ* 'children of different families'
b.	*biruʔ* 'book'	*<u>bə</u>~biruʔ* 'books'	*<u>bərə</u>~biruʔ* 'books of various kinds'
c.	*hozil* 'dog'	*<u>hə</u>~hozil* 'dogs'	*<u>həzə</u>~hozil* 'dogs of various kinds'
d.	*qulih* 'fish'	*<u>qə</u>~qulih* 'fish (pl)'	*<u>qələ</u>~qulih* 'fish of various kinds'

(7') a. i. *aras <u>lə</u>~laqiʔ gaʔ!*
bring RED~child PART
'Bring the children (of the same family)!'

ii. *aras <u>ləqə</u>~laqiʔ gaʔ!*
bring RED~child PART
'Bring the children (of different families)!'

a'. i. *<u>lə</u>~laqiʔ=mu kwaraʔ qani.*
RED~child=1SG.GEN all this
'All these are my children.'

---

reduplication (Lee 2007:65–67, cf. also Li & Tsuchida 2001:22), the present study inclines to consider them as involving total reduplication since full reduplication with final coda can still be observed in Formosan languages (cf. Zeitoun & Wu 2006) as well as Austronesian languages (cf. section 6.9.5.1 in Blust 2009).

ii. * *ləqə~laqiʔ=mu* *kwaraʔ* *qani.*
RED~child=1SG.GEN all this
*‘All these are my children (of different families).’

b’. i. *nanak* *bə~biruʔ* *njux* *kabaŋ=mu.*
only RED~book IPFV.PROX school.bag=1SG.GEN
‘My school bag has all and only books (of one kind).’

ii. *nanak* *bərə~biruʔ* *njux* *kabaŋ=mu.*
only RED~book IPFV.PROX school.bag=1SG.GEN
‘My school bag has all and only books (of various kinds).’

c’. i. *giwaj* *hə~hozil* *ŋasal=njaʔ.*
completely RED~dog house=3SG.GEN
‘His house is full of dogs (of one kind).’

ii. *giwaj* *həzə~hozil* *ŋasal=njaʔ.*
completely RED~dog house=3SG.GEN
‘His house is full of dogs (of various kinds).

d’. i. *ʔusa* *bazij* *qə~qulih!*
Go buy RED~fish
‘Go and buy some fish (of one kind)!’

ii. *ʔusa* *bazij* *qələ~qulih!*
Go buy RED~fish
‘Go and buy some fish (of various kinds)!’

Before clarifying whether the above reduplication patterns truly involve bare consonant reduplication, the present study will temporarily refer to them using two neutral terms, the C-type of reduplication, covering the reduplication pattern in (4), (5), and (7I) and the CC-type of reduplication, covering the reduplication pattern in (6) and (7II).

3 Bare consonant reduplication or full syllable copying?

The reduplication patterns illustrated in the previous section are assumed to involve bare consonant reduplication in both Rau (1992) and W. Lin (2004). But a question arises as to whether they truly involve bare consonant reduplication rather than full syllable copying. Bare consonant reduplication is indeed rare in the world’s languages given that reduplication normally involves the copying of prosodic constituents. This tendency is captured by the Prosodic Morphology Hypothesis (McCarthy & Prince 1995b:318) which states that “templates are defined in terms of the authentic units of prosody: mora (μ), syllable (σ), foot (Ft), prosodic word (PrWd).” If reduplication in Squliq Atayal truly involves bare consonant reduplication, as assumed in W. Lin (2004) and Rau (1992), the reduplicant will not belong to any of the prosodic units. Moreover, although bare consonant reduplication can be found in a few languages such as Levantine, Arabic, Zuni, Quileute (Broselow & McCarthy 1983), Temiar (Broselow & McCarthy 1983; Gafos 1998), Semai, Marshallese, Coushatta, Yokuts, and Secwepemc (Hendricks 1999, 2001),

no Formosan language other than Squliq Atayal is reported to display such kind of reduplication (Zeitoun & Wu 2006:125). Most importantly, due to the fact that Squliq Atayal has a rule of vowel reduction that targets pre-penultimate vowels (Egerod 1965, 1980; Li 1980, 1981; Rau 1992; H. Huang 2006b), what appears to be bare consonant reduplication may well be the result of full syllable copying, which later undergoes vowel reduction. Take *ŋasal* 'house' for example. As shown in (8), both bare consonant reduplication (8a) and full syllable copying (8b) can generate the same reduplicated form ***ŋə***~*ŋasal* 'houses'. The reduplicated forms derived from (8a) and (8b) differ only in whether the schwa standing before the base is inserted between the reduplicant and the base to break up the CC cluster or is a copied vowel which later undergoes reduction due to the pre-penultimate vowel reduction rule. Due to the fact they are pronounced exactly the same, evidence needs to be provided to justify whether bare consonant reduplication is truly involved. For the sake of clarity, inter-consonantal inserted schwa and pre-penultimate reduced schwa in reduplication here below are represented as *ə* and *ə̠*, respectively.

(8) Two analyses of the C-type of reduplication,
e.g., *ŋasal* 'house' > ***ŋə***~*ŋasal* 'houses'

a. Bare consonant reduplication:
ŋasal → ***ŋ***~*ŋasal* → ***ŋə***~*ŋasal*
first C copying *V insertion to break up CC cluster*

b. Full syllable copying:
ŋasal → ***ŋa***~*ŋasal* → ***ŋə***~*ŋasal*
first σ copying *pre-penultimate V reduction*

3.1 The C- type of reduplication as bare consonant reduplication

As a matter of fact, Zeitoun & Wu (2006:104) suspect that the C-type of reduplication in Squliq Atayal might have evolved from CV- and C*a*-reduplication (which belong to full syllable copying in the present study),[8] the first pattern conveying meanings such as (i) plural/collective, (ii) continuous/repetitive aspect, or (iii) intensification, and the second denoting other meanings such as (i) predictive/irrealis/future or (ii) reciprocity. Their construal is plausible since both CV- and C*a*-reduplication are common in Formosan languages, the former pattern conveys the meanings mentioned above and the latter pattern, in addition to predictive/irrealis/future and reciprocity, can also convey the meaning of instrument nominalisation (Blust 1998; Liao 2011; Yeh 2009; Zeitoun 2002, 2012). Therefore, examples of the C-type of reduplication in (4a–b) and (5a–c), which denote plurality, and in (5f), which denotes continuity, may be considered as being evolved from CV-reduplication; and those in (4c–d), which denote instrument nominalisation, in (5d–e), which signify future, and in (9), which denotes reciprocity, may be considered as being evolved from C*a*-reduplication.

---

[8] C*a*-reduplication, first coined by Blust (1998) refers to the copying of base initial consonant followed by a fixed vowel *a*. In this paper, C*a*-reduplication is considered to fall into the category of full syllable copying (just as CV-reduplication).

(9)	*bihij*	'to hit'	>	*bə~bihij*	'to hit each other'
	sjaqeh	'to hate'	>	*mə-sə~sjaqeh*	'to hate each other'
	soja?	'to love'	>	*mə-sə~ soja?*	'to love each other'

Zeitoun & Wu briefly mention in a footnote (Zeitoun & Wu 2006: n. 11) that their hypothesis is supported by a dialectal comparison with C'uli' Atayal, though no examples are provided to illustrate their point.

The examples in (10), which come from the C'uli' dialects of Mayrinax, Matabalay, and Mabalay, may seem to serve as supporting evidence to the hypothesis of full syllable copying since in the reduplicated forms, the reduplicants take the form of C*a*- rather than C-.[9]

(10) Reduplication in the C'uli' dialects of Mayrinax, Matabalay, and Mabalay

Based on L. Huang (2000)

a. Mayrinax	*maniq*	'eat'	>	*na~niq-an*	'restaurant' (p. 60)
b. Mayrinax	*patiaq*	'read'	>	*pa~patiaq-an*	'school' (p. 60)
c. Mayrinax	*hihip-an*	'kiss'	>	*ha~hihip-an*	'will kiss (LF)' (p. 114)

Based on W. Lin (2004)

d. Matabalay	*biru?*	'study'	>	*ba~biru?*	'will study' (p. 45)
e. Matabalay	*ruma?*	'house'	>	*ra~ruma?*	'houses' (p. 31)
f. Matabalay	*ʔula?i?*	'child'	>	*ʔa~ʔula?i?*	'children' (p. 31)
g. Mabalay	*h‹am›ili?*	'accuse someone'			

> *ma-ha~hili?*[10] 'make accusations back and forth' (Lambert 1999:26)

However, two things are worth noting. First, though the reduplicants of the C'uli' Atayal examples in (10) take the form of C*a*-, the reduplicated forms (at least some of them) may not truly involve C*a*-reduplication because the examples not only denote meanings of instrument nominalisation (10a–b) and irrealis/future (10c–d), but also the function of plurality (10e–f) and continuity (10g), which are not typically conveyed by C*a*-reduplication (but by CV-reduplication) in the Formosan languages (cf. Zeitoun & Wu 2006:125-126).

Second, though the reduplicated forms as observed in (10) seem to involve C*a*-reduplication, they could also be bare consonant reduplication in disguise. Just as in Squliq Atayal, C'uli' Atayal also disallows consonant clusters and have a rule of inter-consonantal vowel insertion to break up adjacent consonants; but different from Squliq dialects, which choose *ə* as the inserted vowel, C'uli' dialects such as Matabalay and Mabalay tend to break up consonant clusters by inserting vowel *a* (W. Lin 2004:14; Lambert 1999:18–20). This can be illustrated by examples below:

---

[9] As a matter of fact, (10b) is ambiguous. Since the first vowel of the base is *a*, the reduplicant in (10b) can also be considered as taking the form of CV-. However, due to the fact that the function denoted by reduplication in (10a) and (10b) is identical (i.e., instrument nominalisation) and that (10a) cannot be involving CV-reduplication, the reduplicant in (10b) is considered as taking the form of C*a*-.

[10] For the Mabalay data cited from Lambert (1999), the present paper ignores some of the phonetic details and replaces the three segments *ħ*, *ɮ*, and *ɣ* with *h*, *l* and, *g* for simplicity.

(11) Insertion in the C'uli' dialects to break up CC cluster
Based on W. Lin (2004)

a.	Matabalay	/hmkaŋi/	[h**a**mkaŋi]	'find' (p. 15)
b.	Matabalay	/m-s-jaqih/	[m**a**sjaqih]	'hate each other' (p. 15)
c.	Matabalay	/s-m-p-papak/	[s**a**mp**a**papak]	'very noisy in strong intensity meaning' (p. 16)

Based on Lambert (1999:20)

d.	Mabalay	/pgi-an/	[p**a**gian]	'dry (clothes)
e.	Mabalay	/spuh-i/	[s**a**puhi]	'sweep'

Given the fact that the C'uli' dialects have a rule of inter-consonantal *a* epenthesis, C*a*-reduplication in (10) could also come from bare consonant reduplication, which later on undergoes inter-consonantal vowel insertion, as illustrated below:

(12) Ca-reduplication in C'uli' Atayal as bare consonant reduplication in disguise
e.g., *ruma?* 'house' > ***ra***~*ruma?* 'houses'

ruma?	→	*r*~*ruma?*	→	*ra*~*ruma?*
	first C copying		*V insertion to break up CC cluster*	

In sum, the C'uli' Atayal examples in (10) may not truly involve C*a*-reduplication and, therefore, do not serve as strong evidence in support of the hypothesis of full syllable copying. It would be out of the scope of the present paper to provide an in-depth examination of reduplication in C'uli' Atayal. Even if C'uli' Atayal truly incorporates CV- and C*a*-reduplication (i.e., full syllable copying), it doesn't mean the same must be true in Squliq Atayal. As H. Huang (2006b:497) points out, cross-dialectal comparison cannot serve as strong evidence because "different dialects could develop into different patterns in the course of time." Therefore, it would not be surprising if C'uli' dialects incorporate full syllable reduplication while Squliq dialects undergo bare consonant reduplication.

Based on language internal evidence, this paper argues for the bare consonant hypothesis for the C-type of reduplication in Squliq Atayal. As shown in (8), for the C-type of reduplication, bare consonant reduplication and full syllable copying would derive the same output. But this is true only when the root undergoing reduplication is disyllabic or longer. This is because the vowel reduction rule in Squliq Atayal targets pre-penultimate vowels;[11] only when the root undergoing the C-type of reduplication is disyllabic or longer will the vowel of the reduplicant be located in pre-penultimate

[11] The examples below clearly demonstrate that vowel reduction targets pre-penultimate but not penultimate vowels when affixation takes place. The roots in (a) and (b) are disyllabic, while those in (c) and (d) are monosyllabic. When suffixed with a monosyllabic affix, the first vowel of the roots in (a) and (b) are shifted to pre-penultimate syllable and undergo reduction. On the other hand, the first (and only) vowel of the roots in (c) and (d), though also shifted leftward due to suffixation, does not undergo reduction because it is situated in the penultimate syllable.

(i)	a.	/huziq, an/	>	*hziqan*	[h**ə**žiqan]	'wet (LF)' (H. Huang 2006b:491)
	b.	/betaq, an/	>	*btaqan*	[b**ə**taqan]	'to stab (LF)' (H. Huang 2006b:492)
	c.	/si(ʔ), an/	>	*sian*	[s**i**an]	'to put (LF)' (Egerod 1999:280)
	d.	/kat, un/	>	*katun*	[k**a**tun]	'to bite (PF)' (Egerod 1999:106)

position and be the target of the vowel reduction rule. Therefore, examining the C-type of reduplication that targets monosyllabic root can help clarify whether bare consonant or full syllable copying is involved. As shown in (13), bare consonant reduplication and full syllable copying will yield different results when the root undergoing the C-type of reduplication is monosyllabic.

(13) Bare consonant reduplication and full syllable copying yield different results when the root is monosyllabic

a. Full syllable copying will predict **a full vowel** in the reduplicant

zik → ***zi***~*zik* → ***zi***~*zik*

first σ copying *pre-penultimate V reduction*

NOT APPLICABLE!!!

b. Bare consonant reduplication will predict **a schwa vowel** in the reduplicant

zik → ***z***~*zik* → ***zə***~*zik*

first C copying *V insertion to break up CC cluster*

If a monosyllabic root undergoes CV- or C*a*-reduplication, when no suffixes are added, the reduplicated form will be disyllabic in length and the reduplicant will be standing at the penultimate syllable and free from the rule of vowel reduction that targets pre-penultimate syllables (ref. (13a)). In other words, if it is full syllable copying that is involved (i.e., CV- or C*a*-reduplication), the reduplicant vowel must be a copy of the root vowel or a fixed *a* when the root is monosyllabic. On the other hand, if the reduplicant vowel is a schwa, the schwa can only come from insertion and what is copied has to be a bare consonant (ref. (13b)).

Reduplication that applies on monosyllabic root is rare in Squliq Atayal due to the fact that the most canonical root in Squliq Atayal is disyllabic (Rau 1992:26). Nonetheless the examples in (14a) cited from Zeitoun & Wu (2006:105) as well as the examples in (14b) collected firsthand by the author clearly show that the vowel of the reduplicant corresponding to the monosyllabic root is always a schwa rather than the copy of the root vowel or a fixed vowel *a*.

(14) The C-type of reduplication applying on monosyllabic root

a. Zeitoun & Wu (2006:105)

i.	*zik*	'below' >	*zə*~*zik*	'very deep'	**zi*~*zik*
ii.	*mə-kut*	'to cut' >	*mə-kə*~*kut*	'to cut each other'	**mə-ka*~*kut*[12]

b.

i.	*kə‹m›at*	'to bite' >	*mə-kə*~*kat*	'to bite each other'	**mə-ka*~*kat*
ii.	*rə‹m›aw*	'to help' >	*mə-rə*~*raw*	'to help each other'	**mə-ra*~*raw*
iii.	*tə‹m›ux*	'to bark' >	*mə-tə*~*tux*	'to bark at each other'	**mə-ta*~*tux*
iv.	*mit*	'goat' >	*mə*~*mit*	'goats'	**mi*~*mit*

[12] Zeitoun (2002, 2012) has reconstructed PAn *ma-Ca~ as one of the reciprocal markers based on Formosan languages. In the synchronic grammar of Squliq Atayal, the reciprocal marker (which takes the surface form of [mə-Cə~]) is reflected as /ma-C~/. This is because while [mə-] before the reduplicant can be considered as derived from /ma-/ due to the pre-penultimate vowel reduction rule, the schwa between the copied consonant and the base cannot come from the reduction of the fixed vowel /a/ since it is situated at the penultimate syllable.

(14') a. *cjux* ***k‹əm›at*** *ŋjaw* *hozil* *qasa.*
IPFV.DIST bite<AF> cat dog that
'The dog is biting the cat.'

a'. *cjux* ***mə-kə~kat*** *hozil* *qasa.*
IPFV.DIST AF-RED~bite dog that
'The dogs are biting each other.'

b. *m-usa=saku?* ***r‹əm›aw*** *mə-qumah* *hija?* *suxan.*
AF-go=1SG.NOM <AF>help AF-work 3SG.NEU tomorrow
'I will work for him tomorrow.'

b'. ***mə-rə~raw****=sami.*
AF-RED~help=1PLEXCL.NOM
'We help each other.'

c. *cjux* ***t‹əm›ux*** *hozil* *qasa.*
IPFV.DIST <AF>bark dog that
'The dog is barking.'

c'. *cjux* ***mə-tə~tux*** *hozil* *qasa.*
IPFV.DIST AF-RED~bark dog that
'The dogs are barking at each other.'

d. *ciŋaj* ***mə~mit***
many RED~goat
'many goats'

Take (14ai) and (14aii) for example. In (14ai), the vowel before the monosyllabic root is a schwa rather than a copy of the root vowel (i.e., not undergoing CV- reduplication) even though the reduplicated form conveys the meaning of intensity. In (14aii), on the other hand, the vowel before the monosyllabic root is a schwa rather than the fixed vowel *a* (i.e., not undergoing C*a*-reduplication) even though the reduplicated form conveys the meaning of reciprocity. In all of the examples in (14), since the schwa vowels before the roots are not situated at pre-penultimate syllables, they cannot be derived from vowel reduction. Therefore, reduplication of monosyllabic root in (14) clearly shows that the C-type of reduplication truly involves bare consonant reduplication rather than CV-reduplication or C*a*-reduplication which latter undergoes vowel reduction.

The C-type of reduplication, being justified as involving bare consonant reduplication, is referred to as C-reduplication hereafter. In other words, in the synchronic grammar of Squliq Atayal, neither CV- nor C*a*-reduplication is active. The functions commonly denoted by CV- and C*a*-reduplication in other Formosan languages (i.e., plural/collective, continuous/repetitive aspect or intensification in the former and predictive/irrealis/future, reciprocity or instrument nominalisation in the latter) are conveyed by C-reduplication in the synchronic grammar of Squliq Atayal. The present study in no way attempts to refute the possible existence of CV- or C*a*-reduplication in the diachronic grammar of Squliq Atayal; rather it simply tries to show that neither CV- nor C*a*-reduplication is active in the synchronic grammar of Squliq Atayal.

3.2 The CC-type of reduplication as full syllable copying

How about the CC-type of reduplication? Unlike C-reduplication, there is no clear language internal evidence supporting the CC-type of reduplication as involving bare consonant reduplication. This is because the CC-type of reduplication involves the copying of the initial consonants of the first two syllables of the base (W. Lin 2004:32); therefore, the stem that undergoes such reduplication must be bigger than two syllables and the reduplicant, which is prefixal, will always be situated at pre-penultimate position and be the target of vowel reduction. As the vowels in the reduplicant can only surface as *ə*, there is no way to determine whether it comes from insertion, if the reduplicant constitutes bare consonants (15a), or from reduction, if the reduplicant contains two full syllables (15b).

(15) Two analyses of the CC-type of reduplication,
e.g., *laqiʔ* 'child' > ***ləqə***~*laqiʔ* 'children of different families'

a. Bare consonant reduplication

laqiʔ → *lq*~*laqiʔ* → *ləqə*~*laqiʔ*

copying of the first two consonants; *V insertion to break up CC cluster*

b. Full syllable copying

laqiʔ → *laqi*~*laqiʔ* → *ləqə*~*laqiʔ*

copying of the first two syllables; *pre-penultimate vowel reduction*

Of course, categorising the CC-type of reduplication as involving bare consonant reduplication just like C-reduplication has the advantage of making the two reduplication patterns similar. However, since functionally distinct reduplication patterns do not need to be alike in terms of form (cf. Kennedy 2008, H-s. Lin 2010, Urbanczyk 2006), the fact C-reduplication involves bare consonant reduplication does not necessarily mean the CC-type of reduplication must be, too. Though the present paper agrees with Rau (1992) and W. Lin (2004) in categorising the C-type of reduplication as undergoing bare consonant reduplication, the paper argues that the CC-type of reduplication involves full syllable copying (i.e., (15b)) based on the following reasons. First of all, bare consonant reduplication is universally rare and is against the observation that reduplicants tend to copy prosodic units (Prosodic Morphology Hypothesis, McCarthy & Prince 1995b:318). On the other hand, disyllabic reduplication is quite common, especially in Formosan languages (Zeitoun & Wu 2006).[13] Therefore, it is better to categorise the CC-type of reduplication as disyllabic reduplication rather than adopting the bare consonant analysis, unless there is strong evidence supporting the latter approach. In addition, pre-penultimate vowel reduction is a productive rule in Squliq Atayal; surface CC-reduplicant can be naturally derived from disyllabic reduplication which later undergoes vowel reduction (cf. (15b)). Most importantly, though bare consonant reduplication of two bare consonants can be found in languages like Semai, which is an Austro-Asiatic language spoken in West Malaysia, the contents of the reduplicants in

[13] According to Zeitoun & Wu (2006), disyllabic reduplication is productively found in the Formosan languages, along with C*a*-reduplication, CV-reduplication, and rightward reduplication.

Semai are different from that in Mrqwang Atayal. The comparison of examples of the CC-type of reduplicant in Mrqwang Atayal in (6) and (7II) with CC-reduplication in Semai in (16) reveals that while the two consonants of the reduplicants in Mrqwang come from *the first two consonants* of the base, those in Semai come from *the peripheral edges* of the base.

(16) Bare consonant reduplication in Semai (Diffloth 1976:252)

pn~payaɲ	'appearance of being disheveled'
ct~cʔɛːt	'sweet'
kc~kmrʔɛːc	'short, fat arms'

The pattern as observed in Semai is not accidental. Hendricks (1999) examines several languages that involve bare consonant reduplication, including Semai, Marshallese, Coushatta, Yokuts, and Secwepemc; all of them copy segment(s) from the peripheral edge(s) of a sting to which the reduplicant corresponds. (17) cited from Hendricks (2001:300, cf. also Hendricks 1999:264) clearly shows that the CC-reduplicants that do not come from the peripheral edges of the base are impossible.

(17) Possible and impossible CC-reduplicants (Hendricks 2001:300)

Input stem	Reduplicants	
C_1VC_2	C_1C_2	
C_1VC_2 VC_3	C_1C_3	*C_1C_2, C_2C_3
C_1VC_2 VC_3 VC_4	C_1C_4	*C_1C_2, C_1C_3, C_2C_3, C_2C_4, C_3C_4

Thus, typologically speaking, the CC-type of reduplication in Squliq Atayal cannot be involving bare consonant reduplication since the two consonants of the reduplicant do not come from the edges of the base. Namely, the CC-type of reduplication actually copies two syllables and will be referred to as disyllabic reduplication hereafter.

In sum, it has been shown that C-reduplication and disyllabic reduplication, which are functionally distinct, involve different reduplication mechanisms, the former copying a bare consonant while the latter, two full syllables. The following section proposes OT analyses to the two reduplication patterns.

4 OT analyses to C- reduplication and disyllabic reduplication

This section provides OT analyses to C-and disyllabic reduplication. (18) and (19) respectively list the generalisations of C-reduplication and disyllabic reduplication to be accounted for.

(18) Generalisations of C-reduplication
- a. The reduplicant is a prefix.
- b. The left edge of the reduplicant matches the left edge of the base.
- c. A single consonant is copied.
- d. A weak vowel is inserted between the copied consonant and the base. (That is, the *ə* before the base comes from inter-consonantal insertion.)

(19) Generalisations of disyllabic reduplication
- a. The reduplicant is a prefix.
- b. The left edge of the reduplicant matches the left edge of the base.

c. The reduplicant is disyllabic in shape.
d. The vowels in the reduplicant are reduced to schwa. (That is, the *ə*'s before the base are copied vowels that undergo reduction.)

4.1 Constraint ranking shared by C- and disyllabic reduplication

C-reduplication and disyllabic reduplication are identical in two respects: The reduplicant is prefixal and its left edge matches the left edge of the base. The prefixal nature of the reduplicant in the two reduplication patterns suggests that ALIGN-RED-L, which requires the reduplicant (RED) to appear at the beginning of the word, must dominate ALIGN-ROOT-L, which asks the root to appear at the left edge of the word, as exemplified in (22) and (23).[14]

(20) ALIGN-RED-L: Align the left edge of a RED with the left edge of the word.

(21) ALIGN-ROOT-L: Align the left edge of the root with the left edge of the word.

(22) C-reduplication: ALIGN-RED-L » ALIGN-ROOT-L
bə~*bətunux* 'stones' (← *bətunux* 'stone(s)')
/RED, btunux/
bə~*bətunux* > *bə*~*bə*~*tunux*

(23) Disyllabic reduplication: ALIGN-RED-L » ALIGN-ROOT-L
ləqə~*laqiʔ* 'children of different families' (← *laqiʔ* 'child')
/RED, laqiʔ/
ləqə~*laqiʔ* > *ləqə*~*laqi*~*ʔ*

On the other hand, the edge-matching of the reduplicant and the base at the left edge in C- and disyllabic reduplication is easily accounted for by the family of anchoring constraint (McCarthy & Prince 1993, 1995a). (25) and (26) illustrate how ANCHOR-BR-L predicts the BR edge-matching of C- and disyllabic reduplication,

---

[14] Notice that though bare consonant reduplication is prefixal, the reduplicant is not in absolute initial position, as shown in examples like *mə-rə*~*raw* 'to help each another' (< *rə‹m›aw* 'to help'). In the example, it is preceded by the prefix *mə-*. This can be accounted for by ranking constraints requiring the prefixes such as *mə-* to be left aligned with the word (e.g., ALIGN-mə-L) above ALIGN-RED-L. Assuming alignment constraints are categorical (cf. discussion in section 4.2), the ALIGN-RED-L constraint needs to be redefined as "the RED is not preceded by more than a syllable within the word." (in the data, the prefix that precedes the reduplicant is always one syllable in size; cf. McCarthy 2003:94). The tableau below illustrates how the constraint ranking ||ALIGN-mə-L » ALIGN-RED-L » ALIGN-ROOT-L|| makes the correct prediction.

mə-rə~*raw* 'to help each another' (← *rə‹m›aw* 'to help')

/mə-, RED, raw/	ALIGN-*mə*-L	ALIGN-RED-L	ALIGN-ROOT-L
a. rə~‹mə›raw	*!		*
b. mə-ra~ra~w		*!	*
☞c. mə-rə~raw			*

respectively. The constraint is never violated and is dominant in both reduplication patterns.

(24) ANCHOR-BR-L: The left peripheral element of a RED corresponds to the left peripheral element of the base.

(25) C-reduplication: ANCHOR-BR-L
ŋə~ŋasal 'houses' (← *ŋasal* 'house')
/RED, ŋasal/
ŋə~ŋasal > *sə~ŋasal*

(26) Disyllabic reduplication: ANCHOR-BR-L
ləqə~laqiʔ 'children of different families' (← *laqiʔ* 'child')
/RED, laqiʔ/
ləqə~laqiʔ > *qəʔə~laqiʔ*

The constraint ranking for C- and disyllabic reduplication proposed so far are summarised in (27).

(27) Constraint ranking shared by C-reduplication and disyllabic reduplication

ANCHOR-BR-L, ALIGN-RED-L
|
ALIGN-ROOT-L

4.2 Ranking differences between C- and disyllabic reduplication

Now we shall consider the differences between the two reduplication patterns. There are two features that distinguish C- and disyllabic reduplication: the reduplicant size and the nature of the schwa vowels before the base. For the reduplicant size, what is copied in C-reduplication is a single segment but that of disyllabic reduplication is a bi-syllable. In the literature, templatic constraints (McCarthy & Prince 1993) such as RED = PCat, which ensure the mapping of reduplicant to prosodic categories, is commonly used to account for the reduplicant shape. Nonetheless, templatic constraints are known to have the problem of predicting non-existing patterns (McCarthy & Prince 1995a, 1995b, Hendricks 1999, 2001; Crowhurst 2004). Besides, since the C-reduplicant is a bare consonant and does not belong to any of the prosodic categories, templatic constraint simply cannot work. Hendricks (1999) examines several languages that undergo bare consonant reduplication and proposes a Compression Model that properly accounts for the phenomenon (cf. also Hendricks 2001). Based on the Compression Model, the shape of the C-reduplicant in Squliq Atayal can be easily analysed as the result of the competition between edge-matching faithfulness (ANCHOR) and root/reduplicant alignment; the ranking ||ALIGN-RED-L » ALIGN-ROOT-L|| proposed above in (22) can help compress the C-reduplicant to its minimal size. This is because the reduplicant is squeezed between the left edge of the word and that of the root and any additional segment standing between the reduplicant and the root will cause the root to be further away from the left edge of the word and incurs additional violations in ALIGN-ROOT-L (compare (29d) with (29c)). The ANCHOR-BR-L constraint proposed above in (24), on the other hand, functions to predict that if there is only a single consonant copied, it will

be the base initial segment (compare (29d) with (29a)). (29) also shows that ALIGN-ROOT-L must outrank MAX-BR, which favors total reduplication, to ensure the minimal copying (compare (29d) with (29b)).

(28) MAX-BR: Every segment in the base has a correspondent in the reduplicant.

(29) C-reduplication: ANCHOR-BR-L, ALIGN-RED-L » ALIGN-ROOT-L » MAX-BR
ŋə~ŋasal 'houses' (← *ŋasal* 'house')

/RED, ŋasal/	ANCHOR-BR-L	ALIGN-RED-L	ALIGN-ROOT-L	MAX-BR
a. sə~ŋasal	*!		**	****
b. ŋasal~ŋasal			***!**	
c. ŋəs~ŋasal			***!	***
☞d. ŋə~ŋasal			**	****

Notice that for the Compression Model to work, the alignment constraints must be evaluated gradiently. In (29), the root of all of the candidates is not aligned to the left edge of the word. (29d) wins over (29b) and (29c) simply because its root is closer to the left edge of the word, incurring fewer violations in ALIGN-ROOT-L. Had ALIGN-ROOT-L been evaluated categorically, all candidates in (29) will incur a violation in the constraint and MAX-BR will wrongly select (29b) as the optimal candidate.

Nonetheless, McCarthy (2003) has argued that alignment constraints should be categorical because gradient alignment constraints are not only unnecessary but may predict typologically unattested patterns. Thus, here below I will propose an alternative analysis in which the alignment constraints are evaluated categorically. In the new analysis, the constraint ranking ||ALIGN-RED-L » ALIGN-ROOT-L|| merely functions to predict the prefixal status of the reduplicant in both C- and disyllabic reduplication. The minimisation of the reduplicant to its smallest size is attributed to DEP-OO-SEG (Benua 1997), which is an OO-faithfulness constraint that requires every segment in the reduplicated form to have a correspondent in its corresponding unreduplicated form (Gouskova 2003, 2004). By ranking it above MAX-BR, the reduplicant will be limited to a bare consonant, as illustrated in (31).

(30) DEP-OO-SEG (Benua 1997; Gouskova 2003, 2004)
Every segment in the reduplicated form has a correspondent in its corresponding unreduplicated form.

(31) C-reduplication: ANCHOR-BR-L, ALIGN-RED-L » ALI-ROOT-L, DEP-OO-SEG » MAX-BR
ŋə~ŋasal 'houses' (← *ŋasal* 'house')

/RED, ŋasal/	ANCHOR-BR-L	ALIGN-RED-L	ALIGN-ROOT-L	DEP-OO-SEG	MAX-BR
a. sə~ŋasal	*!		*	**	****
b. ŋasal~ŋasal			*	***!**	
c. ŋəs~ŋasal			*	***!	***
☞d. ŋə~ŋasal[15]			*	**	****

[15] Three potential candidates that might beat or tie with the attested output based on the current constraint ranking are (a) *ŋ~ŋasal*, (b) *ŋəᵢ~ŋaᵢsal*, and (c) *ŋa~ŋasal*. Candidate (a) contains

The reduplicant in disyllabic reduplication, on the other hand, is disyllabic in shape. Therefore, the constraint ranking ||DEP-OO-SEG » MAX-BR||, which helps to compress the C-reduplicant to its minimal size, will make the wrong prediction for disyllabic reduplication. Some additional constraints are necessary to predict the disyllabic shape of the reduplicant in disyllabic reduplication. The disyllabic shape of the reduplicant can be accounted for by RED-PRWD-L and RED-PRWD-R, which respectively requires the reduplicant to start and end with a prosodic word (cf. Crowhurst 2004; H-s. Lin 2010, 2012). By assuming that a prosodic word must dominate a foot (i.e., with undominated HEADNESS constraint) and that a foot is disyllabic (i.e., FTBIN), a reduplicant will be minimally disyllabic.

(32) RED-PRWD-L: Align the left edge of a RED with the left edge of a prosodic word.

(33) RED-PRWD-R: Align the right edge of a RED with the right edge of a prosodic word.

The analysis based on RED-PRWD-L and RED-PRWD-R is atemplatic since RED-PRWD-L and RED-PRWD-R together only require the left and the right edges of the reduplicant to be aligned with the left and right edges of some prosodic word. Therefore, candidates such as [(σσ)]-BASE (where […] = prosodic word, (…) = foot), [(σσ)][(σσ)]-BASE, and [(σσ)][(σσ)][(σσ)]-BASE all satisfy both constraints even though the reduplicants contain a different number of prosodic word. The exact disyllabic size of the reduplicant can be compressed by ranking DEP-OO-SEG below RED-PRWD-L and RED-PRWD-R. This is because DEP-OO-SEG invites the reduplicant to be as small as possible and will compress the reduplicant to its minimal (disyllabic) size. (34) illustrates how ||RED-PRWD-L, RED-PRWD-R » DEP-OO-SEG|| works to predict the disyllabic size of the reduplicant.

(34) Disyllabic reduplication: RED-PRWD-L, RED-PRWD-R » DEP-OO-SEG
qələ~*qulih* ‘fish of various kinds’ (← *qulih* ‘fish’)

/RED, qulih/	RED-PRWD-L	RED-PRWD-R	DEP-OO-SEG
a. [(qələ)]hə~qulih		*!	******
b. [(qələh)]~qulih			*****!
☞c. [(qələ)]~qulih			****

Notice that just as in C-reduplication, DEP-OO-SEG must outrank MAX-BR to rule out candidates involving total reduplication, as illustrated in (35).

(35) Disyllabic reduplication: DEP-OO-SEG » MAX-BR
qələ~*qulih* ‘fish of various kinds’ (← *qulih* ‘fish’)
/RED, qulih/
[(*qələ*)]~*qulih* > [(*qələh*)]~*qulih*

---

complex syllable margins and will be shown in (38) to be ruled out by *COMP-M. Candidate (b) contains a reduced vowel which decreases base-reduplicant identity and will be shown in (41) to be ruled out by IDENT-BR[V-PLACE]. Finally, candidate (c) is against the pre-penultimate reduction rule in the language and will be shown in n. 18 to be ruled out by the PL-MARKEDNESS constraint.

Unlike in disyllabic reduplication, RED-PRWD-L and RED-PRWD-R in C-reduplication must be dominated by DEP-OO-SEG to correctly predict the reduplicant shape, as illustrated below:

(36) C-reduplication: DEP-OO-SEG » RED-PRWD-L, RED-PRWD-R
ŋə~*ŋasal* 'houses' (← *ŋasal* 'house')

/RED, ŋasal/	DEP-OO-SEG	RED-PRWD-L	RED-PRWD-R
a. [(ŋəsə)]~ŋasal	***!*		
☞b. ŋə~ŋasal	**	*	*

The rankings of the constraints proposed so far for C- and disyllabic reduplication are summarised in (37).

(37) Constraint rankings for C- and disyllabic reduplication
C-reduplication:
HEADNESS, FTBIN, ANCHOR-BR, ALIGN-RED-L
ALIGN-ROOT-L, ***DEP-OO-SEG***
MAX-BR ***RED-PRWD-L /R***

Disyllabic reduplication:
HEADNESS, FTBIN, ANCHOR-BR, ALIGN-RED-L, ***RED-PRWD-L /R***
ALIGN-ROOT-L, ***DEP-OO-SEG***
MAX-BR

The other feature that distinguishes C-reduplication from disyllabic reduplication is the nature of the schwa before the base. For C-reduplication, which involves bare consonant reduplication, the schwa before the base is not part of the reduplicant but comes from inter-consonantal epenthesis (e.g., *ŋasal* > *ŋ*~*ŋasal* > *ŋə*~*ŋasal*). In section 4.1, the ranking ||DEP-OO-SEG » MAX-BR|| is proposed to predict the size of the C-reduplicant. However, the ranking will not only compress the reduplicant to its minimal size but also prevent any segment from intervening between the copied segment and the base, which is contrary to the fact that there is always a schwa standing between. Since what is copied is a bare consonant in C-reduplication, if the copied segment is adjacent to the base, complex onset would result in. As Squliq Atayal does not tolerate complex syllable margins, *COMP-M,[16] which plays a dominant role in the language, must crucially outrank DEP-OO-SEG, as illustrated in (38).

(38) C-reduplication: *COMP-M » DEP-OO-SEG
ŋə~*ŋasal* 'houses' (← *ŋasal* 'house')
/RED, ŋasal/
ŋə~*ŋasal* > *ŋ*~*ŋasal*

[16] As Squliq Atayal tolerates nucleus to branch, *COMP-NUC must rank relatively low (cf. H. Huang 2006a, 2006b).

However, though the ranking ||*COMP-M » DEP-OO-SEG|| helps to predict the existence of the schwa before the base, it does not guarantee the schwa must come from inter-consonantal insertion rather than pre-penultimate reduction. As a matter of fact, as (39) illustrates, the MAX-BR constraint, though very lowly ranked, would predict the schwa to be a copy of the base (i.e., (39b)). This is because the more segments copied, the fewer violations of MAX-BR.

(39) MAX-BR wrongly predicts the schwa to come from vowel reduction
ŋə~ŋasal 'houses' (← *ŋasal* 'house')

/RED, ŋasal/	*COMPLEX-M	DEP-OO-SEG	MAX-BR
a. ŋə~ŋa$_i$sal		**	****!
☞ b. ŋə$_i$~ŋa$_i$sal		**	***

A reduced *ə* corresponding to a full vowel in the base would violate IDENT-BR[V-PLACE] since a schwa and a full vowel (such as *a* in (39)) differ featurely in many respect, especially place of articulation. An inserted *ə*, on the other hand, is free from IDENT-BR[V-PLACE] violation since an inserted segment has no base correspondent. Since the schwa vowel in the reduplicant in C-reduplication comes from insertion, IDENT-BR[V-PLACE] must dominate MAX-BR, as illustrated in (41).

(40) IDENT-BR[V-PLACE]: Corresponding vowels in the reduplicant and the base have identical place features.

(41) C-reduplication: IDENT-BR[V-PLACE] » MAX-BR
ŋə~ŋasal 'houses' (← *ŋasal* 'house')
/RED, ŋasal /
ŋə~ŋa$_i$sal > *ŋə$_i$~ŋa$_i$sal*

On the other hand, the schwa vowels before the base in disyllabic reduplication do not come from insertion but are copied vowels that are reduced at the pre-penultimate position. But why pre-penultimate vowels need to undergo vowel reduction requires explanation. H. Huang (2006a:12) proposes that the reason why vowel reduction targets all vowels except for the last two is that the last two vowels are grouped into a foot and vowels in the foot domain are protected from vowel reduction. Since stress in Squliq Atayal generally falls on the last syllable, the foot that stands at the right edge of the word should constitute the main stress foot; namely, the head foot. McCarthy (2003) proposes two End Rule constraints to predict the location of the head foot. The two constraints ER-L and ER-R are categorical alignment constraints that respectively require the head foot to stand at the left and right edge of the prosodic word. In Squliq Atayal, since the head foot is aligned to the right edge of the word, the ER-R constraint is crucial. Thus, the footing as proposed in H. Huang can be easily predicted by the constraint ranking ||ER-R, FTBIN » PARSESYLL||, as illustrated in (44). The head foot is marked in bold here below.

(42) PARSESYLL: Syllables must parse into foot.

(43) ER-R: The head foot is not followed by another syllable within the prosodic word. (McCarthy 2003)

(44) ||ER-R, FTBIN » PARSESYLL|| predicts the head foot to be right aligned
/paqut-un/ → *pəqutun* 'ask (PF)'

/paqut-un/	ER-R	FTBIN	PARSESYLL
☞a. [pə**(qutun)**]			*
b. [**(paqutun)**]		*!	
c. [**(paqu)**tən]	*!		*

As to explain why vowels standing at the head foot do not undergo vowel reduction, the present paper adopts the HEAD(PCat)-IDENT(F) constraint proposed in Alderete (1995), which is a faithfulness constraint that regulates faithfulness between the segment in an output prosodic head (such as the main stress foot or the head syllable of the prosodic word) and its input counterpart.

(45) HEAD(PCat)-IDENT(F) (Alderete 1995)
Correspondent segments in prosodic heads PCat agree in value for feature [F].
If PCat is a prosodic head, PCat contains β, and αℜβ, then α and β agree in the value of F.

Thus, the lack of vowel reduction in the main stress foot can be explained by the HEAD(FOOT)-IDENT(F)-IO constraint, as illustrated in (47).

(46) HEAD(FOOT)-IDENT(F)-IO: Input-Output correspondent segments in the head foot agree in value for feature [F].

(47) Disyllabic reduplication: HEAD(FOOT)-IDENT(F)-IO
/qulih/ 'fish'
(qulih) > **(qəlih)**

Vowels outside the domain of the head foot are not protected by HEAD(FOOT)-IDENT(F)-IO. Nonetheless, they don't have to change unless to satisfy some markedness constraints. It is proposed that the change of vowel outside the head foot to schwa is due to the markedness constraint of PLACE-MARKEDNESS HIERARCHY in (48).

(48) PLACE-MARKEDNESS HIERARCHY (PL-MARKEDNESS)
(Prince & Smolensky 1993[2004], Alderete et al. 1999)
*PL/DORS, *PL/LAB » *PL/COR » *PL/PHAR

The PL-MARKEDNESS constraint is usually used to predict the place markedness of the consonants. But Clement & Hume (1995) propose that vowels and consonants bear the same place features; therefore, back vowels would violate *PL/DORS; round vowels, *PL/LAB, front vowels; *PL/COR and low vowels, *PL/PHAR. The hierarchy predicts that *ɜ* is the least marked because it is placeless. [17] When outranked by HEAD(FOOT)-IDENT(F)-IO, the PLACE-MARKEDNESS HIERARCHY can function to ensure vowels outside, but not inside, the head foot will undergo vowel reduction, as illustrated in (49).

[17] Notice that in the present paper, we limit the PLACE-MARKEDNESS HIERARCHY to vowels only to prevent consonants standing outside of the foot head from surfacing as *ʔ*.

(49) HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS
/paqut-un/ → pəqutun 'ask (PF)'
pə**(qutun)** > pə**(qətən)**

Coming back to vowel reduction in disyllabic reduplicant, though the reduplicant itself also forms a foot (cf. (34)), due to the fact it is not the head foot, the vowels in the reduplicant are not protected by HEAD(FOOT)-IDENT(F)-IO. As a consequence, ||HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS|| correctly predicts the reduction of vowels in the reduplicant, as illustrated in (50). The two newly introduced constraints are ranked the same way in C-reduplication (i.e., ||HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS||) to predict the lack of vowel reduction in the head foot of the base.[18]

(50) Disyllabic reduplication: ||HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS|| predicts vowel reduction in the reduplicant
qələ~qulih 'fish of various kinds' (< *qulih* 'fish')

/RED, qulih/	HEAD(FOOT)-IDENT(F)-IO	PL-MARKEDNESS	IDENT-BR [V-PLACE]
a. (qəᵢləⱼ)~**(qəᵢləⱼh)**	*!*		
b. (quᵢliⱼ)~**(quᵢliⱼh)**		***!*	
☞c. (qəᵢləⱼ)~**(quᵢliⱼh)**		**	**

Tableau (50) also shows that PL-MARKEDNESS must dominate IDENT-BR[V-PLACE], which requires BR identity, to ensure the reduplicant vowels can be reduced even though their corresponding vowels in the base are not.

---

[18] The tableau in (i) illustrates how the constraint ranking works to predict C-reduplication.
(i)||HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS|| makes the correct prediction for C-reduplication
ŋə~ŋasal 'houses' (← *ŋasal* 'house')

/RED, ŋasal/	HEAD(FOOT)-IDENT(F)-IO	PL-MARKEDNESS
a. ŋə~**(ŋəsəl)**	*!*	
b. ŋa~**(ŋasal)**		***!
☞c. ŋə~**(ŋasal)**		**

Notice that HEAD(FOOT)-IDENT(F)-IO merely functions to preserve identity of segments in the head foot that have input correspondents. Therefore, while it bans schwa vowels derived from reduction, it does not prohibit those that are inserted since epenthesised vowels have no input correspondents. Consequently, the constraint ranking ||HEAD(FOOT)-IDENT(F)-IO » PL-MARKEDNESS|| also makes the correct prediction for C-reduplication targeting monosyllabic root, even though the vowel standing between the copied consonant and the base (which is situated in the head foot) is a schwa. Consider (ii) below.

(ii) PL-MARKEDNESS does not prohibit epenthesised schwa from occurring in the head foot
zə~zik 'very deep' (← *zik* 'below')

/RED, zik/	HEAD(FOOT)-IDENT(F)-IO	PL-MARKEDNESS
a. **(zi~zik)**		**!
☞b. **(zə~zik)**		*

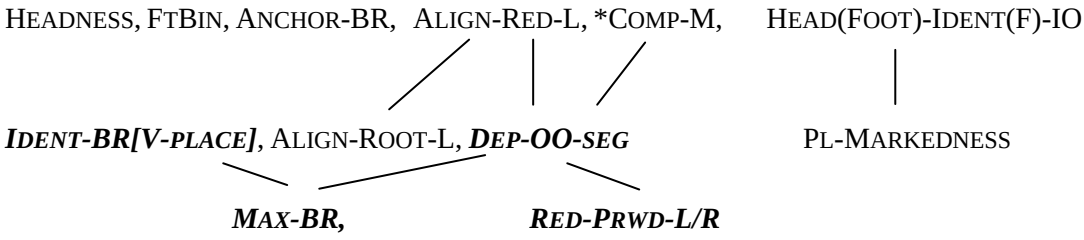
Finally, to ensure the vowels in the reduplicant in disyllabic reduplication are reduced rather than inserted, IDENT-BR[V-PLACE], which prefers the schwa in the reduplicant to come from insertion, must be outranked by MAX-BR, which prefers the schwa to come from reduction. The ranking ||MAX-BR » IDENT-BR[V-PLACE]|| is the reverse of that proposed for C-reduplication, whose *ə* before base comes from insertion (cf. 41).

(51) Disyllabic reduplication: ||MAX-BR » IDENT-BR[V-PLACE]|| predicts the schwa vowels to come from reduction
qələ~qulih 'fish of various kinds' (← *qulih* 'fish')

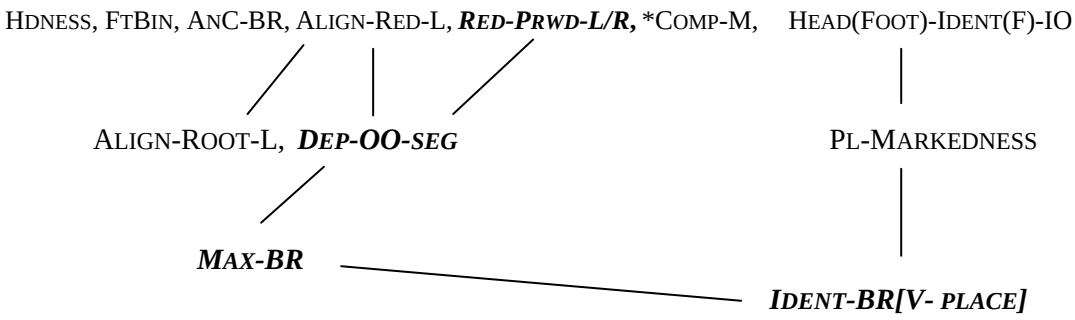
/RED, qulih/	MAX-BR	IDENT-BR[V-PLACE]
a. qələ~qu$_i$li$_j$h	**!*	
☞b. qə$_i$lə$_j$~qu$_i$li$_j$h	*	**

The final constraint rankings for C-reduplication and disyllabic reduplication are summarised below:

(52) The final constraint rankings for C-reduplication and disyllabic reduplication
C-reduplication:



Disyllabic reduplication:



5 Concluding remarks

In this paper, we have re-examined the C-type and the CC-type of reduplication in Squliq Atayal. Previous studies simply assume the two reduplication patterns involve bare consonant reduplication. Based on examples of the C-type of reduplication that applies on monosyllabic root, the paper argues that the C-type of reduplication truly involves bare consonant reduplication (i.e., C-reduplication) because the vowel before the monosyllabic root (and after the copied consonant) surfaces as schwa even when it is in the penultimate syllable. On the other hand, the paper argues that the CC-type of

reduplication involves copying of two syllables (i.e., disyllabic reduplication) because the two consonants copied do not come from the peripheral edges of the base as true bare consonant reduplication in other languages always does. The forms of C-reduplication and disyllabic reduplication differ in two respects: the size of the reduplicant and the nature of the schwa vowel before the base. The differences are shown to be nicely predicted by the slight re-ranking of certain universal constraints. The first difference is captured by the re-ranking of two constraints, DEP-OO-SEG and RED-PRWD-L/R; ||DEP-OO-SEG » RED-PRWD-L/R|| predicts C-reduplication to copy a bare consonant and the reverse ranking predicts the reduplicant in disyllabic reduplication to be disyllabic. The second feature that differentiates C- and disyllabic reduplication, on the other hand, is captured by the re-ranking of two BR correspondence constraints, MAX-BR and IDENT-BR[V-PLACE]; ||IDENT-BR[V-PLACE] » MAX-BR|| predicts the schwa vowel before the base in C-reduplication to come from insertion while the reverse order predicts that before the base in disyllabic reduplication to come from reduction.

References

Alderete, John. 1995. Faithfulness to prosodic heads. Paper presented at Tilburg Conference on The Derivational Residue in Phonology. Massachusetts: University of Massachusetts Amherst. (ROA #94)

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy and Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30.3:327–364.

Benua, Laura. 1997. Transderivational identity: Phonological relations between words. PhD dissertation. Massachusetts: University of Massachusetts Amherst.

Blust, Robert. 1998. *Ca*-reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37.1:29–64.

Blust, Robert. 2009. *The Austronesian languages*. Pacific Linguistics 602. Canberra: The Australian National University.

Broselow, Ellen, and John McCarthy. 1983. A theory of infixing reduplication. *The Linguistic Review* 3:25–98.

Clements, G. N., and Elizabeth Hume. 1995. The internal organization of speech sounds. *The handbook of phonological theory*, ed. by John Goldsmith, 245–307. Oxford: Blackwell.

Crowhurst, Megan. 2004. Mora alignment. *Natural Language and Linguistic Theory* 22:127–177.

Diffloth, Gérard. 1976. Expressives in Semai. *Austroasiatic studies I*: 249–264. Oceanic Linguistics, Special Publication no. 13. Honolulu: University of Hawai‘i Press.

Egerod, Søren. 1965. Verb inflextion in Atayal. *Lingua* 15:251–282.

Egerod, Søren. 1966. A statement on Atayal phonology. *Artibus Asiae Supplementum* XXIII (Essays offered to G. H. Luce by his colleagues and friends in honour of his 75th birthday) 1:120–130.

Egerod, Søren. 1980. *Atayal-English dictionary*. Scandinavian Institute of Asian Studies Monograph Series No. 35. London and Malmö: Curzon Press.

Egerod, Søren. 1999. *Atayal–English dictionary*. 2nd ed. by Jens Østergaard Petersen. Historisk–filosofiske Skrifter 20. Copenhagen: The Royal Danish Academy of Sciences and Letters.

Gafos, Diamandis. 1998. A-templatic reduplication. *Linguistic Inquiry* 29.3:515–527.

Gouskova, Maria. 2003. *Deriving economy: Syncope in optimality theory*. PhD dissertation. Massachusetts: University of Massachusetts Amherst.

Gouskova, Maria. 2004. Minimal reduplication as a paradigm uniformity effect. *The proceedings of the 23rd West Coast Conference on Formal Linguistics*, ed. by Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez, and Benjamin Schmeiser, 265–278. Somerville, MA: Cascadilla Press.

Hendricks, Sean Q. 1999. Reduplication without template constraints: A study in bare-consonant reduplication. PhD dissertation. Tucson: University of Arizona.

Hendricks, Sean Q. 2001. Bare-consonant reduplication without prosodic templates: Expressive reduplication in Semai. *Journal of East Asian Linguistics* 10.4:287–306.

Huang, Hui-chuan J. 2006a. Resolving vowel clusters: A comparison of Isbukun Bunun and Squliq Atayal. *Language and Linguistics* 7.1:1–26.

Huang, Hui-chuan J. 2006b. Squliq Atayal syllable onset: Simple or complex? *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang and Dah-an Ho, 489–505. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

Huang, Lillian M. 1995. The syntactic structure of Wulai and Mayrinax Atayal: A comparison. *Bulletin of National Taiwan Normal University* 40:261–294.

Huang, Lillian M. 2000. *taiyayu cankao yufa* [*A reference grammar of Atayal*]. Formosan Language Series 1. Taipei: Yuan-Liou. [In Chinese]

Kennedy, Robert. 2008. Evidence for morphoprosodic alignment in reduplication. *Linguistic Inquiry* 39.4:589–614.

Lambert, Wendy Mae. 1999. Epenthesis, metathesis, and vowel-glide alternation: Prosodic reflexes in Mabalay Atayal. MA thesis. Hsinchu: National Tsing Hua University.

Lee, Amy Pei-jung. 2007. A typological study of reduplication in Formosan languages. PhD dissertation. Colchester: University of Essex.

Li, Paul Jen-kuei. 1980. The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 51.2:349–405.

Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52.2:235–301.

Li, Paul Jen-kuei. 1985. Linguistic criteria for classifying the Atayalic dialect groups. *Bulletin of the Institute of History and Philology, Academia Sinica* 56.4:699–718.

Liao, Hsiu-chuan. 2011. Some morphosyntactic differences between Formosan and Philippine languages. *Language and Linguistics* 12.4:845–876.

Lin, Hui-shan. 2010. Disyllabic verbal reduplication in Pazih─leftward or rightward? *Language and Linguistics* 11.4:679–733.

Lin, Hui-shan. 2012. Variations in Kavalan reduplication. *Language and Linguistics* 13.6:1051–1093.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih Dictionary.* Language and Linguistics Monograph Series A-2. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Lin, Wan-Ying. 2004. Vowel epenthesis and reduplication in Squliq and C'uli' Atayal dialects. MA thesis. Hsinchu: National Tsing Hua University.

McCarthy, John. 2003. OT constraints are categorical. *Phonology* 20.1:75–138.

McCarthy, John and Alan Prince. 1993. Prosodic morphology I: Constraint interaction and satisfaction. Ms., University of Massachusetts Amherst, and Rutgers University, New Brunswick, N.J.

McCarthy, John and Alan Prince. 1995a. Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249–384. University of Massachusetts Occasional Papers in Linguistics 18. Amherst, MA: Graduate Linguistics Student Association.

McCarthy, John and Alan Prince. 1995b. Prosodic morphology. *The handbook of phonological theory*, In John A. Goldsmith (ed.), 318–366. Blackwell handbooks in Linguistics 1. Cambridge, MA: Blackwell.

Prince, Alan and Paul Smolensky. 1993[2004]. *Optimality Theory: Constraint interaction in generative grammar.* Malden, MA: Blackwell.

Rau, Der-Hwa Victoria. 1989. Language contact and language choice in the Philippines. Paper presented at the SEASSI conference 28–30 July 1989. Honolulu: University of Hawai‘i at Mānoa.

Rau, Der-Hwa Victoria. 1992. *A grammar of Atayal*. Taipei: Crane Publishing Co., Ltd.

Urbanczyk, Suzanne. 2006. Reduplicative form and the root-affix asymmetry. *Natural Language and Linguistic Theory* 24.1:179–240.

Van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Publications in Language Sciences 27. Dordrecht: Foris Publications.

Yeh, Marie M. 2009. C*a*-reduplication in Formosan Languages. *Grazer Linguistische Studien* 71:135–156.

Zeitoun, Elizabeth. 2002. Reciprocals in Formosan languages. Paper presented at the 9th International Conference on Austronesian Linguistics, 8–11 January 2002. Canberra: The Australian National University.

Zeitoun, Elizabeth. 2012. Towards the reconstruction of reciprocal prefixes in PAn based on a comparative study of Formosan languages. Paper read at A Typological study of Austronesian languages in Taiwan and their revitalization, 1–2 December 2012. Taipei: Academia Sinica.

Zeitoun, Elizabeth and Chen-huei Wu. 2006. An Overview of reduplication in Formosan languages. *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 97–142. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

5 *Wordhood and event integration in Tsou serial verb constructions*[*]

GUJING LIN

1 Introduction

Tsou is an Austronesian language currently spoken in southwestern Taiwan. It embraces a variety of verb sequencing structures, most of which are characterised by a contiguous series of verbs with concordant marking of voice contrast, as shown in the AV-AV marking in (1). Apart from this concordant serialisation, the language also has three less studied verb sequences that exhibit non-concordant voice marking to specify event participants of instruments, locations, and associatives. One of these non-concordant structures with PV-AV marking is illustrated in (2).[1]

(1) *mi=’o* ***m-ici*** ***bonu*** *to* *tacumu.*[2]
AV.REAL=1SG AV-want eat.AV NTOP banana
‘I want to eat bananas.’ (concordant series; H. Chang 2005:7)

(2) *i=’o* ***haf-a*** ***uh*** *to* *taipahu* *’o* *naau.*
NAV.REAL=1SG take-PV go.AV NTOP Taipei TOP Naau
‘I took Naau to Taipei.’ (non-concordant series)

At first glance, both structures manifest formal properties comparable to contiguous serial verb constructions in other languages: The verb sequence forms a single clause

* This paper has benefited tremendously from the comments and suggestions of Matt Shibatani, Linda Lanz, three anonymous reviewers, and the editors Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu. Sincere thanks go to them, and also to the Tsou elders for sharing their knowledge of the language with me.

1 This paper follows the Leipzig Glossing Rules in glossing examples, with the following additions: AV, actor voice; COS, change of state; EVID, evidential; HAB, habitual; LV, locative voice; NAV, non-actor voice; NTOP, non-topic nominal; PV, patient voice; REAL, realis; RED, reduplication; RV, referential voice; SUB, subordinator.

2 This paper is based on data from the author’s fieldwork and the narratives recorded by Tung (1964) and Szakos (1994). Data from the narratives are referenced first by story number and then by the order in the text. For instance, the reference ‘Tung I-29:005’ indicates the fifth sentence (by his segmentation) in the twenty-ninth story of section 1 in Tung’s book, and ‘Szakos 34:37’ indicates the thirty-seventh sentence in the thirty-fourth story in Szakos’s PhD dissertation.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 101–129
Asia-Pacific Linguistics, 2015

with only one value of modality and pronominal markings (indicated on the initial auxiliaries *mi*, in (1), and *i*, in (2)), and within the sequence there is no overt marker of coordination, subordination, or syntactic dependency of any other sort. While comparable formal properties lend support to subsuming both sequences under the label serial verb construction (SVC), the surface similarities of contiguous verb series hide deeper differences in structural compression and event integration. It is argued that a concordant SVC consists of multiple words and shows a weaker structural binding than a non-concordant SVC, whose components together constitute one cohesive word in the serial context. This difference in structural compression further corresponds with the way event information is bundled together in the joint predicate. A one-word, non-concordant SVC expresses more closely-knit event information than a multi-word, concordant SVC, even though the pattern of voice marking appears to suggest otherwise.

The paper is structured as follows: Section 2 presents a brief overview on the structural and semantic diversity embraced by SVCs, focusing on the variation of wordhood and its relevance to the semantic integration of serialised verbs. Section 3 provides a sketch of Tsou clause structure as well as the wordhood and functional scope of simplex predicates, setting the stage for the investigation of SVCs later. In section 4 I explore the formal properties of both concordant and non-concordant SVCs, paying attention to verb types, voice marking, specification of modality and pronominal markings, and contiguity of component verbs. Section 5 examines the wordhood of both concordant and non-concordant SVCs and explores the event relations that enter each type of serial structure. Section 6 discusses the correlation between semantic integration and structural compression (one-word vs multi-word). Section 7 concludes the present study.

2 Continuum of structural compression and semantic integration

A serial verb construction is often described as a construction in which two or more verbs, due to event integration, function together as a single predicate without any marker of non-finiteness (cf. Foley & Olson 1985; Bisang 1995; Durie 1997; Crowley 2002; Bril 2004; Aikhenvald 2006a, to name a few).[3] Any analysis of SVCs therefore naturally touches on the issue of predicatehood and, in particular, the way multiple predicational elements jointly constitute the function of a single predicate. Even though a predicate is typically understood as a verb phrase representing what is said of the subject NP (Trask 1993; Matthews 1997), in most syntactic theories the head of a predicate is normally an X^0-level category and a word. The correspondence between a multi-headed form and a mono-headed function yields a variety of variation along which predicatehood and wordhood may interrelate. An often-repeated observation of complex predication, under which SVCs are often subsumed together with other multi-headed structures such as causative predicates and light verbs, states that the composition of multiple heads may

[3] Aikhenvald (2006a:1) notes: "A serial verb construction is a sequence of verbs which act together as a single predicate, without any overt marker of coordination, subordination, or syntactic dependency of any other sort. Serial verbs describe what can be conceptualised as a single event. They are monoclausal; their intonational properties are those of a monoverbal clause, and they have just one tense, aspect, and polarity value. Serial verbs may also share arguments and obliques. Each component of an SVC must be able to occur on its own."

appear in one-word and multi-word structures as well:[4] SVCs are realised in one-word structures in Yimas (Foley 1991, 1997), Olutec (Zavala 2006), and Toqabaqita (Lichtenberk 2006), to name just a few, but in multi-word sequences in Mwotlap (François 2006), Dumo (Ingram 2006), and Watam (Foley 2010). Coexistence of both one-word and multi-word structures within a language is not unusual, as has been reported for Tariana (Aikhenvald 2006b) and Lakota (de Reuse 2006; see Aikhenvald 2006a:37–40 for a detailed overview of wordhood variation in SVCs). One thing to be noted for the discussion of wordhood is that the concept of word can be defined at multiple levels of linguistic representation, and the different senses of word may not necessarily converge, following Matthews (1991, 2002), Mohanan (1994, 1997), Dixon & Aikhenvald (2002), and Shibatani (2009). In the studies of SVCs, it has been reported in multiple languages that an SVC may meet the criteria of one morphological word but constitute two (or more) phonological words; alternatively, components inside an SVC may jointly form one phonological word but remain individual words on morphological grounds (cf. the collection of papers in Aikhenvald & Dixon (2006)). For present purposes, this study adopts morphological criteria for evaluating lexical independence, leaving the detail of phonological wordhood for further research.

Having adopted morphological criteria for assessing structural compression, let's return to wordhood variation in SVCs and its implication for interpreting the different ways component verbs are bound together. As mentioned earlier, even though wordhood does not determine the status of an SVC, wordhood variation has been associated with the way events of component verbs are bundled together in the principle of iconic motivation. There is a tendency for one-word SVCs to exhibit more of the semantic fusion found in a single, tightly-bound event denoted by a monoverb predicate, and for multi-word SVCs (especially non-contiguous ones) to allow for a series of (sub)events connected within a macro-event, albeit in loose binding (cf. Aikhenvald 2006a:50–51). The boundary between a single, tightly-bound event and a sequence of subevents linked within a macro-event is mostly hazy and gradient, but the difference in integration, as suggested by Foley (2010), can be revealed by observing the scope effects of negation markers and adverbial elements. Component verbs in the Thai multi-word SVC (Diller 2006:169), for instance, may be placed under different scopes of negation, and the separate pattern is particularly apt when the component subevents are conceptually separable.[5] Separate polarity values are also possible in multi-word SVCs in Watam (Foley 2010:86–107) when the component subevents are not implicatively linked. The verb sequence in (3) requires a shared polarity value between the final two verbs, which together designate a motion event with a specified manner 'throw' and a downward trajectory, but allows the first verb to stay outside of the negation scope. The separate polarity value, according to Foley (ibid.:101–102), is possible because the event of acquiring an egg does not entail throwing it down, but throwing an egg generally entails a downward trajectory due to gravity.

---

[4] Take causative predicates in Chichewa and Catalan for examples (Alsina 1997). A causative predicate forms a single morphological word in the former but two words in the latter, even though in both languages the causative predicate involves an identical unified argument configuration.

[5] Separate polarity values are particularly apt for verb sequences like *no':n lap*^{3} 'recline sleep' (Diller 2006:169), in which the desirable sequence may not be realised.

(3) Watam

namot	*i*	*yor*	*i*	*aŋji-r*	*ba-pika-r*	*ba-irik-tap.*
man	a	egg	a	get-REAL	NEG-throw-REAL	NEG-go.down-NEG

'A man got an egg but didn't throw it down.' (Foley 2010:102)

The correlation between structural compression (one-word vs multi-word) and event integration has implications for the assessment of the two types of Tsou verb sequencing structures, in particular regarding whether both structures should be analysed as SVCs in the face of their structural diversity and semantic differences. Such a challenge has been raised by H. Chang (2009), who considers the non-concordant verb series in Tsou an instance of SVCs but argues against a similar interpretation of the concordant series in view of the dependency between the component verbs. His argument is based on the observation that a number of initial verbs in the concordant series do not occur on their own (see (4b)). They must combine with the following verb for which they provide semantic modification, as in (4a).

(4) a.

mo	***na'no***	***nac'o***	*'o*	*paic~~u~~.*
AV.REAL	very.AV	sad.AV	TOP	Paic~~u~~

'Paic~~u~~ is very sad.' (concordant series, H. Chang 2009:445–447)

b. *

mo	***na'no***	*'o*	*paic~~u~~*	*ho*	*mi=ta*	***nac'o.***
AV.REAL	very.AV	TOP	Paic~~u~~	SUB	AV.REAL=3SG	sad.AV

intended (Lit.) 'When Paic~~u~~ is sad, the degree is intense.'

Even though concordant sequences such as *na'no nac'o* 'very sad' in (4a) thus fail to meet the conventional requirement that each component of SVCs must be able to occur on its own (cf. n. 3), the failure in no way suggests that concordant sequences are necessarily non-SVCs. To begin with, despite examples like (4a), a fair number of initial verbs in concordance sequences are able to occur independently outside the serial context, a point H. Chang (ibid.:443) also acknowledges but does not elaborate. In (5a) and (6a) below, the initial verbs *bumeemeal~~u~~* 'carefully, AV' and *tosvo* 'stop, AV' each provide semantic modification for the following verbs with which they form a concordant verb sequence; in (5b) and (6b), they appear as an independent predicate inside the clause of which they form a part.

(5) a.

os='o	*eainc-a:*	***bumeemeal~~u~~***	***pono.***
NAV.REAL=1SG	say-PV	carefully.AV	shoot.AV

'I told them: (one should) shoot carefully.' (concordant series, Szakos 34:37)

b.

a'~~u~~mt-a	*na'n-a*	*ko'sosmoeo-a*	*ho*	*i=he=cu*
really-PV	very-PV	frighten-PV	and	NAV.REAL=3PL=COS
apopovah-a	*ho*	***bumeemeal~~u~~***	*ho*	*la*	*eahioa*
act.properly-PV	and	be.careful.AV	SUB	HAB	work.AV
ho	*mo*	*m~~u~~ch~~u~~.*			
SUB	AV.REAL	rain.AV			

'(the lightning) was very frightening, so they had to act properly and be careful when it starts raining during work' (independent verb, Szakos 10:78)

(6) a.

tena	*akameos~~u~~*	***tosvo***	***mongsi.***
IRR	immediately.AV	stop.AV	cry.AV

'(he) stopped crying immediately.' (concordant series, Szakos 6:19)

b. *mi=mza=cu* ***tosvo*** *ho* *svove-i* *'e* *eamasiana.*
AV.REAL=1PL=COS stop.AV and turn.toward-LV TOP Eamasiana
'We stopped and turned toward Eamasiana (place name).' (independent verb, Szakos 33:77)

A comparison of (4), (5), and (6) illustrates that verb sequences with concordant voice marking are heterogeneous insofar as independent occurrence of component verbs is involved, but in no way does this suggest that concordant voice marking should be taken as diagnostic of a non-SVC status. Doing so would assign undue emphasis to a non-SVC interpretation (as inferred from (4)), while ignoring evidence that does not support it (as (5) and (6)). In certain instances, components inside a concordant sequence may display dependency in elicited data but are found to allow independent occurrence in conversation. A case in point is the aspectual initial verb *i'vaho* 'again, AV', which combines with the following verb in the elicited sentence in (7a) but is found to stand alone in an excerpt of a dialogue in (7b).

(7) a. *mi=ta* ***i'vaho*** ***m-imo*** *to* *emi.*
AV.REAL=3SG again.AV AV-drink NTOP wine
'He drank wine again.' (concordant sequence, H. Chang 2009:447)

b. *mi='o=cu* *el~~u~~* *ta* *te='o* *phin-i* *ci*
AV.REAL=1SG=COS find.AV NTOP IRR=1SG buy-LV REL
te='o *la* *sapie-i.*
IRR=1SG HAB wear-LV
'B: I have found what I will buy and wear.'
te='o=n'a ***i'vaho;*** *pano* *la* *sapie-i* *no*
IRR=1SG=PROG again.AV exist HAB wear-LV NTOP
o'oko *no* *la* *'o* *o'oko* *no* *t‹m›ops~~u~~*?
children NTOP HAB TOP children NTOP <AV>study
'Now I want to see whether there are shoes for school children as well.'
(Lit.) 'I will do (the asking) again: are there shoes for school children?' independent verb, Szakos 39:25–26)

The non-uniform complexity within concordant sequences, as between (4) vs (5) and (6), and the gradual variation in between, as illustrated in (7), calls into question the assumption of a clear-cut SVC/non-SVC opposition by the single criterion of independent occurrence of component verbs. Thus, we have reason not to take this criterion as a necessary and sufficient condition for 'genuine' SVCs; instead, it is viewed as a markedness principle—failure to display such a feature is instructive of a less-exemplary status with regard to prototypical SVCs, but it alone is not deterministic of a binary SVC/non-SVC opposition.

It may be tempting at this point to suggest that some concordant verb sequences can be treated as SVCs, while others cannot. However, as illustrated earlier, assuming abrupt discontinuities for empirically nuanced transitions would limit our ability to address the features of concordant sequences collectively. Even worse, if we thereby exclude concordant sequences all together from the discussion of SVCs, we bypass the properties of a monoclausal configuration, lack of overt marking of syntactic dependency, and a single specification of modality and pronominal markings, all of which are found in concordant sequences (and non-concordant ones as well, see section 4) and suggest their

similarity to SVCs in other languages. As such, the term SVC still proves useful as a convenient descriptive label for the convergence of these properties in Tsou. Besides, subsuming both concordant and non-concordant sequences under the broad label SVC further allows us to explore how verb sequences with similar properties may reflect the different ways event information is bundled together and compressed into the functional domain of a single predicate, i.e., a single clause.

3 A sketch of Tsou grammar

3.1 Clause structure

Most Tsou clauses begin with an auxiliary and a predicate, with nominals following immediately behind (Tung 1964; Szakos 1994; Zeitoun 1993, 2000, 2005; M. Chang 2004, among others). Every nominal is preceded by a particle indicating the dependency relation of the nominal to its licensing predicate. The pre-nominal particle illustrates a two-way contrast, referred to as the TOPIC[6] vs NON-TOPIC distinction in the present study. A clause may contain multiple nominals coded as the NON-TOPIC, but only one nominal can be selected for the TOPIC, as shown in examples (8)–(11) below (see also Zeitoun 2000:93–94).

(8) *mo=Ø* *mo-si* *to* *ca'hʉ* *to* *pooyoyo* *'o* *amo.*
AV.REAL=3SG AV-put NTOP chair NTOP pants TOP father
'Father put pants on a/the chair.' (actor=TOP, AV verb, AV AUX *mo*)

(9) *i=si* *si-a* *to* *ca'hʉ* *to* *amo*
NAV.REAL=3SG put-PV NTOP chair NTOP father
'o *pooyoyo.*
TOP pants
'Father put the pants on a/the chair.' (patient=TOP, PV verb, NAV AUX *i*)

(10) *i=si* *si-eni* *to* *pooyoyo* *to* *amo* *'o* *oko.*
NAV.REAL=3SG put-RV NTOP pants NTOP father TOP child
'Father put aside pants for the child.' (beneficiary=TOP, RV verb, NAV AUX *i*)

---

[6] I follow Schachter (1976, 1977), Shibatani (1988, 1991), and Dryer (1997) and associate the term 'topic' with a clausal constituent which, in Philippine-type languages, often stays semantically distinct from the actor and syntactically manifests a fair number of properties exhibited by subjects in English. The choice of such labelling over conventional case terms such as 'nominative' and 'oblique' is made on the consideration of the non-demoting nature of the voice alternation in Tsou (and in Philippine-type languages in general). In English, the passive agent is relegated to an adjunct and the nominative-marked patient manifests a preponderance of subject properties; this, however, is not the case for Tsou NAV constructions, in which subject properties are divided between the 'non-topic' actor and the 'topic' nominal. I understand that in discourse analysis, the term 'topic' often designates a constituent which is referentially prominent but syntactically may not be an obligatory clausal element (but see Givón's (1979) analysis of sentential subjects as grammaticalised discourse topic). To emphasise the distinctiveness of this use, throughout this paper I mark the term TOPIC in small capitals.

(11) *i=si* *si-i* *to* *pooyoyo* *to* *amo* *'o* *ca'hʉ.*
NAV.REAL=3SG put-LV NTOP pants NTOP father TOP chair
'Father put pants on the chair.' (location=TOP, LV verb, NAV AUX *i*)

The verbal marking system in Tsou displays the Philippine-type voice contrasts (also known in the Austronesian literature as a 'focus system'). Depending on the nominal that bears the TOPIC marking, a verb may appear in at most four different voice forms: Actor Voice, Patient Voice, Referential Voice, and Locative Voice (cf. Zeitoun 2000, 2005; H. Huang & Huang 2007, among others).[7] The examples in (8)–(11) illustrate the four voice forms of the lexical root *si* 'put': *mo-si* (AV), *si-a* (PV), *si-eni* (RV), and *si-i* (LV). The alignment of the four-way voice distinction with the TOPIC vs NON-TOPIC contrast establishes grammatical relations possible in a monoclausal configuration of Tsou (e.g., the ACTOR-TOPIC relation in (8) and the PATIENT-TOPIC relation in (9))

Verbs are not the only place where voice marking is indicated. A characteristic feature of Tsou is that voice marking is simultaneously registered on the co-occurring auxiliary in realis mood. An AV verb requires an AV auxiliary, but verbs marked in the three NON-ACTOR-VOICE (NAV henceforth) forms take a NAV auxiliary instead, as shown in (12) and (13) below. Aside from voice and mood, the Tsou auxiliary also exhibits pronominal marking, which invariably coreferences the actor of the clause, as shown in the first person singular marking *='o* in (12) and (13).

(12) *mi='o* *c‹m›uhu* *to* *moatʉ'nʉ.*
AV.REAL=1SG <AV>butcher NTOP goat
'I butchered a goat.' (AV AUX *mi*, actor=TOP)

(13) *i='o* *chu-a* *'o* *moatʉ'nʉ.*
NAV.REAL=1SG butcher-PV TOP goat
'I butchered the goat.' (NAV AUX *i*, patient=TOP)

3.2 Simplex predicates: wordhood and its functional scope

This section examines the morphological wordhood and functional scope of Tsou simplex predicates. The results provide a basis for the investigation of wordhood and event relations inside the composite predicate in both types of Tsou SVCs in section 5.

3.2.1 Morphological integrity of a word

In theories that subscribe to the lexical integrity principle (Bresnan & Mchombo 1995) and segregate a language's lexicon from its syntax, the word as the output of the lexicon is susceptible to morphological processes. However, its internal parts are invisible to syntactic processes such as coordination, nor can they be split up. The following

[7] Of the four voice categories in Tsou, the three NAV categories are invariably marked by specific suffixes, whereas the AV category may be marked by a prefix (cf. (8)), an infix (cf. (12)), or a zero affix (compare (25a) and (25b)). The presence of the AV zero affix will not be explicitly marked in this paper. Note also that Referential Voice has been referred to at various times as 'Instrumental/Benefactive Voice' (cf. Zeitoun 2005) and as 'Benefactive Voice' (cf. H. Huang & Huang 2007).

description shows that a Tsou simplex predicate meets these three criteria and exhibits the morphological integrity expected of a word.

(i) Input to derivational processes

Words can undergo further morphological processes, but a syntactically derived structure cannot (see Alsina 1997; Harris 2000, among others). In Tsou, a simplex predicate can combine with derivational affixes to create new words, as shown in (14a) by the attachment of the circumfix *ho....na* 'season' to the predicate *mu'ei* 'hot'. A simplex predicate can also undergo reduplication, as shown in (15a). A simplex predicate therefore behaves like a word with respect to morphological derivations. The grammaticality of (14a) and (15a) contrasts with the behavior of syntactically conjoined phrasal structures such as *maeo ho bonʉ* 'take and eat': (14b) shows that the phrasal structure cannot combine with the *ho...na* circumfix; (15b) illustrates that the conjoined structure as a whole cannot undergo reduplication.

(14) a. *mu'ei* 'hot' + *ho...na* 'season' > *ho-mu'ei-na* 'summer' (Tung 1964:474)
b. *mu'ei* 'hot' + *vovoezʉ* 'dry' > * *ho-mu'ei-ho-vovoezʉ-na*
intended 'hot-and-dry season'

(15) a. *maeo* 'take', *ma~maeo* 'take many things' (Tung 1964:170)
b. *maeo ho bonʉ* 'take and eat', * *ma~*[*maeo ho bonʉ*]
intended 'take and eat many things'

(ii) Coordination

It is generally assumed that only two morphologically independent words can be conjoined, as in (16a); parts of words cannot, as in (16b) (see Alsina 1997; Harris 2000; Shibatani 2009, among others).

(16) a. Elise outran and outwitted Russell. (Harris 2000:603)
b. * Elise out-ran and -witted Russell.

In Tsou, simplex predicates can be conjoined using the coordination marker *ho*, as shown in (17a). Internal parts of a simplex predicate, such as **papas* and **an* in (17b), cannot be a conjunct in a coordinate structure. The ungrammaticality of (17b) is a consequence of the fact that a simplex predicate behaves as a morphologically independent word.

(17) a. *i=si=cu* *papas-a* *ho* *an-a* *to*
NAV.REAL=3SG=COS chop-PV and eat-PV NTOP
naau 'o fou.
Naau TOP meat
'Naau chopped and ate the meat.'

b. * *i=si=cu* [*papas* *ho* *an*]-*a* *to*
NAV.REAL=3SG=COS chop and eat-PV NTOP
naau 'o fou.
Naau TOP meat
intended 'Naau chopped and ate the meat.'

(iii) Non-interruptability

The word as a cohesive unit entails that its internal parts cannot be split up by material whose occurrence implies word edges (see also Alsina 1997; Booij 2009;

Mohanan 1997; Harris 2000).[8] In Tsou, the negation particle *o'te* and the evidential marker *nana* (for information based on hearsay evidence) typically precede the verb they modify, as in (18a) and (18b). Examples (19a) and (19b) show that these two elements cannot occur inside and thus split up a simplex predicate *papas-a* 'cut'. The ungrammaticality shows that a simplex predicate comprises an inseparable word on morphological grounds.

(18) a. *hoci o'te fozu na chumu...*
if NEG accumulate TOP water
'If water is not accumulated,…' (Tung I-2:12)

b. *mi=cu nana nocmuu ho eainc-a.*
AV.REAL=COS EVID come.near.AV and say-PV
'He reportedly came near and said.' (Tung I-45:029)

(19) a. * *i=si papas-o'te-a to naau 'o fou.*
NAV.REAL=3SG chop-NEG-PV NTOP Naau TOP meat
intended 'Naau did not chop the meat.'

b. * *i=si papas-nana-a to naau 'o fou.*
NAV.REAL=3SG chop-EVID-PV NTOP Naau TOP meat
intended 'Naau reportedly chopped the meat.'

3.2.2 A monoclausal configuration and a single event

A tacitly accepted assumption underlying most syntactic theories is that a predicate names the event described by a clause of which it forms a part (Foley & Olson 1985; Bresnan 2001; Levin & Rappaport Hovav 2005); the functional domain of a predicate is therefore monoclausal and encodes a single event. Inside a monoclausal structure of Tsou, grammatical relations and argument roles are not to be duplicated; there is only one TOPIC relation, one actor, etc. In (20), the monoclausal configuration of the simplex predicate 'butcher' is made evident by the single specification of pronominal marking on the auxiliary, i.e., the bound pronoun *=si*, which cross-references the actor of the clause. In contrast, a bi-clausal construction as illustrated in (21) allows two pronominal clitics, one for the matrix clause (the third person singular *=si*) and the other for the lower clause (the third person plural *=hin'i*).

(20) *i=si chu-a to ino 'o tyoua.*
NAV.REAL=3SG butcher-PV NTOP mother TOP chicken
'Mom (already) butchered the chicken.'

(21) *i=si*$_i$ *cucufn-i to ino*$_i$ *'e 'o'oko*$_j$
NAV.REAL=3SG urge-LV NTOP mother TOP children
ho te=hin'i$_j$ *bon~~u~~ to f'ue.*
SUB IRR=3PL eat.AV NTOP sweet.potato
'The mother urged the children to have sweet potatoes.'

---

[8] The non-interruptability test (based on the placement of negation particles and adverbial elements) is suggested as a diagnostic for wordhood status in Harris (2000:601), and Alsina (1997:225–227). The fundamental principle of this test, as noted in Alsina (ibid.:225), states that materials that occur outside of the word's domain cannot appear inside the word.

The monoclausal configuration of a Tsou predicate is also coupled with a single specification of modality marking and temporal reference, which assigns a temporal/modal grounding for the event depicted by the predicate, as seen in (20) for the use of the realis auxiliary *i* and in (22a) for the temporal reference 'in the past'. Assigning multiple, conflicting temporal references to a single predicate is not allowed, as illustrated in (22b) for the joint use of 'in the past' and 'today'. Therefore, if a target structure contains a single specification of modality marking and/or temporal reference, it follows that the structure encodes a single event.

(22) a. *ine* *noanao* *moso* *nana* *la* *mav'ov'o*
NTOP past AV.REAL EVID HAB various.AV
na *moso* *la* *tmucou.*
TOP AV.REAL HAB transform.to.human.AV
'In the past, there were various kinds of shape shifters.' (Tung I-26:001)

b. * *ine* *noanao* *moso* *nana* *la* *mav'ov'o*
NTOP past AV.REAL EVID HAB various.AV
na *moso* *la* *tmucou* *maitan'e.*
TOP AV.REAL HAB transform.to.human.AV today
intended 'In the past, there were various kinds of shape shifters today.'

4 Formal properties of Tsou SVCs

This study adopts a working definition of Tsou SVCs as a sequence of two (or more) verbs which act together within a monoclausal configuration and share a single specification of modality and pronominal markings. Components within the sequence each display the regular voice marking seen in autonomous verbs, and as such are expected to stand alone outside the serial context, even though the failure to display independent occurrence is not itself diagnostic of non-SVC status considering the fair number of gradual variations seen in (5)–(7). Following this definition, two types of SVCs can be distinguished in Tsou: concordant and non-concordant. In the former, the serialised verbs agree in voice marking (e.g., AV-AV, see (25a) below). In the latter, however, no voice concord can be established.

Description of these two types of SVCs is organised according to the four properties: (i) verb types that enter SVCs, (ii) pattern of voice marking, (iii) specification of modality and pronominal markings, and (iv) contiguity of constituent verbs. Of the four properties, (i) and (ii) are relevant to the exploration of event relations inside SVCs, and (iii) and (iv) are related to the assessment of monoclausality and wordhood. They will be further incorporated in the discussion of structural compression and semantic interpretation in section 5.

4.1 Concordant SVCs

(i) Verb types

Inside a concordant SVC, verbs come in an ordered sequence depending on their semantics. While the verb sequence can consist of up to four verbs (H. Chang 2009; Yeh & Huang 2009), the last verb in the sequence always specifies participants involved in the depicted event and their relation to the event (in terms of 'who does what to whom').

All the preceding verbs modify the depicted event by providing evaluative attributes including participants' attitude, epistemic inference, deontic reasoning, manner, frequency, degree, and aspectual phases (see H. Chang 2009; Yeh & Huang 2009; Lin 2010, among others).[9] Examples (23)–(24) illustrate the use of the attitude verb 'love', the manner verb 'together', and the frequency verb 'twice'.[10] For ease of reference, the modifying verb, even if there is more than one in a sentence, is labelled 'first verb/V1' and the participant-denoting verb is labelled 'second verb/V2' hereafter.

(23) *os='o* ***kaeb-a*** ***an-a*** *'o* *fou.*
NAV.REAL=1SG like-PV eat-PV TOP meat
'I like to eat the meat.' (participants' attitude)

(24) *tena=c'u* *boevovei* *to* *taico* *to* *eoeasva* *ho*
IRR=COS retreat.AV NTOP middle NTOP yard and
toehungu ***eipopsoh~~u~~*** ***paehai****.*
together.AV twice.AV shout.AV
'(They) retreat to the middle of the yard and together shout twice.' (Tung I-31:013, manner and frequency)

(ii) Concordant voice marking

In a concordant SVC, the constituent verbs each receive voice marking, just like a syntactically autonomous verb in an independent clause, as shown in (25a). H. Chang (2005, 2009:465–468) brings to our attention that constituent verbs in the concordant serial context maintain voice concord along the AV vs NAV contrast, not the four-way distinction typically seen in a syntactically autonomous predicate (i.e., AV/PV/RV/LV, see L. Huang 1997 for voice concord in other Formosan languages).[11] An AV-marked V1 is followed by an AV-marked V2, but a PV-marked V1 may be followed by a PV-marked V2, as in (25b), an RV-marked V2, as in (25c), or an LV-marked V2, as in (25d).

(25) a. *moh=cu* ***ason~~u~~*** ***e~~u~~mo~~u~~m~~u~~****.*
AV.REAL=COS probably.AV enter.AV
'Probably (they) entered (the house).' (Tung I-24:040, AV-AV)

b. *o=he=cu* ***ason-a*** ***opcoz-a*** *homio.*
NAV.REAL=3PL=COS probably-PV kill-PV at.the.time
'They probably killed (him) at that time.' (Tung I-44:030, PV-PV)

---

[9] H. Chang (2009:473) points out that modifying verbs (his 'adverbial verbs') are rigidly ordered and structured in Tsou. Note that in his analysis, modifying verbs are analysed as functional heads that take a participant-denoting verb (his 'lexical verb') as the complement.

[10] Tsou does not serialise motion sequences such as 'go' and 'see', even though motion verbs are among the types of expressions often expressed in SVCs on a cross-linguistic scale. Verb sequences such as **uh b-aito* 'go$_{AV}$ see$_{AV}$' and **us-a b-aito* 'go$_{PV}$ see$_{AV}$' are deemed ungrammatical by consultants. The absence of 'go-see' verb sequences suggests that SVCs in one language does not necessarily translate into SVCs in another language. For more details, see Lin (2010:377, 411–417).

[11] Both H. Chang (2009) and Lin (2010) ascribe the two-way voice concord to the restricted voice marking of V1, which only has two voice possibilities: AV and PV. H. Chang further attributes the restriction to the nature of V1 (his 'adverbial verb') as a functional head, not a lexical verb.

c. *siho* *la=c'u* *a~~u~~l~~u~~* *s~~u~~c'~~u~~h~~u~~,* *la* ***ake-a***
when HAB=COS really.AV arrive.AV HAB a.little-PV
teoteoc-neni *to* *eono* *'o* *poeave.*
cut-RV NTOP sacred.tree TOP knife
'When (they) actually reach (the tree), (they) cut the Sacred Tree with the sword a little bit.' (Tung I-31:011; PV-RV)

d. *tena* *ahoz-a* *tutv-a* *'o* *otofnana* *ho*
IRR begin-PV beat-PV TOP poisonous.weed and
aasb~~u~~t-a ***tfu-i*** *to* *chumu.*
sometimes-PV soak-LV NTOP water
'(one) first pounds the poisoning-weed and sometimes soaks (it) in water.' (Tung I-14:003; PV-LV)

Restrictions on voice marking raise two questions regarding the status of V1 as a verb and the status of a concordant SVC as a genuine SVC. To begin with, the AV/NAV distinction of V1 resembles the reduced AV/NAV marking on realis auxiliaries (see section 3.1 for details). If the two-way voice contrast on the V1 verb is an indication that the element in question is an auxiliary, the claim that verb sequences such as 'sometimes-soak' in (25d) are SVCs is open to question. It is open to question because SVCs are typically characterised as a juxtaposition of two verbs, not an auxiliary and a verb.

However, a close inspection of the other properties of V1 indicates that it is behaviorally distinct from an auxiliary. First, a Tsou auxiliary attracts pronominal marking (such as the third person plural *=he* in (25b) and aspectual markers (such as the perfective marker *=cu* in (25b)), but the V1 verb does not. Second, the AV/NAV contrast on the auxiliary is further reduced/neutralised in irrealis mood, but such neutralisation is not detected in the V1 verb. Examples (24) and (25d) illustrate that the same irrealis auxiliary *tena* is used for both AV and PV-marked verbs, but the relevant V1 alternates between AV and PV forms even in irrealis mood (that is, the AV/PV alternation on the V1 verb agrees with the voice marking of the V2 verb). Third, unlike auxiliaries, the semantic contribution of V1 verbs is not limited to facilitating the modality interpretation of V2 verbs. Rather, it may contribute information relevant to the argument structure of the joint predicate. In (26), the V1 verb *ah~~u~~e~~u~~* 'should' expresses the speaker's attitude that the V2 event 'go over' be obligatory. This sense of obligation carries with it an implicit requirement of an actor, which is evident from the selection of V2 verbs according to their valency. Combining the deontic V1 with an actor-less seismological verb (e.g. **ah~~u~~e~~u~~* 'should' + *y~~u~~sk~~u~~* 'landslide') is considered unacceptable. The ungrammaticality makes clear the semantic contribution of V1 verbs to the argument structure of the joint predicate, supporting the view that V1 verbs are not to be treated as auxiliaries.

(26) *ho* *ta='u* *la* ***[ah~~u~~e~~u~~*** ***m-eecunu]*** *no* *emoo=su.*
SUB IRR=1SG HAB should.AV AV-go.over NTOP house=2SG
'Oh, (I wish) I should go (with you) to your house.' (Tung I-43:010)

Judging by the three features listed above, the V1 verb still displays features of a verb and is less of an auxiliary; a concordant SVC therefore should not be treated as an auxiliary-verb sequence.

(iii) A single specification of modality and pronominal markings

A concordant SVC contains only one specification of modality and pronominal marking. The realis NAV auxiliary *o* in (27a) indicates that the two serial verbs both occur in realis mood; the third person plural invisible *=he* encodes the actor shared by both verbs. It is impossible for the individual verbs in the series to each take an independent auxiliary and an independent pronominal clitic, as shown in (27b). Considering that only one actor nominal and one modality specification are allowed, a concordant SVC arguably comprises a single clause (cf. section 3.2.2).

(27) a. *o=he* *nana* ***a~~u~~mt-a*** ***opcoz-a*** *na* *nia* *ngohoo.*
NAV.REAL=3PL EVID indeed-PV kill-PV TOP late Ngohoo
'They indeed killed the late Ngohoo.' (Tung I-8:007)

b. * *o=he* *a~~u~~mt-a* *o=he* *opcoz-a* *na* *nia*
NAV.REAL=3PL indeed-PV NAV.REAL=3PL kill-PV TOP late
ngohoo.
Ngohoo
intended 'They indeed killed the late Ngohoo.'

(iv) Contiguity of components

At first glance, a concordant SVC appears to form a contiguous sequence insofar as the constituent verbs cannot be separated by the conjunction marker *ho*, as in (28).

(28) a. *mi='o* ***aothom~~u~~*** ***m-aavo*** *to* *phingi.*
AV.REAL=1SG try.AV AV-open NTOP door
'I tried to open a door.'

b. * *mi='o* ***aothom~~u~~*** *ho* ***m-aavo*** *to* *phingi.*
AV.REAL=1SG try.AV and AV-open NTOP door
intended 'I tried to open a door.'

The contiguity, however, is only superficial considering that the negation marker *o'te* 'not', the adverbial particle *c'o* 'only', and the hearsay marker *nana* can intervene between the constituent verbs, as shown in (29), (30), and (31), respectively. Insertion of these particles indicates that V1 and V2 can be separated and put under different modification scopes. The separability of component verbs has consequences for assessing wordhood of the concordant SVC, as will be illustrated in section 5.1.

(29) a. *mi='o=cu* ***ahoi*** *o'te* ***bon~~u~~*** *ta* *fou.*
AV.REAL=1SG=COS start.AV NEG eat.AV NTOP meat
'I have started not eating meat.' (H. Chang 2005:15)

b. *os='o=cu* ***ahoz-a*** *o'te* ***an-a*** *'o* *fou.*
NAV.REAL=1SG=COS start-PV NEG eat-PV TOP meat
'I have started not eating the meat.' (H. Chang 2005:15)

(30) *i='o* ***uci-a*** *c'o* ***an-a*** *'o* *tac~~u~~m~~u~~.*
NAV.REAL=1SG want-PV only eat-PV TOP banana
'I wanted to just eat the bananas.'

(31) *o=si=cu* *nana* *eainc-a:* ***su-a***
NAV.REAL=3SG=COS EVID say-PV dream-PV
s'o *nana* ***ait-i*** *na* *mo* *cihi* *ci* *hahocng~~u~~.*
just EVID see-LV TOP AV.REAL one.AV REL man
'She said she dreamed that she saw a man.' (Szakos 40:4)

4.2 Non-concordant SVCs

Like concordant SVCs, non-concordant SVCs allow a single specification of modality and pronominal marking and arguably comprise a monoclausal configuration. However, they differ from their concordant counterparts in verb types, pattern of voice marking, and contiguity of component verbs.

(i) Verb types
Three types of non-concordant SVCs can be reliably identified in Tsou. Depending on the nature of the first verb in the series, the three types of SVCs are labelled 'Instrumental SVC', 'Locational SVC', and 'Associative SVC'. Examples (32)–(34) below illustrate the three types of SVCs, respectively.

(32) *i='o* ***tith-eni*** ***m-apaso*** *to* *fou* *'o* *poyave.*
NAV.REAL=1SG use-RV AV-chop NTOP meat TOP knife
'I used the knife to chop meat.' (Instrumental SVC)

(33) *i='o* ***yon-i*** ***m-apaso*** *to* *fou* *'o* *oyonapei'i.*
NAV.REAL=1SG stay-LV AV-chop NTOP meat TOP kitchen
'I stayed in the kitchen chopping meat.' (Locational SVC)

(34) *i='o* ***haf-a*** ***uh*** *to* *taipahu* *'o* *naau.*
NAV.REAL=1SG take-PV go.AV NTOP Taipei TOP Naau
'I took Naau to Taipei.' (Associative[12] SVC)

In non-concordant SVCs, the semantics of the second verb is relatively unrestricted, except for meteorological and seismological verbs such as *y~~u~~sk~~u~~* 'landslide', as shown in (35a). This restriction is related to the pattern of argument sharing in non-concordant SVCs, which requires individual verbs to share the same actor. For example, the participant who used the knife in (32) is understood to be coreferential with the one who chopped meat in the serial context. However, due to the inherent actor-less nature of

[12] The labels 'instrumental', 'locational', and 'associative' are used here in a mnemonic sense to indicate the semantic contribution of V1 to the joint predicate. They are not intended to directly reflect the voice marking of V1 (although in the first two cases they would appear so). The case in point concerns the use of 'associative' for the 'take-go' verb series as in (34), in which the PV marking on the V1 verb is typically aligned with an affected patient participant in the Formosan literature. However, what the 'take-go' series introduces to the whole construction is not a patient affected by the motion 'go', as is clear from the free translation. While the use of 'associative' for a PV-marked SVC may at first appear puzzling, this is understandable if we accept the view that SVCs may impose non-compositional, idiomatic attributes to component verbs (Durie 1997:322–324; Aikhenvald 2006a:21–37).

meteorological and seismological verbs,[13] as shown in (35b), the requirement of a shared actor does not obtain in the putative verb sequence 'stay-landslide', motivating the non-occurrence of this verb sequence.

(35) a. * ***yon-i*** ***yʉskʉ*** *'o* *kalen.*
stay-LV landslide.AV TOP Hualien
intended 'There is a landslide in Hualien.' (Locational SVC)

b. *te* *yʉskʉ* *hohucma*?
IRR landslide.AV tomorrow
'Will there be a landslide tomorrow?' (Lin 2010:170)

(ii) Non-concordant voice marking

A non-concordant SVC is composed of verbs of non-agreeing voice forms. The second verb in the verb series always occurs in the AV form,[14] whereas the first verb always occurs in the appropriate NAV form depending on the semantics of the construction. An associative SVC involves the PV-marked first verb *haf-a* 'take' serialised to an AV-marked second verb, as shown in the 'take-go' series in (36). A locational SVC requires the AV-marked first verb *yon-i* 'stay' serialised to an AV-marked second verb, as in (37). An instrumental SVC involves the RV-marked first verb *tith-eni* 'use' serialised to an AV-marked second verb, as in (38).[15] Other logically possible combinations, such as LV-LV, are not allowed, as in (39).

(36) *o=si=cu* *nana* ***haf-a*** ***uh*** *tan'e* *na* *oko=si.*
NAV.REAL=3SG=COS EVID take-PV go.AV here TOP child=3 SG
'It reportedly took its child to this place.' (Tung I-36:004, Associative SVC)

(37) *os='o* ***yon-i*** ***tufku*** *to* *yʉsʉ* *'o* *coca.*
NAV.REAL=1SG stay-LV wash.AV NTOP clothes TOP yard
'I stayed in the yard washing clothes.' (Locational SVC)

(38) *os='o* ***tith-eni*** ***bonʉ*** *to* *fou* *'o* *poyave.*
NAV.REAL=1SG use-RV eat.AV NTOP meat TOP knife
'I used the knife to eat meat.' (Instrumental SVC)

(39) * *i='o* ***yon-i*** ***tufku-i*** *to* *yʉsʉ* *'o* *coca.*
NAV.REAL=1SG stay-LV wash-LV NTOP clothes TOP yard
intended 'I stayed in the yard washing clothes.'

The lack of voice concord may appear to suggest that the component verbs each retain their lexical and syntactic autonomy and a non-concordant SVC comprises a multi-word structure. This autonomy interpretation, however, fails to account for the voice pattern on

---

[13] In Tsou, meteorological and seismological verbs do not subcategorise any obligatory argument, although locational or temporal adjuncts may be included in the clause structure. Readers are referred to Lin (2010:169–179) for more details.

[14] The invariant AV marking of V2 is a widely attested feature among verb sequencing structures in other Formosan languages, regardless of the voice marking of V1 (cf. L. Huang 1997 and H. Chang 2006). I thank Elizabeth Zeitoun for bringing this to my attention.

[15] The first verb in the Instrumental SVC may appear in the PV form *tith-a* 'use'. Consultants report that the PV form and the RV form (*tith-eni*) can be used interchangeably.

the co-occurring auxiliary, which should co-register the voice marking of the clausal verb (cf. section 3.1). I will turn to this point immediately below.

(iii) A single specification of modality and pronominal marking

A non-concordant SVC contains a single specification of modality and person information, thereby manifesting features comparable to those of a single clause. In (40a), the occurrence of the realis NAV auxiliary *i* indicates that the two serial verbs both occur in realis mood; the occurrence of the first person singular clitic *='o* encodes the shared actor between the two verbs. It is impossible for individual verbs to each take an independent auxiliary and an independent pronominal clitic, as in (40b).

(40) a. *i='o* ***yon-i*** ***bon~~u~~*** *to* *f'ue* *'o* *oyonapei'i.*
AV.REAL=1SG stay-LV eat.AV NTOP sweet.potato TOP kitchen
'I stayed in the kitchen eating sweet potatoes.'

b. * *i='o* *yon-i* *mi='o* *bon~~u~~* *to*
NAV.REAL=1SG stay-LV AV.REAL=1SG eat.AV NTOP
f'ue *'o* *oyonape'i.*
sweet.potato TOP kitchen
intended 'I stayed in the kitchen eating sweet potatoes.'

c. *i='o* ***haf-a*** ***uh*** *to* *takau*
NAV.REAL=1SG take-PV go.AV NTOP Kaohsiung
'o *naau.*
TOP Naau
'I took Naau to Kaohsiung.'

Despite the shared value of modality and person information, it is to be noted that the auxiliary remains invariably in NAV marking and therefore maintains voice concord only with V1 in a non-concordant SVC, not with V2. Example (41) indicates that an AV-marked auxiliary is not allowed in a non-concordant SVC, while (40a) illustrates that the NAV auxiliary maintains voice concord with the LV-marked first verb but not with the AV-marked second verb.[16] Recall that in an independent clause (see section 3.1), an AV-marked verb is accompanied by an AV auxiliary but a NAV-marked verb is accompanied by a NAV auxiliary, respectively. This difference in ability to maintain voice concord with the auxiliary thus makes evident that the V1 verb displays more of the finiteness features expected of an autonomous verb. By contrast, the V2 verb remains invariably in AV marking, which in the serial format fails to agree with the NAV-marked auxiliary and does not reflect grammatical relations in the monoclausal configuration (LOCATION-TOPIC in (40a)). Judging by the voice pattern on the auxiliary, it becomes clear that the non-concordant voice marking does not reflect lexical and syntactic autonomy but instead indicates reduced and fossilised inflectional possibilities in a non-concordant SVC.[17] This reduced voice contrast has consequences for the interpretation of wordhood and will be dealt with in section 5.3.

---

[16] Tsou manifests a four-way voice contrast on the verb (AV-PV-LV-RV) but a two-way (AV-NAV) voice contrast on the auxiliary (if the auxiliary occurs in realis mood).

[17] Reduced or fossilised inflectional possibilities are not entirely unusual in SVCs if we consider that serial verbs often serve as a pathway to lexicalisation and grammaticalisation. See n. 19.

(41) * *mi='o* [*yon-i* *bonu*] *to* *f'ue*
AV. REAL=1SG stay-LV eat.AV NTOP sweet.potato
'o *oyonapei'i.*
TOP kitchen
intended 'I stayed in the kitchen eating sweet potatoes.'

(iv) Contiguity of components

Constituents in a non-concordant SVC form a strongly contiguous sequence that cannot be broken up by the conjunction marker *ho*, as in (42). Nor can the constituents be separated by intervening particles such as the negation marker *o'te* or the hearsay marker *nana*, as shown in (43a) and (44a), respectively.

(42) * *i='o* ***yon-i*** *ho* ***m-apaso*** *to* *fou* *'o* *oyonapei'i.*
NAV.REAL=1SG stay-LV and AV-chop NTOP meat TOP kitchen
intended 'I stayed in the kitchen chopping meat.' (Locational SVC)

(43) a. * *te='o* ***yon-i*** *o'te* ***bonu*** *to* *f'ue*
IRR=1SG stay-LV NEG eat.AV NTOP sweet.potato
'o *oyonapei'i.*
TOP kitchen
intended 'I will not stay in the kitchen eating sweet potatoes.'

b. *te='o* *o'te* ***yon-i*** ***bonu*** *to* *f'ue*
IRR=1SG NEG stay-LV eat.AV NTOP sweet.potato
'o *oyonapei'i.*
TOP kitchen
'I will not stay in the kitchen eating sweet potatoes.'

(44) a. * *i=si* ***yon-i*** *nana* ***bonu*** *to*
NAV.REAL=3SG stay-LV EVID eat.AV NTOP
kamcia *to* *naau* *'o* *oyonatmopsu.*
candy NTOP naau TOP school
intended 'Naau stayed at school reportedly eating (some) candy.'

b. *i=si=cu* *nana* ***haf-a*** ***maine'e*** *ho*
NAV.REAL=3SG=COS EVID take-PV go.home.AV and
tuocos-i *na* *nte* *himooko.*
ask-LV TOP IRR provide.care
'He reportedly took (the child) back and asked who might be the caregiver (of the child).' (Tung I-42:019)

If the negation particle and the hearsay marker are to occur in a non-concordant SVC, they must precede the entire verb series, with scope over both V1 and V2, as in (43b) and (44b) above. In the case of the negation particle, for instance, it is impossible to negate the second verb while asserting the truth of the first, as shown in (43a). The distribution of these two markers provides evidence that the two serial verbs form an inseparable whole and arguably act like a single morphological word. I will return to this issue in section 5.3.

Table 1 summarises the formal properties of concordant and non-concordant SVCs, together with their semantic contribution to the joint predicate ('participant' for

information relevant to participants involved in an event; 'modifying' for information relevant to how an event is performed).

Table 1: Concordant and non-concordant SVCs: Structural properties

Structural Properties	Concordant SVCs	Non-Concordant SVCs
verb types	V1: unrestricted <modifying>	V1: restricted <participant>
	V2: unrestricted <participant>	V2: unrestricted <participant>
voice marking	concordant	non-concordant
modality and person marking	one specification	one specification
contiguity	weakly contiguous	strongly contiguous

5 Wordhood and event relations of concordant and non-concordant SVCs

As noted in section 2, SVCs may come in one-word or multi-word structures, encoding a single, tightly integrated event or a package of events conceptualised as being connected. Utilising diagnostics for evaluating wordhood and event integration, I argue that a concordant SVC of Tsou forms a sequence of separate words (section 5.1) and expresses a package of two (or more) events loosely connected within a macro frame (section 5.2). A non-concordant SVC, by contrast, comprises a single word (section 5.3) and displays features of tighter event integration (section 5.4).

5.1 Concordant SVCs: Wordhood

On initial inspection, component contiguity and voice concord may suggest that a concordant SVC forms a one-word structure. Evidence from its ability to undergo derivational processes, coordination, and non-interruptability, however, suggests otherwise.

(i) Input to derivational processes

As mentioned in section 3.2.1, a simplex verb in Tsou can combine with derivational affixes such as *ho....na* 'season' to create new words (*mu'ei* 'hot' vs *ho-mu'ei-na* 'summer'). A concordant SVC, however, cannot combine with this circumfix even when the combination appears semantically appropriate. Attempts to combine the concordant SVCs *asng~~u~~cs~~u~~ mu'ei* 'persistently hot' and *conino mepeipi* 'usually fog (up)' with the *ho...na* affix are considered ungrammatical, as shown in (45) and (46) below. The results indicate that a concordant verb series does not form a morphological word.

(45) *asng~~u~~cs~~u~~* 'continuously' + *mu'ei* 'hot' 'persistently hot'
**ho-asng~~u~~cs~~u~~-mu'ei-na* intended 'a season that remains persistently hot'

(46) *conino* 'usually' + *mepeipi* 'fog (up)' 'usually fog (up)'
**ho-conino-mepeipi-na* intended 'a season in which fog usually forms'

(ii) Coordination

Words, phrases, and clauses in Tsou can be conjoined by the conjunction marker *ho*, as seen in (47a) below by the conjoining of two verbs 'scold' and 'beat'. Parts of a concordant verb series can be conjoined, too, as illustrated in (47b)–(47c) where the V1 component 'ever' has scope over the coordination of two verbs in V2 position. The

possibility of conjoining elements indicates that a concordant SVC does not form a morphological word but constitutes a phrasal structure composed of separate words.

(47) a. *i=si* [*koic-a* *ho* *tiusn-a*] *'o* *oko=si.*
NAV.REAL=3SG scold-PV and beat-PV TOP child=3SG
'He scolded and beat his child.'

b. *o='u=s'a=n'a* ***aaht-a*** ***koic-a*** *'o* *oko=si.*
NAV.REAL=1SG=SA=PROG ever-PV scold-PV TOP child=3SG
'I have never scolded the child.' (concordant SVC)

c. *ahoi* *ne* *moh=ta=n'a* *yayo*
begin.AV NTOP AV.REAL=3SG=PROG be.born.AV
ac'~~u~~h~~u~~ *maitan'e,* *o'amocu*
continue.AV today never
o='u=s'a=n'a ***aaht-a*** **[*koic-a*** *ho* ***tiusn-a*]**
NAV.REAL=1SG=SA=PROG ever-PV scold-PV and beat-PV
'e *oko.*
TOP child
'I have never scolded nor beaten the child from his birth till now.'

(iii) Non-interruptability

In a concordant SVC, the juxtaposition of constituent verbs cannot be broken up by the conjunction marker *ho* (see (26a-b)), but the constituents can be separated by the negation particle *o'te*, as in (48), and the adverbial *c'o* 'only', as in (49). These examples establish that a concordant SVC does not comprise a single morphological word; rather, it is a sequence of morphological words allowing intervening particles.

(48) *os='o=cu* ***ahoz-a*** *o'te* ***an-a*** *'o* *fou.*
NAV.REAL=1SG=COS start-PV NEG eat-PV TOP meat
'I have started not eating the meat.' (H. Chang 2005:15)

(49) *hoci* *eno* *o'te* *ahtu* *m~~u~~ch~~u~~,* ***asng~~u~~c~~u~~*** *c'o* ***vovoez~~u~~.***
if then NEG ever.AV rain.AV continuously.AV only dry.AV
'If it has never rained, then, it (the weather) would remain dry.' (Tung I-1:9)

5.2 Concordant SVCs and internal event relations

In section 4.1, it is shown that a concordant SVC comprises a single clause insofar as the verb series has a single specification of modality and pronominal marking. Monoclausality, however, does not entail that information contributed by component verbs is necessarily bundled into a single, closely-knit event, often considered a landmark feature of a single predicate. While the precise boundary of a single event and a package of (sub)events linked within a macro-event often varies by culture and is not easy to define, several lines of evidence indicate that the component (sub)events inside a concordant series are not closely integrated but conceptually dissociable. In (50), the concordant series is separated by the adverbial element *c'o* 'only', which modifies the verb to the right independently from the desire verb to the left. Separate scope of modification is apt in this case for it is possible to crave bananas but not other types of fruits.

(50) *i=’o* ***uci-a*** *c’o* ***an-a*** *’o* *tac~~u~~m~~u~~.*
NAV.REAL=1SG want-PV only eat-PV TOP banana
‘I wanted to just eat the bananas.’

Conceptual independence between event information contributed by component verbs is also discernable in the scope properties of negation particles. Example (51a) shows that the concordant series can be separated by an intervening negation particle, which negates the event denoted by the V2 verb (‘flood’) but not the information contributed by the V1 verb (‘eventually’). That is, the truth value of the flooding event is conceptually separable from details specifying the long-standing expectation (‘eventual’) of its happening. Independent negation for information denoted by V1 is also possible. Example (51b) shows that negation can have scope only over the V1 but not the V2—it is possible to assert the truth of going somewhere but refute the details regarding how the motion event is performed.

(51) a. *moh=cu* *la* ***petoh~~u~~e~~u~~*** *o’te* ***et~~u~~p~~u~~*** *’e* *hp~~u~~hp~~u~~ng~~u~~.*
AV.REAL=COS HAB eventually.AV NEG flood.AV TOP world
‘The world finally was not flooded.’ (Tung I-30:004)

b. *’ote* ***n~~u~~th~~u~~*** ***maine’e*** *ianan’ou* *maine’e.*
NEG together.AV go.home.AV separately.AV go.home.AV
‘Don’t go home together! Go home separately.’

As an interim summary, a concordant SVC displays features expected of a sequence of separate words: it is unable to undergo further derivational processes, its internal parts can be conjoined, and it allows intervening particles (except for the conjunction marker *ho*). Semantically, a concordant SVC allows conceptual independence between component verbs in terms of the information they contribute, for it is possible to confirm the occurrence of a specific event but deny details on how the event is performed or specification regarding the speaker’s attitude toward the event. In section 6 the possibility of allowing conceptual independence in a concordant SVC will be contrasted with the event integration seen in a non-concordant SVC.

5.3 Non-concordant SVCs: Wordhood

At first blush, the lack of voice concord may argue for lexical autonomy of component verbs inside a non-concordant SVC and thus a multi-word interpretation of the sequencing structure. Evidence from auxiliary marking (section 4.2 under (40)-(41)), however, shows that the lack of voice concord in fact indicates reduced and fossilised voice contrast and does not determine wordhood status. Instead, the following three lines of evidence indicate that a non-concordant SVC forms an inseparable word:[18] 1) it can

[18] On initial inspection, the non-concordant voice marking may render suspicious the claim of the one-word structure. However, if we consider that SVCs often serve as a pathway to lexicalisation and grammaticalisation and the meaning of the whole is not necessarily equal to the sum of the parts in the serial context, it is no longer unexpected for components inside a one-word SVC to retain features or markings reminiscent of their status as otherwise independent verbs—even if these features/markings may fail to be in accord in certain cases. In the Toqabaqita SVC below, the one-word verb series requires a count noun for the subject, but the modifying verb *sukani* ‘be of little degree’ takes a mass noun when used outside of

undergo derivational processes, 2) its internal parts cannot be conjoined, and 3) the constituent verbs cannot be separated for hosting the negation particle *o'te* or the hearsay marker *nana*.

(i) Ability to undergo derivational processes

A non-concordant SVC such as *tith-eni bonʉ* 'use-eat' can undergo category changing derivational processes and be used as a noun when preceded by a prenominal particle, as shown in (52a). When the 'use-eat' series functions as a nominal, it can be reduplicated for plural marking, as shown in (52b). The reduplicated form can be further attached with the prefix *ma-* for expressing the collective notion of 'all kinds of', as shown in (52c). Even though both reduplication and the *ma-* prefixation attach to the initial verb, they have the entire SVC in the scope. In (52c), for instance, the entire verb series 'use eat' is nominalised for a collective expression; the structure does not mean 'all kinds of using' plus 'eat'.

(52) a. *'o* *tith-eni* *bonʉ*
TOP use-RV eat.AV
'the dining-ware' (Lit. the use-eat)

b. *'o* *ti~tith-eni* *bonʉ*
TOP RED~use-RV eat.AV
'the dining-ware (PL)'

c. *'o* *ma-ti~tith-eni* *bonʉ*
TOP MA-RED~use-RV eat.AV
'all kinds of dining-ware'

(ii) Coordination

Unlike concordant SVCs, parts of a non-concordant verb series cannot be conjoined, as illustrated by the attempt to conjoin 'dip' and 'drink' inside the SVC in (53). The impossibility of coordination indicates the integrity of non-concordant serial verbs as a morphological word, like a simplex predicate.

(53) a. *te='o* [*yo'u* *ho* *m-imo*] *to* *chumu.*
IRR=1SG dip.AV and AV-drink NTOP water
'I will dip water and drink it.' (coordination of simplex predicates)

b. * *te='o* *tith-eni* [*yo'u* *ho* *m-imo*] *to*
IRR=1SG use-RV dip.AV and AV-drink NTOP
chumu *'o* *hopi.*
water TOP ladle
intended 'I will use the ladle to dip water and drink it.'

c. *te='o* [*tith-eni* *yo'u*] *ho* [*tith-eni* *m-imo*]
IRR=1SG use-RV dip.AV and use-RV AV-drink

---

serialisation (Lichtenberk 2006:259). Therefore, inconcordant values of grammatical categories are not hard evidence to the wordhood status of an SVC.

(i) *nau* *ku* [*sukani* *mataqi*].
1SG 1SG:NFUT be.of.little.degree be.sick
'I'm a little sick.' (Toqabaqita, Lichtenberk 2006:259)

to *chumu* *'o* *hopi.*
NTOP water TOP ladle
'I will use the ladle to dip water and drink it.'

(iii) Non-interruptability

Component verbs inside a non-concordant SVC are inseparable, as has been illustrated in section 4.2 by insertion of negation particles (cf. (54a) below). Insertion of the hearsay marker *nana* returns a similar result, as shown in (55). If negation particles and the hearsay marker are to occur, they must precede the entire verb series, as shown in the use of *o'te* in (54b) and the use of *nana* in (55b), respectively.

(54) a. * *te='o* *yon-i* *o'te* *bon~~u~~* *to* *kamcia* *'o* *hopo.*
IRR=1SG stay-LV NEG eat.AV NTOP candy NTOP bed
intended 'I will stay in bed not eating candy.'

b. *te='o* *o'te* ***yon-i*** ***bon~~u~~*** *to* *kamcia* *'o* *hopo.*
IRR=1SG NEG stay-LV eat.AV NTOP candy TOP bed
'I will not stay in bed eating candy.'

(55) a. * *i=si* *yon-i* *nana* *bon~~u~~* *to*
NAV.REAL=3SG stay-LV EVID eat.AV NTOP
kamcia *to* *naau* *'o* *oyonatmops~~u~~.*
candy NTOP naau TOP school
intended 'Naau stayed at school reportedly eating (some) candy.'

b. *i=si=cu* *nana* ***haf-a*** ***maine'e*** *ho*
NAV.REAL=3SG=COS EVID take-PV go.home.AV and
tuocos-i *na* *nte* *himooko.*
ask-LV TOP IRR provide.care
'He reportedly took (the child) back and asked who might be the caregiver (of the child).' (Tung I-42:019)

5.4 Non-concordant SVCs and internal event relations

In non-concordant SVCs, the lack of voice concord may appear to advocate that the component verbs are loosely juxtaposed and their event information is weakly connected (section 4.2). Scope properties of negation particles and adverbial elements, however, suggest that the component verbs are closely integrated and cannot be negated or modified separately (as in (54)-(55) above). One more piece of supporting evidence for event integration comes from the scope of temporal expressions. For example, in (56) below, the temporal expression 'today' has scope over all component (sub)events: both the administering of the herbal medicine and the killing are understood to occur concurrently sometime today. It is semantically infelicitous that only the killing occurred today, with the administering of the herbal medicine taking place prior to this time.

(56) *i=si* ***tith-eni*** ***opcoi*** *to* *hang~~u~~* *to*
NAV.REAL=3SG use-RV kill.AV NTOP enemy NTOP
pasuya *'o* *s'os'o* *maitan'e.*
Pasuya TOP herbal.medicine today
'Pasuya used the herbal medicine to kill an enemy today.'

The requirement of temporal concurrence makes evident the strong event integration inside the non-concordant SVC, which contrasts with the weaker integration seen in the concordant structure (cf. (50)–(51)). How the difference in meaning should be analysed vis-à-vis the structural properties is a point I will return to in the next section.

6 Correlation between event integration and structural compression

The investigation in section 5.2 and section 5.4 shows that concordant and non-concordant SVCs differ in the degree of event integration they manifest. The difference raises questions regarding whether both types of verb series fit the common characterisation of SVCs as comprising a single event, and how the notion of 'single event' should be defined relative to 'multiple events' on the one hand, and 'multiple (sub)events connected within a macro-event' on the other. To start with, in both concordant and non-concordant SVCs there is only one specification of modality marking, which holds for the entire verb series. The single marking indicates that the verb series as a whole is assigned one conceptual representation to which speakers express belief on its occurrence (cf. the irrealis auxiliary *tena* in (57a) and the realis auxiliary *i* in (57b)).

(57) a. *tena* *ahoz-a* *tutv-a* *'o* *otofnana* *ho*
IRR begin-PV beat-PV TOP poisonous.weed and
aasbʉt-a ***tfu-i*** *to* *chumu.*
sometimes-PV soak-LV NTOP water
'(one) first pounds the poisoning-weed and sometimes soaks (it) in water.' (Tung I-14:003; same as (25d))

b. *i='o* ***yon-i*** ***m-apaso*** *to* *fou* *'o* *oyonapei'i.*
NAV.REAL=1SG stay-LV AV-chop NTOP meat TOP kitchen
'I stayed in the kitchen chopping meat.' (same as (33))

In as much as modality marking is concerned, both types of SVCs express a single event, but this in no way suggests that a single event is necessarily a simple, mono-faceted event. Talmy (2000) and Bohnemeyer et al. (2007) note that across many different languages, sequences of (sub)events can be conceptualised as connected to each other and be linguistically encoded into one macro situation/event. The scope effects of negation and adverbial elements observed in the two types of Tsou SVCs illustrate the need to adopt a macro-event interpretation for verb serialisation as a whole but separate concordant from non-concordant series for the degree of event integration involved. Inside the macro-event of a concordant SVC (see (57a) and section 5.2), the V1 component provides evaluative attributes specifying the way in which the V2 event unfolds, and examples such as (58) show that negation and adverbial modification can hold for one component but not the other. Insofar as scope effects are concerned, a concordant SVC displays an arguably weaker degree of integration, for the unfolding of the V2 event does not have an implicative force over the attainment of the V1 attribute, and vice versa. In (58), it is perfectly possible for an anticipated flooding not to take place, despite the longstanding expectation ('eventual') attributed to its happening. As such, conceptual independence between an event and evaluative attributes imparted to the

event underlines the loose integration inside a concordant SVC, even though the overall macro-event is bound by a single modality value.

(58) *moh=cu* *la* ***petoh**u**eu*** *o'te* ***etupu*** *'e* *hpuhpungu.*
AV.REAL=COS HAB eventually.AV NEG flood.AV TOP world
'The world finally was not flooded.' (Tung I-30:004, same as (51a))

In a non-concordant SVC, by contrast, the component events are related in a different way, with the V1 component naming a specific participant conceptualised as involved in the V2 event. In (52b), if one chops something, it can be reasonably expected that the chopping takes place somewhere; in (59), travelling with companions matches a recognizable concatenation of events common in the speaker's experience. The strong, although not necessarily binding, association between the specified participant and the unfolding of the V2 event underlines the close integration seen in a non-concordant SVC, whose components cannot be isolated for separate scopes of negation and modification, as in (60) (cf. section 4.2).

(59) *i='o* ***haf-a*** ***uh*** *to* *taipahu* *'o* *naau.*
NAV.REAL=1SG take-PV go.AV NTOP Taipei TOP Naau
'I took Naau to Taipei.' (same as (34))

(60) * *te='o* *yon-i* *o'te* *bonu* *to* *kamcia* *'o* *hopo.*
IRR=1SG stay-LV NEG eat.AV NTOP candy TOP bed
intended 'I will stay in bed not eating candy.' (same as (54a))

Concordant and non-concordant SVCs evidently occupy different places on the continuum of loose vs tight integration within a macro-event, and the semantic difference ties in with the results from wordhood tests. The loose event integration seen in a concordant SVC corresponds with its multi-word status, in which the component verbs remain separate words in as much as the verb sequence as a whole cannot undergo derivational processes, the components can be conjoined, and the sequence can be separated for hosting intervening negation and adverbial elements. By contrast, the tight event integration inside a non-concordant SVC correlates with the strong structural compression holding between its components, which together comprise a single word following the results from the same diagnostics applied to a concordant SVC. The correlation between event integration and structural compression tallies well with the principle of iconicity (Haiman 1985; see also Aikhenvald (2006a:50–51) for reference to SVCs), under which structural independence between words reflects the conceptual distance between the information they represent.

7 Conclusion

Depending on the pattern of voice marking, Tsou SVCs divide into concordant and non-concordant types. The opposition in marking often leaves the impression that the former comprises a tightly-bound unit, whereas the latter is but loosely juxtaposed. Results from wordhood tests show that the impression is misleading. A concordant SVC comprises a loosely-juxtaposed multi-word structure, whereas a non-concordant SVC exhibits properties of a tightly-bound unit that comprises a single word. The structural difference is not random but tied to the kind of event information bundled together within

a macro-event. Inside a concordant SVC, the conceptual independence between the unfolding of an event and evaluative attributes assigned to the event corresponds to the weak structural compression between the component verbs, each of which remains a separate word. Inside a non-concordant SVC, on the contrary, the close association between the unfolding of an event and the naming of a participant involved in the unfolding ties in well with the one-word structure. To conclude, concordant and non-concordant SVCs tease apart distinct aspects of a conceptual macro-event, and wordhood as well as degree of semantic integration jointly reflects that distinction.

References

Aikhenvald, Alexandra Y. 2006a. Serial verb constructions in typological perspective. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 1–68. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2006b. Serial verb constructions in Tariana. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 178-201. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. and R.M.W. Dixon. (ed.) 2006. *Serial verb constructions: A cross-linguistic typology*. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Alsina, Alex. 1997. A theory of complex predicates: Evidence from causatives in Bantu and Romance. In *Complex predicates*, ed. by Alex Alsina, Joan Bresnan, and Peter Sells. 203–246. CSLI Lecture Notes 64. Stanford, CA: CSLI Publications.

Alsina, Alex, Joan Bresnan, and Peter Sells. 1997. Complex predicates: Structure and theory. In *Complex predicates*, ed. by Alex Alsina, Joan Bresnan, and Peter Sells, 1–12. CSLI Lecture Notes 64. Stanford, CA: CSLI Publications.

Bisang, Walter. 1995. Verb serialization and converbs: Differences and similarities. In *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms—Adverbial participles, gerunds*, ed. by Martin Haspelmath and Ekkehard König, 137–188. Empirical Approaches to Language Typology 13. Berlin: Mouton de Gruyter.

Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke, and Felix K. Ameka 2007. Principles of event segmentation in language: The case of motion events. *Language* 83.3:495–532.

Booij, Geert. 2009. Lexical integrity as a formal universal: A constructionist view. In *Universals of language today*, ed. by Sergio Scalise, Elisabetta Magni, Antonietta Bisetto, 83–100. Studies in Natural Language and Linguistic Theory 76. Dordrecht: Springer.

Bresnan, Joan and Sam A. Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natual Language and Linguistic Theory* 13.2:181–254.

Bresnan, Joan 2001. *Lexical-Functional syntax.* Blackwell textbooks in linguistics 16. Oxford: Blackwell.

Bril, Isabelle. 2004. Complex nuclei in Oceanic languages: Contribution to an areal typology. In *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness*, ed. by Isabelle Bril and Françoise Ozanne-Rivierre, 1–48. Empirical Approaches to Language Typology 29. Berlin: Mouton de Gruyter.

Chang, Henry Y. 2005. Restructuring in Tsou. Paper presented at the 12th Annual Conference of Austronesian Formal Linguistics Association. Los Angeles: UCLA, 30 April – 2 May 2005.

Chang, Henry Y. 2006. The guest playing host: Adverbial modifiers as matrix verbs in Kavalan. In *Clause structure and adjuncts in Austronesian languages*, ed. by Hans-Martin Gärtner, Paul Law, and Joachim Sabel, 43–82. Studies in Generative Grammar 87. Berlin: Mouton de Gruyter.

Chang, Henry Y. 2009. Adverbial verbs and adverbial compounds in Tsou: A syntactic analysis. *Oceanic Linguistics* 48.2:439–476.

Chang, Melody Ya-yin. 2004. Subjecthood in Tsou grammar. PhD dissertation. Hsinchu: National Tsing Hua University.

Crowley, Terry 2002. *Serial verbs in Oceanic: A descriptive typology*. Oxford: Oxford University Press.

Diller, A.V.N. 2006. Thai serial verbs: cohesion and culture. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 160–177. Oxford: Oxford University Press.

Dixon, R.M.W. and Alexandra Y. Aikhenvald. 2002a. Word: A typological framework. In *Word: A cross-linguistic typology*, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald, 1–41. Cambridge: Cambridge University Press.

Dixon, R.M.W. and Alexandra Y. Aikhenvald, (ed.). 2002b. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.

Dryer, Matthew S. 1997. Are grammatical relations universal? In *Essays on language function and language type: Dedicated to T. Givón*, ed. by Joan Bybee, John Haiman, and Sandra A. Thompson, 115–143. Amsterdam and Philadelphia: John Benjamins.

Durie, Mark. 1997. Grammatical structures in verb serialization. In *Complex predicates*, ed. by Alex Alsina, Joan Bresnan, and Peter Sells, 289–354. CSLI Lecture Notes 64. Stanford, CA: CSLI Publications.

Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford, CA: Stanford University Press.

Foley, William A. 1997. Polysynthesis and complex verb formation: The case of applicatives in Yimas. In *Complex predicates*, ed. by Alex Alsina, Joan Bresnan, and Peter Sells. 355-395. CSLI Lecture Notes 64. Stanford, CA: CSLI Publications.

Foley, William A. 2010. Events and serial verb constructions. In *Complex predicates: Cross-linguistic perspectives on event structure*, ed. by Mengistu Amberber, Brett Baker, and Mark Harvey, 79–109. Cambridge: Cambridge University Press.

Foley, William and Mike Olson. 1985. Clausehood and verb serialization. In *Grammar inside and outside the clause: Some approaches to theory from the field*, ed. by Johanna Nichols and Anthony C. Woodbury, 17–60. Cambridge: Cambridge University Press.

François, Alexandre. 2006. Serial verb constructions in Mwotlap. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and

R.M.W. Dixon, 223–238. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Haiman, John. 1985. *Natural syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.

Harris, Alice C. 2000. Where in the word is the Udi clitic? *Language* 76.3:593–616.

Huang, Lillian M. 1997. Serial verb constructions in some Formosan languages: A typological view. Paper presented at the Eighth International Conference on Austronesian Linguistics, Academia Sinica, Taipei, 28–30 December 1997.

Huang, Huei-ju and Shuanfan Huang. 2007. Lexical perspectives on voice constructions in Tsou. *Oceanic Linguistics* 46.2:424–455.

Ingram, Andrew. 2006. Serial verb constructions in Dumo. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 202–222. Explorations in Linguistic Typology 2. Oxford: Oxford University Press

Levin, Beth and Malka Rappaport Hovav. 2005. *Argument realization.* Research Surveys in Linguistics. Cambridge: Cambridge University Press.

Lichtenberk, Frantisek. 2006. Serial verb constructions in Toqabaqita. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 254–272. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Lin, Gujing. 2010. Argument structure of Tsou: Simplex and complex predicates. PhD dissertation. Houston: Rice University.

Matthews, P. H. 1991. *Morphology*, 2nd edition. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

Matthews, P. H. 1997. *Oxford concise dictionary of Linguistics.* Oxford: Oxford University Press.

Matthews, P. H. 2002. What can we conclude? in *Word: A cross-linguistic typology*, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald, 266–281. Cambridge: Cambridge University Press.

Mohanan, Tara. 1994. *Argument structure in Hindi.* Dissertations in linguistics. Stanford: CSLI Publications.

Mohanan, Tara. 1997. Multidimensionality of representation: NV complex predicates in Hindi. In *Complex predicates,* ed. by Alex Alsina, Joan Bresnan, and Peter Sells, 431–471. CSLI Lecture Notes 64. Stanford: CSLI Publications.

de Reuse, Willem J. 2006. Serial verbs in Lakota (Siouan). In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 301–318. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. *Subject and topic*, ed. by Charles N. Li, 491–518. New York: Academic Press.

Schachter, Paul. 1977. Reference-related and role-related properties of subjects. In *Syntax and Semantics, Vol 8: Grammatical relations*, ed. by Peter Cole and Jerrold M. Sadock, 279–306. New York: Academic Press.

Shibatani, Masayoshi. 1988. Voice in Philippine languages. In *Passive and voice*, ed. by Masayoshi Shibatani, 85–142. Amsterdam and Philadelphia: John Benjamins.

Shibatani, Masayoshi. 1991. Grammaticalization of topic into subject. In *Approaches to grammaticalization, II: Focus on types of grammatical markers*, ed. by Elizabeth Closs Traugott and Bernd Heine, 93–133. Typological Studies in Language 19.2. Amsterdam and Philadelphia: John Benjamins.

Shibatani, Masayoshi. 2009. On the form of complex predicates: toward demystifying serial verbs. In *Form and function in language research: Papers in honour of Christian Lehman*, ed. by Johannes Helmbrecht, Yoko Nishina, Yong-Min Shin, Stavros Skopeteas, and Elisabeth Verhoeven, 255–282. Trends in linguistics: Studies and monographs 210. Berlin/New York: Mouton de Gruyter.

Szakos, József. 1994. Die Sprache der Cou. Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. PhD dissertation. Bonn: The University of Bonn.

Talmy, Leonard. 2000. *Toward a cognitive semantics* (two volumes). Cambridge, MA: MIT Press.

Trask, R. L. 1993. *A dictionary of grammatical terms in linguistics*. London: Routledge.

Tung, T'ung-ho, with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng, and Margaret M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publications 48. Taipei: Academia Sinica. Reprinted in 1991 with notes and errata prepared by Paul Jen-kuei Li.

Yeh, Maya Yuting and Shuanfan Huang. 2009. A study of triple verb serialization in four Formosan Languages. *Oceanic Linguistics* 48.1:78–110.

Zavala, Roberto. 2006. Serial verbs in Olutec (Mixean). In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 273-300. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Zeitoun, Elizabeth. 1993. A semantic study of Tsou case markers. *Bulletin of the Institute of History and Philology, Academia Sinica* 64.4:969–989.

Zeitoun, Elizabeth (Qi, Li-sha). 2000. *zouyu cankao yufa* [*A reference grammar of Tsou*]. Formosan Languages Series 7. Taipei: Yuan-Liou Publishing Co. [In Chinese]

Zeitoun, Elizabeth. 2005. Tsou. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 259–290. Routledge Language Family Series 7. London: Routledge.

6 *An atypical SVC? A study of BLAQ UV qu' NP construction in Atayal**

MAYA YUTING YEH

1 Introduction

Squliq, a dialect of Atayal, (henceforth Squliq Atayal) is spoken in the north central region of Taiwan.[1] Like most Formosan languages, this language is predicate-initial and has a voice system that makes a four-way distinction.[2] Squliq Atayal exhibits a peculiar construction, named "BLAQ UV *qu'* NP construction", in which an evaluative verb in its Actor Voice (AV) form, i.e., BLAQ,[3] and an Undergoer Voice (UV)[4] verb are tightly juxtaposed to each other as an idiomatic unit; the two verbs further share a nominative argument (i.e., *qu'* NP).[5] A typical example of this construction can be seen in (1):

* It is my great pleasure to contribute this paper to a festschrift dedicated to Professor Lillian M. Huang. Prof. Huang has always encouraged young researchers to devote themselves to the study of Formosan languages. She is also a leading figure who has dedicated herself to social service for the revitalisation of Formosan languages. Such a vocation has inspired many indigenous people, including me. I am grateful to two anonymous reviewers for their valuable comments on an earlier draft of this paper. I would also like to show my deepest appreciation to the editors of the volume, Elizabeth Zeitoun, Stacy F. Teng, and Joy J. Wu, for helpful suggestions in earlier versions of this paper. Of course, I am solely responsible for any remaining errors.

1 Atayal consists of two major groups, i.e., Squliq and C'uli'. Each group has several dialects. The Squliq dialects are considered more prestigious and innovative, whereas the C'uli' dialects are more conservative and divergent from each other. Detailed research on the comparison between Squliq and C'uli' can be obtained from Tsuchida (1980, 1983) and Li (1985).

2 Please refer to Li (1980, 1985, 2000), Huang (1993), and Rau (1992) for more background information and sketch grammars of Squliq Atayal (Rau 1992; Huang 1993).

3 BLAQ is a cover term for evaluative verbs like *blaq* 'good', *yaqih* 'bad', *lokah* 'strong' and so on in Squliq Atayal. This capitalised form means that the slot in the construction in question can be filled by different evaluative verbs in actual expressions.

4 UV is a cover term for PV (patient voice), LV (locative voice), and CV (conveyance voice).

5 Juxtaposed verbs, including serial verb constructions, in the examples are underlined.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 131–155
Asia-Pacific Linguistics, 2015

(1) *blaq* *niq-un* *qu'* *mami'* *qa'*.[6]
good[AV] eat-PV NOM rice DEM[7]
'This (type of) rice tastes good.'

Verb juxtaposition makes constructions of the sort superficially similar to serial verb constructions (hereafter SVCs) and commentative complement clause constructions (abbreviated as commentative CCCs), as in (2) and (3) respectively:

(2) a. *cyux* *t-hngyang* *m-ngilis* *qu'* *ciwas.*
PROG.REM EMP-loud AV-cry NOM PN
'Ciwas is crying very loudly.'

b. *wal=nya'* *suq-un* *maniq* *kwara'* *qu'* *mami'.*
PFV=3SG.GEN finish-PV eat.AV all NOM rice
'He ate up all rice.'

(3) a. *blaq=mu* *m-wah* *qu'* *ciwas* *kira'.*
good[AV]=1SG.GEN AV-come NOM PN later
'I think it is good that Ciwas will come later.'

b. *blaq=mu* *lngay-an* *qu'* *gamu'* *qani'.*
good[AV]=1SG.GEN watch-LV NOM TV.show this
'I enjoy watching this TV show.'
(Lit. I feel good when watching this TV show.')

The most conspicuous difference between these constructions is the manifestation of voice in the verb sequences, i.e., AV-UV for the BLAQ UV *qu'* NP construction, AV-AV or UV-AV for SVCs, and AV-AV or AV-UV for commentative CCCs. Other significant differences will be discussed in sections 4–5.

The goal of the present study is to identify the structural category of the BLAQ UV *qu'* NP construction and to explain why the following verb of BLAQ (henceforth V2) appears in a UV form in this construction. Due to the superficial similarity outlined above, my discussion will focus on the relationship of juxtaposed verbs in the target construction, SVCs, and commentative CCCs by examining their respective morphosyntactic behaviors and functional features. I argue that (i) the BLAQ UV *qu'* NP construction is a fusion of SVCs and commentative CCCs, (ii) in terms of function, this construction is categorised as an evaluative construction, and (iii) the UV form of V2 is pragmatically driven, namely, the construction is designated for evaluating an undergoer participant in discourse, otherwise V2 will appear in an AV form when the evaluated

[6] The data used in the present study come from two sources. The first batch of data was drawn from several consultations with my two consultants between 2012 and 2014. The two consultants are Ms. Ciwas Batu' (female; born in 1937) and Ms. Hama' Ihil (female; born in 1944). Both are fluent speakers of a dialectal variant of Jianshi Squliq spoken in Hsinle, a village located in Jianshi Township, Hsinchu County. The other source comes from the Academia Sinica Formosan Language Archive, abbreviated as "Sinica Archive" in this study.

[7] Morpheme glosses follow the Leipzig Glossing Rules, with the following additions: AV, actor voice; CV, conveyance voice; EMP, emphatic; EXT, existential; FILL, filler; GV, goal voice; HORT, hortative; IMM, immediate; LV, locative voice; MOD, modal; NEU, neutral; PCN, place name; PN, personal name; PROS, prospective aspect; PV, patient voice; REM, remote; STAT, stative; UV, undergoer voice.

target is an actor participant. That is, BLAQ and UV will be shown to form a syntactic whole, just as the two verbs do in SVCs; this results in a structural equivalent between the two construction types. By contrast, the connection between the BLAQ UV *qu'* NP construction and commentative CCCs is built on their functional equivalence and the point that V2 in a UV form is acceptable in commentative CCCs.

The remainder of this paper is structured as follows. Section 2 reviews previous studies on the BLAQ UV *qu'* NP construction. Section 3 discusses the function of BLAQ and its grammatical status. Section 4 is a comparison between SVCs and commentative CCCs, both with BLAQ as the initial verb (V1), from the viewpoint of morphosyntax. Section 5 focuses on the morphosyntactic and semantic aspects of the BLAQ UV *qu'* NP construction. Section 6 presents a brief conclusion.

2 A review of previous studies

Two studies, Tsukida (2005) and Yeh (2013), have touched upon the equivalent of the BLAQ UV *qu'* NP construction in Formosan languages. In Tsukida (2005), one Seediq example is found displaying the same template as the BLAQ UV *qu'* NP construction in Squliq Atayal. Her example is given in (4):

(4) *malu* *qeta-an*[8] *ka* *degiyaq* *teruku.*
good.AV see-GV2 NOM mountain PN
'The Truku mountains are beautiful.' (Lit. The Truku mountains are good to be seen.') (Tsukida 2005:309 (120), underlining mine)

Tsukida identifies (4) as an SVC and provides a simple explanation for why V2 is in a UV (GV2 in her terms) form, namely, "the goal argument of the second is co-referential with the actor/theme argument of the first verb. So the second verb is in GV form" (Tsukida 2005:309–310). However, arguing the notion of argument co-referentiality as a straightforward factor in determining a verb's voice form seems rather abrupt because argument co-referentiality is a matter of pragmatics. I believe that at least in Squliq Atayal, voice form selection for verbs in serialization is associated with syntax. For example, it is usually V1, i.e., the **main** verb in an SVC, that governs the morphosyntactic representation of the whole clause or construction, so it can be either an AV or a UV form, but V2 is restricted to an AV form. In other words, a more formal account might provide a better explanation for this construction.

The second study comes from Yeh (2013), whereby many examples of the BLAQ UV *qu'* NP construction are given for the purpose of verb classification in Squliq Atayal. Yeh proposes that if an NP in the named construction is a verb's intrinsic undergoer, the selected undergoer voice form will indicate the class of the verb.[9] See (5):

---

[8] GV2 in Seediq corresponds to LV in Squliq Atayal (cf. comparison between (4) and (5b)).

[9] In Yeh (2013), the notion of "class" of a verb is determined by the "primary UV form selection" (I owe this term to Rik De Busser), in which, a certain UV form of a verb is selected for highlighting the verb's intrinsic undergoer in a realis event, as the *-un* form of *qaniq* 'eat' for *lingo' qa'* 'that apple' in (i):

(i) *wal=mu* *niq-un* *qu'* *lingo'* *qa'* *la'.*
PFV=1SG.GEN eat-PV NOM apple DEM PART
'I have eaten/ate that apple.'

(5) a. *blaq* *niq-un* *qu'* *lingo'* *qa'*.
good[AV] eat-PV NOM apple DEM
'This (type of) apple tastes good.'

b. *blaq* *kt-an* *qu'* *rgyax* *lintongsan.*
good[AV] see-LV NOM mountain PCN
'The Lintongsan mountain is beautiful.'
(Lit. 'The Lintongsan mountain is good to be seen.')

c. *blaq* *s-buling* *qu'* *syup* *qa'*.
good[AV] CV-throw.away NOM trash DEM
'It is easy to throw away the trash (because it is not heavy).'

Based on (5), the respective classes of the three verbs, 'eat', 'watch', and 'throw away' are recognised. That is, *qaniq* 'eat' belongs to the *-un* verb class, *kita'* 'see', the *-an* verb class, and *buling* 'throw away', the *s-* verb class. In other words, Yeh (2013) considers that this construction constitutes a reliable test for her "verb classification".

Yeh's (2013) analysis is not without defect. First, she adopts a slightly different name, i.e., BLAQ UV *qu'* O construction, to reflect the fact that she decomposes the construction into two portions. The two components are BLAQ and UV *qu'* O. In her analysis, she does not take the syntactic status of BLAQ into account, so assigning the nominative argument an O role is problematic. Besides, she does not propose any syntactic tests to determine the relationship between the verb BLAQ and the following UV verb, either. Assigning the argument in question an S role may be acceptable, since it is normally the voice type of V1, i.e., BLAQ, determining the case-marking of arguments and BLAQ is an AV verb. However, solution like this is not convincing enough as an explanation for why V2 in this construction does not appear in an AV form as that one does in SVCs in Squliq Atayal. As a result, using the term NP is much more general and can characterise better the nominal element in such a schematic template, so I now adopt BLAQ UV *qu'* NP for this type of construction.[10] The problems outlined above show that this construction deserves a more detailed analysis than the one proposed in Yeh (2013) and this paper intends to fill this gap.

Based on Yeh's argument, the verb, *qaniq* 'eat', is then classified into the *-un* verb class. Note, however, that the same verb can have other UV forms for other situations. For example, the *-an* form of *qaniq* 'eat', *niq-an*, can be used to convey a part-whole reading, as in (ii):

(ii) *wal=mu* *niq-an* *qu'* *lingo'* *qa'* *la'*.
PFV=1SG.GEN eat-LV NOM apple DEM PART
'I have eaten/ate that apple.'

Based on (ii), it can be said that *qaniq* 'eat' also belongs to the *-an* verb class. Such a classification is therefore problematic and the "class" might not be appropriate and needs to be amended in further research.

[10] There can be another assumption, that is, the NP in the construction in question cannot be assigned any grammatical relation; instead, the NP refers to the discourse topic in a current utterance. This assumption implies a reassessment of the marker *qu'* in Squliq Atayal. I leave this issue for further investigation.

3 BLAQ: A general account on its function and grammatical status

The goals in this section are twofold. The first is to discuss the function of BLAQ and to explain why I choose *blaq* 'good' instead of other evaluative words like *yaqih* 'bad' in the naming of the BLAQ UV *qu'* NP construction. The second is a justification of the grammatical status of BLAQ. Evaluative verbs in the BLAQ UV *qu'* NP construction like *blaq* 'good', *yaqih* 'bad', *lokah* 'strong', etc. are stative verbs in AV forms. This discussion is very important, since in some cases, we may question the AV voice status of such verbs (as in (8) and (9) in section 3.2).

3.1 BLAQ's function and an explanation for *blaq* 'good' in construction naming

In Hare's (1952) terms, BLAQ can be also named a "value word". According to Hare (1952:91), "value-terms have a special function in languages, that of commending; and so they plainly cannot be defined in terms of other words which themselves do not perform this function". He identifies 'good', 'right', and 'ought' as typical value words in English.

In Squliq Atayal, evaluative verbs that have a similar function, i.e., evaluating,[11] and can enter into the BLAQ UV *qu'* NP construction are listed in Table 1:

Table 1: A list of common evaluative verbs in Squliq Atayal

Squliq Atayal	Gloss
blaq	'good'
yaqih	'bad'
lokah	'strong'
tnaq	'enough'
'llaw	'easy'
zihung	'difficult'

The evaluating function can be observed from examples (1) and (5) of *blaq* 'good' given above. Examples of other evaluative verbs are illustrated in (6):

(6) a. *yaqih* *niq-un* *qu'* *lingo'* *qani'*.
bad[AV] eat-PV NOM apple this
'This (type of) apple tastes bad.'

b. *lokah* *niq-un* *qu'* *t-mmy-an*[12] *qani'*.
strong[AV] eat-PV NOM do-rice-LOCNMLZ this
'This (piece of) pickled meat tastes chewy.'

c. *tnaq* *niq-un* *qu'* *t-mmy-an* *qani'*.
enough[AV] eat-PV NOM do-rice-LOCNMLZ this
'The (amount of) pickled meat is enough to eat.'

[11] Note in this study, I use the term "evaluative use", instead of Hare's "commendatory use", to refer to the function of these words. My reason for this naming is simple. "Evaluative" seems more general than "commendatory" because it contains not only words with positive value, but also others with negative value, like *yaqih* 'bad' and *zihung* 'difficult' in Squliq Atayal; in contrast, "commendatory" does not seem to involve negative value.

[12] *t-mmy-an* means 'pickled pork'.

d. *'llaw* *hluzy-un* *qu'* *snyuw* *qani'*.
easy[AV] pull-PV NOM rope this
'It is easy to pull this rope (to climb because it is rough)'

e. *zihung* *hluzy-un* *qu'* *snyuw* *qani'*.
difficult[AV] pull-PV NOM rope this
'It is difficult to pull this rope (to climb because it is rough)'

Among these words, I choose *blaq* 'good' as the representative in naming the construction in question. Such a choice is based on a hierarchical structure, as in Figure 1:

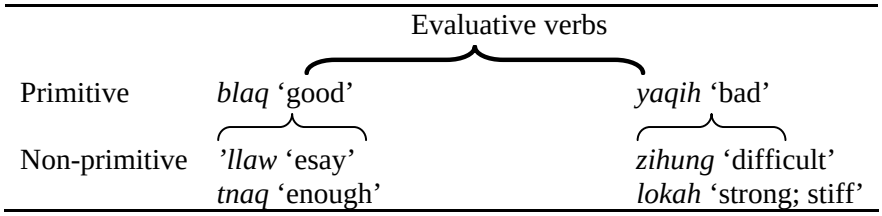


Figure 1: Classification of evaluative verbs in Squliq Atayal: A hierarchical model

First, in Figure 1, *blaq* 'good' and *yaqih* 'bad' can be separated from the rest of the evaluative verbs on the basis of a distinction between primitives or non-primitives. The notion of semantic primitives is proposed by Wierzbicka (1972).[13] Primitives are innately understood, and their meanings cannot be paraphrased in simpler terms; in contrast, other evaluative concepts or words can be defined through the primitives GOOD or BAD by either paraphrase or entailment.

On top of this, there is a contrast between *blaq* 'good' and *yaqih* 'bad'. One is morphosyntactic, and the other, cultural. Morphosyntactically speaking, *blaq* 'good' has a PV alternative form, i.e., *bliq-un* 'good', which has the same, evaluative use; in contrast, such a corresponding form of *yaqih* 'bad' is absent from this language. That means, compared with *blaq* 'good', *yaqih* 'bad' is marked because of the restricted syntactic environment it occurs. In terms of culture, in some situations, for example, when one is in a bad state (e.g., illness), Atayal people prefer using *ini' k-blaq* 'not good' rather using *yaqih* 'bad' to describe such a state.[14] Based on these two reasons, I choose *blaq* 'good' as one component in the named "BLAQ UV *qu'* NP construction". Ultimately, my selection of *blaq* 'good' reflects a notion of (un)markedness, i.e., *blaq* 'good' is unmarked and is indeed a universal primitive, too.

3.2 Grammatical status of BLAQ

BLAQ is a verbal element, since it can be negated by *ini'*. It can be further considered a stative verb, since the stative marker *k-* is obligatorily present in negative constructions (Huang 2000; Zeitoun & Huang 2000). The relevant examples are given in (7):

[13] 'Primitives' is also called "semantic primes" (Goddard & Wierzbicka 1994; Wierzbicka 1994, 1996). The definition or function of 'semantic primes' is not very different from that of the original term, i.e., 'primitives'. For this reason, I use 'primitives' in this study.

[14] Wierzbicka (1996:53) also points out this idea.

(7) a. *ini' k-blaq qu' squliq qasa'.*
NEG STAT-good[AV] NOM person that
'That man is not nice.'

b. *ini' k-lokah qu' in-lung-an=nya'.*
NEG STAT-strong[AV] NOM PFV-think-LOCNMLZ=3SG.GEN
'His heart is not hard.'

c. *ini' k-zihung qu' zyaw qasa'.*
NEG STAT-difficult[AV] NOM thing that
'That thing is hard (to solve).'

Their AV form is zero-marked. Take *blaq* 'good' for example, as seen in (8):

(8) a. *blaq qu' squliq qasa'.*
good[AV] NOM person that
'That man is nice.'

b. * *m-blaq qu' squliq qasa'.*
AV-good NOM person that

Besides, as seen from (7) and (8), these evaluative verbs have only one core argument, the nominative NP. In other words, all sentences in (7) and (8) are intransitive.

But, BLAQ can also take a genitive pronoun denoting the experiencer in commentative CCCs, as illustrated in (9):

(9) a. *blaq=mu balay niq-un qu' lingo' qa'.*
good[AV]=1SG.GEN true eat-PV NOM apple DEM
'I am in a good mood when eating the (kind of) apples.'

b. *zihung=mu yax-an qu' blihun qa'.*
hard[AV]=1SG.GEN open-LV NOM door DEM
'It is hard for me to open the door (because I have tried to open it).'

In (9), the first person bound genitive pronoun is encliticised to an AV verb. This seems very odd, because an AV verb normally does not take a genitive pronoun. However, the encliticisation of a genitive pronoun to an AV verb also occurs in other types of complement clause constructions. See (10):

(10) a. *yan=maku' m‹n›aniq qu' yumin la.*
resemble[AV]=1SG.GEN AV<PFV>eat NOM PN PART
'I thought as if Yumin had eaten (his meal).'

b. *siy=nya' suq-un maniq ni' yumin*
think[AV]=3SG.GEN finish-PV eat.AV GEN PN
kwara' qu' mami' la.
all NOM rice PART
'He thought Yumin ate up all the rice.'

Likewise, in (10), *yan* 'resemble' and *siy* 'think' are complement-taking verbs and they are in their AV form[15] co-occurring with a genitive clitic pronoun (i.e., *=maku'* 'first singular genitive pronoun' and *=nya'* 'third singular genitive pronoun').

I do not gain a full understanding about such special composites yet. At this stage, I take them as formulaic expressions exclusive with complement clause constructions in Squliq Atayal. The logic behind such composites is left for future study. Most importantly, I maintain the analysis that BLAQ is an AV verb in the BLAQ UV *qu'* NP construction.

4 BLAQ as V1 in two different constructions, SVCs and commentative CCCs: A comparison

As briefly introduced in section 1, SVCs and commentative CCCs exhibit the same verb juxtaposition as the BLAQ UV *qu'* NP construction. This similarity may lead us to suspect that there is some link between our target construction and SVCs and commentative CCCs.

In this section, I will make a comparison between SVCs and commentative CCCs by respectively examining them with four morphosyntactic parameters: (A) voice manifestation, (B) actor sharing, (C) TAM (tense/aspect/mood) sharing, and (D) polarity sharing. A brief summary of their differences is given in Table 2:

Table 2: Juxtaposed verbs' relationship in SVCs and commentative CCCs in Squliq Atayal: A comparison

Construction type Parameter	SVCs	Commentative CCCs
(A) Manifestation of voice in verb sequences	(a) AV UV (b) UV AV	(a') AV UV (b') AV UV
(B) Sharing of arguments	Obligatory	Optional
(C) Sharing of TAM information	Obligatory	It depends
(D) Sharing of polarity value	Obligatory	Optional

The findings in Table 2 show that the relationship between V1 and V2 is tighter in an SVC than in a commentative CCC. Furthermore, I will argue that (non-)simultaneity of events respectively encoded by juxtaposed verbs is an important functional factor triggering the morphosyntactic differences between SVCs and commentative CCCs. My discussion in this section will further outline the structural category of the BLAQ UV *qu'* NP construction.

4.1 A short sketch of SVCs in Squliq Atayal

From a typological perspective, the most important characteristic for the identification of SVCs is twofold: two or more verbal components which form an SVC

[15] The negative form of *yan* 'resemble' is *k-yan*.

act together as **a single predicate** and the construction is then **conceptualised as one event** (Sebba 1987; Crowley 1987, 2002; Foley & Olson 1985; Lord 1993; Aikhenvald & Dixon 2006, among others). This characteristic can be evidenced in terms of formal properties, as in (I):

(I) a. no intervening coordinator, subordinator marker of coordination, subordination, or syntactic dependency of any sort between verbs in SVCs, including pause
b. sharing of arguments
c. sharing of TAM
d. sharing of polarity value

These properties are also syntactically critical in differentiating SVCs from multiclausal structures (e.g., coordination, complement clauses, subordinate clauses, etc.). These characteristics also occur in Squliq Atayal. In this language, the crucial piece of syntactic evidence for them is as displayed in (II):

(II) The initial verb (V1) in a series of verbs is the main verb, which governs the morphosyntactic representation of the whole clause or construction by carrying the information of voice, TAM, polarity value etc. The most remarkable representation is that V1 may be either an AV or a UV verb, while non-initial verbs are restricted to an AV form.

This has been attested in Yeh & Huang (2009).

The following discussion only focuses on the properties exhibited in certain SVCs, i.e., the ones with BLAQ as V1. I begin with a brief discussion on the function of BLAQ to the meaning of a whole construction.

4.1.1 BLAQ as V1 in SVCs

BLAQ is a manner verb used to describe the way the other action encoded by a succeeding verbal component is performed. This categorises SVCs of the sort as manner SVCs. (11) illustrates this type of SVCs:

(11) a. *cyux* *lokah* *m-t-zyaw* *qu'* *yumin.*
PROG.REM strong[AV] AV-do-thing NOM PN
'Yumin is working diligently.' (Lit. Yumin is hard-working.')

b. "*blaq* *m-t-zyaw* *ru,* *teta'=simu'* *ini'*
good[AV] AV-do-thing CONJ only.then=2PL.NOM NEG
iy *ini'* *iy* *k-k'uy* *yal* *m-t-zyaw.*"
FILL NEG FILL STAT-weary very AV-do-thing
' "(If you (PL)) work well, you won't be tired out by work." '
(Sinica Archive:29-015-c, 29-015-d)

In (11a), *lokah* 'strong' is used to refer to the way Yumin performs a working event. Similarly, in (11b), *blaq* 'good' is translated into 'well' in the situation of being used to modify an action of working. Recall our discussion on BLAQ's function in section 3.1. *Lokah* 'strong' in (11a) and *blaq* 'good' in (11b) remain evaluative verbs, since using

stative verbs of the sort as V1 in SVCs can be analogous to an act of attributives assigned to the action encoded by V2 in the same constructions.

4.1.2 Manifestation of voice in verb sequences in SVCs

In Squliq Atayal, verbal components of an SVC can be in either an AV-AV (11a–b) or UV-AV sequence (12).

(12) *si'* *ga'*, *bliq-un=mamu'* *m-lahang* *qu'*
must TOP good-PV=2PL.GEN AV-take.care NOM
p‹in›-ts-yag-an=mamu' *ka* *sinuw* *qani'*.
CAUS<PFV>-do-thing-LOCNMLZ=2PL.GEN FILL sago this
'You have to look after the sago you planted carefully.' (Lit. 'You have to look after the sago you planted in a good way.') (Sinica Archive:29-009-d)

Note that V2 always appears in an AV form in SVCs.

Other voice sequences, i.e., the AV-UV in (13a) and the UV-UV sequence in (13b), are prohibited in SVCs in this language:

(13) a. * "*blaq* *p-ts-yag-un* *ru,* *teta'=simu'*
good[AV] CAUS-do-thing-PV CONJ only.then=2PL.NOM
ini' *iy* *ini'* *iy* *k-k'uy* *yal* *m-t-zyaw.*"
NEG FILL NEG FILL STAT-weary very AV-do-thing

b. * *si'* *ga'*, *bliq-un=mamu'* *khang-an* *qu*
must TOP good-PV=2PL.GEN take.care-LV NOM
p‹in›-ts-yag-an=mamu' *ka* *sinuw* *qani'*.
CAUS<PFV>-do-thing-LOCNMLZ=2PL.GEN FILL sago this

Sentences like (13a–b) are not grammatical.

4.1.3 Sharing of arguments in SVCs

Sharing of arguments is a definitional property of all serial verbs in Squliq Atayal. This can be seen from (12), in which the two verbs *bliq-un* 'good' and *m-lahang* 'take care' share the same actor and undergoer argument.

It is worth noticing that when the shared argument is an enclitic pronoun, it always attaches to V1 (i.e., *bliq-un* 'good' in (12)). In contrast, V2 is not allowed for an enclitic pronoun. This can be illustrated in the following ungrammatical sentence:

(14) * *si'* *ga'*, *bliq-un* *m-lahang=mamu'* *qu'*
must TOP good-PV AV-take.care=2PL.GEN NOM
p‹in›-ts-yag-an=mamu' *ka* *sinuw* *qani'*.
CAUS<PFV>-do-thing-LOCNMLZ=2PL.GEN FILL sago this

4.1.4 Sharing of TAM information in SVCs

In Squliq Atayal, tense, aspect, and mood categories are marked once in an SVC. These categories can be realised as affixes attaching to V1 or as an independent marker placed immediately before V1:

(15) a. *p-lokah* *m-t-zyaw* *kira'* *qu'* *yumin!*
FUT-strong[AV] AV-do-thing later NOM PN
'Yumin will work hard later.'

b. *kya* *nyux* *iy* *m‹in›blaq* *maqut.*
probably EXT.IMM FILL AV<EMP>good ask.AV
'(People) probably had ever cautiously interrogated (him) to a certain degree.' (Lit. '(People) probably had ever cautiously interrogated (him) in a good way.') (Sinica Archive:17-018-c)

c. *nanu'* *ana'* *ga',* *ziray=ta'* *raral*
what no.matter TOP age=1PI.GEN in.the.past
hya' *ga',* *bliq-aw=simu'* *k‹m›al*
3SG.NEU top good-PV.HORT=2PL.NOM <AV>speak
ga', *nanu'* *balay* *qutux* *qu'* *yanay*
TOP what true one NOM brother.in.law
qani' *ga',* *ini'* *bax-i'* [*hmc-i'*]
this TOP NEG give.up-PV.IMP [at.will-PV.IMP]
c'abaw *k‹m›al.*
indelicate[AV] <AV>speak
'In our time, it was the brother-in-law whom you should speak with properly and couldn't speak indelicately.' (Sinica Archive:19-006-a)

In (15a), a future tense marker *p-* is attached to V1 in root form; in (15b), the free aspectual marker *nyux* appears before V1, but not before V2; in (15c), V1 is marked with a hortative marker *-aw*.

In contrast to V1, non-initial verbs like V2 in SVCs never convey TAM information, so the constructions in (16) are ungrammatical:

(16) a. * *blaq=simu'* *kyal-aw* *ga',*
good[AV]=2PL.NOM speak-PV.HORT TOP

b. * *kya* *iy* *blaq* *nyux* *maqut.*
probably FILL good[AV] EXT.IMM ask.AV

4.1.5 Sharing of polarity value in SVCs

In Squliq Atayal, there can be only one negator per SVC. The negator is placed immediately before V1, and its scope is over the whole SVC, as in (17a):

(17) a. *ini'* *iy* *k-lokah* *mgey* *qu'* *kneril.*
NEG FILL STAT.NEG-strong run.away.AV NOM female
'(Because) women were not good at fleeing.' (Sinica Archive:16-013-c)

(17b) is an ungrammatical sentence, since a negator cannot be placed between verbal components in an SVC in this language:

(17) b.	*	*k-lokah*	*ini'*	*pgey*	*qu'*	*kneril.*
		STAT.IMP-strong	NEG	run.away	NOM	female

4.1.6 A discussion of results from an examination of SVCs with BLAQ as V1 using four parameters

Thus far, I have examined formal properties of SVCs with BLAQ as V1, via four parameters. Two points are conspicuous. First, TAM and polarity values are respectively marked once per SVC. Second, it is always V1 that determines the case-marking of arguments and carries TAM and polarity value in the whole construction.

The first point can be explained in terms of Aikhenvald's (2006) argument, namely, "no independent choice or contrast in any of these categories is possible for the individual components of an SVC…This makes them all different from multiclausal structures" (Aikhenvald 2006:8). This means that BLAQ and its following verb in SVCs in question together form a syntactic whole. This point can be justified by means of reconsidering the construction in an opposite way. That is, rewriting an SVC like (11b) by means of placing two identical aspectual markers before *blaq* 'good' and *m-t-zyaw* 'work (AV)' respectively, we get a pragmatically bizarre sentence, since two juxtaposed clausal structures have no sequential relationship in discourse, as in (18):

(18) ?	"*musa'*	*blaq*	*musa'*	*m-t-zyaw*	*ru,*
	PROS	good[AV]	PROS	AV-do-thing	CONJ
	teta'=simu'	*ini'* *iy*	*ini'* *iy*	*k-k'uy*	*yal*
	only.then=2PL.NOM	NEG FILL	NEG FILL	STAT-weary	very
	m-t-zyaw."				
	AV-do-thing				

Besides, the first point also implies the simultaneity of subevents. Subevents here refer to the state encoded by the verb BLAQ (V1) and the action encoded by V2. Simultaneity of the sort is a characteristic of SVCs in Squliq Atayal.

The second point implies the division of labor between V1 and V2 in SVCs. V1 carries most morphosytactic weight, while V2 is assigned conveying the semantic relationship of argument(s).[16] In other words, the absence of any one of the two verbal components results in an incomplete sentence, both semantically and syntactically, because we cannot decompose (12) into two separate sentences, i.e., (19a) and (19b):

(19) a.	*	*si'*	*ga',*	*bliq-un=mamu'*	*qu'*	
		must	TOP	good-PV=2PL.GEN	NOM	
		p‹in›-ts-yag-an=mamu'		*ka*	*sinuw*	*qani'.*
		CAUS<PFV>-do-thing-LOCNMLZ=2PL.GEN		FILL	sago	this

[16] For example, in (12), we can only rely on V2 (i.e., *m-lahang* 'take care') to realise there are two participants (i.e., an actor and a care-receiver or benefactor) in a caretaking event, instead of V1.

b. * *si'* *ga'*, *m-lahang* *qu'*
must TOP AV-take.care NOM
p‹in›-ts-yag-an=mamu' *ka* *sinuw* *qani'*.
CAUS<PFV>-do-thing-LOCNMLZ=2PL.GEN FILL sago this

(19a) is semantically odd. (19b) is unacceptable for a grammatical reason: *m-lahang* 'take care' is a semantically transitive verb, so there should be a nominative actor and an oblique undergoer argument in this single clause; however, both the two arguments are absent.

Our discussion on SVCs in this section can be recapitulated in (III):

(III) a. When BLAQ is the main verb of SVCs, SVCs as such are categorised as manner serial verb constructions.

b. TAM and polarity values are marked once per SVC.

c. Syntactic weight regarding the voice type, TAM, and polarity value of a whole construction is always loaded on V1.

d. (b) and (c) together imply verbal components in an SVC together form a syntactic whole; in other words, the construction is monoclausal.

e. The subevents separately encoded by verbal components occur simultaneously; this evidences why V2 is always in an AV form, despite the fact that semantically the verb is intransitive or transitive.

4.2 BLAQ as V1 in commentative CCCs

I now turn to the semantic and formal relationships between juxtaposed verbs in commentative CCCs in Squliq Atayal. BLAQ is the first verb in the constructions, so it is symbolised as V1.

4.2.1 A sketch of complement clause constructions in Squliq Atayal

Since complement clause constructions in Squliq Atayal has never been dealt with in previous studies, it is necessary to briefly outline their schema.

Based on two typological studies, i.e., Noonan (1985) and Dixon (2006), complement clause constructions can be said to have the following main characteristics, respectively in three different aspects, as in (IV):

(IV) a. Functional aspect:
The constructions are designated for people to state what a proposition which is seen, believed, known etc is. This proposition can be a fact, an activity, or a potential state, etc.

b. Formal aspect:
i. They refer to a syntactic situation in which a clause is a core argument of a higher clause; the clause functioning as a core argument is called a "complement clause".

ii. The complement clause has its own internal constituent structure.
iii. In general, there is a complementiser used to link a higher clause and a complement clause. The complementiser can be a word, particle, clitic or affix. For example, in English, the complementiser is *that*, as in (20).

c. Semantic aspect:
Complement constructions can be divided into a number of classes based on the semantics of complement-taking verbs, including utterance verbs (e.g., *tell*, *say*, *promise*), propositional attitude verbs (e.g., *believe*, *think*, *suppose*), commentative verbs (e.g., *regret*, *be sorry*, *good*), knowing verbs (e.g., *know*, *discover*, *realize*), and so on.

(20) Nelson regrets that Perry got the nod. (Noonan 1985:117 (337))

In Squliq Atayal, two formal aspects of complement clause constructions diverge from the definition above: (IVbi) and (IVbiii). These distinct behaviours can be respectively rewritten as (Va) and (Vb):

(V) a. Since no case marker is used to introduce a complement clause, the clause cannot be analysed as a core argument of the complement-taking verb of a complement clause construction.

b. No complementiser is found in Squliq Atayal.

Regarding (Va), even though a complement clause is not a core argument in this language, the clause is used to embody the proposition obligatorily required by the semantics of the complement-taking verb. (Vb) means that verb juxtaposition occurs in complement clause constructions. The two points can be illustrated in (21):

(21)

kong-un=mu	*m-wah*	*qu'*	*yumin.*
afraid-PV=1SG.GEN	AV-come	NOM	PN

'I am afraid that Yumin will come.'

In the above discussion, I have provided a sketch of complement clause constructions in Squliq Atayal. In what follows, I turn my focus to commentative CCCs. Constructions of the sort are very similar to our target construction, the BLAQ UV *qu'* NP construction, both formally and functionally.

4.2.2 BLAQ as V1 in commentative CCCs

As pointed out by Hare (1952), the primary function of the word *good* is for the speaker or the higher clause subject to commend on someone or something. The same idea is proposed in Noonan (1985). *Good* is further categorised as a commentative complement-taking verb in Noonan's typological study. (22) displays two examples for evaluative complement clause constructions from Noonan (1985):

(22) a. I regret that Sally left Herman. (Noonan 1985:119 (347))

b. Irish

Is	*maith*	*dhó*	*í*	*a*	*theacht*
be	good	to.him	her	COMP	come.NMLZ

'It's good for him that she came.' (Noonan 1985:124 (378); glosses mine)

In (22), the two predicates, *regret* and *maith* 'good', introduce their respective clause, *Sally left Herman* and *í a theacht* 'she came'.

In Squliq Atayal, as mentioned in 3.1, BLAQ verbs such as *blaq* 'good', *yaqih* 'bad', *lokah* 'strong' and so on have the same function. Now let us discuss the relationship between BLAQ (as V1) and V2 of a commentative CCC in terms of the aforementioned four parameters, as repeated in (VI):

(VI) a. Manifestation of voice in verb sequences
b. Sharing of arguments
c. Sharing of TAM information
d. Sharing of polarity value

It will be shown that commentative CCCs are quite distinct from SVCs, though both take BLAQ as their respective V1. This difference can have a semantic interpretation, i.e., the notion of (non-)simultaneity of events; in other words, V1 and V2 in commentative CCCs respectively specify two different events. Moreover, we observe that non-simultaneity of events is a good reason for why V2 appears in a UV form.

4.2.2.1 Manifestation of voice in verb sequences in commentative CCCs

When taking a complement clause, the voice manifestation of BLAQ and the verb following it may either be AV-AV or AV-UV, as in (23):

(23) a. *yaqih* *busuk=su.*
bad[AV] drunk[AV]=2SG.NOM
'It is bad (for me) that you are drunk!' (Sinica Archive:11-018-d)

b. *yaqih* *q‹n›ihul=nya'* *balay* *panga'*
bad[AV] <UV.PFV>force=3SG.GEN true carry.on.back
payat *waya'.*
four basket
'It is bad that he (i.e., the boss) really forced (the ass) to carry four baskets of salt on back (at a time).' (Sinica Archive:01-022-a)

When BLAQ acts as V1, in both SVCs and CCCs, a sequence of AV-AV is allowed. However, only SVCs allow a UV-AV sequence, and only CCCs allow an AV-UV sequence.

4.2.2.2 Sharing of arguments in commentative CCCs

A complement-taking verb and the predicate of a complement clause do not necessarily share identical arguments, as in (24):

(24) a. *aki'=mu* *balay* *blaq* *m-wah* *qu'* *ciwas.*
MOD=1SG.GEN true good[AV] AV-come NOM PN
'I really hope Ciwas can come.'
(Lit. 'It is good for me that Ciwas will come.')

b. *blaq=mu* *niq-un* *qu'* *t-mmy-an=nya'*.
good[AV]=1SG.GEN eat-PV NOM do-rice-LOCNMLZ=3SG.GEN
'I enjoy tasting his pickled pork.'
(Lit. 'It is good for me to taste his pickled pork.')

In (24a), the complement-taking verb *blaq* 'good' and the complement predicate *m-wah* 'come' have separate arguments. In contrast, in (24b), the complement-taking verb *blaq* 'good' and the complement predicate *niq-un* 'eat' share the first genitive bound pronoun *=mu*.

4.2.2.3 Sharing of TAM information in commentative CCCs

In Squliq Atayal, a commentative complement-taking verb does not share tense, aspect, and mood with its complement. Moreover, we even cannot assign a fixed tense and/or aspect value to the information expressed in such a complement clause. For example, in (23b), the complement clause expresses a past-tense event, in (24a), the coming event has not happened yet, and in (24b), the tense value described in the complement clause can be ambiguous. This can be explained in terms of the notion of grounding; that is, the difference between the two events respectively encoded by BLAQ and its complement clause can be effectively perceived.

Based on Hopper & Thompson (1980:280), the material which supplies the main points of the discourse is known as foregrounded, while what merely assists, amplifies, or comments on it, is referred to as backgrounded. Following this, information given in terms of the word BLAQ is foregrounded, and a complement proposition accommodates a backgrounded interpretation. Noonan (1985:117) has similarly proposed that, since evaluated states or events are normally taken to be real, the proposition embodied in a complement clause construction is backgrounded to any current utterance moment. Note that, in Squliq Atayal, a habitual event still has a factive interpretation, though it does not occur again at the utterance moment. As a result, the relationship between the complement-taking verb (V1) and the complement predicate (V2) can be realised in terms of the following grounding template, as in Table 3:

Table 3: Grounding realisation of BLAQ and its complement proposition

Information conveyed by	BLAQ (V1)	The complement proposition (V2)
Type of grounding	Foregrounded	Backgrounded

4.2.2.4 Sharing of polarity value in commentative CCCs

In contrast with SVCs involving BLAQ, the higher clause of a commentative CCC does not have to share its polarity value with its complement clause, as illustrated in (25):

(25) a. *blaq* *niq-un=nya'* *qu'* *t-mmy-an=mu*.
good[AV] eat-PV=3SG.GEN NOM do-rice-LOCNMLZ=1SG.GEN
'It is good that he ate my pickled pork.'

b. *blaq* *ini'=nya'* *niq-i'* *qu'*
good[AV] NEG=3SG.GEN eat-PV.NEG NOM
t-mmy-an=mu.
do-rice-LOCNMLZ=1SG.GEN
'It is good that he didn't eat my pickled pork.'

c. *ini'=mu* *k-blaq* *niq-un* *qu'*
NEG=1SG.GEN STAT-good[AV] eat-PV NOM
t-mmy-an=mu.
do-rice-LOCNMLZ=1SG.GEN
'It is not good for me that (I) ate my own pickled pork.'

d. *ini'=mu* *k-blaq* *ini'* *niq-i'* *qu'*
NEG=1SG.GEN STAT-good[AV] NEG eat-PV.NEG NOM
t-mmy-an=mu.
do-rice-LOCNMLZ=1SG.GEN
'It is not good for me that (I) didn't eat my own pickled pork.'

4.2.2.5 A discussion of results from an examination of commentative CCCs with BLAQ as V1 using four parameters

Thus far, I have examined the relationship between a complement-taking verb (i.e., BLAQ) and the verbal predicate of a complement clause using four morphosyntactic parameters. Since the two verbs do not need to share argument(s), TAM information, and polarity, the two verbs encode two different events and it is better not to interpret them as a syntactic whole, either. As a result, any commentative CCC can be decomposed into two temporally sequential clauses. For example, (25a) can be paraphrased as (26):

(26) *niq-un=mu* *qu'* *t-mmy-an=nya'* *lga',*
eat-PV=1SG.GEN NOM do-rice-LOCNMLZ=3SG.GEN PART:TOP
blaq *qu'* *in-lung-an=maku'.*
good[AV] NOM PFV-think-LOCNMLZ=1SG.GEN
'After eating his pickled pork, I am/was in a good mood.'

(26) consists of an adverbial clause and a main clause. The two are separated by a composite form made of a particle and a topic marker, *lga'*. The two clauses have a cause-effect relationship as in the English sentence *Eating his pickled pork makes me happy*. In brief, the two events in question do not occur simultaneously; moreover, *blaq* 'good' and *niq-un* 'eat' then have no division of labour in semantic and morphosyntactic representations. What *niq-un* 'eat' heads is **a syntactically complete clause**. Depending on the context, the event it encodes can be expressed not only in an AV form, but also a UV form (cf. Table 2 in section 4.2.2.3). However, note that, since an event of this sort has a factual reading, the voice form of V2 is determined by the thematic role of NOM NP. A semantically transitive verb is encoded in a UV form (e.g., (23b), (24b), and (25)), while a semantically intransitive verb, in an AV form (e.g., (23a) and (24a)). In consequence, the subject of a UV verb refers to an undergoer argument, while that of an AV verb refers to an actor argument. The difference of the subject's thematic role is a

mere side effect of the use of a commentative CCC in Squliq Atayal. This can also be observed from our discussion on the BLAQ UV *qu'* NP construction. There remains a difference: it is the thematic role of the subject in a complement clause that triggers the use of a UV form for V2 in the BLAQ UV *qu'* NP construction.

5 BLAQ UV *qu'* NP construction

In this section, I concentrate on our target construction, i.e., the BLAQ UV *qu'* NP construction. Like commentative CCCs, this construction has an evaluative use. Two points need to be accounted for. First, I need to recognise the structural category of the constructions in question. Second, it is necessary to explain why V2 in the constructions is a UV form.

For the first question, I use the same pattern as our discussion on SVCs and commentative CCCs and examine the formal relationship between juxtaposed verbs. The findings, along with those on SVCs (in section 4.2) and commentative CCCs (in section 4.3), are provided in Table 4:

Table 4: The relationship between the juxtaposed verbs in the BLAQ UV *qu'* NP construction, SVCs, and commentative CCCs in Squliq Atayal: A comparison

Construction type Parameter	SVCs	BLAQ UV *qu'* NP construction	Commentative CCCs
(A) Manifestation of voice in verb sequences	(a) AV AV (b) UV AV	AV UV	(a') AV AV (b') AV UV
(B) Sharing of arguments	Obligatory	Obligatory	Optional
(C) Sharing of TAM information	Obligatory	Obligatory	Depends
(D) Sharing of polarity value	Obligatory	Obligatory	Optional

Among these four parameters, the BLAQ UV *qu'* NP construction and SVCs share three parameters; only the voice manifestation of the BLAQ UV *qu'* NP construction is identical to that of commentative CCCs.

I will argue that the BLAQ UV *qu'* NP construction is syntactically very similar to SVCs, but in terms of function (i.e., the evaluation of something in a realis event), we can see this type of construction is associated with commentative CCCs.

A simple answer to the second question is that the use of a UV form in V2 in the BLAQ UV *qu'* NP construction implies that the event the verb encodes is designated for taking an undergoer participant as the highlighted in contrast to the case of V2 in AV, whereas the event the AV verb encodes requires an actor participant as the highlighted argument.

5.1 AV-UV as the only sequence in BLAQ UV *qu'* NP construction

The juxtaposed verbs in the BLAQ UV *qu'* NP construction are in an AV-UV sequence, as in (27a):

(27) a. *blaq* *niq-un* *qu'* *mami'* *qa'*.
good[AV] eat-PV NOM rice DEM
'The rice tastes good.'

Other possible sequences delivering the same meaning like *something tasting good* as AV-AV in (27b–c) and UV-UV in (27d) are prohibited in Squliq Atayal:

(27) b. * *blaq* *maniq* *qu'* *mami'* *qa'*.
good[AV] eat.AV NOM rice DEM

c. * *blaq* *maniq* *squ'* *mami'* *qa'*.
good[AV] eat.AV OBL rice DEM

d. * *bliq-un* *niq-un* *qu'* *mami'* *qa'*.
good-PV eat-PV NOM rice DEM

5.2 Sharing of arguments in BLAQ UV *qu'* NP construction

Juxtaposed verbs share a nominative-marked argument, and this argument is placed in the rightmost position of a whole verbal sequence:

(28) a. *blaq* *niq-un* *qu'* *mami'* *qa'*.
good[AV] eat-PV NOM rice DEM
'The rice tastes good.'

It cannot be placed in between the verbs, as in (28b):

(28) b. * *blaq* *qu'* *mami'* *qa'* *niq-un*.
good[AV] NOM rice DEM eat-PV

If V1 (BLAQ) carries a bound pronoun, the new sentence is a complete complement clause construction. It is a commentative CCC, as discussed in section 3.2 (cf. (9a)) and 4.2.2.2. Such a sentence is not our concern in this section. See (28c):

(28) c. *blaq=mu* *niq-un* *qu'* *mami'* *qa'*.
good[AV]=1SG.GEN eat-PV NOM rice DEM
'I enjoy tasting the (type of) rice.' (Lit. 'I am in a good mood when tasting the (type of) rice.')

5.3 Sharing of TAM information in BLAQ UV *qu'* NP construction

Juxtaposed verbs in the BLAQ UV *qu'* NP construction share identical TAM information, and a TAM marker is placed in the leftmost position of a whole construction, as in (29a):

(29) a. *musa'* *blaq* *niq-un* *qu'* *mami'* *qa'*.
PROS good[AV] eat-PV NOM rice DEM
'The (type of) rice will taste good.'

Worth noticing is that, in (29a), though the eating event has not happened yet, it is treated as realis, since the speaker cannot assign a value with no basis of habitual experience.

The same argument has been proposed in our discussion on commentative CCCs (see section 4.2.2.3).

However, placing a TAM marker in between *blaq* 'good' and *niq-un* 'eat (PV)' in the constructions is prohibited from this language, as in (29b):

(29)	b.	*	*blaq*	*musa'*	*niq-un*	*qu'*	*mami'*	*qa'*.
			good[AV]	PROS	eat-PV	NOM	rice	DEM

5.4 Sharing of polarity value in BLAQ UV *qu'* NP construction

A negator is placed immediately before V1, which means that its scope is also over V2, as in (30a):

(30)	a.	*ini'*	*k-blaq*	*niq-un*	*qu'*	*mami'*	*qa'*.
		NEG	STAT-good[AV]	eat-PV	NOM	rice	DEM
		'This (type of) rice doesn't taste good.'					

It cannot appear between the two verbs in the BLAQ UV *qu'* NP construction, as in (30b):

(30)	b.	*	*blaq*	*ini'*	*niq-i'*	*qu'*	*mami'*	*qa'*.
			good[AV]	NEG	eat-PV.NEG	NOM	rice	DEM

5.5 A discussion on results from an examination on BLAQ UV *qu'* NP construction via four parameters

Thus far, I have used four parameters to examine the relationship between an evaluative verb and its following UV verb. These results, which have been tabulated in the beginning of the present section (see Table 4), can be analysed in terms of two different viewpoints that determine the structural category of the BLAQ UV *qu'* NP construction.

From one viewpoint, the construction in question can be analysed as an SVC. This argument can be straightforwardly evidenced by three tests. They are: Argument sharing, TAM sharing, and polarity sharing. A TAM marker or a negator is placed in the leftmost position of a whole construction, while a nominative subject, the rightmost position. The same representation occurs in SVCs, but not in commentative CCCs. As a result, like verbal components in SVCs, we can take BLAQ and its following UV verb together as a syntactic whole. Along the same line of SVCs (cf. (12) in section 4.1.2 vs (19a-b) in section 4.1.6), a BLAQ UV *qu'* NP construction like (28a) cannot be paraphrased into a composite of two separate clauses, as in (31a) and (31b):

(31) a. *blaq*.
good[AV]
'It is good (that…)'

b.	*	*niq-un*	*qu'*	*mami'*	*qa'*.
		eat-PV	NOM	rice	DEM

In (31a), the proposition the speaker evaluates is absent, so this clause is not complete from a discourse perspective. (31b) is not a grammatical sentence, either. Paraphrases of

this kind do work in commentative CCCs (cf. (25b) in section 4.2.2.4 vs (26) in section 4.2.2.5). In brief, the BLAQ UV *qu'* NP construction is syntactically equivalent to SVCs.

The other viewpoint is on manifestation of voice in verb sequences. We can see very clearly that the constructions in question may be associated with commentative CCCs because both construction types display an AV-UV sequence in the manifestation of voice. As mentioned, this kind of sequence is impossible for SVCs. Now our target construction appears identical to commentative CCCs. Of course, this is not the reality.

Thus far, based on morphosyntactic representation, the reality is that the BLAQ UV *qu'* NP construction is a fusion of SVCs and commentative CCCs.

But, recall the notion of (non-)simultaneity. Does the conclusion conflict with this notion? If two events respectively encoded by juxtaposed verbs in the same construction occur simultaneously, the whole construction can be an SVC rather than a commentative CCC. In fact, in the BLAQ UV *qu'* NP construction, 'BLAQ' describes a foregrounded state, while UV, a backgrounded event. More precisely, when the speaker uses the construction in question to evaluate something in some event, the evaluating act is performed at the current speech time, whereas the evaluated event along with the participant(s) there is realised as a realis one, no matter it has happened (at the time earlier than the current speech time) or has not happened (at the time later than the current speech time). A grounding distinction occurs between BLAQ and the verbal predicate of BLAQ's complement clause, too (cf. Table 3 in section 4.2.2.3). Namely, the two events encoded by BLAQ and UV in the BLAQ UV *qu'* NP construction do not occur at the same time. But why can the two verbs be interpreted as a syntactic whole?

A possible explanation is that SVCs and commentative CCCs work together in the BLAQ UV *qu'* NP construction. SVCs have syntactic imports on the constructions in question, while commentative CCCs, semantic imports. Since three morphosyntactic parameters show a link between the constructions in question and SVCs, there is no problem with the syntactic imports of SVCs. The question then moves to commentative CCCs. The relationship between a complement-taking verb and its complement clause under the same commentative CCCs can be analogous as a cause-effect relationship. Namely, the speaker performs an evaluative act in terms of a complement-taking verb based on the event encoded in its complement clause. This can further imply some kind of temporally sequential relationship between the act and the event, no matter the event is expressed in an AV or UV complement clause in commentative CCCs. Moreover, the BLAQ UV *qu'* NP construction serves the function of evaluation, too. In other words, in terms of semantics, it is a cause-effect or temporally sequential relationship stemming from commentative CCCs triggering BLAQ and UV in the same constructions as a syntactic whole. As a result, the argument for the BLAQ UV *qu'* NP construction as a fusion of SVCs and commentative CCCs is unproblematic. In terms of SVCs, the target construction is a syntactic whole, or an idiomatic unit. In terms of commentative CCCs, the construction is designated for an evaluative act.

5.6 V2 in UV: Exclusive to undergoer

The other research question of this study is: Why does the second verb in the BLAQ UV *qu'* NP construction appear in the UV form? It appears that thematic role is driving voice selection here, as briefly explained in the beginning of section 5.

Recall the function of the construction in question. It is designated for the speaker or an experiencer to evaluate some participant in some event. Since there is only one slot left for the evaluated subject, there should be a rule of the use of the construction, as shown in (VII):

(VII) One participant (category) evaluated per BLAQ UV *qu'* construction:
In Squliq Atayal, it is impossible to use the same BLAQ UV *qu'* construction to evaluate more than one participant in different categories.

That is, two roles opposite to each other like an actor and an undergoer cannot co-occur as two evaluated targets in the same construction. Otherwise, we will get an ungrammatical sentence as in (32):

(32) *	*blaq*	*niq-un*	*qu'*	*yumin*	*qu'*	*mami'*.
	good[AV]	eat-PV	NOM	PN	NOM	rice

Instead, one thematic role is evaluated at a time. Different roles require different constructions. The BLAQ UV *qu'* NP construction is for an undergoer participant, while constructions like the BAQ AV *qu'* NP construction. (33) displays instances of undergoer being evaluated, while (34) is for the evaluation of an actor participant:

(33) a. *blaq* *niq-un* *qu'* *mami'* *qa'*.
good[AV] eat-PV NOM rice DEM
'The (type of) rice tastes good.'

b. * *blaq* *maniq* *qu'* *mami'* *qa'*.
good[AV] eat.AV NOM rice DEM

c. * *blaq* *maniq* *sa'* *mami'* *qa'*.
good[AV] eat.AV OBL rice DEM

d. * *baq* *niq-un* *qu'* *mami'* *qa'*
can[AV] eat-PV NOM rice DEM

(34) a. *baq* *maniq* *qu'* *yumin.*
can[AV] eat.AV NOM PN
'Yumin is good at eating.'

b. *baq* *m-t-zyaw* *qu'* *yumin.*
can[AV] AV-do-thing NOM PN
'Yumin is good at working.'

c. * *blaq* *maniq/m-t-zyaw* *qu'* *yumin.*
good[AV] eat.AV/AV-do-thing NOM PN

d. * *baq* *niq-un* *qu'* *yumin.*
can[AV] eat-PV NOM PN

e. * *blaq* *niq-un* *qu'* *yumin.*
good[AV] eat-PV NOM PN

f. * *blaq* *niq-un* *na'* *yumin.*
good[AV] eat-PV GEN PN

Highlighting an undergoer drives the second verb to appear in a UV form, as in (33a), but never in an AV form (33b). The evaluated undergoer is nominative-marked (33a), but not oblique-marked (33c). *Baq* 'can' implies the concept of ability, applicable to actor role, but not to undergoer (33d).

In (34), it can be seen that when evaluating an actor, Squliq Atayal speakers use other evaluative verb *baq* 'can' for the purpose of highlighting the actor's considerable ability in exerting his force on other participant(s) (i.e., the undergoer) in an event. The evaluated event is realised in a neutral AV form.[17] In other words, *blaq* 'good' is not recruited in actor being evaluated (34c/34e), nor a UV form is for encoding the event an actor is in (34d/34e). In (34f), since there is no subject, the sentence there is apparently ungrammatical.

Based on this discussion, it is clear that the use of a UV form in the BLAQ UV *qu'* NP construction is thematic-information driven.

6 Conclusion

In the foregoing discussion, I have defined the structural category of the BLAQ UV *qu'* NP construction. This construction is a fusion of SVCs and commentative CCCs. I end this study by summarising the characteristics of this construction, as in (VIII):

(VIII) a. The juxtaposed verbs, BLAQ and its following UV verb, act together as a syntactic whole; this says a link between the constructions in question and SVCs.

b. Both the BLAQ UV *qu'* construction and commentative CCCs have an AV-UV sequence in the manifestation of voice and the two constructions have an identical function, i.e., evaluation.

c. The representation of V2 in a UV form is pragmatically-driven, since the form is exclusive to an evaluated undergoer in a realis event.

[17] In Squliq Atayal, a neutral verbal form, despite its voice type, is used to describe a factual event or state.

References

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 1–68. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. and R.M.W. Dixon. (ed.) 2006. *Serial verb constructions: A cross-linguistic typology*. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Crowley, Terry. 1987. Serial verbs in Paamese. *Studies in Language* 11.1:35–84.

Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. Oxford: Oxford University Press.

Dixon, R. M. W. 2006. Complement clauses and complementation strategies in typological perspective. In *Complementation: A cross-linguistic typology*, ed. by R. M. W. Dixon and Alexandra Y. Aikhenvald, 1–48. Explorations in Linguistic Typology 3. Oxford: Oxford University Press.

Foley, William A. and Mike Olson. 1985. Clausehood and verb serialization. In *Grammar inside and outside the clause: Some approaches to theory from the field*, ed. by Johanna Nichols and Anthony C. Woodbury, 17–60. Cambridge: Cambridge University Press.

Goddard, Cliff and Anna Wierzbicka. 1994. *Semantic and lexical universals: Theory and empirical findings*. Amsterdam and Philadelphia: John Benjamins.

Hare, R. M. 1952. *The language of morals*. Oxford: Clarendon Press.

Hopper, Paul J. 1979. Aspect and foregrounding in discourse. In *Syntax and semantics,Vol. 12*: *Discourse and syntax*, ed. by Talmy Givón, 213-41. New York: Academic Press.

Hopper, Paul J. (ed.). 1982. *Tense-Aspect: Between semantics and pragmatics*. Typological Studies in Language 1. Amsterdam: John Benjamins.

Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56.2:251–299.

Huang, Lillian M. 1993. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364–390.

Li, Paul Jen-kuei. 1980. The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 51.2:349–405.

Li, Paul Jen-kuei. 1985. Linguistic criteria for classifying the Atayalic dialect groups. *Bulletin of the Institute of History and Philology, Academia Sinica* 56.4:699–718.

Li, Paul Jen-kuei. 2000. Formosan languages: The state of the art. In *Austronesian Taiwan: Linguistics, history, ethnology and prehistory*, ed. by David Blundell,

45–67. Berkeley: Phoebe A. Hearst Museum of Anthropology, University of California; Taipei: Shung Ye Museum of Formosan Aborigines.

Lord, Carol. 1973. Serial verbs in transition. *Studies in African Linguistics* 4.3:269–296.

Noonan, Michael. 1985. Complementation. In *Language typology and syntactic description, Volume 2: Complex constructions*, ed. by Timothy Shopen, 42–140. Cambridge: Cambridge University Press.

Rau, Der-Hwa Victoria. 1992. *A grammar of Atayal*. Taipei: Crane Publishing Co., Ltd.

Sebba, Mark. 1987. *The syntax of serial verbs: An investigation into serialisation in Sranan and other languages*. Creole Language Library 2. Amsterdam and Philadelphia: John Benjamins.

Tsuchida, Shigeru. 1980. Linguistic position of Sikikun and Manawyan: Linguistic bases of subgrouping Sqolyeq and Ts'ole dialects in Atayal. Unpublished Ms.

Tsuchida, Shigeru. 1983. Austronesian languages in Taiwan (Formosa). In *Language atlas of the Pacific area*, ed. by S. A. Wurm and Shiro Hattori. Canberra: The Australian National University.

Tsukida, Naomi. 2005. Seediq. *In The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 291–325. Routledge Language Family Series. London: Routledge.

Wierzbicka, Anna. 1972. *Semantic primitives*. Linguistische Forschungen 22. Frankfurt: Athenäum.

Wierzbicka, Anna. 1994. Semantic primitives across languages: A critical review. In *Semantic and lexical universals: Theory and empirical findings*, ed. by Cliff Goddard and Anna Wierzbicka, 445–500. Studies in language companion series 25. Amsterdam and Philadelphia: John Benjamins.

Wierzbicka, Anna. 1996. *Semantics: Primes and universals*. Oxford: Oxford University Press.

Yeh, Maya Yuting. 2013. Event conceptualization and verb classification in Squliq Atayal. PhD dissertation. Taipei: National Taiwan University.

Yeh, Maya Yuting and Shuanfan Huang. 2009. A study of triple verb serialization in four Formosan languages. *Oceanic Linguistics* 48.1:78–110.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391–414.

7 *Nominal aspect in Tsou**

HENRY Y. CHANG

1 Introduction

It is generally held that temporal information is characteristic of verbal categories and normally marked on them. However, recent studies have shown that temporal marking is also possible and productive with nominal categories across many genetically unrelated languages (Nordlinger & Sadler 2004, 2008; Tonhauser 2007, 2008, among others). A grammatical analog in familiar languages like English is that adjectival modifiers which specify the temporal dimension of their head nouns, as in *the late president* and *a frequent flyer*, are paired up with tense/aspect projections in nominal syntactic structures (Cinque 2005). Along this line of thought, this paper investigates nominal aspect in Tsou, with a special focus on the nominal temporal marking particle *nia*.

Before entering into a detailed discussion of Tsou nominal temporal marking, a clarification on the term 'nominal temporal marking' is necessary. In this paper, the term refers to the temporal marking on a plain noun instead of a nominal derived from a verb. In the latter case, temporal marking is much expected and widespread cross-linguistically, given the verbal source of the nominal. In the Formosan language Mayrinax Atayal, for example, a derived nominal can carry mood/aspectual affixes: *paquriq* 'thief' is made up of the irrealis prefix *pa-* and the verb root *quriq* 'steal'; *tinahuq* 'cooked dish' is made up of the perfective infix ‹*in*› and the verb root *tahuq* 'cook' (L. Huang 2002:211, 216).

In the Tsou literature, there is no systematic study of nominal temporal marking; neither is there any serious work done on the semantic functions and syntactic structure of the nominal temporal particle *nia*. Previous studies either leave *nia* undocumented or simply gloss it without any justification. The present paper attempts to fill the gap.

Tsou is an Austronesian language spoken in the southwest highlands of Taiwan with a population of around 4,610 people (up to August 2014). Like most of the other Formosan languages, Tsou is a predicate-initial language—the predicate of the sentence precedes

* This paper makes a substantial revision of H. Chang (2012b) and replaces H. Chang's tense analysis with an alternative aspect analysis. For this significant improvement, I acknowledge the critical input from Jo-Wang Lin and Jackson Sun. I also thank Niina Zhang and Guglielmo Cinque for their helpful suggestions. I am indebted to my principal language consultants Mo'e Peongsi (Mo'o Peongsi), Mo'e Yakumangana (Mo'o Yakumangana), Sayung'e Yulunana (Sayungu Yulunana). Thanks are also due to my research assistants Chiafen Wu, Terry T. Lee, and Maochang Ku for their assistance of various kinds. I am equally grateful to three anonymous reviewers and the editors of this volume for their useful comments. The usual disclaimer applies.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, *157–182*
Asia-Pacific Linguistics, 2015

the associated arguments and adjuncts; functional categories precede the lexical categories they associate with. Grammatical transitivity is marked both on a lexical verb and its preceding realis auxiliary, as illustrated in Table 1.

Table 1: Markers of grammatical transitivity in Tsou (H. Chang 2012a)

	Auxiliary	Lexical Verb
Intransitive (INTR)	*mi-, mo, moh-, moso*	‹*m*›/*Ø*
Transitive (TR)	*i-, os-, oh-*	*-a* (plain transitive, TR) *-i* (TR/High applicative, LA) *-*(*n*)*eni* (Instrumental applicative, IA)

Readers are referred to Zeitoun (2005) and H. Chang (2012a-b, forthcoming) for further information about Tsou basic grammar. Given the fact that this paper is concerned with temporality in Tsou noun phrases, a brief introduction to its temporal marking system and noun phrase structure is provided below.

Let's consider Tsou temporal marking system first. In Tsou, temporal/mood information is encoded on the sentence-initial auxiliary rather than on the lexical verb following it (Zeitoun 1996; Weng 2000; S. Huang & Huang 2003). Auxiliaries distinguish between realis and irrealis mood.[1] In realis mood, the auxiliary additionally differentiates grammatical transitivity. The temporal/mood system is summarised in Table 2 below.

Table 2: The temporal, mood and aspectual system in Tsou[2]

Reality	Realis		Irrealis			
Transitivity	INTR	TR	INTR/TR			
Proximity						
Immediate to speech time	*mi-, mio, mo*	*i-*	Habitual —	Predictive *te-, tena, ta-*	Hypothetical *nte*	Counterfactual —
Remote from speech time	*mo*(*h*)*- moso*	*o*(*h*)*-*	*la-*	—	—	*nto*(*h*)*-, ntoso*

Let's move on to look at Tsou noun phrase structure briefly. As in other Formosan languages, every argument noun phrase is normally preceded by a case marker in Tsou. It is worth noting that in addition to their grammatical functions, Tsou case markers also encode pragmatic and cognitive functions; specifically, they can indicate (i) whether the nominal argument they mark is visible at speech context, (ii) how far the nominal

[1] In S. Huang & Huang (2003), the realis-irrealis distinction is treated as a distinction between a future tense and a nonfuture tense. In this view, Tsou would be a tense language, an account different from ours.

[2] This table is based on Zeitoun (2005:279), with the substitution of the term 'grammatical transitivity' for the term 'voice'. For reasons why 'voice' markers are analysed as markers of grammatical transitivity, we refer to H. Chang (2011).

argument is from the speaker, and (iii) how the nominal argument is perceived by the speaker, as summarised in Table 3 below.

Table 3: The case-marking system in Tsou[3]

Visibility	Proximity/Perceptibility	Ergative/Genitive	Oblique	Absolutive
Visible	proximal	*ta*	*ta*	*'e*
	intermediate	*ta*	*ta*	*si*
	distal	*ta*	*ta*	*ta*
Invisible	witnessed	*to*	*to*	*'o*
	unwitnessed	*no*	*no*	*na*
	perceptible	—	*nca*	*co*
	locative	—	*ne*	—

Within a case-marked nominal constituent, a demonstrative typically immediately follows the noun phrase it refers to, as in (1a-b).[4]

(1) Demonstrative and noun phrase

a. *i-he=n'a* *nookai* *ne hucma* *'e* ***emoo*** ***eni***.
TR.REAL-3PL.ERG=DT move.into yesterday ABS house this
'They just moved into this house yesterday.'

b. *os-'o* *toveuc-i* *si* ***bunuvhu*** ***sico***.
TR.REAL-1SG.ERG pick-LA ABS plum that
'I picked up plums from that tree.

In contrast, a numeral/adjectival modifier precedes the head noun, with a relativiser *ci* coming between the modifier and the modifiee, as in (2a-b).

(2) Numeral/Adjectival modifier and noun phrase

a. *mi-'o* ***aut'ucu*** *to* ***mo*** ***tuyu*** ***ci*** ***feu'u***.
INTR.REAL-1SG.ABS raise[INTR] OBL INTR.REAL three REL pig
'I raise three pigs.'

b. *zou* ***muululu*** ***ci*** ***yaefuefa*** *'o* *mo'o.*
EMP handsome REL young.man ABS PN
'Mo'o is a handsome young man.'

In this regard, a numeral/adjectival modifier patterns with a relative clause. As in (3), a relative clause *mo m'out'ut'u* 'dozed off' precedes its head noun *oko* 'child', with *ci* intervening between them:

(3) Relative clause and noun phrase

i-si *p'-eac'-a* *'o* ***mo*** ***m'out'ut'u*** ***ci*** ***oko***
TR.REAL-3SG.GEN CAUS-stand-TR ABS INTR.REAL doze.off[INTR] GEN child
'He asked the child who dozed off to stand up.'

---

[3] For typographic convenience, this paper substitutes *ng* for the velar nasal, and *'* for the glottal stop.

[4] This paper follows the Leipzig Glossing Rules, with the following amendments: APPL.H, high applicative; COS, change of state; D, deictic; DT, downtoner; EMP, emphatic marker; EVI, evidential; EXP, experiential perfect aspect; HAB, habitual; IA, instrumental applicative; LA, locative applicative; NPST, nonpast; PLN, place name; PN, personal name; REAL, realis; and SUB, subordinator.

The paper is organised as follows. Section 2 explores the semantic functions of *nia*. Section 3 identifies the grammatical status of *nia* and section 4 presents the syntactic structure of *nia*-phrases. Section 5 concludes the paper by discussing the typological and theoretical implications following from the proposed analysis as well as pointing out some directions for future study.

2 The functions of *nia*

In Tsou, nouns can be temporally marked by the morpheme *nia*, which Tung (1964) treats as a lexical item meaning 'passed-away' or 'ancient'. However, the semantic functions of *nia* are much wider than previously thought. In what follows, I summarise the semantic properties of *nia*.

2.1 Discontinuous existence (DE)

The most prominent function of *nia* is to encode an entity that used to exist prior to the utterance time. In (4a), *nia* is associated with a person described by the kinship term *amo* 'father', who passed away prior to the utterance time. Likewise, in (4b), *nia* marks the proper name *beoku* and indicates that he is no longer alive.

(4) Kinship term and proper name

a. *kuhcu* *to* ***nia*** *amo-'u* (*na*) *eni*.
fur.clothing GEN NIA father-1SG.GEN ABS this
'This is my deceased father's fur clothing.'

b. *moso* *la* *nana* *aacni* *yuso* *na* *la-si*
INTR.REAL HAB reportedly always two ABS HAB-3SG.GEN
ke~~u~~pu *no* ***nia*** *beoku*.
bag GEN NIA PN
'It is said that the late Beoku's bags used to be always two (in a pair).'

This is the sense which Tung (1964) glosses as 'passed-away/ancient', a sense comparable to the English temporal adjective *late*, as indicated in the above translations. I label this function as discontinuous existence (DE), which is intended for a state of existence that holds prior to the speech time but discontinues from it.

The DE function of *nia* is also observed with terms of occupations and artifacts, as in (5).

(5) Terms of occupation and artifacts

a. *mi=cu* *kuzo* *'o* ***nia*** *icangaya* *ta* *tapang~~u~~*.[5]
INTR.REAL=COS bad ABS NIA chief GEN PLN
'The chief of the Tapang~~u~~ village passed away.'

b. *i-si* *haf-a* *to* *chumu* *'o* ***nia***
TR.REAL-3SG.ERG bring-TR ERG water ABS NIA

[5] The predicate *kuzo* 'bad' is an euphemism for 'die' in this circumstance.

hiapeoza ne eovokutana.
bridge LOC PLN
'The Eovokutana bridge was washed away by the flood.'

It even applies to temporal expressions, as in (5c). In (5c), the *nia*-phrase refers to a temporal point that passes by before the utterance time:

(5) Temporal expressions s

c. *mi-'o*	*maine'e*	*to*	***nia***	***taseona-si***		*ne*
INTR.REAL-1SG.ABS	return[INTR]	OBL	NIA	morning-3SG		SUB.PST
moso	*meesi*.					
INTR.REAL	rite					

'I went home in the morning on the day of war rite.'

Note that Tung's characterisation does not work for (5b–c), though it holds true of (5a). A washed-away bridge is neither 'dead' nor 'ancient' in the usual sense; a past moment does not refer to its death either. Meanwhile, the entities encoded by *nia* in (5a–c) share the property of change of existence—they used to exist but are gone prior to the speech time.

2.2 Discontinuous identity (DI)

In this function, which I label as discontinuous identity (DI), *nia* refers to an entity that used to hold a certain identity/post but has lost it prior to the utterance time. In (6a), *nia* is concerned with an individual who used to be Mo'o's wife but broke up with him prior to the utterance time; in (6b), *nia* describes an individual who used to be a township chief but has left the post before the speech time.

(6) Terms of relation/occupation

a. *'a*	***nia***	(*la*)	*vcong~~u~~*	*to*	*mo'o*	(*na*)	*taini*.
EVI	NIA	HAB	spouse	GEN	PN	ABS	3SG.ABS

'She is indeed Mo'o's ex-wife.'

b. *'a*	***nia***	(*la*)	*ngocoo*	(*na*)	*taini*.
EVI	NIA	HAB	township.chief	ABS	3SG.ABS

'He is indeed an ex-township chief.'

In this usage, *nia* functions like the English temporal prefix *ex-*. The relevant individuals are still alive; they are neither dead nor ancient. Tung's glossing leaves this function unexplained.

2.3 Discontinuous possession (DPoss)

In this function, *nia* describes an entity that used to belong to some individual but the possessive relation has ceased to exist prior to the speech time. I label this function as

discontinuous possession (DPoss). In (7a), the DPoss applies to a building; in (7b), it applies to an animal.[6]

(7) Artifacts and animals

a. ***nia*** *la* *emoo-'u* (*na*) *eni*.
NIA HAB house-1SG.GEN ABS this
lit. 'This is my former/old house.'
'This house used to be mine.'

b. *tonoi* *'o* ***nia*** *av'u-'u*.
that ABS NIA dog-1SG.GEN
lit. 'That is my former/old dog.'
'That dog used to be mine.'

Semantically, *nia* in this function is roughly equivalent to the English adjectives *former* and/or *old*. Tung's glosses ('ancient/'passed-away') fall short in accounting for this function. The *nia*-phrase does not refer to an ancient building in (7a); neither does it denote a dead dog in (7b).

2.4 Discontinuous appearance (DA)

In this usage, *nia* refers to an entity that used to be in good shape but broke down before the utterance time. I label this function as discontinuous appearance (DA). Note that the DA function usually applies to artifacts and natural kinds, as illustrated in (8a–c) and (9), respectively.

(8) Artifacts

a. *mi=cu* *aemo'~~u~~* *si* ***nia*** *ca'h~~u~~*.
INTR.REAL=COS fall.apart[INTR] ABS NIA chair
'The chair already fell apart.'
Or 'The broken chair fell apart.'

---

[6] There is a family name called *niahosa* in Tsou, which is arguably derived from *nia* plus *hosa* 'village'. In this case, *nia* seems to be in its COP function, meaning literally 'old'—*niahosa* means 'old village'. This analysis receives empirical support from the following examples. In (i), the two words *nia hosa* are kept separate and intended for the reading 'old village'. In contrast, in (ii), they are fused together into a compound and intended for a family name.

(i) *oh* *la* *yon-i* *no* ***nia*** *takupuyan~~u~~* *'o* ***nia***
TR.REAL HAB stay-LA GEN NIA PN ABS NIA
hosa *ne* *sinvi.*
village LOC PLN
'The Takupuyan~~u~~ family used to live at the Sinvi's old village.'

(ii) *na* ***nia*** *atuhcu* *mameoi* *ta* ***niahosa*** *zou* *bania*
D NIA mainly elderly GEN PN EMP PN
na *ongko-si*.
ABS name-3SG.GEN
'As for the Niahosa's main ancestor, his name was Bania.'

According to my language consultants, the Niahosa family was the early settlers of the old village and hence named after it.

b. *i-ta* *tuapuz-a* *si* *mi=cu* *kuzo* *ci*
TR.REAL-3SG.ERG burn-TR ABS INTR.REAL=COS bad REL
nia *ca'h~~u~~.*
NIA chair
'He burned the broken chair.'

c. *mi=cu* *k~~u~~htos~~u~~* *si* ***nia*** *ufi.*
INTR.REAL=COS harden[INTR] ABS NIA rice cake
'The rice cake has become stiff.'

(9) Natural kinds
mi=cu *tmuchumu* *si* ***nia*** *yuho.*
INTR.REAL=COS turn.into.water[INTR] ABS NIA snow
The snow has turned into water.'

Unlike the other functions discussed above, *nia* in this function finds no lexical equivalents in familiar languages like English. In the contexts comparable to those like (8)–(9), English normally does not employ any nominal temporal modifier. It is more natural for English speakers to refer to the artifact in (8a) as 'the chair' instead of 'the former chair' ('the late/ex- chair' is even worse). The same comment also applies to (8b–c). Likewise, English typically refers to the natural object in (9) as 'the snow' rather than 'the former snow'. Another two potential translations, 'the late snow' and 'the ex-snow', are utterly nonsense.

As an anonymous reviewer points out, *nia* may not contribute any semantic import to the sentence it patterns with in this function. The example in (9) is a good case in point. The DA reading may derive from the achievement verb *tmuchumu* 'turn into water' plus the marker of change of state *cu* rather than from *nia*. The function of *nia* may be to resonate with, rather than to convey, the DA sense here. This is exactly the reason why I use the term 'function' rather than 'meaning' in the cross-class description of *nia* and why I adopt a grammatical instead of a lexical approach to *nia*, as will become clearer in the subsequent sections. In the meantime, it should be noted that *nia* is not always "meaningless" in this function. Otherwise, we cannot explain why the *nia*-phrase in (8a) is ambiguous between 'a chair' and 'a broken chair'. On the second reading, *nia* does indeed express a DA sense.

2.5 Discontinuous function (DF)

As in its DA function, *nia* in a discontinuous function (DF) also applies to artifacts and natural kinds. In (10a), the bed is not broken down; instead, it has ceased to function as a bed prior to the speech time; in (10b), the place has changed its function from a tea garden to something else. Likewise, in (10c), the tree may not change its appearance, but it must involve a change of function—a change from a natural object to a building material.

(10) Artifacts, locations, and natural kinds
a. *os-'o* *si-i* *to* *macucuma* *'o* ***nia*** *hopo-'u.*
TR.REAL-1SG.ERG put-LA OBL things ABS NIA bed-1SG.GEN
'I put things on my (former) bed.'

b. *te-'o* *uh* *to* ***nia*** ***'o~'ochia-'u***.
IRR-1SG.ERG get.to[INTR] OBL NIA RED~tea-1SG.GEN
'I will go to a place which used to be my tea garden.'

c. *os-'o* *teai* *teova* *'o* ***nia***
TR.REAL-1SG.ERG make.TR hut ABS NIA
evi *to* *voyu*.
wood GEN PN
'I built a hut from Voyu's wood.'

A similar DF also applies to parts of animal. In (11a), the bearskin has ceased to be the fur-covered skin of a bear, instead, it has been used for a fur coat; in (11b), there is a change from goat flesh to roasted meat/food.

(11) Parts of animals

a. *mi-ta* *maceof~~u~~* *to* ***nia*** *feo'u* *no* *cmoi.*
INTR.REAL-3SG.ABS wear.INTR OBL NIA skin GEN bear
'He wears a fur coat made of bearskin.'

b. *i-ta=cu* *chu-a* *'o* ***nia*** *fou-moat~~u~~'n~~u~~*.
TR.REAL-3SG.ERG=COS roast-TR ABS NIA meat-goat
'He has roasted the goat meat.'

In this function, *nia* indicates that something used to have a typical function but has lost it prior to the utterance time. As in its DA function, *nia* in this function is typologically rare. It does not have a lexical equivalent in familiar languages like English.

2.6 Summary

We have shown that *nia* can involve various states of affairs, including existence, identity, possession, appearance, and function, as summarised in Table 4 below.

Table 4: The functions of *nia*

Function	Reference	Application	English analogs
DE	an entity that used to exist but ceases to be so prior to the speech time	people (proper name), animal parts, artifacts	*late* (for people only), *old* (for artifacts), *last* (for time/event), *former* (for animal parts)
DI	an entity that used to hold a certain identity/post but loses it prior to the speech time	terms of relation/occupation	*ex-*, *former*
DPoss	an entity that used to belong to an individual but ceases to be so prior to the speech time	artifacts, animals	*former*, *old*
DA	an entity that used to be in good shape but ceases to be so prior to the speech time	artifacts, natural kinds	(Unattested)
DF	an entity that used to hold a typical function but loses it prior to the speech time	artifacts, natural kinds, parts of animal	(Unattested)

Despite their apparent interpretive diversity, the above-mentioned five functions of *nia* should have something in common. Among the five functions, the DI is most straightforward. As is well known, a nominal is semantically a predicate indicating its identity. For instance, a wife is a female partner of a male individual. The DI function of *nia* is arguably to terminate the identity of the female partner or her partnership with the male individual prior to the speech time. Similarly, the DPoss function is also predictable. Note that a possessive relation between an owner and an entity must involve the identity of the entity. For example, a dog of mine refers to a dog that bears the identity of being possessed by me. The DPoss function of *nia* is presumably to discontinue the identity of being possessed by me prior to the speech time. In contrast, the DE, DA, and DF functions of *nia* are less obvious. A nominal does not assert existence by itself, though it may implicate it. Neither does a nominal assert appearance nor function by itself. It is likely that like the DPoss function, the DE, DA, and DF functions are derived from the DI function, in conjunction with other elements in the discourse. Specifically, a reading like 'my deceased/late father', as observed with the expression *nia amo'u* in (4a), may involve the following derivations: (i) the nominal *amo* 'father' asserts its identity of being a father of some individual; (ii) the genitive pronoun *'u* 'my' introduces a possessor, relating it to the nominal; (iii) *nia* is responsible for disrupting the identity of the nominal prior to the speech time, and by transitivity, the disruption of the identity gives rise to that of the possessive relation; (iv) the disruption of the identity/relation leads to an inference of the entity described by the nominal as being deceased, as the reading 'he is not my father any more at the present time' is perceived as 'he (my father) is no longer alive'. The same logic may carry easily over to the derivations of the DA and DF functions as well.

It appears that the above five functions have a shared grand function of disrupting the identity of an entity that has held during some time leading up to the speech time, taking the DI function as their common denominator. Taken analytically, the shared grand function can be schematised by two functional features in the grammar, namely, [+discontinuous] and [+past], with the understanding that the speech time is the present time in relation to an entity, hence the qualification 'prior to the speech time' equals to the past. These two functional features are both temporal in nature, thereby pointing to the conclusion that *nia* must be identified as a temporal marking particle. Important questions immediately arise: Does *nia* represent a lexical item, as suggested in the above English translations, or a functional category? If it is a functional category, does it encode aspect or tense? In the next section, I address these questions from a cross-linguistic perspective.

3 The categorical status of *nia*

3.1 *nia* as nominal

In addition to being temporal, *nia* should be nominal. As in the above examples (4)–(11), *nia* occurs exclusively in noun phrases. This is particularly evident in those

examples where the *nia*-phrases are led by case markers.[7] The presence of case markers indicates that *nia* indeed falls within a nominal phrase. Note also that the *nia*-phrase and the head noun cannot be intervened by the relativiser *ci*. Take (8a) and (9a) as examples. The insertion of *ci* between *nia* and the head noun in these examples yields ungrammaticality, as shown in (12a–b).

(12) Artifacts

a. *	*mi=cu*	*aemo'ʉ*	*si*	***nia***	*ci*	*ca'hʉ*.
	INTR.REAL=COS	fall.apart[INTR]	ABS	NIA	REL	chair

b. *	*mi=cu*	*tmuchumu*	*si*	***nia***	*ci*	*yuho*.
	INTR.REAL=COS	turn.into.water[INTR]	ABS	NIA	REL	snow

This rules out the possibility that *nia* occurs either as a verb or an adverbial and the possibility that the *nia*-phrase functions as a relative clause. What it suggests is that the *nia*-phrase is a simple rather than a complex nominal phrase.

Negative evidence also lends support to the nominal analysis. On the one hand, *nia* differs significantly from a lexical verb or an adverbial verb in its distribution. A lexical verb or an adverbial verb can pattern with a clausal temporal/mood auxiliary, as in (13a) and (14a), but *nia* cannot, as in (13b) and (14b).

(13) *nia* vs a lexical verb/an adverbial verb

a.	***mi***-*ta*	*i'vaho*	*m-imo*	*to*	*emi*.
	INTR.REAL-3SG.ABS	again[INTR]	INTR-drink	OBL	wine

'He drinks wine again.' (H. Chang 2009:447)

b. *	***mi***-*ta*	***nia***	*m-imo*	*to*	*emi*.
	INTR.REAL-3SG.ABS	NIA	INTR-drink	OBL	wine

(14) *nia* vs an adverbial verb

a.	***la***-*ta*	*asngʉcʉ*	*baito*	*to*	*tposʉ*.
	HAB-3SG.ABS	often[INTR]	see.INTR	OBL	book

'He reads books often.' (H. Chang 2009:449)

b. *	***la***-*ta*	***nia***	*baito*	*to*	*tposʉ*.
	HAB-3SG.ABS	NIA	see.INTR	OBL	book

---

[7] In (6a–b), repeated below as (ia)–(ib), the *nia*-phrases are not case-marked; instead, *nia* occurs in a phrase that looks like a verbal projection. Nonetheless, a closer look at the structure indicates that the phrase in question functions as a nominal predicate of the sentence.

(i) a.

'a	***nia***	(*la*)	(**ci*)	*vcongʉ*	*to*	*mo'o*	(*na*)	*taini*.
EVI	NIA	HAB	REL	spouse	GEN	PN	ABS	3SG.ABS

'She is indeed Mo'o's ex-wife.'

b.

'a	***nia***	(*la*)	(**ci*)	*ngocoo*	(*na*)	*taini*.
EVI	NIA	HAB	REL	township.chief	ABS	3SG.ABS

'He is indeed an ex-township chief.'

If *nia* serves as an adverbial or adjectival expression modifying the head noun in these sentences, it would be expected that a relativiser *ci* can come between them. This expectation is, however, not borne out. This indirectly supports the nominal predicate analysis.

On the other hand, *nia* cannot be replaced by a clausal auxiliary that performs a similar temporal function. Among the clausal auxiliaries, *moso* is the auxiliary that has the closest function to *nia*. However, *moso* cannot be in the place of *nia*. Compare:

(15) *nia* vs *moso*

a. *'a* ***nia***/*<u>*moso*</u> (*la*) *vcong~~u~~* *to* *mo'o* (*na*) *taini*.
EVI NIA/INTR.REAL HAB spouse GEN PN ABS 3SG.ABS
'She is indeed Mo'o's ex-wife.'

b. *'a* ***nia***/*<u>*moso*</u> (*la*) *ngocoo* (*na*) *taini*.
EVI NIA/INTR.REAL HAB township.chief ABS 3SG.ABS
'He is indeed an ex-township chief.'

It is evident that despite their shared functional property, *nia* and *moso* play a different grammatical role: the former is used for a past discontinuous property/relation in the nominal domain, whereas the latter serves to locate a remote past event/state in the verbal domain.

Another piece of evidence comes from the fact that a verb should be nominalised when it co-occurs with *nia*. As in (16a–b), the verb *mayasvi* 'to perform a war rite' and *esv~~u~~t~~u~~* 'decide' must take the syntactic nominaliser *hia* once it patterns with *nia*.

(16) *nia* and nominalisation

a. *'o* ***nia*** **(**hia**)* *mayasvi* *ne-noaa'o* *o'a* *mo*
D NIA NMLZ perform.rite long.time.ago NEG INTR.REAL
totiski *ta* *hia* *mayasvi* *maitan'e*.
same OBL NMLZ perform.rite today
'The way Mayasvi (war rite) was performed long time ago is not the same as the way it was performed nowadays.'

b. *'o* ***nia*** **(**hia**)* *esv~~u~~t~~u~~* *ne* *moso* *auyu*
D NIA NMLZ decide[INTR] SUB.PST INTR.REAL first
to~~u~~sv~~u~~sv~~u~~t~~u~~, *o'a* *i-to* *hioa*.
discuss[INTR] NEG TR.REAL-1PL.ERG work.TR
'As for the decision made at the last discussion, we did not implement it.'

This follows straightforwardly from the nominal analysis—as a nominal temporal marker, *nia* categorically selects (c-selects) a nominal complement.

Moreover, the nominal analysis accounts for a particular restriction on the distribution of *nia*. As in (17), *nia* cannot go with a temporal expression like *taseona* 'morning' if led by the particle *ne*.

(17) *nia* incompatible with a subordinate clause

* *'a* *mo-'u=cu* *bon~~u~~* ***ne*** ***nia*** ***taseona***.
EVI INTR.REAL-1SG.GEN=COS eat.INTR SUB.PST NIA morning
Intended for 'I have eaten this morning.'

It is not the case that *nia* is incompatible with *taseona*. It was illustrated in (5), repeated below as (18), that *nia* has no difficulty in patterning with *taseona*.

(18) Temporal expressions

mi-'o		*maine'e*	*to*	***nia***	***taseona-si***
INTR.REAL-1SG.ABS		return[INTR]	OBL	NIA	morning-3SG.GEN
ne	*moso*	*meesi*.			
SUB.PST	INTR.REAL	rite.INTR			

'I went home in the morning on the day of rite.'

What is wrong in (17) should be a conflict between the grammatical status of *ne* and that of *taseona*. In Tsou, temporal expressions like *taseona* might surface as a verbal predicate.[8] Thus, they are eligible for taking a clausal temporal/modal auxiliary without any lexical support and heading a subordinate clause, as in (19).

(19) *nia* replaced by *mo*

a.

'a	*mo-'u=cu*	*bonu*	*ne*	***mo***
EVI	INTR.REAL-1SG.ABS=COS	eat.INTR	SUB.PST	INTR.REAL

taseona maitan'e.
morning today
'I have eaten this morning.'

b.

mi-'o	*uh*	*ne*	*tfuya*	*ne*	***mo***	*taseona*.
INTR.REAL-1SG.ABS	get.to[INTR]	LOC	PLN	SUB.PST	INTR.REAL	morning

'I went to Tfuya in the morning.'

As a verbal predicate, *taseona* is ineligible for patterning with the nominal *nia*. This is presumably the reason why (17) is ungrammatical. Note, however, that (18) is different from (17) in that in (18), *taseona* is nominalised, as evidenced by the fact that *taseona* is marked for the impersonal genitive *-si* in this case and the fact that the whole constituent is led by an oblique case marker *to*. Accordingly, *nia* is allowed to appear in this construction. In (17) and (19), the particle *ne* is treated as a subordinator rather than as a locative case marker on grounds that (i) in this function, *ne* heads a clausal modifier instead of a noun phrase (witness the presence of the clausal temporal/modal auxiliary *mo* and the absence of the relativiser *ci* in (19)); (ii) *ne* bears a past reading here, a property missing in its locative function; (iii) unlike its locative function, the subordinate *ne* contrasts with *ho* in temporal interpretation, as in (20), but *ho* does not occur as a case marker at all (see Table 3 for more details).

(20) *ne* vs *ho* =past vs nonpast

a.

'a	*mo-'u=cu*	*bonu*	***ne***	***mo***
EVI	INTR.REAL-1SG.ABS=COS	eat.INTR	SUB.PST	INTR.REAL

taseona maitan'e.
morning today
'I have eaten this morning.'

b.

te-to	*e'ohu*	***ho***	*taseona*.
IRR-1PL.ABS	leave[INTR]	SUB.NPST	morning

'We shall leave tomorrow morning.'

---

[8] Readers are referred to Pan (2010) for a more comprehensive description of temporal expressions in Tsou.

c. *la* *muni* *'o* *teo'ua* ***ho*** *mo* *taseona*.
HAB crow[INTR] ABS chicken SUB.NPST INTR.REAL morning
'Cocks crow in the morning.'

Based on the above observations, it is concluded hat *nia* is a nominal temporal marker, signalling a past discontinuous property/relation in the nominal domain. The next question to ask is: Is *nia* an adjectival/adverbial modifier or a marker of tense/aspect?

3.2 Adjectival/adverbial or tense/aspectual?

In familiar languages like English, there are three ways to encode a past discontinuous property/relation within a nominal. The first way is through a temporal adjectival such as the English *late* in (21a). Alternatively, this notion can be represented by a relative clause that contains either a temporal adverbial such as the English *formerly* in (21b) or a past habitual perfect auxiliary such as the English *used to* in (21c).

(21) Ways of encoding a past change of property/relation
a. the late president (adjectival)
b. the person who was formerly a president (adverbial)
c. the person who used to be a president (tense/aspectual)

In this section, I argue that *nia* is a particle of tense/aspect but not exactly of the third option.

In section 1, I have demonstrated that in Tsou an adjectival modification is done through relativising the adjectival verb, as already shown in (2b), repeated below as (22). In (22), the concept 'a handsome young man' literally means 'a young man who is handsome', as evidenced by the presence of the relativiser *ci* coming between the modifier and the head noun.

(22) Adjectival modifier as a relative clause
zou ***muululu*** ***ci*** ***yaefuefa*** *'o* *mo'o*.
EMP handsome REL young.man ABS PN
'Mo'o is a handsome young man.'

Unlike an adjectival modifier, *nia* and the head noun cannot be intervened by *ci*, as in (23).

(23) *nia* incompatible with *ci*
a. * *kuhcu* *to* ***nia*** *ci* *amo-'u* (*na*) *eni*.
fur.clothing GEN NIA REL father-1SG.GEN ABS this
Intended for 'This is my deceased father's fur clothing.'

b. * *'a* ***nia*** (*la*) *ci* *vcongu* *to* *mo'o* (*na*) *taini*.
EVI NIA HAB REL spouse GEN PN ABS 3SG.ABS
Intended for 'She is indeed Mo'o's ex-wife.'

This suggests that *nia* does not occur as an adjectival modifier like the English *late* and that *nia* is not an expression of the first means.

Meanwhile, *nia* does not surface as an adverbial modifier either. In Tsou, adverbial modifiers that express semantic notions such as frequency, manner, degree, and the like typically surface as verbs—they are inflected for transitivity and occur right after the temporal/modal auxiliary, as in (24).

(24) *nia* not as an adverbial modifier

a. *mi-ta* *i'vaho* *m-imo* *to* *emi.*
INTR.REAL-3SG.ABS again[INTR] INTR-drink OBL wine
'He drinks wine again.' (H. Chang 2009:447)

b. *la-ta* *asng~~u~~c~~u~~* *baito* *to* *tpos~~u~~.*
HAB-3SG.ABS often[INTR] see.INTR OBL book
'He reads books often.' (H. Chang 2009:449)

However, *nia* is morphologically invariant and cannot pattern with a clausal temporal/modal auxiliary, as already discussed in section 3.1. Furthermore, unlike an adverbial modifier, *nia* precedes rather than follows the habitual particle *la*, as already shown in (6), repeated below as (25) (see (24b) for a comparison).

(25) *nia* preceding *la*

a. *'a* ***nia*** *la* *vcong~~u~~* *to* *mo'o* (*na*) *taini.*
EVI NIA HAB spouse GEN PN ABS 3SG.ABS
'She is indeed Mo'o's ex-wife.'

b. *'a* ***nia*** *la* *ngocoo* (*na*) *taini.*
EVI NIA HAB township.chief ABS 3SG.ABS
'He is indeed an ex-township chief.'

Obviously, *nia* does not serve as an adverbial modifier like the English *formerly* and is thus not of the second option.

On the other hand, there are reasons to believe that *nia* encodes tense or aspect. As summed up in section 2, *nia* instantiates two functional features, namely, [+discontinuous] and [+past]. These two functional features are canonical features of tense/aspect. Recall also that *nia* precedes a habitual aspect particle, which suggests that it is situated above an outer/viewpoint aspectual projection. As will become clearer in section 4, this is a position of tense/aspect projection.

In what follows, I shall first determine whether *nia* represents nominal tense or nominal aspect and then return to the issue of its syntactic relation.

3.3 Nominal tense or aspect?

Given the past denotation of *nia*, it is very tempting to analyse it as a marker of past tense. However, the tense analysis leaves the other key functional feature of *nia* [+discontinuous] unaccounted for. A past tense does not assert discontinuity, as evidenced by the fact that a past state of affairs can extend over to the present (Tonhauser 2007:839; Lin 2007:248). Compare:

(26) Arthur was sick, but he is still sick today. (Tonhauser, 2007:839)

(27) John was swimming one hour ago and is still swimming now. (Lin, 2007:248)

In this regard, *nia* does not behave like a past tense marker at all.

Conceptually speaking, a tense analysis is disfavoured as well. As pointed out in Zeitoun (1996), Tsou is not a grammatical tense language. It is by no means reasonable to argue for a nominal tense analysis in a language without a verbal tense. After all, nominal tense is grammatically more marked than its verbal counterpart.[9]

Further evidence from the grammatical behaviour of *nia* also casts doubt on the past-tense analysis. Note that *nia* is not always obligatory. In its DA function, *nia* can be left out without affecting the meaning and grammaticality of the sentences, as in (28a–c).

(28) Optional *nia*

a. *mi=cu* *aemo'ʉ* *si* (***nia***) *ca'hʉ.*
INTR.REAL=COS fall.apart[INTR] ABS NIA chair
'The chair already fell apart.'

b. *i-ta* *tuapuz-a* *si* *mi=cu* *kuzo*
TR.REAL-3SG.ERG burn-TR ABS INTR.REAL=COS bad[INTR]
ci (***nia***) *ca'hʉ.*
REL NIA chair
'He burned the broken chair.'

c. *mi=cu* *kʉhtosʉ* *si* (***nia***) *ufi.*
INTR.REAL=COS harden[INTR] ABS NIA rice.cake
'The rice cake has become stiff.'

This differentiates *nia* from a marker of grammatical tense. As noted in Dahl (1985:24), Dahl & Velupillai (2005:266), and Haude (2011:17), a grammatical tense marking is normally obligatory. Take English, a tense language, as an example. As an uncompromising rule, every English sentence is required to make a grammatical tense distinction. Thus, even a nominal predicate requires an expletive linking verb to carry the tense-marking, as in (29):

(29) Obligatory tense-marking in English
Mary *(<u>was</u>) a good teacher.

From a grammatical system perspective, a nominal tense analysis is also doubtful. Note that the non-past counterpart of *nia* is *tena*, which has been proven to be a marker of irrealis mood instead of tense in the literature (Zeitoun 1996; Weng 2000) (See also Table 2). Compare:[10]

(30) *nia* vs *tena*=Past vs Irrealis

a. *'a* ***nia*** *la* *vcongʉ-'u* (*na*) *taini.* (Past)
EVI NIA HAB spouse-1SG.GEN ABS 3SG.ABS
'She/he indeed is my ex-wife/-husband.'

---

[9] As a matter of fact, it is questionable whether nominal tense really exists in natural language, as addressed in Tonhauser (2007, 2008).

[10] In the meantime, note that *nia* can co-occur with an irrealis auxiliary in the clausal domain, as will be exemplified in (49) in section 5. This also dismisses the analysis of *nia* as a past tense marker.

b. *zou* *taini* *'o* ***tena-'u*** *vcong~~u~~.* (Irrealis)
EMP 3SG ABS IRR-1SG.GEN spouse
'He/She is my fiancé/fiancée.'

(31) *nia* vs *tena*=Past vs Irrealis

a. *'o* ***nia*** *la* *kingatu* *zou* *mo'o.* (Past)
D NIA HAB chief EMP PN
'Mo'o is the ex-chief.

b. *'o* ***tena*** *la* *kingatu* *zou* *mo'o.* (Irrealis)
D IRR HAB chief EMP PN
'Mo'o is the chief-to-be.'

Besides, a tense marker typically occurs as a bound morpheme (Bybee et al. 1994). However, *nia* is a free particle. This is evidenced by the fact that *nia* and the head noun can be intervened by a free habitual particle *la*, as already illustrated in (30a) and (31a) above.

Accordingly, it is concluded that *nia* does not encode nominal tense. We are now left with the last option—*nia* as a marker of nominal aspect. In what follows, I shall argue that an aspect analysis is indeed on the right track.

Let's first examine one by one how an aspect analysis copes with the challenges that trouble a tense analysis. First, an aspect analysis faces no conceptual problem. As in other Formosan languages, aspect is attested in the Tsou verbal domain (Zeitoun 1996; Weng 2000). The realisation of nominal aspect in the language would be no surprise.

Second, an aspect analysis correctly predicts the [+discontinuous] feature of *nia*. It has been observed in the literature that continuity/discontinuity is an essential property of aspect (Comrie 1976; Dahl 1985; Chung & Timberlake 1985; Yeh 1996; Lin 2007, among others).

Also, an aspect analysis evades the obligatory requirement problem. Unlike a grammatical tense, a grammatical aspect is not always required in the surface syntax.[11] A well-known contrast is that in English, a completive aspect can be left unmarked, as in the a-example of (32), whereas its tense counterpart is obligatory throughout (32a–c).

(32) a. Eva **closed** the door. (Klein 2009:40)

b. Eva **has closed** the door.

c. Eva **had closed** the door.

The aspect analysis is consistent with the fact that *nia* can be optional.

Finally, let's look at the challenge of morphological reduction. Unlike a grammatical tense marker, a grammatical aspect marker can be either free or bound. In English, for instance, the progressive aspect form is periphrastic, while the past tense form is suffixal. The aspect analysis is compatible with the fact that *nia* is morphologically free.

Readers may ask how the aspect analysis copes the problem concerning the past denotation of *nia*, a property which is generally believed to be associated with tense

[11] This contrasts with a lexical aspect, which in some languages such as Russian may be obligatory, as pointed out by an anonymous reviewer.

rather than aspect. In the next section, I propose an experiential perfect analysis, which I believe can tackle the problem nicely.

3.4 Perfect *nia*

Perfect is of the aspect category, not of the tense category, as indicated by the fact that a perfect can be further specified by a tense, hence the name past, present, and future perfect. While a perfect aspect is composed of several subtypes, what concerns us here is the experiential perfect aspect, as defined in (33) and illustrated in (34) below.[12]

(33) Experiential perfect (Based on Comrie 1976:58)
The experiential perfect indicates that a given situation has held at least once during some time leading up to a reference time.

(34) Bill has been to America. (Comrie 1976:59)

In this view, an experiential perfect must refer to a state of affairs holding prior to a reference time but discontinuing from it. If the reference time is the present, the state of affairs asserted by an experiential perfect will be anchored in the past. This is exactly the situation with (34). In uttering (34) *Bill has been to America*, the speaker asserts that (i) Bill visited America some time in the past; (ii) Bill is not in America at the present time.

The same observation holds of the experiential perfect aspect *guo* in Mandarin Chinese. As noted in Iljic (1990), Yeh (1996), and Lin (2007), *guo* indicates that the situation it associates with must take place prior to a reference time but does not hold up to it. Accordingly, sentences like (35) are infelicitous since there arises a semantic contradiction: The continuation of the swimming event throughout the present asserted by the progressive aspect marker *zai* in the second proposition contradicts the discontinuity from the present asserted by the experiential perfect *guo* in the first proposition.

(35) Mandarin (Lin 2007:248; glosses mine)

*	約翰	一個	小時	前	就	游過	泳
*	Yuehan	yi-ge	xiaoshi	qian	jiu	you-guo	yong,
	John	one-CLF	hour	ago	JIU	swim-EXP	swim
	現在	還 在	游.				
	xianzai	hai zai	you.				
	now	still PROG	swim				

'John swam one hour ago and now is still swimming.'

In other words, an experiential perfect carry two functional features at a time, namely, [+discontinuous] and [+anterior].

The first functional feature of *nia* matches that of an experiential perfect. The other functional feature [+past] of *nia* can be subsumed under the functional feature [+anterior]

---

[12] Despite its reference to the past, a past perfect is irrelevant here. According to Comrie (1976:53), a past perfect, as in *She had fell asleep when the phone rang last night*, relates a past state (*she fell asleep*) to an earlier situation (*the phone rang*); in other words, both situations in a past perfect are temporally located in the past. Note, however, that *nia* has a relevance at the present time.

of an experiential perfect in that [+past] must be [+anterior]: A past state of affairs must hold prior to a specific reference time—the present time. In light of the shared properties, I propose that *nia* be analysed as a marker of experiential perfect in the nominal domain, as analogous to the English *has been* and the Mandarin *-guo* in the verbal domain (with their morphological difference aside). Just like its verbal counterparts, the nominal experiential *nia* asserts a discontinuity of state of affairs from the present in the nominal domain. Note, however, that unlike an experiential aspect marker in the verbal domain, *nia* always sets its reference time at the present. Thus, *nia* realises a present perfect but not a past or future perfect.

4 The syntax of *nia*

This section discusses the syntactic status/position of *nia* and its relation to the head noun.

4.1 *nia* as a syntactic head

We have established in the foregoing sections that *nia* represents an experiential perfect aspect. In this section, I argue that *nia* occurs as a syntactic head of an experiential perfect project (abbreviated as ExpP).

Evidence for the head analysis of *nia* comes from the fact that *nia* can attract the aspectual down-toning clitic *=n'a*, as in (36).[13]

(36) *nia* as a clitic host

mi=cu	*aemo'ʉ*	*'o*	***nia=n'a***	*imucu*.
INTR.REAL=COS	fall.apart[INTR]	ABS	EXP=DT	PLN

'The Imucu people with a small population have scattered around.'

In this respect, *nia* behaves like a typical syntactic head such as a temporal/modal auxiliary (arguably representing T), as in (37a) or a causative verb (representing *v*), as in (37b).

(37) *=n'a* and its syntactic heads

a. ***te-ko=n'a*** *tuocos-neni* *a'o* *no*
IRR-2SG.ERG=DT ask-IA 1SG.ABS OBL
huphina-si *pohe-taini*.
price-3SG.GEN corn-3SG.GEN
'Would you please ask the price of his corn for me?'

b. ***poa=n'a*** *peteonav'-eni* *to* *av'u* *'o* *fuzu*.
CAUS=DT chase.for a long time-IA OBL dog ABS wild.pig
'They kind of sent dogs to chase the wild pig for a long time.'

Therefore, I argue that *nia* occurs as a syntactic head. Taking its functions into the picture, I further argue that *nia* heads an experiential perfect projection (written as ExpP).

---

[13] The aspectual downtoner *=n'a*, which originates as an aspectual adverbial meaning 'still', serves to decrease the quantity of the people here.

4.2 *nia* above HabP

In terms of linear word order, *nia* precedes the habitual aspect particle *la*, which in turn precedes the head noun, as already exemplified in (25a–b), repeated below as (38a–b).

(38) *nia* preceding *la*

a. *'a* ***nia*** *la* *vcong~~u~~* *to* *mo'o* (*na*) *taini*.
EVI EXP HAB spouse GEN PN ABS 3SG.ABS
'She is indeed Mo'o's ex-wife.'

b. *'a* ***nia*** *la* *ngocoo* (*na*) *taini*.
EVI EXP HAB township.chief ABS 3SG.ABS
'He is indeed an ex-township chief.'

This means that *nia* is merged above *la* and the head noun syntactically, given that Tsou is predicate-initial and that it is a configurational language. On the assumption that *la* heads a habitual aspect projection (labelled as HabP) and that *nia* heads an ExpP, the *nia*-phrases in sentences like (38a–b) can be graphically represented as follows:

(39) [$_{\text{ExpP}}$ *nia* [$_{\text{HabP}}$ *la* [$_{\text{NP}}$ vcong~~u~~/ngooco]]]

Note that *nia* and *la* combine directly with the head nouns without any relativiser in (40). This brings us back to the observation made in section 3.2 that the relation of *nia* to its head noun is neither of adjectival modification nor of relativisation. It seems that like a perfect auxiliary in the verbal domain (e.g., *have*-perfect in English), *nia* specifies the temporal aspect of the head noun in a c-commanding configuration. This amounts to the conclusion that *nia* is periphrastic.

Incidentally, the syntactic representation in (40) accords with what has been known as the T/A/A-architecture, where a perfect aspect tops all of the other viewpoint aspects (including a habitual aspect), as diagrammed below (Alexiadou et al. 2003:x) (the symbol '>' indicates 'dominate').

(40) Tense > Perfect > Aspect > VP

4.3 *nia* under D

Attentive readers might have noticed that *nia*-phrases can be marked by case markers of various kinds, as summed up in the following examples:

(41) *nia* and case markers

a. *i-ta=cu* *chu-a* ***'o*** ***nia*** *fou-moat~~u~~'n~~u~~*.
TR.REAL-3SG.ERG=COS roast-TR ABS EXP meat-goat
'He has roasted the goat meat.'

b. *kuhcu* ***to*** ***nia*** *amo-'u* (*na*) *eni*.
fur.clothing GEN EXP father-1SG.GEN ABS this
'This is my deceased father's fur clothing.'

c. *mi-ta* *maceof~~u~~* ***to*** ***nia*** *feo'u* *no* *cmoi*.
INTR.REAL-3SG.ABS wear[INTR] OBL EXP fur GEN bear
'He wears bearskin clothing.'

Given that Tsou case markers are deictic in nature, indicating the visibility, proximity, and perceptibility of the noun phrase they mark with respect to the spatial/cognitive deixis of the speaker (see Table 3), it is very likely that they contain a deictic ingredient, which heads a deictic phrase (DP).[14] Apart from the semantic argument, there is an additional piece of syntactic evidence in favour of the DP-analysis. A case-marked noun phrase can be replaced by a demonstrative pronoun, with or without the case-marking by a particular case marker *na*, as in (42).

(42) Demonstratives' replacement of case-marked NPs

a. *i-he=n'a* *nookai* *ne hucma* *'e* ***emoo*** ***eni***.
TR.REAL-3PL.ERG=DT move.into[INTR] yesterday ABS house this
'They just moved into this house yesterday.'

a'. *i-he=n'a* *nookai* *ne hucma* ***(na) eni***.
TR.REAL-3PL.ERG=DT move.into[INTR] yesterday ABS this
'They just moved into this yesterday.'

b. *os-'o* *toveuc-i* *si* ***bunuvhu*** ***sico***.
TR.REAL-1SG.ERG pick-LA ABS plum that
'I picked up plums from that tree.'

b'. *os-'o* *toveuc-i* ***(na) sico***.
TR.REAL-1SG.ERG pick-LA ABS that
'I picked up plums from that.'

If a demonstrative pronoun represents a DP, the constituent it substitutes for should be a DP as well.

In the DP-analysis, *nia* should be situated under the D head, since *nia* follows, rather than precedes the case marker. Under this view, a DP containing *nia* can be schematised as follows.

(43) $[_{DP}$ D $[_{PerfP}$ *nia* …

This syntactic representation becomes more graphic in constructions where the DP containing *nia* occurs as a topic and is led by a deictic particle homophonous to a case marker. As in (44a–b), *nia* follows the deictic particle *'o* and *na*, which presumably originate as an absolutive case marker (compare Table 3).

(44) *nia*-phrase as a Topic

a. ***'o*** *nia* *la* *kingatu* *zou* *mo'o*. (Past)
D EXP HAB chief EMP PN
'Mo'o is the ex-chief.'

b. ***na*** *nia* *atuhcu* *mameoi* *ta* *niahosa* *zou*
D EXP mainly elderly GEN PN EMP
bania *na* *ongko-si*.
PN ABS name-3SG.GEN
'As for the Niahosa's main ancestor, his name was Bania.'

[14] It should be noted that the *D* stands for *deictic* rather than the usual *determiner* here.

It is noteworthy that *nia* has nothing to do with referentiality and hence does not occupy a determiner position. This is evidenced by the fact that in addition to definite DPs, *nia* can also appear in indefinite ones, as in (45a–b).

(45) *nia* and indefinites

a. *pan* *to* ***nia*** *la* *huyu* *no* *fuzu* *tan'e*.
there OBL EXP HAB trail GEN wild.pig here
'There used to be trails of wild pigs here.'

b. *pan* *to* ***nia*** *ca'h~~u~~* *tan'e*.
there OBL EXP chair here
'There used to be a(n) old/broken chair here.'

It becomes obvious that the experiential perfect *nia* is merged above a habitual projection and under a deictic phrase.

5 Conclusion

We have shown that *nia* occurs as a periphrastic auxiliary of experiential perfect aspect in the nominal domain, indicating a discontinuous identity of a nominal from the present time within a simple nominal phrase and that *nia* heads an ExpP, which is merged above a HabP and an NP but in the complement of a deictic head D. These findings have far-reaching typological and theoretical implications.

First, the findings lead to the conclusion that Tsou is a language with nominal aspect. In this regard, Tsou differs typologically from familiar languages like English. English does also have expressions of experiential perfect aspect in the nominal domain. Throughout the paper, we have seen a bunch of English adjectival temporal expressions (e.g., *former, late, old, last*, etc.) which translate Tsou *nia*-phrases. From a cartographic viewpoint (Cinque 1999, 2006), these adjectival modifiers may project a nominal ExpP. But what really distinguishes English from Tsou lies in the grammatical realisation of their nominal experiential perfect. In Tsou, *nia* has developed into an all-purpose grammatical auxiliary of experiential perfect in the nominal domain, whereas no such element is attested in English. This explains (i) why *nia* may appear as a dummy like expletive, e.g., in its DA function, where its appearance makes no semantic contribution, and (ii) why *nia* finds no comparable equivalents in English in some cases (e.g., in its DA and DF functions). In the Formosan literature, there has been no report of nominal aspect thus far. This article thus represents the very first attempt of its kind. It is desirable to investigate other Formosan languages and look into the question of whether they behave like Tsou in exhibiting nominal aspect.

Second, there has been a heated debate in the literature on whether nominal tense is attested in natural languages. Nordlinger & Sadler (2004, 2008) argue for a positive answer, while Tonhauser (2007, 2008) argues for a negative response. Our findings seem to be in favour of Tonhauser's position. With its past denotation, *nia* looks like a nominal marker of past tense at first glance. However, along with other facts, its discontinuous property renders the past-tense analysis untenable.

The existence of nominal aspect may revolutionise our understanding of grammatical aspect. It challenges the generally held view that grammatical aspect applies only to

verbal categories. As a matter of fact, a nominal can entertain grammatical aspect as well. More specifically, the periphrastic *nia* specifies the perfect aspect of the head noun in much the same way as the periphrastic perfect *have* specifies the temporal aspect of the head verb in English. In addition, recall that the temporal auxiliaries *tena* and *la* that apply to verbs are also realised in the nominal domain in a periphrastic manner, a fact that further confirms the existence of nominal aspect. An anonymous reviewer raised an important question whether *nia* is really discontinuous. He gave a comparable English sentence to illustrate his concern. As in (46), it seems that the *nia*-like experiential perfect modifier *former* may be continuous.

(46) His former wife is actually his current wife.

Note, however, that as the anonymous reviewer admits himself, there must be a disruption of the previous relation, i.e., a divorce, occurring between the two times of a husband-and-wife relation in (47). This amounts to saying that the first state of affairs asserted by *former* turns out to be still discontinuous. The same observation holds true of *nia* as well. Meanwhile, the same observation can carry over to sentences of paradox where a nominal indicating a future state co-occurs with another nominal indicating a past state. For instance:

(47) Her fiance is actually her ex-husband.

Again, there are two times of a husband-and-wife relation involving here and there occurs a discontinuity of the relation in between. In this sense, the effect of a nominal experiential perfect particle under discussion is restricted to the nominal phrase where it appears—its scope of modification does not go beyond the nominal domain. Borrowing Nordlinger & Sadler's (2004) classification of nominal tense, the experiential perfect aspect in question can be labelled as an independent nominal aspect in the sense that it is independent either of the clausal aspect or the aspect specified in another nominal. This characterisation accounts for an additional fact in connection with *nia*: apart from a realis auxiliary, a *nia*-phrase can also go either with an irrealis auxiliary in the clausal domain. Compare:

(48) *nia* as an independent nominal aspect

a. *mi=cu* *aemo'~~u~~* *si* ***nia*** *ca'h~~u~~.*
INTR.REAL=COS fall.apart[INTR] ABS EXP chair
'The chair already fell apart.'

b. *mi=cu* *k~~u~~htos~~u~~* *si* ***nia*** *ufi.*
INTR.REAL=COS harden[INTR] ABS EXP rice.cake
'The rice cake has become stiff.'

(49) *nia* as an independent nominal aspect

a. ***teav'a*** *payo'-a* *'o* ***nia*** *hia* *to~~u~~sv~~u~~sv~~u~~t~~u~~.*
NEG.IMP[TR] forget-TR ABS EXP NMLZ discuss
'Don't forget the resolution of the last meeting.'

b. ***ta-'u*** *tupuz-a* *si* *mi=cu* *kuzo* *ci* ***nia*** *ca'h~~u~~.*
IRR-1SG.ERG burn-TR ABS INTR.REAL=COS bad[INTR] REL EXP chair
'I will burn the broken chair.'

In this respect, *nia* contrasts significantly with the subordinators *ne/ho*. As noted in Pan (2010), *ne* encodes past but *ho* nonpast, as exemplified in (50a–b):

(50) *ne* vs *ho* = past vs nonpast

a. *mi-'o* *uh* *ne* *maibay~~u~~* ***ne*** (*mo*) *taseona.*
INTR.REAL-1SG.ABS get.to[INTR] LOC PLN SUB.PST INTR.REAL morning
'I went to Chiayi this morning.'

b. *ta-'u* *uh* *ne* *maibay~~u~~* ***ho*** (*mo*) *taseona.*
IRR-1SG.ABS get.to[INTR] LOC PLN SUB.NPST INTR.REAL morning
'I will go to Chiayi in the morning.'

In the spirit of Nordlinger & Sadler (2004), *ne/ho* can be treated as markers of propositional tense/aspect in the sense that they are interpreted in the propositional level. Compare the ungrammatical (51a-b) indicated below with their grammatical counterparts in (50a-b). It is evident that *ne* is required to co-occur with a realis auxiliary but *ho* with an irrealis auxiliary.

(51) Temporal concord constraint

a. * ***ta-'u*** *uh* *ne* *maibay~~u~~* ***ne*** (*mo*) *taseona*.
IRR-1SG.ABS get.to[INTR] LOC PLN SUB.PST INTR.REAL morning

b. * ***mi-'o*** *uh* *ne* *maibay~~u~~* ***ho***
INTR.REAL-1SG.ABS get.to[INTR] LOC PLN SUB.NPST
(*mo*) *taseona*.
INTR.REAL morning

The independent-propositional temporal distinction is an important issue and deserves a more in-depth investigation. I leave it for future study.

Another issue I leave open is concerned with the diachronic source of *nia*. Paul Li (pers. comm., 21 January 2015) suggests that *nia* can be decomposed into *ni-* plus *-a*, with *ni-* as a reflex of the proto-Austronesian perfective marker *‹*in*›*. This is an interesting suggestion, but short of direct empirical support. On the one hand, *ni-* is not attested elsewhere in the language, although it is indeed found in the so-called Tsouic languages Kanakanavu and Saaroa (H. Chang 2006). On the other hand, there is no reasonable account for the occurrence of *-a*, which sounds like the transitive suffix in the language, in this context.

References

Alexiadou, Artemis, Monika Rathert and Arnim von Stechow (eds.). 2003. *Perfect explorations*. Interface Explorations 2. Berlin: Mouton de Gruyter.

Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Chang, Henry Y. 2009. Adverbial verbs and adverbial compounds in Tsou: A syntactic analysis. *Oceanic Linguistics* 48.2:439–476.

Chang, Henry Y. 2011. Tsou case marker revisited: Visibility, proximity and beyond. *Diversity of languages: Papers in honor of Professor Feng-fu Tsao on the occasion of his retirement*, eds. by Raung-fu Chung et al., 91–123. Taipei: Crane Publishing Co., Ltd.

Chang, Henry Y. 2012a. A phase-based account of extraction asymmetries in Tsou and Chinese. Paper presented at the 20th Annual Conference of the International Association on Chinese Linguistics (IACL-20). 29–31 August. Hong Kong: The Hong Kong Polytechnic University.

Chang, Henry Y. 2012b. Nominal tense in Tsou: *Nia* and its syntax/semantics. *UCLA working papers in linguistics* 17:43–58, ed. by Thomas Graf, Denis Paperno, Anna Szabolcsi, and Jos Tellings (also available at semanticsarchive.net).

Chang, Henry Y. Forthcoming. Tsou. *Encyclopedia of Chinese language and linguistics*. Leiden: Brill.

Chang, Melody Ya-yin. 2002. Nominalization in Tsou. *Language and Linguistics* 3.2:335–348.

Chung, Sandra and Alan Timberlake. 1985. Tense, aspect, and mood. *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon*, ed. by Timothy Shopen, 202–258. Cambridge: Cambridge University Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford Studies in Comparative Syntax. New York: Oxford University Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36:315–332.

Cinque, Guglielmo. 2006. *The cartography of syntactic structures, volume 4: Restructuring and functional heads*. Oxford studies in comparative syntax. Oxford; New York: Oxford University Press.

Cinque, Guglielmo. 2011. A (partial) map of nominal functional structure. Invited talk at the Mini-workshop on Syntax. Hsinchu: National Tsing Hua University.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Comrie, Bernard. 1985. *Tense*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Basil Blackwell.

Dahl, Östen and Viveka Velupillai. 2005. Tense and Aspect. *The world atlas of language structures*, eds. by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie, 266–281. Oxford: Oxford University Press.

Huang, Huei-ju. 2010. The syntax and pragmatics of clausal constituents in Tsou discourse. PhD dissertation. Taipei: National Taiwan University.

Huang, Lillian M. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3.2:197–225.

Haude, Katharina. 2011. Tense marking on dependent nominals in Movima. *Tense across languages*, eds. by Renate Musan and Monika Rathert, 189–206. Linguistische Arbeiten 541. Berlin and Boston: Mouton de Gruyter.

Huang, Shuanfan and Huei-ju Huang. 2003. On the status of reality marking in Tsou. *Taiwan Journal of Linguistics* 1.2:1–33.

Iljic, Robert. 1990. The verbal suffix *-guo* in Mandarin Chinese and the notion of recurrence. *Lingua* 81.4:301–326.

Klein, Wolfgang. 1994. *Time in language*. London: Routledge.

Klein, Wolfgang. 2009. How time is encoded. *The expression of time*, eds. by Wolfgang Klein and Ping Li, 39–81. Berlin and New York: Mouton de Gruyter.

Lin, Jo-Wang. 2007. Predicate restriction, discontinuity property, and the meaning of the perfective marker *guo* in Mandarin Chinese. *Journal of East Asian Linguistics* 16.3:237–257.

Nordlinger, Rachel and Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80.4:776–806.

Nordlinger, Rachel and Louisa Sadler. 2008. When is a temporal marker not a tense? Reply to Tonhauser 2007. *Language* 84.2:325–331.

Pan, Chia-jung. 2010. *The grammatical realization of temporal expressions in Tsou*. LINCOM Studies in Austronesian Linguistics 7. Muenchen: Lincom Europa.

Tonhauser, Judith. 2007. Nominal tense? The meaning of Guaraní nominal temporal markers. *Language* 83.4:831–869.

Tonhauser, Judith. 2008. Defining crosslinguistic categories: The case of nominal tense (Reply to Nordlinger and Sadler). *Language* 84.2:332–342.

Tung, T'ung-ho, with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng, and Margaret M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publications 48. Taipei: Academia Sinica. Reprinted in 1991 with notes and errata prepared by Paul Jen-kuei Li.

Weng, Cui-xia. 2000. zouyu he shaoyu de shi, tai, mao xitong zhi bijiao yu yanjiu [A contrastive study of tense, mood and aspect systems in Tsou and Thao]. MA thesis. Chiayi: National Chung University. [In Chinese]

Yeh, Meng. 1996. An analysis of the experiential *guo*$_{exp}$ in Mandarin: A temporal quantifier. *Journal of East Asian Linguistics* 5.2:151–182.

Zeitoun, Elizabeth. 1996. The Tsou temporal, aspectual and modal system revisited. *Bulletin of the Institute of History and Philology, Academia Sinica* 67.3:503–532.

Zeitoun, Elizabeth. 2005. Tsou. *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 259–290. London and New York: Routledge.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual, and modal systems of some Formosan languages: A typological perspective. *Oceanic linguistics* 35.1:21–56.

8 *Outside-in or inside-out functional uncertainty? An LFG analysis on relativisation in Jianshi Squliq Atayal*[*]

ADLAY KUN-LONG LIU

1 Introduction

Atayal, spoken in northern Taiwan, is one of the Austronesian languages. From the perspective of historical linguistics, the Atayalic branch descends either from the highest level of the Austronesian family—Proto Austronesian (PAn) (Blust 1977, 1995, 1999), or from Nuclear Austronesian, one of the four daughters of PAn (Ross 2009). This language includes two dialects, Squliq and C'uli', the former of which is more homogeneous while the latter of which preserves more archaic phonological and morphological features (Li 1985, 1996; Rau 2004). The target language of this paper, Jianshi Atayal, is one of the regional variants of the Squliq dialect. It is spoken in Jianshi Township, Hsinchu County.

Both Squliq and C'uli' Atayal are predicate-initial, with the rigid word order Verb-Object/Oblique-Subject. In addition, like most of the other Formosan languages, the main verb in a clause must agree with the thematic role of the subject, forming a multiple voice system composed of four voices—actor voice (AV), patient voice (PV), locative voice (LV) and circumstantial voice (CV).[1] PV, LV, and CV are also collectively

[*] The data of Squliq Atayal presented here are mostly from my field notes. I am greatly indebted to my informants, Yakaw Bowtu (male, born in 1944), Umas Hakaw (female, born in 1950), Masa Pisuy (female, born in 1942), Kumay Silan (male, born in 1935), and Leci Mbin Tana (female, born in 1952), for their generous help. My sincere thanks also go to the anonymous reviewers of this paper and the editors of this book. Their constructive comments and suggestions were greatly helpful in the revision process.

[1] LV and CV are analysed as applicatives in Chen (2007). Whether her analysis holds is beyond the scope of this paper. Thus, we just follow the convention established by Egerod (1965, 1966, 1980, 1999), Huang (1991, 1993, 1995b), and Rau (1992) and continue calling the verbal markers "voices". In addition, following Huang (1993), this paper adopts the term "circumstantial voice" to replace "instrumental/beneficiary voice".

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 183–210
Asia-Pacific Linguistics, 2015

called non-actor voice (NAV).[2] The above two morphosyntactic properties can be demonstrated in the following examples of Jianshi Atayal from Liu (2004a:27) (boldface added):

(1) ***m**-aniq* *qulih* *qu'* *tali'*.[3]
AV-eat fish NOM Tali
'Tali eats fish.'

(2) *niq-**un*** *na'* *tali'* *qu'* *qulih* *qasa.*
eat-LV OBL Tali NOM fish that
'That fish is eaten by Tali.'

(3) *niq-**an*** *na'* *tali'* *qulih* *qu'* *ngasal* *qasa.*
eat-LV OBL Tali fish NOM house that
'That house is the place where Tali eats fish.'

(4) ***s**-qaniq* *na'* *tali'* *qulih* *qu'* *qway.*
CV-eat OBL Tali fish NOM chopsticks
'The chopsticks were used by Tali to eat fish.'

The grammatical subjects in the above sentences—*tali'*, *qulih qasa* 'that fish', *ngasal qasa* 'that house', and *qway* 'chopsticks'—are located in sentence-final position, preceded by the omissible nominative case marker *qu'* while other arguments intervene between the subject and the sentence-initial main verb. These sentence-initial main verbs are affixed with voice markers *m-*, *-un*, *-an*, and *s-*, which vary with the thematic roles of grammatical subjects, that is, an actor, a patient, a locative, and an instrumental respectively.

This paper delves into the issue on what syntactic mechanisms are behind syntactic computation in Jianshi Atayal. As will be introduced soon in the literature review, this language exhibits two kinds of relative clauses (RCs), externally-headed relative clauses (EHRCs) and internally-headed relative clauses (IHRCs). According to Lexical-Functional Grammar (LFG), there are two possible syntactic mechanisms behind RCs: One is the outside-in functional uncertainty and the other is inside-out functional uncertainty. Do EHRCs and IHRCs employ the same mechanism? Or do they employ different mechanisms? The answers will be given in the subsequent discussions.

2 Literature review

This section will introduce previous studies on two issues related to this paper. The first one is about RCs in Jianshi Atayal, presenting the basic properties differentiating EHRCs and IHRCs. The second one is about how the functional uncertainty works in LFG, showing the debates surrounding the outside-in and inside-out licensing.

---

[2] The general grammar of Squliq Atayal can be referred to in Huang (1993) and Rau (1992), and that of C'uli' Atayal can be referred to in Huang (1995a).

[3] The orthography of Squliq Atayal in this paper is based on Li (1991). The abbreviations used in the glosses of this paper follow the Leipzig Glossing Rules except for the following: AV: actor voice; CV: circumstantial voice; LV: locative voice; PAR: particle; PV: patient voice; Ø: null element.

2.1 RCs in Jianshi Atayal

According to Liu (2004a, 2004b, 2005), Jianshi Atayal has both EHRCs and IHRCs. There are several morphological, syntactic and semantic differences between them. One of the most remarkable differences is the (non-)existence of the complementiser *ka'*, which has to intervene between the head noun (HN) and the EHRC but does not occur between the HN and the IHRC since IHRCs have their own HNs inside themselves. This is illustrated in the following examples, where the HN is *kneril* 'woman' and the RC is *minihiy yumin* 'has beaten Yumin':

(5) [*m‹in›ihiy* *yumin* ***ka'***] *kneril* *ga'* *cyux* *m-'uyay* *la.*
AV<PRF>beat Yumin COMP woman TOP PROG AV-hungry PAR
RC HN
'As for the woman who has beaten Yumin, (she) is hungry.'

(6) *kneril* [***ka'*** *m‹in›ihiy* *yumin*] *ga'* *cyux* *m-'uyay* *la.*
woman COMP AV<PRF>beat Yumin TOP PROG AV-hungry PAR
HN RC

(7) [*m‹in›ihiy* *yumin* ***Ø*** *kneril*] *ga'* *cyux* *m-'uyay* *la.*
AV<PRF>beat Yumin Ø woman TOP PROG AV-hungry PAR
RC HN
'As for the woman, who has beaten Yumin, (she) is hungry.'

(8) [*kneril* ***Ø*** *m‹in›ihiy* *yumin*] *ga'* *cyux* *m-'uyay* *la.*
woman Ø AV<PRF>beat Yumin TOP PROG AV-hungry PAR
HN RC

The examples in (5) and (6) contain EHRCs, and the examples in (7) and (8) contain IHRCs. It is shown in the English translations that the relative orders between the HN and the RC in (5) and (6) and those in (7) and (8) are flexible, having nothing to do with semantic difference. To be more accurate, the distinction between prenominal RCs, and postnominal RCs only makes sense to EHRCs since IHRCs contain their HNs and the flexibility of the relative order merely manifests the flexibility of the HN position inside IHRCs. That is to say, the symbol *Ø* in (7) and (8) has no morphosyntactic status, merely highlighting the contrast to the complementiser *ka'* in EHRCs. Besides, the above English translations show that EHRCs and IHRCs are also different in semantics: EHRCs are restrictive RCs while IHRCs are non-restrictive. Since semantics is not what this paper focuses, this difference will not be pursued in this paper.

The evidence on the internal-headedness of the RCs without the complementiser *ka'* comes from the position of temporal adverbials. The temporal adverbial of an RC cannot be located outside the RC itself if the RC is an EHRC, as exemplified in (9). The temporal adverbial *hira'* 'yesterday' follows the HN *ngarux* 'bear', which leads to the ungrammaticality. On the contrary, the example in (10) admits of the separation of the temporal adverbial *hira'* 'yesterday' from the predicate *minaniq sayun* 'ate Sayun', which shows that the HN *ngarux* 'bear' is still located in the RC and that the RCs without complementisers are internally headed. Brackets are added on the following examples cited from Liu (2004a) to mark the boundary of RCs.

(9) ?* [*m‹in›aniq* *sayun* *ka'*] *ngarux* ***hira'*** *ga'*
AV<PRF>eat Sayun COMP bear yesterday TOP
kut-unna' *tali'*.
behead-PV OBL Tali
'As for the bear which ate Sayun yesterday, (it) will be beheaded by Tali.'
(Liu 2004a:57)

(10) [*m‹in›aniq* *sayun* *Ø* *ngarux* ***hira'***] *ga'* *kut-un*
AV<PRF>eat Sayun Ø bear yesterday TOP behead-PV
na' *tali'*.
OBL Tali
'As for the bear, which ate Sayun yesterday, (it) will be beheaded by Tali.'
(Liu 2004a:57)

Moreover, for both EHRCs and IHRCs, only subject-relativisation is permitted.[4] This complies with the accessibility hierarchy proposed by Keenan & Comrie (1977), in which subject is the easiest to be relativised among all grammatical functions. It is illustrated in the following examples:

(11) ***m‹in›ihiy*** *yumin* (*ka'*) ***kneril*** *ga'* *cyux*
AV<PRF>beat Yumin COMP woman TOP PROG
m-'uyay *la.*
AV-hungry PAR
'As for the woman(,) who has beaten Yumin, (she) is hungry.'
'#As for the woman(,) who Yumin has beaten, (she) is hungry.' (Liu 2004a:42)

(12) *wal* *bhy-**an*** *na'* *yumin* (*ka'*) ***kneril*** *ga'*
AUX.PST beat-PV OBL Yumin COMP woman TOP
cyux *m-'uyay* *la.*
PROG AV-hungry PAR
'As for the woman(,) who was beaten by Yumin, (she) is hungry.' (Liu 2004a:43)

The RC in (11) is an AV clause, which takes an actor as its grammatical subject, so the HN *kneril* 'woman' can only be interpreted as the actor of the RC rather than the patient, as indicated in the English translations. If we would like to make the HN *kneril* interpreted as the patient of the RC, then we have to alter the voice marking from AV to PV, as shown in (12). Recall that grammatical voice marking agrees with the thematic role of the subjects. Thus, once the voice marking of this RC is PV, the HN *kneril* is interpreted as a patient. What does not change is the grammatical function of the HN *kneril*—it is still the subject. In other words, the multiple voice system facilitates the relativisation of different arguments by promoting them into subject position.

[4] Most of the Formosan languages (e.g., Mayrinax Atayal (Huang 1995a, 2000), Paran Seediq (H. Chang 1997), Tsou (M. Chang 1998, 2004), Amis (Wu 1995, 1996, 1997), and Nanwang Puyuma (Teng 1997)) are also subject-sensitive in the formation of RCs.

2.2 LFG and Functional Uncertainty

LFG is a model of grammar composed of several parallel modules. Each module stands for one linguistic factor behind the superficial phenomena observed from human language, such as f(unctional)-structure, which represents grammatical functions/syntactic roles, a(rgument)-structure, which represents thematic/semantic roles, i(nformation)-structure, which represents discourse-pragmatics, and p(rosodic)-structure, which represents phonology (cf. Bresnan 2001; Butt et al. 1999; Dalrymple 2001; Falk 2001). Phrase structure, which plays the dominant role in Mainstream Generative Grammar (MGG) (cf. Chomsky 1995, 2005)), is merely one of the projections, i.e., c(onstituent)-structure, in LFG.[5] These modules interact with each other via constraints of correspondence.

Relativisation is characterised by long-distance dependencies, where an element at the front of an outer clause is understood as filling a gap within an embedded clause, as exemplified in (13).

(13) He is the man [**to whom**$_i$ I wonder [who John told [which book to give **Ø**$_i$]]]. (slightly modified from Chomsky 1986:50)

In the above example, the fronted element (filler) *to whom* is bold-faced and its gap is marked with a bold-faced symbol of null element *Ø*. What is more important is that the filler and the gap do not exist in the same clause, with the filler located in the outer clause. How to associate the filler with its gap several clauses away is one of the most important issues in syntactic theories. MGG achieves this association via movement, which can be further dismantled into smaller operations like merge, remerge, copy and deletion (Chomsky 1995; Radford 2009; Zhang 2004, among others), and movement has been the main core of MGG. LFG as a syntactic theory without movement, however, must resort to some other mechanism, that is, functional uncertainty. Note that although long-distance dependencies are related with other constructions, nothing other than RCs will be discussed in this paper.

Plainly speaking, functional uncertainty means that the content of an element is undetermined within the f-structure of the clause where it exists. It needs a constraint, which is realised as an equation, to stipulate that its content must be identical with that of another element in another clause. Thus, the potential content are infinite since there may be an infinite number of another element in other outer/embedded clauses.

Kaplan & Maxwell (1988) and Kaplan & Zaenan (1989) proposed the outside-in functional uncertainty, as (14), which identifies the value (i.e., content) of the filler in the outer clause with that of a gap in the embedded clause. To be more specific, if there exists a particular word with this constraint, usually a word bearing a discourse function (e.g., DF in (14)), like a relative pronoun, our internal cognitive operation will automatically fill it into a gap bearing a grammatical function (e.g., GF in (14)) in the same clause or in an embedded clause (indicated by multiple closed complement clause symbol COMP* in

---

[5] In other words, syntax is not completely autonomous, which means syntactic phenomena rely not only on syntax-internal explanations but also on syntax-external ones. This gives LFG more flexibility and makes it compatible with the linguistic facts presented in other linguistic fields like discourse-pragmatics and sociolinguistics. However, this also increases the complexity of modules in contrast to MGG.

(14)) if there is a verb subcategorising a complement clause.[6] This process of identification is from the outer clause where the filler exists, to the inner clause where the gap exists, so it is the outside-in functional uncertainty. Take (15) for example. The c-structure rule and the f-structure of it are illustrated in (16) and (17) respectively:

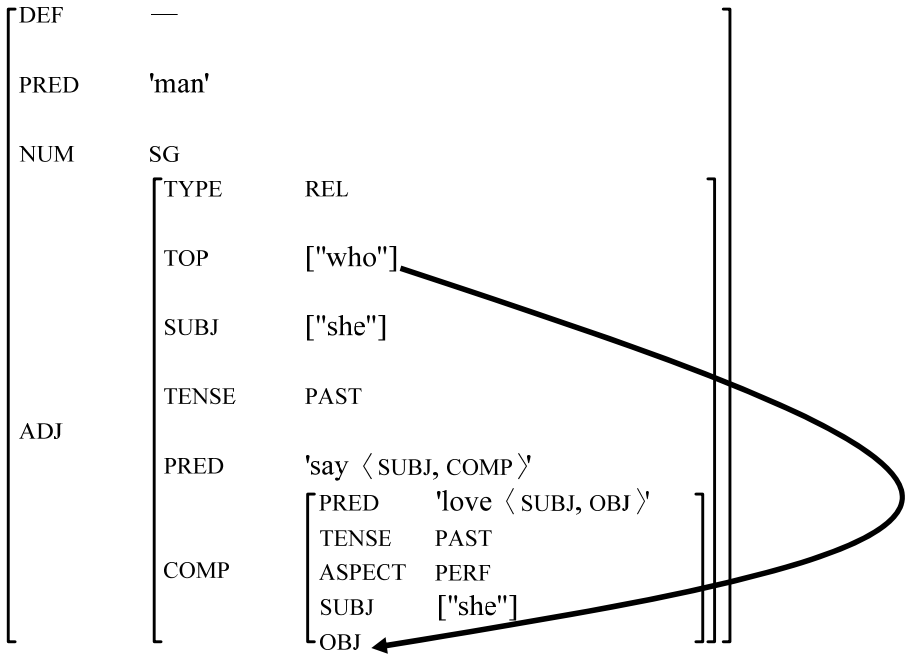
(14) (↑ DF) = (↑ COMP* GF)

(15) a man who she said that she had loved

(16)[7]

CP	→	DP	C'
		(↑ TOP) = ↓	↑ = ↓
		(↑ TOP) = (↑ COMP* GF)	

(17) f-structure



The c-structure rule in (16) shows that the specifier of CP, i.e., the relative pronoun, is annotated with two equations, (↑ TOP) = ↓ and (↑ TOP) = (↑ COMP* GF). The former indicates that the specifier is also the grammaticised topic of this RC, and the latter is the outside-in functional uncertainty constraint, which stipulates that the grammaticised topic is identical with one of the grammatical functions in the inner clause. In (17), there is a line showing the process of identification with the direction of its arrowhead. Following the outside-in functional uncertainty constraint, this line goes across the boundary of a clause, entering the embedded matrix of COMP, pointing to the missing object inside.

---

[6] Kleene star '*' means 'any number of', including 'none of'. It is borrowed from mathematics and commonly used in LFG studies.

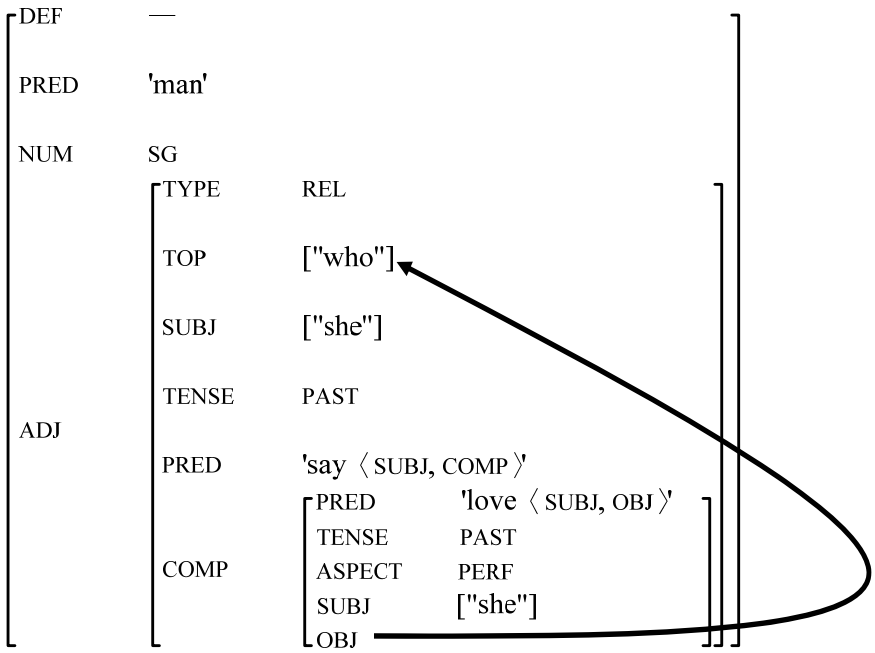
[7] The up arrow is a metavariable referring specifically to the f-structure of the mother node, while the down arrow is a metavariable referring to the f-structure of the current node.

On the other hand, Bresnan (1995, 2001) proposed the inside-out functional uncertainty, as the constraint in (18), which requires that the value of the gap in the embedded clause be identical with that of its filler in the outer clause. If there is an invisible grammatical function (e.g., GF in (18)) in an embedded clause, which native speakers can tell from the argument structure of the predicate, our internal cognitive operation will search the outer clause (indicated by (COMP*) in (18)) for a discourse function e.g., DF in (18)). Under this analysis, the example (15) has the c-structure rule in (19) and the f-structure in (20).

(18) (↑ GF) = ((COMP* ↑) DF)

(19) VP → V *e*
↑ = ↓ (↑ OBJ) = ↓
(↑ OBJ) = ((COMP* ↑) TOP)

(20) f-structure



In (19), the c-structure rule has an empty object (indicated by (↑ OBJ) = ↓), which is annotated with the inside-out functional uncertainty constraint stipulating that its semantic content must be identical with a grammaticised topic outside this clause (indicated by ((COMP *) TOP)). Again, in order to illustrate the direction of functional uncertainty, the bolded line of identification is added with an arrowhead in (20), which displays that the cognitive process of identification goes upwards across the clause boundary to the grammaticised topic in the outer clause.

Falk (2001:159) adopted a mixed analysis that "[o]utside-in licensing is constrained to cases where the gap is SUBJ [i.e., subject], while inside-out licensing involves (for English at least) an empty category in c-structure. Languages that only allow extraction of SUBJ only allow outside-in licensing of long-distance dependencies". His proposal is based on the fact that subject is an overlay function, which serves a clause-external role,

with the empirical support of cross-linguistically valid subject extraction and the *that*-trace effect.

In the subsequent sections, I will study what syntactic mechanisms Jianshi Atayal RCs employ.

3 EHRCs

In this section, we will argue for the operator function of the complementiser *ka'* first and then for the application of the outside-in functional uncertainty to EHRCs.

3.1 Complementiser

As stated earlier, the existence of the complementiser *ka'* is one of the most distinct characteristics differentiating EHRCs from IHRCs in Jianshi Atayal since the complementiser *ka'* must intervene between the HN (the modifiee) and its RC (the modifier). The word *ka'* is indeclinable, like the complementiser *that* in English, so it is more preferable to treat it as a complementiser than as a relative pronoun (Liu 2004a).[8] EHRCs do not make use of relative pronouns. In addition, as *that*-RCs in English, EHRCs do not allow pied-piping, as presented in (21).

(21) * *kneril* [***ka'*** ***roziq*** *siy-on* *na'* *sayun*] *qasa* *ga'*
woman COMP eye like-PV OBL Sayun that TOP
blaq *kt-an.*
good see-LV
'The woman whose eyes Sayun likes is beautiful.'

In the above example, there is a body-part relationship between the HN *kneril* 'woman' and the word *roziq* 'eye' inside the RC. This semantic association does not trigger any pied-piping in syntax, as born out in the ungrammaticality of (21). If *ka'* is analysed as a complementiser, then the explanation of no pied-piping is more straightforward since a complementiser does not trigger pied-piping. Cross-linguistically, pied-piping co-occurs with relative pronouns, and in some languages relative pronouns are only used with pied-piping (Krifka & Musan 2012).

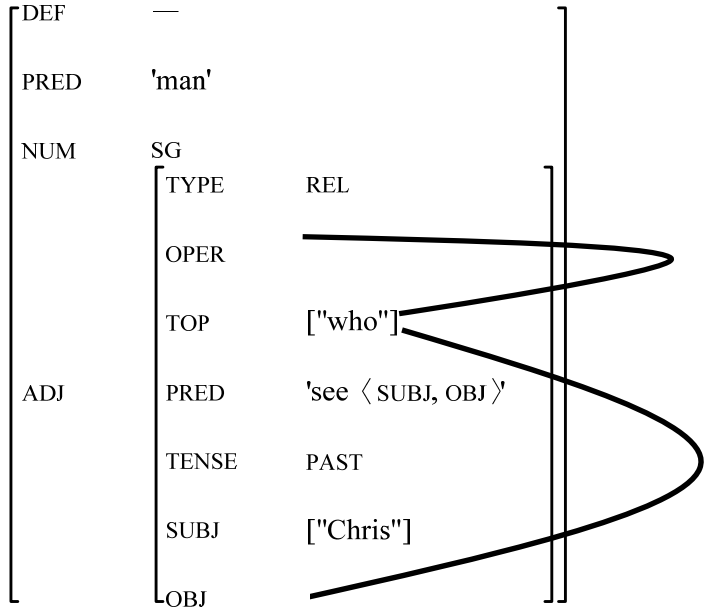
Then, what is the complementiser *ka'*? Before answering this question, we have to follow Dalrymple (2001) and Falk (2001, 2010) and make a distinction between the function Operator (OPER) and the function Topic (TOP) in the f-structure. The former serves as "an intermediate step in the flow of information between the in-clause and out-of-clause portions of the relative construction" (Falk 2010:220), while the latter serves as a grammaticised discourse topic in a clause (Bresnan & Mchombo 1987). The differentiation between OPER and TOP helps explain the phenomenon of pied-piping in English. Let's look at the following two examples in (22) and (24) from Dalrymple (2001:401–402) and their f-structures (modified from Dalrymple (2001:401–402)):

---

[8] Despite the lack of diachronic evidence, it may have been developed from a case marker, which is a common phenomenon in Formosan languages (Tang 2008).

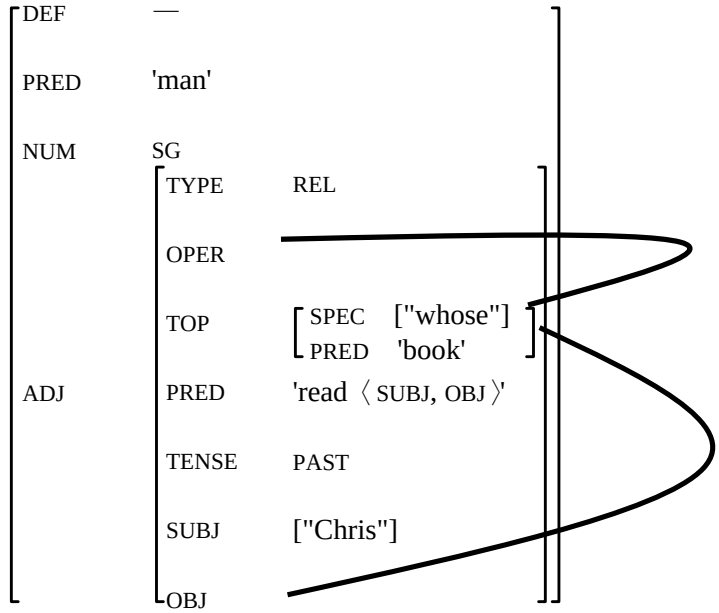
(22) a man who Chris saw

(23)



(24) a man whose books Chris read

(25)



There is no pied-piping in (22), so the *wh*-relative *who* functions as both OPER and TOP, as illustrated in (23). Example (24) is an RC with pied-piping, which has not only a *wh*-relative *whose* but also its complement *books* in the clause-initial position. Here the

wh-relative no longer has duality of function: The *wh*-relative itself only serves as OPER, whereas the whole fronted phrase within which the *wh*-relative exists serves as TOP and the filler of the gap in the object position, as displayed in (25).

Falk (2001) analysed *wh*-less RCs in English as containing an unexpressed filler bearing the function OPER but not bearing the function TOP because only *wh*-less RCs contain the existential construction, which introduces new information. Compare the two examples in (26).

(26) a. * the book which there is on the table (slightly modified from Falk 2001:168)

b. the book that there is on the table (slightly modified from Falk 2001:168)

Similarly, the existential construction can form an EHRC in Jianshi Atayal, like *wh*-less RCs in English. This is exemplified in (27). In line with Falk's analysis, it is proposed that RCs in Jianshi Atayal do not contain the grammatical function TOP in their f-structures. What they have in the f-structure is OPER.

(27)

kneril	*ka'*	*m-aki'*	*ngasal*	*qasa*	*ga,*	*sayun*	*la.*
woman	COMP	AV-exist	house	that	TOP	Sayun	PAR

'As for the woman that exists in that house, (she) is Sayun.'

TOP was used to analyse relativisation for a long time since Bresnan (1982), one of the earlier LFG studies on relativisation. However, it is proposed in latest studies like Alsina (2008) and Asudeh (2012) that what really matters to relativisation or unbounded dependencies is not TOP but OPER (i.e., OP in Alsina (2008) and UDF in Asudeh (2012)). TOP only works on the cases of pied-piping, unavailable in cases without pied-piping, while OPER is available in all RCs. Even though it can be analysed that there is the stipulation OPER = TOP in languages without pied-piping, the term TOP should be avoided in long-distance dependencies since it complicates syntactic generalisations that are neutral to discourse-pragmatics, as suggested in Asudeh (2012). Therefore, unlike the analyses of the functional uncertainty presented in section 2.2, where TOP is used, the discourse function (DF) of the functional uncertainty in my analysis is OPER.

Furthermore, I propose that the complementiser *ka'* in Jianshi Atayal plays the role as OPER in the f-structure. It is a lexical property marked in the lexicon, as shown in (28). This analysis is tenable for the following reasons. First, in contrast to *that* in English, which occurs in several clause types, *ka'* is a complementiser specialised for EHRCs in this language. [9] Its (obligatory) occurrence intuitively marks the existence of relativisation for native speakers. Compared with the proposal of the existence of a null operator, it is closer to the empirical evidence to analyse the specialised complementiser *ka'* as the operator of EHRCs. Second, the proposal of the null operator is faced with "the problem of eternal silence", as pointed out by McCloskey (2006). That is, there are no cases in Jianshi Atayal where the postulated operator is other than null. Cross-linguistically, at least to my limited knowledge, it is much more common that RCs have either a complementiser or an overt operator than it is that RCs have both of them

[9] It is possible to term *ka'* in Squliq Atayal as a "relativiser". This paper does not adopt this term since it is still unknown to me what the lexical category of relativisers is, a lexical category independent from complementisers or merely a subcategory of complementisers. Therefore, this paper conservatively claims that *ka'* is a complementiser like *that* in English.

simultaneously. McCloskey (2006) even claimed that there is no language which has an RC with both of a complementiser and an overt operator at the same time. However, it still remains dubious why overt operators and complementisers are always in complementary distribution. Maybe in most of cases, as proposed here, complementisers alone carry the function OPER and there is no need for an RC to have an overt operator. At least for Jianshi Atayal, a language with a complementiser specialised for EHRCs but without any overt operator, this proposal is more tenable. Third, from the perspective of theorisation, it is more economic for a syntactic framework not to postulate a null operator in the lexicon of Jianshi Atayal. What element can function as an operator, a phenomenon which varies from language to language and even from construction to construction, should be attributed to more fundamental mechanisms, like lexical properties. Metaphorically speaking, since some people in this world have six fingers on one hand, for a more economic model of human beings, it is more preferable to resort the explanation of this rare case to fundamental mechanisms like genes than to the proposal of the existence of the covert sixth finger for all human beings.

Once we realise that the assignment of OPER can be lexical—that is to say, it can be assigned to different lexical categories—why relative pronouns allow pied-piping is not mysterious: They are determiners in nature, which makes themselves possible to be (part of) the topicalised constituent.[10] For example, the *wh*-relative *who* in English is specified in lexicon for its operator function, being part of the topicalised constituent, being either nominative or accusative, and being human animate, as shown in (29).

(28) *ka'*: C
($\uparrow$ OPER) = $\downarrow$
$\downarrow \in$ ($\uparrow$ ADJ TYPE REL)

(29) who: D
($\uparrow$ OPER) = $\downarrow$
$\downarrow \in$ ($\uparrow$ TOP)
($\downarrow$ CASE) = NOM $\cup$ ACC
($\downarrow$ HUMAN) = +
($\downarrow$ ANIM) = +

For any linguist familiar with the transformational approach, this proposal may look absurd at first glance because the filler (an operator) and its gap may not belong to the same lexical category. For instance, the operator of RCs in Jianshi Atayal is a complementiser *ka'* while the gap left in RCs is a nominal. It is impossible to claim that there is any movement between them.[11] If we put aside the concept of movement for the

[10] Under the framework of MGG, Tsai (1997, 1999a, 1999b) proposed that languages vary with the base site of their null operators: three typological types include D/N in English,CP/IP in Mandarin Chinese, and PP/DP in Japanese. Despite the difference in our theoretical models, the proposal of this paper is close to that of Tsai's works. Converting his typological classification into ours, we may hypothesise that C in Mandarin Chinese bears OPER, like Jianshi Atayal while case marker (K) in Japanese bears OPER. It needs more future research whether this expedient conversion really works.

[11] My thanks go to one of the reviewers for reminding me that movement is not the only strategy available for *wh*-dependencies. The base-generation strategy is another possible strategy, as attested in Adger & Ramchand (2005) and Tsai (1997, 1999a), among others.

moment, it will be easy to understand that for the issue on syntactic dependency, what LFG really studies is not how one element has two different grammatical functions but how two different grammatical functions are mutually linked. This can be manifested in movement paradoxes, where there is a mismatch between a seemingly moved element and the position it is supposed to be moved from (cf. Bresnan 2001; chap. 2), as exemplified below (slightly modified from Bresnan (2001:17)):

(30) a. [That he was sick]$_{CP}$ we talked about Ø for days.

b. * We talked about [that he was sick]$_{CP}$ for days.

c. We talked about [the fact that he was sick]$_{DP}$ for days.

In the a-example of (30), the topicalised phrase is a CP and it seems to move from the object of *about* (*Ø*). However, the sentence will be ungrammatical after the CP is put back into the object position, as attested in the b-example. On the contrary, the grammatical sentence in the c-example has a DP as the object of *about*. Thus, in the a-example, the dependency between the filler (CP) and the gap (DP) is not based on the identical syntactic category but on the association between the two functions TOP and OBJ. Then, look back on the phenomenon in Jianshi Atayal whereby there is a dependency between the filler (C) and the gap (DP), and we will not find it absurd anymore. Most important of all, we would like to know how OPER and other grammatical functions are linked.

This section points out the importance of the complementiser *ka'* in EHRCs, especially about the function OPER specified in its lexicon.[12] After understanding the grammatical function it bears in the f-structure, we can discuss what syntactic mechanism EHRCs employ in Jianshi Atayal.

3.2 Outside-in functional uncertainty

In the Atayalic languages, including Jianshi Atayal, the equivalent of complement clauses in English is sentential subjects. Put differently, when a predicate subcategorises an event as its argument in the a-structure, the event will become a finite clause in the c-structure having the function SUBJ in the f-structure. The predicate is in the NAV, where the agent or experiencer of the event will be an oblique argument. A sentential subject has no complementiser but an optional nominative case marker *qu'* preceding the whole clause (Huang 1993; Rau 1992), as exemplified below:

(31)	*baq-un=maku'*	***(qu')***	*siy-on*	*temu'*	*sayun.*
	know-PV=1SG.OBL	NOM	like-PV	Temu	Sayun

'That Sayun is liked by Temu is known by me.'

Because Jianshi Atayal is a predicate-initial language, the part before *qu'* is composed of the main verb of the matrix clause *baqun* 'know', suffixed with the PV marker, and of the theme *maku'* 'I', an oblique clitic, while the rest after *qu'* is the sentential subject, which has its own verb *siyon* 'like', oblique argument *temu'*, and subject *sayun*.

---

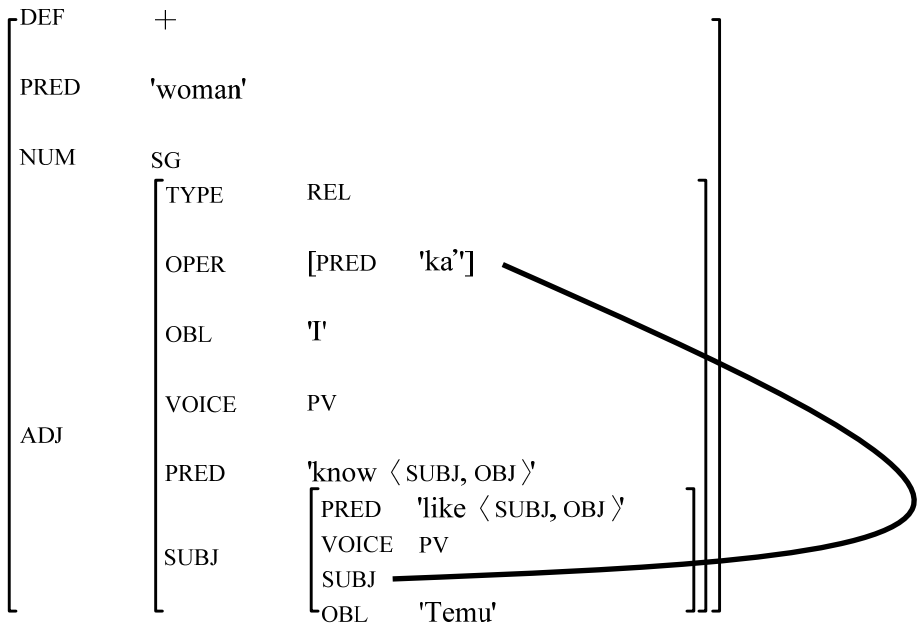
[12] Although it is not mentioned clearly in section 3, the existence of OPER means that the licensing between the filler/HN and the gap is mediated. The details about the contrast between the mediated analysis and the unmediated one will be discussed later in section 4.

Even though there are no real "complement" clauses in this language, long-distance dependencies between the filler/operator and its gap still exist, as presented in (32).

(32) *kneril* *[ka'* *baq-un=maku'* [*qu'* *siy-on* *na'*
woman COMP know-PV=1SG.OBL NOM like-PV OBL
temu' *Ø*]]
Temu Ø
'the woman who I know Temu likes'
'for the woman x, that x is liked by Temu is known by me'

The above example has a long RC which is composed of two parts: the predicate of the matrix clause and the embedded clause as the subject of the matrix clause. The former is *baqun=maku'* 'known by me/I know' while the latter is *siyon na' temu'* 'be liked by Temu'. Both the relative complementiser *ka'* and the optional nominative case marker *qu'* are located in the initial position of their respective clause. What is more important, the gap of the whole RC is deeply embedded inside the sentential subject, serving as the grammatical subject of it, as marked with *Ø* in (32). That is to say, the dependency between *ka'* (OPER) and *Ø* (SUBJ) is long in that they are not within the same clause: OPER is in the matrix/outer clause while SUBJ is in the embedded/inner clause. As substantiated in the f-structure (33), the filler OPER licenses its gap SUBJ across the clause boundary into the matrix of the sentential subject. Note that sentential subjects have the same grammatical function as phrasal subjects, as proposed in Berman (2003) and Kroeger (1993), among others.

(33)

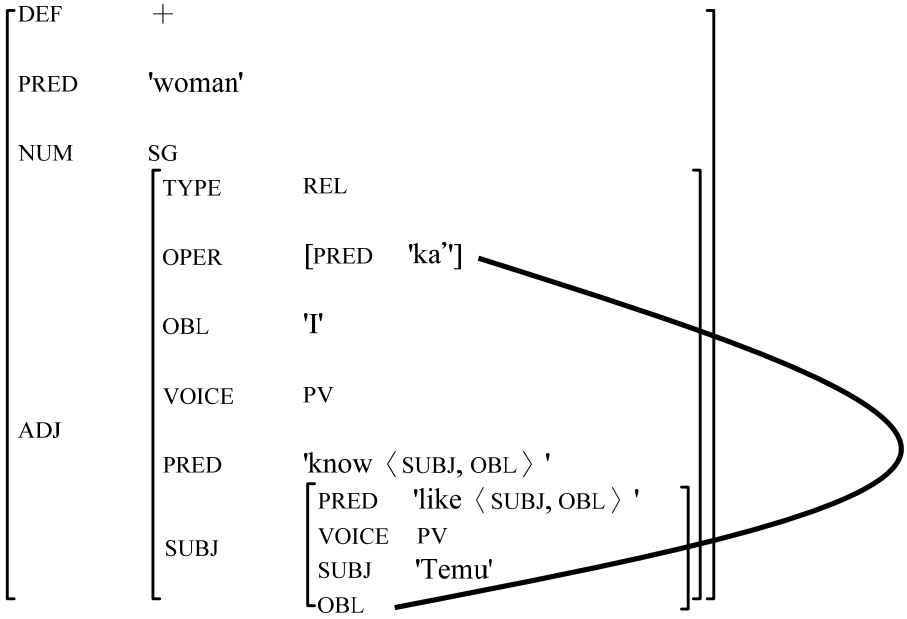


Recall that only grammatical subjects can be relativised in Jianshi Atayal (cf. section 2.1). This property still exists even in long-distance RCs, as exemplified in (34). This sentence is ungrammatical because the gap deeply embedded in the sentential subject is an object within this embedded clause, as illustrated in the f-structure in (35). We can tell the object

function from the verb *smoya* 'like', the AV infix of which indicates the missing argument must be the object.

(34) * *kneril* [*ka'* *baq-un=maku'* [*qu'* *s‹m›oya* *Ø*
woman COMP know-PV=1SG.OBL NOM like<AV>like *Ø*
temu']]
Temu
'the woman who I know Temu likes'
'for the woman x, that Temu likes x is known by me'

(35) *



It is proposed in this paper that relativisation of Jianshi Atayal employs the constraint of the outside-in functional uncertainty in its grammar. The constraint is shown in (36).[13] It stipulates that OPER in the outer clause must equal SUBJ in the inner clause. In comparison with (14) in section 2.2, the function labels in this constraint are more specific and more fitting to Jianshi Atayal: (↑ OPER) in (36) specifies that the discourse function (DF), which connects the information of two sentences, is only OPER, without the possible involvement of TOP; (↑ SUBJ $^{+}$) indicates the SUBJ in the same clause or embedded in one or more than one sentential subject clause (SUBJ). In other words, not only EHRCs with long-distance dependency but also ordinary EHRCs apply this mechanism. The stipulation for SUBJ is not ad hoc since it represents that the inter-clausal connection of information must be directly associated with the old information in the innermost clause so as to string up all clausal information in one RC. SUBJ in Jianshi Atayal plays this crucial role. Constraint (36) is lexical so it is specified in the lexical entry of the complementiser *ka'*, as shown in (37): The underneath OPER equals the

[13] RCs in Tagalog have the same subject-only constraint, so our constraint here is almost the same as that in Kroeger (1993) except that OPER is used in our version.

underneath SUBJ which may be embedded in another SUBJ. The c-structure rule in (38) does not specify the relative order between the daughter nodes NP and CP, as marked with the common ",". In other words, EHRCs can be prenominal or postnominal. The rule which only specifies dominance between a mother node and its daughter nodes is an immediate dominance rule (ID rule) (cf. Gazdar et al. 1985, chap. 3). Accordingly, without *ka'* (OPER), the constraint of the outside-in functional uncertainty will be unavailable in the f-structure of EHRCs.

(36) $(\uparrow \text{OPER}) = (\uparrow \text{SUBJ}^{+})$

(37) *ka'*: C
$(\uparrow \text{OPER}) = \downarrow$
$(\downarrow \text{OPER}) = (\downarrow \text{SUBJ}^{+})$
$\downarrow \in (\uparrow \text{ADJ TYPE REL})$

(38) NP → NP, CP
$\uparrow = \downarrow$ $\downarrow \in (\uparrow \text{ADJ})$

Why did we not postulate that EHRCs employ the constraint of inside-out functional uncertainty, like (39)? The inside-out functional uncertainty is not adopted for two reasons. First, as pointed out in Falk (2001), whether this constraint is annotated in the c-structure rule, like (40), or it is specified in the lexical entry, it is necessary to present an empty subject *e* inside an EHRC, which violates the principle of economy of expression in (41). That is, invisible nodes, including invisible words and phrases, should not exist in our c-structure. C-structure should be as close to the surface structure as possible.

(39) $\uparrow = ((\text{SUBJ}^{+} \uparrow) \text{OPER})$

(40) IP → I' *e*
$\uparrow = \uparrow$ $(\uparrow \text{SUBJ}) = \downarrow$
$\uparrow = ((\text{SUBJ}^{*} \uparrow) \text{OPER})$

(41) **Economy of Expression** (Bresnan 2001:91):
All syntactic phrase structure nodes are optional and are not used unless required by independent principles (completeness, coherence, semantic expressivity).

Second, the inside-out analysis needs two mechanisms to achieve its goal: One is the c-structure rule in (40), and the other is the lexically-specified OPER of the complementiser *ka'*, as illustrated in (28) (cf. section 3.1). Both syntactic and lexical mechanisms are used. Note that although these two mechanisms look like they exist independently, they have the same and the only goal—to yield the licensing between the filler and the gap—instead of working independently in their respective modules and collaboratively once in a while.[14] The inside-out analysis loses its notional ground in Jianshi Atayal even though it does not in the languages where there are a variety of words carrying OPER and different grammatical functions can be relativised, such as English.

[14] In phonology, such functional unity of processes is known as a conspiracy (Kisserberth 1970). The advocators of Optimality Theory (OT) claim that OT is better than the rule-based approach in that the former can capture conspiracies but the latter cannot (Kager 1999).

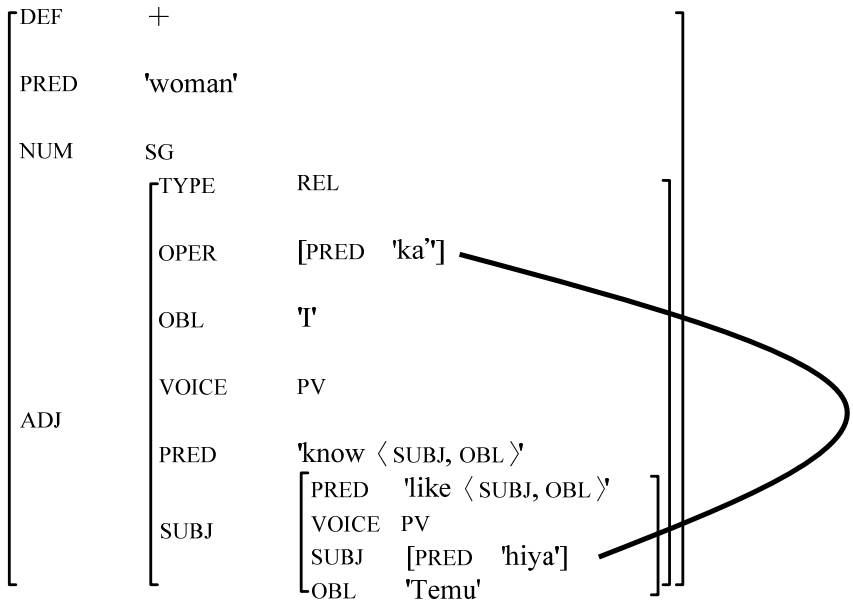
By contrast, the outside-in analysis only makes use of the lexical mechanism. The simplicity of the outside-in analysis wins out over the inside-out analysis once again.

The outside-in analysis however seems to face another challenge coming from resumptive pronouns. As illustrated in (42) and (43), the co-occurrence of the complementiser *ka'* (OPER) and the resumptive pronoun *hiya* (SUBJ) is grammatical. The predicate feature of OPER may clash with that of SUBJ, as illustrated in the f-structure (44) of example (43). In (44), the PRED of OPER is *ka'*, and it cannot license the resumptive pronoun through functional uncertainty since the embedded SUBJ has its own PRED *hiya*.

(42) *kneril* [*ka'* *siy-on* *na'* *temu'* ***hiya***]
woman COMP like-PV OBL Temu 3SG.NOM
'the woman who is liked by Temu'

(43) *kneril* [*ka'* *baq-un=maku'* [*qu'* *siy-on* *na'*
woman COMP know-PV=1SG.OBL NOM like-PV OBL
temu' ***hiya***]]
Temu 3SG.NOM
'the woman who I know is liked by Temu'

(44) *



The outside-in functional uncertainty does not work in (44), and neither does the inside-out one. That is, this is not a problem only for the outside-in analysis but one for both functional uncertainty analyses. Apparently, the EHRCs with resumptive pronouns employ a mechanism different from the functional uncertainty. This can be proved in (45), which shows that the distribution of resumptive pronouns is not all the same as that of the gaps:

(45) * [*siy-on* *na'* *temu'* ***hiya*** *ka'*] *kneril*
like-PV OBL Temu 3SG.NOM COMP woman
'the woman who is liked by Temu'

Unlike (42) and (43), (45) is a postnominal EHRC with a resumptive pronoun preceding the HN *kneril* 'woman'. It is ungrammatical not because postnominal EHRCs are unacceptable (cf. section 2.1) but because the precedence of the resumptive pronoun. Linearity does not affect the functional uncertainty constraints, but it has something to do with anaphoric binding since precedence is an important constraint behind it. We leave the issue on anaphoric binding in Jianshi Atayal to future research. However, it is undoubted that the mechanism behind resumption is anaphoric binding, not the functional uncertainty. Thus, it offers no problem to our outside-in analysis.

In this section, I propose a more specific constraint of the outside-in functional uncertainty for the grammar of Jianshi Atayal. In addition, the analysis proposed here also dovetails Falk's (2001) typological claim that outside-in functional uncertainty applies to subject extraction in EHRCs.

4 IHRCs

In the beginning of this section, I argue for the unmediated licensing between the filler and the gap. Then it is shown with the evidence from IHRCs that the inside-out functional uncertainty is applied even though IHRCs in Jianshi Atayal only permit subject-relativisation.

4.1 Unmediated licensing

In order to account for *that*-RCs in English, Chomsky (1977) proposed that there is a null operator, the invisible equivalent of relative pronouns, in the RCs without any overt relative pronoun. From the perspective of syntactic theories, a null operator is the medium between the HN and its gap either in c-structure (in MGG) or f-structure (in LFG). Thus, this analysis is termed as the (anaphorically) mediated analysis of RCs.[15] Along with the mediated analysis, some syntacticians proposed alternative analysis—the (anaphorically) unmediated analysis in which there is no null operator as the medium between the HN and the gap and the licensing between them is directly established through either movement in MGG or functional uncertainty in LFG (Aldridge 2004; Falk 2010; Kayne 1994; Schachter 1973; Vergnaud 1974, among others).

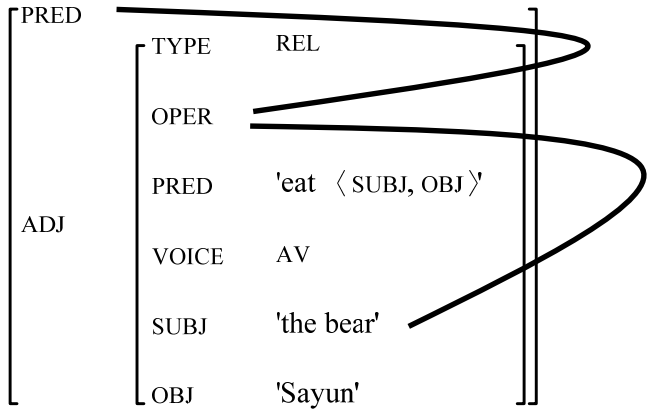
The distinction between the mediated analysis and the unmediated one can be illustrated in (47) and (48) respectively, both of which are the f-structures of the IHRC part bracketed in (46). In this example, the HN *ngarux* 'bear' plays a dual role: It is both the visible subject of the IHRC (i.e., the filler/HN) and the invisible subject of the matrix clause (i.e., the gap). All HNs of IHRCs have this duality. In the mediated analysis, as shown in (47), there is one OPER inside the ADJ carried by the IHRC itself. As illustrated by the two lines, the OPER directly licenses the PRED of the subject in the matrix clause and its content further depends upon the SUBJ of the IHRC. That is, the licensing between the filler SUBJ and the gap PRED is mediated by the OPER. However, the unmediated

[15] The terminology on the distinction between the anaphorically mediated analysis and the anaphorically unmediated one is adopted from Falk (2010). In MGG, the formal is termed the standard/adjunction approach and the latter is termed the raising/Kaynian approach (Alexiadou et al. 2000).

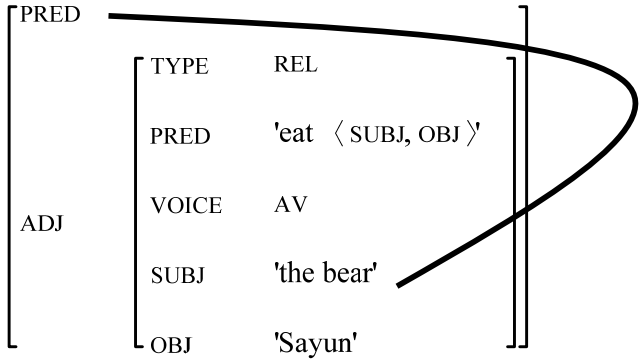
analysis in (48) presents a different f-structure. There is no OPER inside the ADJ (i.e., the IHRC), which further leads to the unmediated licensing between the SUBJ of the IHRC and the gap with uncertain content (i.e., PRED).

(46) *kut-unna'* *tali'* [*qu'* *m‹in›aniq* *sayun* ***ngarux***].
behead-PV OBL Tali NOM AV<PRF>eat Sayun bear
'The bear, which ate Sayun, will be beheaded by Tali.'

(47) the mediated analysis



(48) the unmediated analysis



In this paper, we believe that the unmediated analysis is more appropriate to IHRCs in Jianshi Atayal. There are two main reasons. The first one is theoretical simplicity. Comparing (47) and (48), we can easily perceive that the mediated analysis in (47) is more complex than the unmediated analysis in (48). With everything being equal, the unmediated analysis wins out because it employs only one pre-existing mechanism, not an ad hoc one—that is, functional uncertainty—and it is applied for merely once. If one pre-existing mechanism is enough, why do we bother adopting an analysis with more mechanisms? The idea that there is no operator in RCs is hard to be accepted by some syntacticians since Chomsky's (1977) analysis has been viewed as the "standard" one. However, note that Chomsky (1977) analyses only EHRCs in English. With the emergence of more fresh theoretical analyses and with the discovery of more cross-linguistic evidence, it should be examined equally with other analyses, not viewed

as Cinderella's glass slippers, into which everyone is eager to fit themselves. Anyway, the complexity of the mediated analysis is undoubtedly a weakness.

So far, readers may wonder why the unmediated analysis is not adopted in our previous analysis of EHRCs since it has theoretical simplicity. The answer to this is co-related with the second weakness of the mediated analysis—the conveyer of OPER. As presented in section 3.1, once the grammatical function OPER is proposed, we must claim where it is specified and what its conveyer is. Recall that EHRCs all have the complementiser *ka'*, whose function is to mark relativisation. It corresponds to OPER in f-structure. Nevertheless, as exemplified in (46), IHRCs do not have such a choice since they look like ordinary clauses superficially without any special marker for relativisation. Even if we syntactically annotate the grammatical subjects of IHRCs as OPER, as presented in (49), we will still face the doubt about why SUBJ and OPER are identical here. If so, maybe one of them, especially OPER, does not have the necessity of existence in IHRCs. Once again, the proposal that SUBJ equals OPER is still not theoretically simpler than the unmediated analysis.

(49) IP → I' DP
↑ = ↓ (↑ SUBJ) = ↓
(↑ OPER) = ↓

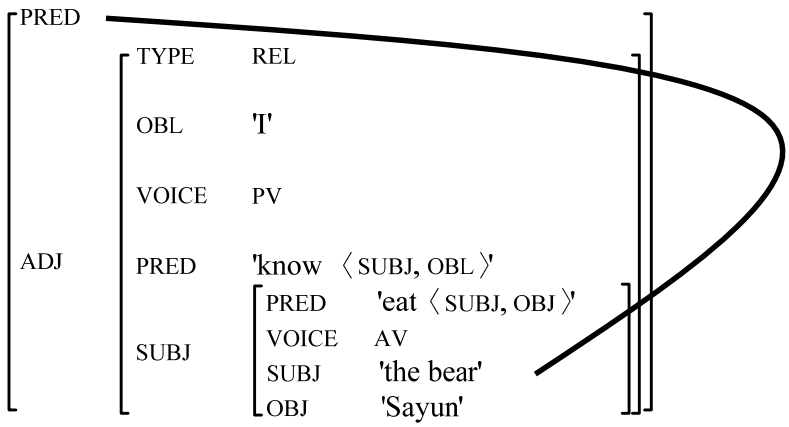
At the end of this section, we have to remind readers that although the unmediated analysis is adopted here for the IHRCs in Jianshi Atayal, we do not assume that it is fitting into all RCs cross-linguistically. Just as we adopt the mediated analysis for (ordinary) EHRCs but the unmediated one for IHRCs, the syntactic mechanisms behind relativisation should vary not just from language to language but also from construction to construction. For instance, the IHRCs in Quechua (an American language) always carry special marking on the verbs (Cole et al. 1982). It is not guaranteed that the unmediated analysis is appropriate to them. Our analysis here has its limitation of extension.

4.2 Inside-out functional uncertainty

Aside from EHRCs, IHRCs in Jianshi Atayal also allow long-distance dependencies. The only difference is the opposite direction of licensing: The HN in a deeply-embedded clause licenses the invisible gap in the outer clause, as shown in (50). Recall that complement clauses in English are realised as sentential subjects in Jianshi Atayal. For instance, the sentence in (50) has one intermediate clause, *baqun maku'* 'known by me', which subcategorises the sentential subject *qu' minaniq sayun ngarux* 'the bear ate Sayun'. Note that the verb *kutun* 'behead' does not subcategorise any sentential subject but it is still followed by a sequence of sentential subjects. Its (invisible) real subject (i.e., the gap) is identical with the grammatical subject of the innermost clause *ngarux* 'bear' (i.e., the filler/HN). That is to say, there is one clause intervening between the HN *ngarux* 'bear' in the innermost clause and the gap in the matrix clause. The f-structure of the IHRC part in (50) is illustrated in (51). The licensing of the gap is across two clause boundaries, one corresponding to SUBJ and the other corresponding to ADJ in the f-structure.

(50) *kut-unna'* *tali'* [*qu'* *baq-un=maku'* [*qu'* *m‹in›aniq*
behead-PV Tali OBL NOM know-PV=1SG.GEN NOM AV<PRF>eat
sayun ***ngarux***]].[16]
Sayun bear
'The bear, which I know that ate Sayun, will be beheaded by Tali.'

(51)



Are the IHRCs in Jianshi Atayal constrained by the outside-in functional uncertainty or the inside-out one? Both of them are technically possible. Take (50) and (51) for example. If the constraint of the outside-in functional uncertainty (↑ PRED) = (↑ ADJ SUBJ^{+}) is annotated on the c-structure rule, the c-structure rule will inevitably contain a null head of the HN since the starting point of the outside-in functional uncertainty is the gap in the case of IHRCs. As illustrated in the c-structure rule (52), the starting point is the PRED of the DP, not being part of the IHRC, so the existence of *e* is obligatory because the constraint cannot be annotated under the adjunct CP. Otherwise, the front part of this constraint (↑ PRED) will refer to the PRED of CP.

(52) DP → *e* CP
(↑ PRED) = (↑ ADJ SUBJ $^{+}$) ↓ ∈ (↑ ADJ)

By contrast, the constraint of the inside-out functional uncertainty, as the annotation (↑ SUBJ) = ((ADJ SUBJ ↑) PRED) in (53), does not require the existence of a null element because the starting point (i.e., the HN) is the grammatical subject of the innermost clause, as indicated in the part (↑ SUBJ). Hence, the constraint can be annotated under the DP of the innermost clause, as shown in (53).

(53) IP → I' DP
↑ = ↓ (↑ SUBJ) = ((ADJ SUBJ* ↑)PRED)

As mentioned in Section 3.2, the economy of expression requires that c-structure should be as close to the surface structure as possible. With everything being equal, the inside-out functional uncertainty is more favorable for the IHRCs in Jianshi Atayal than

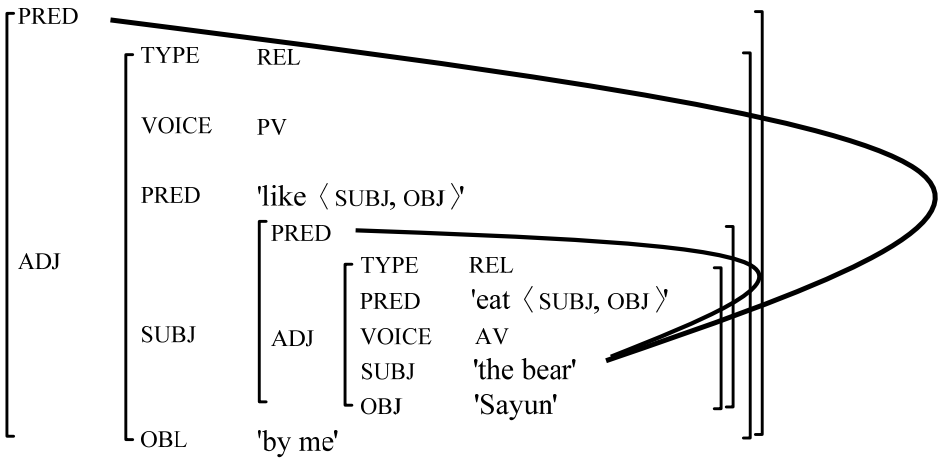
[16] In the cases of deeply-embedded IHRCs, verb types may affect grammaticality. Usually adjective-like stative verbs lead to ungrammaticality. More studies on the correlation between verb types and IHRCs are needed.

the outside-in one. This is counter to Falk's (2001) proposal that languages with only subject-extraction only permit the outside-in functional uncertainty. Because Falk (2001) does not take IHRCs and languages other than English into consideration, we do not find it necessary to follow his stipulation.

In addition to the cases of sentential subject, long-distance dependencies also occur when there are stacked RCs, with more than one RC modifying an identical HN. Example (54) is a case of RC-stacking. Once again, the HN *ngarux* 'bear' is located in the innermost RC, serving as SUBJ, and it is further the HN of the intermediate RC, also serving as SUBJ. Ultimately, the HN is the subject of the matrix clause (i.e., the leftmost clause in (54)). The f-structure of the IHRC part in this example is illustrated in (55). The filler *ngarux* 'bear' licenses the gap in the intermediate RC across the innermost ADJ and the gap in the matrix clause across the innermost ADJ, the intermediate SUBJ and the outmost ADJ.

(54) *kut-unna'* *tali'* [*qu'* *siy-on=maku'* [*qu'* *m‹in›aniq*
behead-PV OBL Tali NOM like-PV=1SG.GEN NOM AV<PRF>eat
sayun ***ngarux***]].
Sayun bear
'The bear, which ate Sayun (and) which is liked by me, will be beheaded by Tali.'

(55)

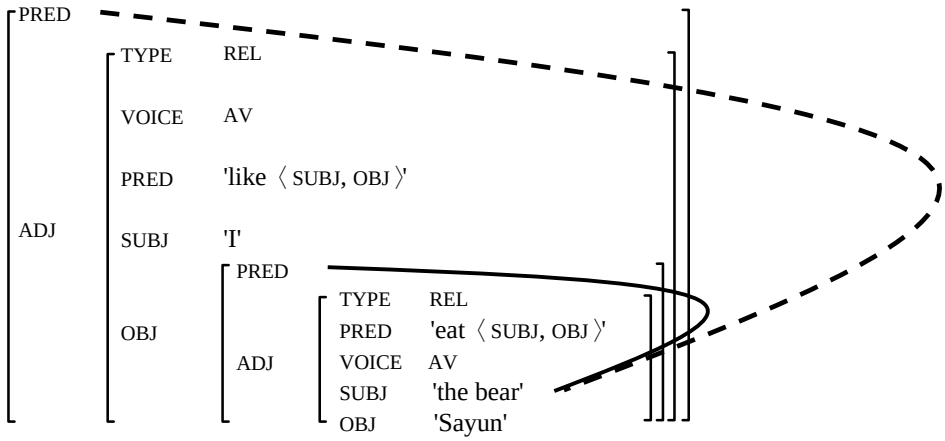


The example (56) below is also a case of RC-stacking, but it is ungrammatical. Comparing (54) and (56) will make us have a better understanding of the nature of long-distance dependencies in this language. The sentence in (56) has the HN/filler *ngarux* 'bear' serving as SUBJ in its innermost clause. The HN eventually is identical with the invisible grammatical subject of the matrix clause. These two points are exactly like what we see in (54). However, in (56), the gap in the intermediate clause conveys the grammatical function OBJ and the grammatical subject in it is *watan*. This is the greatest difference between (54) and (56) and it causes the ungrammaticality of (56). In other words, the grammatical function of the gap in the intermediate RC is crucial: SUBJ in (54) while OBJ in (56). The f-structure of the IHRC part is displayed in (57), in which the dotted line marks the licensing between the filler (i.e., the innermost SUBJ) and the gap (i.e., PRED) across the innermost ADJ, the intermediate OBJ and the outmost ADJ. It

follows that the HN of IHRCs must license the grammatical subject of all stacked RCs. There is no exception even for the intermediate RCs. Of course, the HN does not have to be a subject in the matrix clause because the matrix clause does not modify it.

(56) * *kut-un* *na'* *tali'* [*qu'* *s‹m›oya* [*m‹in›aniq* *sayun*
behead-PV OBL Tali NOM like<AV>like AV<PRF>eat Sayun
ngarux] *qu'* *watan*].
bear NOM Watan
'The bear, which ate Sayun (and) which Watan likes, will be beheaded by Tali.'

(57) *



In the cases discussed above, both sentential subjects and RC-stacking lead to long-distance dependencies. I propose the constraint of the inside-out functional uncertainty in (58). The first part of the equation (↑ SUBJ) indicates that the filler in an IHRC is a grammatical subject, and the part ((ADJ ↑) PRED) in the second part indicates that it licenses the PRED outside of an ADJ, possibly across none or some grammatical subjects or the RCs modifying subjects (as indicated in the part { SUBJ|SUBJ ADJ}*). This constraint can be annotated under the DP inside the IHRC on condition that the f-structure of its mother node (i.e., IP) corresponds to ADJ, as shown in the part ↑ = (ADJ ↑) in the c-structure rule (59). This condition will exclude the cases of non-IHRCs. Recall that the phrase of an EHRC (corresponding to ADJ in the f-structure) is CP instead of IP, as shown in (38), so the constraint of the inside-out functional uncertainty is inapplicable to EHRCs. As shown in (60), the phrase of an IHRC is IP immediately dominated by DP. Since the HN can be placed in the initial position of the IHRC, as shown in (61), the relative order between I' and DP is not fixed in (59). The f-structure of the IHRC part is presented in (62), exactly like the one in (48). In addition, the complementiser *ka'* cannot be inserted in IHRCs since the c-structure rule (60) does not include any CP.

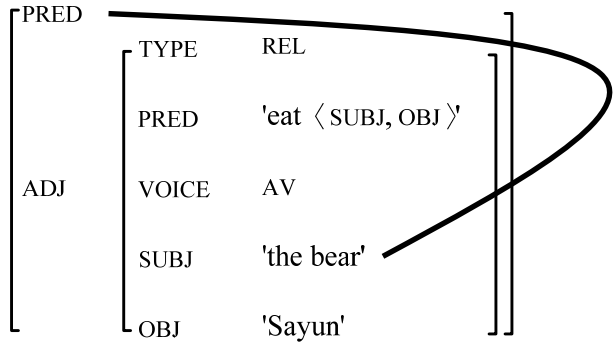
(58) (↑ SUBJ) = ((ADJ { SUBJ|SUBJ ADJ}* ↑)PRED)

(59) IP →I', DP
↑ = ↓ (↑ SUBJ) = ↓
↑ =(ADJ ↑) → (↑ SUBJ) = ((ADJ { SUBJ|SUBJ ADJ}* ↑)PRED)

(60) DP → IP
(↑ ADJ) = ↓

(61) *kut-unna'* *tali'* [*qu'* ***ngarux*** *m‹in›aniq* *sayun*].
behead-PV OBL Tali NOM bear AV<PRF>eat Sayun
'The bear, which ate Sayun, will be beheaded by Tali.'

(62)



I argue in this section that the dependency between the filler and the gap in IHRCs is confined by the inside-out functional uncertainty. Although Falk (2001) claims that subject extraction makes use of the outside-in functional uncertainty, the cases discussed here proves that wrong. The use of what type of functional uncertainty does not simply depend upon what grammatical function is extracted. Generally speaking, it may be determined by the language-particular grammar.

5 Concluding remarks

The paper analyses the two types of RCs in Jianshi Atayal: EHRCs and IHRCs. It is proposed that the outside-in functional uncertainty applies to the former whereas the inside-out functional uncertainty applies to the latter. In our analysis of EHRCs, the licensing between the filler and the gap is mediated with the complementiser *ka'*, which corresponds to the grammatical function OPER in the f-structure. The exceptional cases come from the EHRCs with resumptive pronouns since they apply anaphoric binding. As for the IHRCs, the licensing between the filler and the gap is unmediated, that is, without the existence of OPER. The filler directly licenses the gap through the constraint of the inside-out functional uncertainty. Our analyses are summarised in Table 1.

Table 1: Syntactic mechanisms of EHRCs and IHRCs in Jianshi Atayal

Syntactic Mechanisms	Functional Uncertainty				Anaphoric binding
	Outside-in		Inside-out		
	Mediated (OPER)	Unmediated (OPER-less)	Mediated (OPER)	Unmediated (OPER-less)	
Types	1	2	3	4	5
Jianshi Atayal	EHRCs (ordinary)	--	--	IHRCs	Resumptive EHRCs

Although there are still other possible syntactic analyses, we believe in the economy of expression that syntactic structures should be as close to surface structures and insist on the theoretical simplicity as an important criterion. Moreover, syntactic mechanisms of relativisation vary from construction to construction. No particular analysis is fitting to all RC constructions cross-linguistically or even in the same language.

References

Adger, David and Gillian Ramchand. 2005. Merge and Move: *Wh*-dependencies revisited. *Linguistic Inquiry* 36.2:161–193.

Aldridge, Edith. 2004. Internally headed relative clauses in Austronesian languages. *Language and Linguistics* 5.1:99–129.

Alexiadou, Artemis, Paul Law, André Meinunger and Chris Wilder. 2000. Introduction. *The syntax of relative clauses*, ed. by Artemis Alexiadou et al., 1–51. Amsterdam: John Benjamins.

Alsina, Alex. 2008. A theory of structure-sharing: Focusing on long-distance dependencies and parasitic gaps. *Proceedings of the LFG08 conference*, ed. by Miriam Butt and Tracy Holloway King, 5–25. Stanford: CSLI Publications.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford studies in theoretical linguistics 35. New York: Oxford University Press.

Berman, Judith. 2003. *Clausal syntax of German*. Studies in constraint-based lexicalism. Stanford: CSLI Publications.

Blust, Robert. 1977. The proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *University of Hawai'i at Mānoa Working Papers in Linguistics* 9.2:1–15.

Blust, Robert. 1995. The position of the Formosan Languages: Method and theory in Austronesian comparative linguistics. *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng, 585–650. Symposium series of the Institute of History and Philology, Academia Sinica 3. Taipei: Academia Sinica.

Blust, Robert. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 1. Taipei: Academia Sinica.

Bresnan, Joan. 1982. Control and complementation. *The mental representation of grammatical relations*, ed. by Joan Bresnan, 282–390. Cognitive Theory and Mental Representation. Cambridge, MA: MIT Press.

Bresnan, Joan. 1995. Linear order, syntactic rank, and empty categories: On weak crossover. *Formal issues in Lexical-Functional Grammar*, ed. by Ronald M. Kaplan, John T. Maxwell, Annie Zaenen, 241–274. CSLI Lecture Notes 47. Stanford: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell textbooks in linguistics 16. Oxford: Blackwell.

Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63.4:741–782.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño and Frédérique Segond. 1999. *A grammar writer's cookbook*. CSLI lecture notes 95. Stanford: CSLI Publications.

Chang, Melody Ya-yin. 1998. Wh-constructions and the problem of wh-movement in Tsou. MA thesis. Hsinchu: National Tsing Hua University.

Chang, Melody Ya-yin. 2004. Subjecthood in Tsou grammar. PhD dissertation. Hsinchu: National Tsing Hua University.

Chang, Henry Yung-li. 1997. Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu: National Tsing Hua University.

Chen, Sih-Wei. 2007. *Applicative constructions in Atayal*. MA thesis. Hsinchu: National Tsing Hua University.

Chomsky, Noam. 1977. On Wh-Movement. Proceedings of the 1976 MSSB-UC Irvine Conference on the Formal Syntax of Natural Language, 9–11 June, 1976, Newport Beach, California, ed. by Peter W. Culicover, Thomas Wasow and Adrian Akmajian, 71–132. New York: Academic Press.

Chomsky, Noam. 1986. *Barriers*. Linguistic Inquiry Monographs 13. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. On Phases. Unpublished paper at MIT, Cambridge, MA.

Cole, Peter, Wayne Harbert and Gabriella Hermon. 1982. Headless relative clauses in Quechua. *International Journal of American Linguistics* 48.2:113–124.

Dalrymple, Mary. 2001. *Syntax and semantics Vol. 34: Lexical Functional Grammar*. San Diego: Academic Press.

Egerod, Søren. 1965. Verb inflexion in Atayal. *Lingua* 15:251–82.

Egerod, Søren. 1966. Word order and word classes in Atayal. *Language* 42.2:346–369.

Egerod, Søren. 1980. *Atayal-English dictionary*. Scandinavian Institute of Asian Studies Monograph Series No. 35. London and Malmö: Curzon Press.

Egerod, Søren. 1999. *Atayal–English dictionary*. 2nd ed. by Jens Østergaard Petersen Historisk–filosofiske Skrifter 20. Copenhagen: The Royal Danish Academy of Sciences and Letters.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar: An introduction to parallel constraint-based syntax*. Stanford: CSLI Publications.

Falk, Yehuda N. 2010. An unmediated analysis of relative clauses. *Proceedings of the LFG10 Conference*, ed. by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.

Huang, Lillian M. 1991. The semantics of *s-* in Atayal. *Studies in English Literature and Linguistics* 17:37–50.

Huang, Lillian M. 1993. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995a. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995b. The voice system in Atayal. *Papers from the First International Symposium on Languages in Taiwan*, ed. by Feng-fu Tsao and Mei-hui Tsai, 73-98. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 2000. taiyayu cankao yufa [*An Atayal reference grammar*]. Formosan Language Series 1. Taipei: Yuan-Liou Publishing Co.

Kager, René. 1999. *Optimality Theory*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Kaplan, Ronald M. and John T. Maxwell. 1988. An algorithm for functional uncertainty. Coling Budapest: Proceedings of the 12th international conference on computational linguistics, 22-27 August 1988, ed. by Dénes Vargha, 297–302.

Kaplan, Ronald M. and Annie Zaenen. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. *Alternative conceptions of phrase structure*, ed. by Mark R. Baltin and Anthony S. Kroch. Chicago: University of Chicago Press.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Linguistic Inquiry Monographs 25. Cambridge, MA: MIT Press.

Keenan, Edward L., and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8.1:63–99.

Kisserberth, Charles W. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1.3:291–306.

Krifka, Manfred and Renate Musan. 2012. Information structure: Overview and linguistic issues. *The expression of information structure*, ed. by Manfred Krifka and Renate Musan, 1–43. The Expression of Cognitive Categories 5. Berlin: Mouton de Gruyter.

Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.

Li, Paul Jen-kuei. 1985. Linguistic criteria for classifying the Atayalic dialect groups. *Bulletin of the Institute of History and Philology, Academia Sinica* 56.4:699–718.

Li, Paul Jen-kuei. 1991. *taiwan nandaoyuyan de yuyin fuhao xitong* [*Orthographic systems for Formosan languages*]. Taipei: Ministry of Education, ROC. [In Chinese]

Li, Paul Jen-kuei. 1996. *yilanxian nandao minzu yu yuyan* [*The Formosan Tribes and Languages in Yi-lan*]. Monograph Series of Yi-Lan History, Linguistics 1. Yi-lan: Yi-lan County Government. [In Chinese]

Liu, Adlay Kun-long. 2004a. On relativization in Squliq Atayal. MA Thesis. Hsinchu: National Tsing Hua University.

Liu, Adlay Kun-long. 2004b. Though it's small, it makes a HUGE difference: On *KA'* and relative clauses in Squliq Atayal. Paper presented at 2004 National Conference on Linguistics. 24–25 June. Chiayi: National Chung-Cheng University.

Liu, Adlay Kun-long. 2005. The structure of relative clauses in Jianshi Squliq Atayal. *Concentric: Studies in Linguistics* 31.2:89–110.

McCloskey, James. 2006. Resumption. *The Blackwell companion to syntax (Volume IV)*, ed. by Martin Everaert and Henk van Riemsdijk, 94–117. Blackwell handbooks in linguistics 19. Malden, MA: Blackwell Publishing Ltd.

Radford, Andrew. 2009. *Analysing English sentences: A minimalist approach*. Cambridge Textbooks in Linguistics. New York: Cambridge University Press.

Rau, Der-Hwa Victoria. 1992. *A grammar of Atayal*. Taipei: Crane Publishing Co., Ltd.

Rau, Der-Hwa Victoria. 2004. Lexical similarity, sound change and intelligibility of Atayalic dialects. *Papers in Austronesian subgrouping and dialectology*, ed. by John Bowden and Nikolaus P. Himmelmann, 37-95. Pacific linguistics 563. Canberra: The Australian University.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. by Alexander Adelaar and Andrew Pawley, 295–326. Pacific linguistics 601. Canberra: The Australian National University.

Schachter, Paul. 1973. Focus and relativization. *Language* 49.1:19–46.

Tang, Chih-Chen Jane. 2008. Functional extension vs. grammaticalization: A typological study of modification markers in Formosan nominals. *Language and Linguistics* 9:917–966.

Teng, Stacy Fang-ching. 1997. Complex sentences in Puyuma. MA thesis. Taipei: National Taiwan Normal University.

Tsai, Wei-Tien Dylan. 1997. On the absence of island effects. *Tsing Hua Journal of Chinese Studies* 27.1:125–149.

Tsai, Wei-Tien Dylan. 1999a. *On economizing the theory of A-Bar dependencies*. New York: Garland Publishing, Inc.

Tsai, Wei-Tien Dylan. 1999b. On Lexical Courtesy. *Journal of East Asian Linguistics* 8.1:39–73.

Vergnaud, Jean Roger. 1974. *French relative clauses*. PhD dissertation. Cambridge, MA: Massachusetts Institute of Technology.

Wu, Joy J. 1995. Complex sentences in Amis. MA thesis. Taipei: National Taiwan Normal University.

Wu, Joy J. 1996. Relative Clauses in Amis. *Studies in English Literature and Linguistics* 22:115–146.

Wu, Joy J. 1997. Relative clauses in some Formosan languages. Paper read at the Eighth International Conference on Austronesian languages. December 28-30. Taipei: Academia Sinica.

Zhang, Niina. 2004. Move is remerge. *Language and Linguistics* 5.1:189–209.

9 *Subjecthood and temporal adjuncts in Atayal, Seediq and Tsou**

WEI-TIEN DYLAN TSAI

1 Introduction

In some Formosan languages, Tgdaya Seediq and Tsou in particular, temporal adverbials pattern with nominative arguments in undergoing focus movement, topicalisation, and relativisation. All of them involve A'-movement or A'-dependencies of some sort.[1] The pattern is even more striking in Squliq Atayal, which I will choose as a target language in this study, with supporting evidence from Tgdaya Seediq and Tsou when necessary.

Let's start with a rundown of basic facts about this "temporal-nominative parallel". In Squliq Atayal, nominative *wh*'s cannot stay in situ, as in (1), and they must undergo A'-movement of some sort, as in (2) and (3):[2]

(1) * *m-usa* *Sincik* *suxan* *qu-ima?* (**wh*-in-situ)
AV-go Hsinchu tomorrow NOM-who
Intended for: 'Who will go to Hsinchu tomorrow?'[3]

---

* This article is a tribute to Professor Lillian M. Huang, whose devotion to the cause of Formosan linguistics has set an example for all of us. I am grateful to Henry Y. Chang, Sandra Chung, Hui-chuan J. Huang, Edward Keenan, Paul Law, Paul Jen-kuei Li, David Pesetsky, Eric Potsdam, Joachim Sabel, Stacy F. Teng, Lisa Travis, Joy J. Wu and Elizabeth Zeitoun for their helpful comments in various occasions. The research leading to this article was funded by the National Science Council of Taiwan (NSC 95-2411-H-007-022).

1 This phenomenon has been noted, and studied quite extensively in a series of MA theses and PhD dissertations produced in National Tsing Hua University in Taiwan (C. Chang 1996, H. Chang 1997, M. Chang 1998, 2004).

2 The abbreviations of this paper follow those of the Leipzig Glossing Rules. Other abbreviations not included there are: AV, actor voice; IV, instrumental voice; LV, locative voice; PV, patient voice; REAL, realis mood; REL, relativiser.

3 The basic word order of Squliq Atayal is VOS. The declarative counterpart of (1) is fully grammatical, as shown below:
(i) *m-usa* *Sincik* *suxan* *qu-Temu.*
AV-go Hsinchu tomorrow NOM-Temu
'Temu will go to Hsinchu tomorrow.'

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 211–222
Asia-Pacific Linguistics, 2015

(2) *ima qu-*[$_{DP}$ e_i [$_{CP}$ Op_i [*m-usa Sincik suxan* t_i]]]? (*wh*-pseudo-cleft)
who NOM AV-go Hsinchu tomorrow
'Who is (the person who) go to Hsinchu tomorrow?'[4]

(3) *ima*(*-*ga*) *m-usa Sincik suxan*? (focus movement)
who(*-TOP) AV-go Hsinchu tomorrow
'Who will go to Hsinchu tomorrow?'

In declarative sentences with actor voice (AV), the nominative arguments may either stay at the sentence-final position, as in (4), or undergo A'-construals such as pseudo-cleft, as in (5), and topicalisation, as in (6):

(4) *m-usa Sincik suxan qu-Temu.* (AV declarative)
AV-go Hsinchu tomorrow NOM-Temu
'Temu will go to Hsinchu tomorrow.'

(5) *Temu qu* [$_{DP}$ e_i [$_{CP}$ Op_i [*m-usa Sincik suxan* t_i]]]. (pseudo-cleft)
Temu NOM AV-go Hsinchu tomorrow
'(The person who) goes to Hsinchu tomorrow is Temu.'

(6) *Temu-ga m-usa Sincik suxan.* (topicalisation)
Temu-TOP AV-go Hsinchu tomorrow
'Temu, (he) will go to Hsinchu tomorrow.'

As for non-nominative arguments, (7) shows that the locative object *inu* 'where' must remain in situ, whereas (8) and (9) show that it can not undergo either pseudo-cleft formation nor focus movement:

(7) *m-usa suxan inu qu-Temu*? (*wh*-in-situ)
AV-go tomorrow where NOM-Temu
'Where will Temu go tomorrow?'

(8) * *inu qu* [*m-usa suxan qu-Temu*]? (*wh*-pseudo-cleft)
where NOM AV-go tomorrow NOM-Temu
Intended for: 'Where is (the place) Temu will go tomorrow?'

(9) * *inu m-usa suxan qu-Temu*? (focus movement)
where AV-go tomorrow NOM-Temu
Intended for: 'Where will Temu go tomorrow?'

The declarative counterparts of *inu* 'where' are also blocked from A'-construals, as evidenced by the deviance of (10)–(12):

---

[4] For one thing, (2) may look like an instance of *wh*-movement. But the thing is that there is a nominative case marker in between *ima* 'who' and the rest of the sentence. So it is more likely that (2) is an equational sentence, where the subject is a headless relative. Therefore, it is an empty relative operator that undergoes A'-movement in (2). The same observation applies to its declarative counterpart in (5). In a sense, Formosan languages lend further support to the pseudo-cleft approach to this type of constructions (Paul 2001; Pearson 2001; Potsdam 2004, among others), in that there is a nominative case marker after the *wh*-phrase, serving to mark the rest of the sentence as its subject.

(10) * *Sincik qu* [$_{DP}$ e_i [$_{CP}$ Op_i [*m-usa* t_i *suxan qu-Temu*] (pseudo-cleft)
Hsinchu NOM AV-go tomorrow NOM-Temu
Intended for: '(The place where) Temu will go tomorrow is Hsinchu.'

(11) * *Sincik-ga m-usa suxan qu-Temu.* (topicalisation)
Hsinchu-TOP AV-go tomorrow NOM-Temu
Intended for: 'To Hsinchu, Temu will go tomorrow.'

(12) * [[*m-usa suxan qu-Temu*] *ka slaq*] *ga krahu yal.* (relativisation)
AV-go tomorrow NOM-Temu REL wet.field TOP big very
Intended for: 'As for the wet field where Temu will go tomorrow, (it) is very big.'

How about adverbial expressions? The instrumental *how*, *m-ha-nanu* 'with what', behaves more like a predicate. Sentence (13) shows that it is inflected with voice morphology, and remains in situ in a verbal position:

(13) *m-ha-nanu mhiy ume qu-Temu*? (*wh*-in-situ)
AV-with-what strike.AV plum NOM-Temu
'How will Temu knock down plums?'

Moreover, it resists all kind of A'-construals, as evidenced by the deviance of (14)–(15):

(14) * (*m-*)*ha-nanu qu* [*mihiy ume qu-Temu*]? (*wh*-pseudo-cleft)
AV-with-what NOM strike.AV plum NOM-Temu
Intended for: 'What is (the way) Temu will knock down plums?'

(15) * (*m-*)*ha-nanu-ga mihiy ume qu-Temu*? (focus movement)
AV-with-what-TOP strike.AV plum NOM-Temu
Intended for: 'How will Temu knock down plums?'

Again, its declarative counterparts are not compatible with pseudo-cleft, topicalisation, and relativisation, as exemplified in (16)–(18) respectively:

(16) * *ruma' qu* [*mihiy ume qu-Temu*]. (pseudo-cleft)
bamboo.stick NOM strike.AV plum NOM-Temu
Intended for: '(The instrument with which) Temu will knock down plums is a bamboo stick.'

(17) * *ruma'-ga mihiy ume qu-Temu.* (topicalisation)
bamboo.stick-TOP strike.AV plum NOM-Temu
Intended for: 'With a bamboo stick, Temu will knock down plums.'

(18) * [[*mihiy ume qu-Temu*] *ka ruma'*] *krahu yal.* (relativisation)
strike.AV plum NOM-Temu REL bamboo.stick big very
Intended for: 'As for the bamboo stick with which Temu will knock down plums, (it) is big.'

As a matter of fact, this nominative/non-nominative asymmetry is just an instance of a much broader generalisation concerning the subjecthood of Austronesian languages, often dubbed as "subject sensitivity" (Keenan 1976; Keenan & Comrie 1977; Guilfoyle et al. 1992; Huang 1995; H. Chang 1997, among many others).

Nevertheless, there is an exception to this picture. In Squliq Atayal, temporal expressions are subject to A'-construals, as in (19)–(21):

(19) *suxan* *qu* [*m-usa* *Sincik* *qu-Temu*]. (pseudo-cleft)
tomorrow NOM AV-go Hsinchu NOM-Temu
'(The time when) Temu will go to Hsinchu is tomorrow.'

(20) *suxan-ga* *m-usa* *Sincik* *qu-Temu.* (topicalisation)
tomorrow-TOP AV-go Hsinchu NOM-Temu
'Tomorrow, Temu will go to Hsinchu.'

(21) [[*m-usa* *Sincik* *Temu*] *ka* *riax*] *ga* *byacing-ka-nial.* (relativisation)
AV-go Hsinchu Temu REL time TOP month-of-coming
'The time Temu will go to Hsinchu is the coming month.'

Furthermore, just like *inu* 'where' of (8), (9) and *m-ha-nanu* 'how' of (14), (15), a temporal *wh* may stay in situ, as in (22). But unlike them, *knwan* 'when' may actually undergo A'-movement, as in (23), (24):

(22) *m-usa* *Sincik* *knwan* *qu-Temu*? (*wh*-in-situ)
AV-go Hsinchu when NOM-Temu
'When will Temu go to Hsinchu?'

(23) *knwan* *qu* [*m-usa* *Sincik* *qu-Temu*]? (*wh*-pseudo-cleft)
when NOM AV-go Hsinchu NOM-Temu
'When is (the time) Temu will go to Hsinchu?'

(24) *knwan* *m-usa* *Sincik* *qu-Temu*? (focus movement)
when AV-go Hsinchu NOM-Temu
'When will Temu go to Hsinchu?'

The same pattern is attested in two other Formosan languages Tsou and Tgdaya Seediq as well (cf. C. Chang 1996 and M. Chang 2004).[5] The following table summarises all the facts from the three languages:

(25)

	AV interrogative		AV declarative	
	in situ	A'-movement	in situ	A'-movement
nominative argument	*	ok	ok	ok
temporal adverbial	ok	ok	ok	ok
accusative/oblique argument	ok	*	ok	*
instrumental adverbial	ok	*	ok	*

Putting this temporal-nominative parallel in perspective, I would like to address the following issues. First, what is the common property shared by nominative arguments and temporal adverbials? Second, what factors are responsible for the subject sensitivity effects? Can we provide a coherent account in terms of economy principles? Lastly, why does nominative *wh*'s behave differently from other *wh's*-in-situ?

[5] But see H. Chang (1997) for a marginal contrast between nominative and temporal *wh's* in Tgdaya Seediq.

2 A working hypothesis

To solve the first puzzle, I would like to pursue a line laid out by Pesetsky & Torrego (2001), as stated in (26), where nominative case is analysed as an uninterpretable Tense feature on D:

(26) Nominative case is an instance of an uninterpretable Tense feature (*u*T) on D.

Under this approach, a nominative *wh* in English bears both uninterpretable *wh*- and T-feature, i.e., [*u*Wh, *u*T]. It then raises to check [*u*Wh, *u*T] on C, as illustrated in (27a). As a result, T-to-C movement is blocked, which accounts for the deviance of (27b):

(27) a. Who read the book?
[$_{CP}$ who$_{[\sout{uT}, \sout{uWh}]}$ [$_{C'}$ C$_{[\sout{uT}, \sout{uWh}]}$ [$_{IP}$ __ read the book]]]

b. * Who did read the book?
[$_{CP}$ who$_{[\sout{uT}, \sout{uWh}]}$ [$_{C'}$ did$_{[\sout{uT}]}$ C$_{[\sout{uT}, \sout{uWh}]}$ [$_{IP}$ __ read the book]]]

The head movement is blocked because the derivation of (27a) use lesser number of movement operations than (27b). Namely, (27a) only involves *wh*-movement, while (27b) employs both *wh*-movement and T-to-C movement. As for non-subject *wh*'s which bears only [*u*Wh], T-to-C movement must apply to check off the T-feature on C, as illustrated in (28):

(28) What did Mary read?
[$_{CP}$ what$_{[\sout{uWh}]}$ [$_{C'}$ did$_{[\sout{uT}]}$ C$_{[\sout{uT}, \sout{uWh}]}$ [$_{IP}$ Mary read __]]]

The problem here is that the object *wh* is not the closest XP to C. Pesetsky and Torrego then resort to the principle of Minimal Compliance proposed by Richards (1997), according to which one instance of movement can save another by observing locality principles minimally.[6] As a result, after T-to-C movement applies in (28), the object *wh* is free to check the *wh*-feature on C. Following this dichotomy between subject and object *wh*'s, I would like to propose that temporal adverbials, just like nominative arguments, bear Tense features. The difference is that the T-feature encoded by a nominative case is uninterpretable, whereas the one on a temporal adverbial is interpretable, i.e., *i*T.

It follows that a nominative argument must raise to [Spec, TP] to check off its *u*T, as illustrated below:

(29) [$_{TP}$ [$_{T'}$ T$_{[\sout{uT}]}$ [$_{VP}$ . . .]] DP$_{[\sout{uT}]}$]

By contrast, the interpretable T-feature cannot be checked off, and therefore need not to be checked. Temporal adverbials are therefore not required to raise. If a temporal adverbial did occupy [Spec, TP], then the closest T-feature to C is the *i*T in [Spec, TP]. As a result, there will be no way for a nominative DP to check off its *u*T, which would crash the derivation:

(30) * [$_{TP}$ [$_{T'}$ T$_{[\sout{uT}]}$ [$_{VP}$ . . . DP$_{[uT]}$ . . .]] XP$_{[iT]}$]

---

[6] Following is a simplified version of Minimal Compliance formulated in Pesetsky & Torrego (2001): Once T-to-C movement applies, the attracted XP need not to be the closest one.

This distinction between nominative arguments and temporal adverbials is supported by their distinct syntactic distributions, i.e., while nominative arguments typically occupy a sentence-final position, presumably [Spec, TP], temporal adverbials such as *suxan* 'tomorrow' occur much more freely, as in (31a–c). Note that the only place they cannot occur is in between the predicate and the locative object *Sincik*, as in (31d). We will come back to this issue below.

(31) a. ***suxan*** *m-usa* *Sincik* *qu-Temu.*
tomorrow AV-go Hsinchu NOM-Temu
'Temu will go to Hsinchu tomorrow.'

b. *m-usa* *Sincik* *suxan* *qu-Temu.*
AV-go Hsinchu tomorrow NOM-Temu

c. *m-usa* *Sincik* *qu-Temu* *suxan.*
AV-go Hsinchu NOM-Temu tomorrow

d. * *m-usa* *suxan* *Sincik* *qu-Temu.*
AV-go tomorrow Hsinchu NOM-Temu

Here I follow Pesetsky & Torrego (2001) in assuming that the *u*T on D remains active until the completion of the CP phase. Otherwise, the nominative *wh* would be frozen in [Spec, TP] once its *u*T is checked off. Furthermore, I would like to suggest that sentences involving A'-construals are headed by C with a *u*T, very much in line with Rizzi's (1997) proposal that there is a head expressing finiteness in the CP layer. As a result, we have two options to check off the *u*T on C: One is to raise a nominative argument with the *u*T-feature, and the other is to raise a temporal adverbial with the *i*T-feature.[7]

To deal with the second puzzle, suppose further that there is no T-to-C movement in Squliq Atayal, which seems to be a plausible assumption in view of the fact that a tensed verb cannot raise across a temporal adverbial, as evidenced by the contrast between (31a)

[7] An anonymous reviewer points out a technical issue given Chomsky's (2001) characterisation of feature agreement, where a Probe-Goal relation is established between an uninterpretable feature on the probe and an interpretable feature on the goal. The system is designed as such so that Full Interpretation can be met (that uninterpretable feature can be erased/checked through a matching interpretable one). Since under our approach both C and T carry *u*T, it becomes problematic as to how they pair with *u*T on a nominative argument.

One way to tackle this issue, as proposed by the reviewer, is to adopt Pesetsky & Torrego's (2007) later proposal, which dissociates interpretability from feature valuation. Under the new design, one may assume that C and T both carry interpretable T-feature (or some type of EPP feature), but the T feature of C is not valued. A nominative argument carries *u*T with no value, and a temporal adjunct carries *u*T with a value. When T probes, the nominative DP is the only Goal (under feature complementariness), hence valued after agreement and movement to [Spec, TP]. Due to the valuation, when C probes, both the nominative argument and the temporal adjunct are equally valued (assuming multiple agreement in the sense of Hiraiwa (2005) and Chomsky (2008)), but one of them then raises to [Spec, CP] to value the unvalued T-feature on C.

and (31d).[8] Then it follows from the Minimal Compliance account that non-nominative arguments, as well as instrumental adverbials, are blocked from A'-construals.

Now the last puzzle. Let me first point out that *wh*'s in situ in these languages behave like polarity items. Just like in Chinese, they allow extensive indefinite usages when scoped over by modals, negation, yes-no question, and conditional operators (cf. Tsai 1997, 2003). Therefore, we may safely assume that, in these Formosan languages, *wh*-nominals do not have intrinsic operator features. It follows that they are subject to unselective binding from an implicit Q-morpheme on C (cf. Baker 1970; Pesetsky 1987; Nishigauchi 1990; Tsai 1994), by virtue of being able to introduce a choice function variable in situ (cf. Reinhart 1998), as sketched in the following schema:

(32) $[_{CP}\,[_{C'}\,Q_f\,[_{IP}$ ... f(wh) ...]]]

On the other hand, a nominative *wh* cannot remain in situ, because it must raise to check off the *u*T on C. In which case, the polarity *wh* would raise beyond the scope of the question operator on C, which may well bleed the configuration of unselective binding, as shown below:

(33) * $[_{CP}$ f(wh) $[_{C'}\,Q_f\,[_{IP}$ ... t ...]]]

This results in vacuous quantification. Consequently, the only way to ask a subject *wh*-question in these languages is to employ either pseudo-cleft formation or focus movement. By contrast, a temporal *wh* enjoys the freedom of either remaining in situ, or raising to [Spec, CP].

To sum up, nominative arguments and temporal adverbials pattern together in forming A'-dependencies because both are endowed with T-features. Their distinct distributions are reduced to a difference in the interpretability of their T-features, namely, *u*T vs *i*T.

3 Typological correlations

There are also a couple of typological correlations which might be regarded as the supporting evidence for our proposal. A point of interest has to do with the fact that there is no temporal voice throughout the Formosan languages, whereas there are often voices for patients, instruments, and locations. The following examples are from Squliq Atayal:[9]

(34) *mihiy* *ume* *qu-Temu.* (Actor Voice)
strike.AV plum NOM-Temu
'Temu will knock down plums.'

(35) *bhy-un* *na-Temu* *qu-ume.* (Patient Voice)
strike-PV OBL-Temu NOM-plum
'Plums will be knocked down by Temu.'

---

[8] Another way to look at the impossibility of (31d), as noted by an anonymous reviewer, is to say that the VOS order results from remnant VP raising, hence blocking further construals of the temporal adjunct within.

[9] The verb root for 'strike' is *bihiy*. The various forms in (34)–(37) result from different combinations of the root and voice/aspect affixes.

(36) *b‹in›hy-an* *ume* *na-Temu* *qu-slaq.* (Locative Voice)
<PFV>strike-LV plum OBL-Temu NOM-wet.field
'Plums was knocked down by Temu in the wet field.'

(37) *s-bihiy* *ume* *na-Temu* *qu-ruma'.* (Instrumental Voice)
IV-strike plum OBL-Temu NOM-bamboo.stick
'Plums will be knocked down by Temu with a bamboo stick.'

This means that temporal expressions can never be a subject. The reason is very simple from our perspective: They bear an interpretable T-feature by default, and thus in the same league as nominative arguments in their ability to license A'-construals. We may therefore correlate the absence of temporal voice with the extractability of temporal *wh*-expressions.

Another relevant fact is that temporal *wh*-phrases sometimes carry realis/irrealis morphology. A good example comes from Tsou. There is a contrast between *ne-homna* 'realis when' of (38a) and *ho-homna* 'irrealis when' of (39a), which must agree with their respective tense modals, as evidenced by the deviance of (38b) and (39b) (cf. M. Chang 2004):

(38) a. *m-o* *baito* *ta- mo'o* *ne-homna* *'o- pasuya*?
AV-REAL see.AV OBL-Mo'o REAL-when NOM-Pasuya
'When did Pasuya see Mo'o?'

b. * *m-o* *baito* *ta- mo'o* *ho-homna* *'o- pasuya*?
AV-REAL see.AV OBL-Mo'o IRR-when NOM-Pasuya

(39) a. *te* *baito* *ta-mo'o* *ho-homna* *'o- pasuya*?
IRR see.AV OBL-Mo'o IRR-when NOM-Pasuya
'When will Pasuya see Mo'o?'

b. * *te* *baito* *ta-mo'o* *ne-homna* *'o- pasuya*?
IRR see.AV OBL-Mo'o REAL-when NOM-Pasuya

The same pattern is attested in Seediq, as shown by the contrast between *knuwan* 'irrealis when' of (40a) and *sknuwan* 'realis when' of (41a) (cf. C. Chang 1996):

(40) a. *knuwan* *ka* [*mi-imah* *sino* *tama*]?
when.IRR NOM IRR-drink wine father
'When is (the time) Father will drink wine?'

b. * *sknuwan* *ka* [*mi-imah* *sino* *tama*]?
REAL.when NOM IRR-drink wine father

(41) a. *sknuwan* *ka* [*m‹n›atak* *tunux* *na* *laqi* *ka-payi*]?
REAL.when NOM AV<PFV>cut head of child NOM-grandmother
'When is (the time) Grandmother cut the child's hair?'

b. * *knuwan* *ka* [*m‹n›atak* *tunux* *na* *laqi* *ka-payi*]?
IRR.when NOM AV<PFV>cut head of child NOM-grandmother

Again, they must agree with the tense morphology of the main predicate. This strongly recalls the future morphology marked on nominative DPs in Pittapitta (cf. Hale 1998; Pesetsky & Torrego 2001):

(42) a. *ngapiri-ngu* *thawa* *paya-nha.*
father-FUT kill bird-ACC
'Father will kill the bird (with missile thrown).'

b. *thithi-ngu* *karnta* *pathiparnta.*
elder.brother-FUT go morning
'My elder brother will go in the morning.'

We therefore have a good reason to group temporal adverbials and nominative arguments together in view of their ability to take tense morphology.

4 Concluding remarks

If our analysis is on the right track, it becomes possible to give a more precise characterisation of at least some of the peripheral/EPP features. Our proposal provides yet another reason for relating tense features to A'-construals in the left periphery.

Moreover, subject sensitivity is reduced to checking the *u*T-feature on C. Temporal adverbials form a natural class with nominative arguments in both syntactic and morphological terms: As IP-adverbials, they can be merged *after* the completion of the vP phase. As temporal expressions, they bear interpretable Tense features.

Given the extension of the *u*T analysis to accusative/oblique arguments in Pesetsky & Torrego (2004), the subject sensitivity effects can be accommodated in two ways. Either accusative/oblique arguments check their *u*T with T, or they bear interpretable T-features as a legacy of inherent case-marking. Just like temporal adjuncts merged *before* the completion of the vP phase, they cannot undergo further A'-movement.

References

Baker, C. L. 1970. Notes on the description of English questions: The role of an abstract question morpheme. *Foundations of Language* 6.2:197–219.

Chang, Chung-Liang. 1996. saidekeyu yiwenci de yanjiu [A study of Seediq interrogatives] MA thesis. Hsinchu: National Tsing Hua University. [In Chinese]

Chang, Melody Ya-yin. 1998. Wh-constructions and the problem of wh-movement in Tsou. MA thesis. Hsinchu: National Tsing Hua University.

Chang, Melody Ya-yin. 2004. Subjecthood in Tsou grammar. PhD. dissertation. Hsinchu: National Tsing Hua University.

Chang, Henry Y. 1997. Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu: National Tsing Hua University.

Cheng, Lisa Lai-Shen. 1991. On the typology of wh-questions. PhD dissertation. Cambridge, MA: Massachusetts Institute of Technology.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels and Juan Uriagereka, 89–155. Current studies in linguistics 28. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1–52. Current studies in linguistics 36. Cambridge, MA: MIT Press.

Chomsky, Noam. 2008. On phases. *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta 133–166. Current studies in linguistics 45. Cambridge, MA: MIT Press.

Guilfoyle, Eithne, Henrietta Hung and Lisa Travis. 1992. Spec of IP and spec of VP: two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10.3:375–414.

Hale, Ken. 1998. A note on the Pittapitta nominative case and the future tense. Ms, Massachusetts Institute of Technology.

Hiraiwa, Ken. 2005. Dimensions of symmetry in syntax: agreement and clausal architecture. PhD dissertation. Cambridge, MA: MIT Press.

Holmer, Arthur J. 1996. *A parametric grammar of Seediq*. Travaux de l'Institut de linguistique de Lund 30. Lund: Lund University Press.

Hsiao, Stella I-Ling. 2003. *A preliminary investigation on temporal adverbials in Squliq Atayal*. Ms. Hsinchu: National Tsing Hua University.

Huang, Lillian M. 1995. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Keenan, Edward L. 1976. Remarkable subjects in Malagasy. *Subject and topic*, ed. by Charles N. Li, 247–301. New York: Academic Press.

Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.1:63–99.

Mei, Kuang. 1994. taiyayu (wenshui fangyan) de yuxu, gewei han jiaose duixie [Word order, case, and theta-agreement in Mayrinax Atayal] Paper presented at the First Symposium on Austronesian Languages of Taiwan, Academia Sinica, Taipei, Taiwan. [In Chinese]

Nishigauchi, Taisuke. 1990. *Quantification in the theory of grammar*. Studies in Linguistics and Philosophy Vol. 37. Dordrecht and Boston: Kluwer Academic Publisher.

Paul, Ileana. 2001. Concealed Pseudo-clefts. *Lingua* 111.10:707–727.

Pearson, Matthew. 2001. The clause structure of Malagasy: A minimalist approach. PhD dissertation. Los Angeles: UCLA.

Pesetsky, David. 1987. Wh-in-situ: movement and unselective binding. *The representation of (in)definiteness*, ed. by Eric J. Reuland and Alice G. B. ter Meulen 98–129. Current studies in linguistics studies 14. Cambridge, MA: MIT Press.

Pesetsky, David and Esther Torrego. 2001. T-to-C movement: Causes and consequences. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 355–426. Cambridge, MA: MIT Press.

Pesetsky, David and Esther Torrego. 2004. Tense, case, and the nature of syntactic categories. *The syntax of time*, ed. by Jacqueline Guéron and Jacqueline Lecarme, 495–537. Current Studies in Linguistics Series 37. Cambridge, MA: MIT Press.

Pesetsky, David and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. *Phrasal and clausal architecture: Syntactic derivation and interpretation in honor of Joseph E. Emonds*, ed. by Simin Karimi et al., 262–294. Amsterdam: John Benjamins.

Potsdam, Eric. 2004. Wh-Questions in Malagasy. In *Proceedings of AFLA 11. ZAS Papers in Linguistics 34*, ed. by Paul Law, 244-258. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Reinhart, Tanya. 1998. Wh-in-situ in the framework of the minimalist program. *Natural Language Semantics* 6.1:29–56.

Norwin W. Richards III. 1997. What moves where when in which language? PhD dissertation. Cambridge, MA: Massachusetts Institute of Technology.

Rizzi, Luigi. 1997. The fine structure of the left periphery. *Elements of grammar: Handbook in generative syntax,* ed. by Liliane Haegeman, 281-337. Kluwer International Handbooks in Linguistics Vol. 1. Dordrecht and Boston: Kluwer Academic Publisher.

Tsai, Wei-Tien Dylan. 1994. On economizing the theory of A-bar dependencies. PhD dissertation. Cambridge, MA: Massachusetts Institute of Technology.

Tsai, Wei-Tien Dylan. 1997. taiwan nandaoyu yiwencide wudinyongfa – gamalanyu zouyu ji saidekeyu de bijiao yanjiu [Indefinite *Wh*-construals in Formosan languages: A comparative study of Kavalan, Tsou, and Seediq]. *Tsing Hua Journal of Chinese Studies, New Series* 27.4:381–422. [In Chinese]

Tsai, Wei-Tien Dylan. 2003. Lexical courtesy revisited: Evidence from Tsou and Seediq *wh*-constructions. *Gengo Kenkyu* 2003.123:331–361.

10 *Ditransitive alignment and causative constructions in Seediq*

NAOMI TSUKIDA

1 Introduction

The main purpose of this paper is to demonstrate that in Seediq, a Formosan language with a Philippine type voice system, the patient (P) of monotransitive verbs, the theme (T) of ditransitive verbs and the recipient (R) of ditransitive verbs show secundative alignment in Malchukov et al.'s (2010) terminology (see section 2), and show how secundative alignment is realised (sections 3.1 and 3.2). The alignment is observed in subject-voice assignment. I will then mention four points not observed in languages with no focus system (section 3.3). Lastly, I will touch upon subject-voice assignment in Seediq causative constructions (section 4).

2 Seediq language

A brief summary of Seediq grammar is given below. Please refer to Tsukida (2005, 2009) for more details.

2.1 General information

Seediq, an Austronesian language, is spoken in the central and northeastern part of Taiwan. There are about thirty thousand Seediq people, only some of whom can speak the language fluently. It is thus endangered, as young people and children cannot speak it (Tang 2011). Seediq includes three dialects: Truku, Tgdaya and Toda. This research is based on the Truku dialect.

2.2 Word order

The word order of Seediq is predicate initial and subject final. Arguments other than the subject and adjuncts appear between the predicate and the subject. (In the following sentences, the subject is underlined.)

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 223–251
Asia-Pacific Linguistics, 2015

(1) *mekan* *bu'ut* *ka* *huling*.[1]
eat.AV[2] bone NOM dog
'The dog ate a bone.'

In an NP, modifiers follow the head noun except for quantity expressions.

(2) *deha* *huling* *kumu* *gaga.*
two dog Kumu DIST
'Those two dogs of Kumu's.'

Seediq has some prepositions, but no postpositions.

2.3 Voice system

Seediq is a 'Philippine-type language' with a complex voice system. I will show what this voice system is like in general and how it functions in Seediq.

In this voice system, one of the arguments (including so-called adjuncts) is chosen as the subject. In Seediq, the subject is marked by *ka*, which is analysed as a nominative case marker, and is grammatically salient (Tsukida 2009:311–318). Semantic roles are grouped into three, and verb forms show which one of the three groups is in subject position. These three groups are ACTOR, GOAL and CONVEYANCE. Verb forms are thus in one of three voices: AV (actor voice), GV (goal voice), or CV (conveyance voice). The grouping of semantic roles is the main issue of this paper.

Examples in (3) show the three focus forms of the verb *'usa* 'to go'.

(3) a. *musa* *karingku* *ka* *laqi.*
go.AV Hualien NOM child
'The child goes to Hualien.'

b. *sa-'an* *laqi* *ka* *karingku.*
go-GV2 child NOM Hualien
'A/The child goes to Hualien.'

c. *se-'usa* *laqi* *karingku* *ka* *payi.*
CV-go child Hualien NOM old.woman
'A/The child goes to Hualien for the old woman.'

Example (3a) is an AV sentence. The verb form *musa* indicates that the semantic role of the subject *ka laqi* 'child' is among those which are classified into ACTOR group. Example (3b) is a GV sentence. The verb form *sa-'an* indicates that the semantic role of the subject *ka karingku* 'Hualien' is among those which are classified into GOAL group. Example (3c) is a CV sentence. The verb form *se-'usa* indicates that the semantic role of

---

[1] The data used in this paper based on fieldwork done in Jin-mei village and Cong-de village, Xiu-lin County, Hualien, Taiwan. The main informants are Tian Xin-de and the late Fang A-qi.

[2] The abbreviations which are indicated in the Leipzig Glossing Rules are as follows: AV, actor voice; BN, bound nominative; CNJ, conjunction; CPL, completive; CV, conveyance voice; FN, free nominative; GV, goal voice; INCPL, incomplete; IV, instrumental voice; LIN, linker; NAF, non-agent focus; NFIN, non-finite; NRF, non-referential; O, object; PF, patient focus; RF, referential.

the subject *ka payi* 'old woman' is among those which are classified into CONVEYANCE group. These three voices are summarised in Table 1.

Table 1: Three voices of *'usa* 'to go' and semantic roles

	Semantic role of the subject	Group which the semantic role of the subject is classified into	Voice	Verb form
(3a)	Agent	ACTOR group	AV	*musa*
(3b)	Goal	GOAL group	GV	*sa-'an*
(3c)	Beneficiary	CONVEYANCE group	CV	*se-'usa*

Many so-called adjuncts, such as beneficiary or instrument can also become subject, but there certainly is some limitation. With some verbs, not all the roles are classified into the three groups, and some roles cannot become the subject. With the verb 'to go', for example, instrument cannot be grouped into the CONVEYANCE group or any other role, and cannot become the subject.

Examples in (4) show three voice forms of the verb *kerut* 'to cut'.

(4) a. *k‹em›erut* *babuy* *ngangut* *ka masaw*.
<AV>cut pig outside NOM Masaw
'Masaw slaughters a/the pig outside.'

b. *keret*[3]*-an* *masaw* *ka babuy*.
cut-GV2 Masaw NOM pig
'Masaw slaughters the pig.'

c. *keret-an=na* *babuy* *ka keti'inuh niyi*.
cut-GV2=3SG.GEN pig NOM board PROX
'S/He cuts pig on this board.'

d. *se-kerut* *babuy* *masaw* *ka baki*.
CV-cut pig Masaw NOM old.man
'Masaw slaughter a/the pig for the old man.'

e. *se-kerut* *babuy* *masaw* *ka puting*.
CV-cut pig Masaw NOM knife
'Masaw slaughters a/the pig with the knife.'

In example (4a), *masaw*, the agent, is chosen as the subject and is in nominative case. The verb form is *k‹em›erut* and indicates that the group the semantic role of the subject belongs to is the ACTOR group and thus this is an AV sentence. In example (4b), the patient NP *babuy* 'pig' is chosen as the subject. The verb form *keret-an* indicates that the group the semantic role of the subject belongs to is the GOAL group and this is an GV sentence. Example (4c) is also a GV sentence. The locative NP *keti'inuh niyi* 'this board' is chosen as the subject, which is grouped into the GOAL group. The verb form is *keret-an* and indicates that the group the semantic role of the subject belongs to is the GOAL group. In example (4d), the beneficiary NP *baki* 'old man' is chosen as the subject, which is grouped into the CONVEYANCE group (also including instrument). The verb form

[3] The verb 'to cut' has two allomorphs: One is *keret*, which appears when suffixed, and the other is *kerut*, which appears otherwise.

se-kerut indicates that the group the semantic role of the subject belongs to is the CONVEYANCE group. Example (4e) is also a CV sentence. The noun *putiŋ* 'knife' is chosen as the subject and its semantic role is instrument, which is grouped into the CONVEYANCE group. The verb form *se-kerut* indicates that the group the semantic role of the subject belongs to is the CONVEYANCE group.

Table 2: The voice of *kerut* 'to cut' and the semantic role of the subject

	Semantic role of the subject	Group which the semantic role of the subject is classified into	Voice	Verb Form
(4a)	Agent	ACTOR	AV	*k‹em›erut*
(4b)	Patient	GOAL	GV	*keret-an*
(4c)	Location	GOAL	GV	*keret-an*
(4d)	Beneficiary	CONVEYANCE	CV	*se-kerut*
(4e)	Instrument	CONVEYANCE	CV	*se-kerut*

It may be unexpected for those who are familiar with a four-voice system of other Formosan languages/dialects that I classify √-*an* form, as well as √-*un* form, into GV, and call the latter GV1 and the former GV2. Some explanation seems to be necessary at this point. In Truku Seediq, they mainly differ in terms of tense. When a location, goal, source, recipient or patient is the subject, and when the clause expresses a future situation, then GV1 form appears regardless of the semantic role of the subject. When one expresses a present situation, then GV2 appears quite pervasively, except in limited cases of patient subject, where GV1 appears. Tsukida (2000) analysed this in terms of 'punctualness' of the subject. It seems that GV1 appears when the patient is affected as a whole, whereas the GV2 appears when the patient is partially affected.

Table 3 shows the summary of the grouping of semantic roles according to voice choice.

Table 3: Summary of the grouping of semantic roles according to voice choice

Group	Semantic roles
ACTOR	Agent (example (4a)), Theme (of intransitive verbs) (example (3a))
GOAL	Goal (example (3b)), Patient (example (4b)), Recipient (example (9b)), Location (example (4c)), Source (example (31b))
CONVEYANCE	Conveyed theme (example (9c)), Instrument (example (4e)), Beneficiary (example (4d))

The grouping of recipient, source and conveyed theme will be the main issue of discussion in section 3.

2.4 Case marking system

In Seediq, NPs which are chosen as subjects are marked by *ka*, as is mentioned in section 1.3. This is an independent nominative form. When the subject is first or second person, it appears in the form of a clitic nominative pronoun, and the independent form may be omitted, as in (5a), or may appear, as in (5b), whereas clitic form is obligatory.

(5) a. *m-begay=ku* *laqi=mu* *sunan.*
AV.FUT-give=1SG.NOM child=1SG.GEN 2SG.OBL
'I will give my child to you.'

b. *s‹em›ipaq=ku* *leqi-'an* *ka* *yaku.*
‹AV›hit=1SG.NOM child-OBL NOM 1SG
'I hit a/the child.'

Those NPs which are not chosen as subjects appear between the verb and the subject. The semantic roles of non-subject NPs are indicated by the case of the NP, a prepositional phrase, and partly by word order. There are four cases: Nominative, genitive, direct, and oblique. The non-subject agent appears in genitive case. As for personal pronouns, the genitive case is clitic. Patient, goal, recipient, locative, and beneficiary appear in oblique case. Theme and instrument appear in direct case, which is the citation form. Source appears as a prepositional phrase. How these roles are expressed in a sentence is summarised in Table 4.

Table 4: The manifestation of case roles of the non-subject NPs

Semantic role	Case
Agent (=*na* '=3SG.GEN' in (4c), *masaw* 'Masaw' in (4b), (4d), (4e), *laqi* 'child' in (3b), (3c))	Genitive
Conveyed theme (*laqi* 'child' in (5a)), Instrument	Direct
Patient (*bu'ut* 'bone' in (1), *babuy* 'pig' in (4a), (4c), (4d), (4e), *leqi-'an* 'child-OBL' in (5b)), Goal (*karingku* 'Hualien' in (3c)), Recipient (*sunan* '2SG.OBL' in (5a)), Locative (*ngangut* 'outside' in (4a)), Beneficiary	Oblique
Source	(Prepositional Phrase)

There is case syncretism in Seediq. Only those NPs higher in animacy hierarchy have oblique and genitive distinct from direct case, as shown in table 5. Therefore, word order becomes important when more than two non-subject NPs appear. Such instances, however, are rather rare in natural discourse, though in elicitation one may obtain such sentences.

Table 5: Illustration of Seediq case assignment

	First and second pronoun	Third pronoun	Personal names and some human nouns	Common nouns that express animals and inanimate things
	ex. *yaku* 'I'	ex. *hiya* 'S/he'	ex. *laqi* 'child'	ex. *babuy* 'pig'
GEN	*=mu*	*=na*	*laqi*	*babuy*
Independent NOM	*ka yaku*	*ka hiya*	*ka laqi*	*ka babuy*
Clitic NOM	*=ku*[4]	—	—	—
DIR (=citation form)	*yaku*	*hiya*	*laqi*	*babuy*
OBL	*kenan*	*hiya-'an*	*leqi-'an*	*babuy*

[4] Clitic NOM is obligatory but independent NOM is optional.

3 Indirective and secundative alignment

The main issue of this paper is the grouping of semantic roles, and more particularly the voice/subject assignment in Seediq monotransitive clauses and ditransitive clauses. In Seediq ditransitive clauses, secundative alignment (see Malchukov et al. 2010:3) is observed. Before introducing the Seediq situation in the next section, here we will briefly review the concept of secundative alignment.

Dryer (1986) is one of the early works on indirective and secundative. He used the terms DO/IO and PO/SO, respectively. Malchukov et al. (2010) is a good summary on this topic. They summarise indirective and secundative patterns as follows:

Indirective	T=P≠R
Secundative	T≠P=R

(P = Patient of monotransitive clause; T = Theme of ditransitive clause; R = Recipient of ditransitive clause)

In indirective alignment, the patient of a monotransitive clause (P) and the theme of a ditransitive clause (T) are treated in the same way. The recipient of a ditransitive clause is treated differently from P and T. P and T are called direct object, and R is called indirect object. In secundative alignment, by contrast, the patient of a monotransitive clause (P) and the recipient of a ditransitive clause (R) are treated in the same way. The theme of a ditransitive clause is treated differently from P and R. There are also *neutral alignment* (T=P=R), *horizontal alignment* (T=R≠P), and *tripartite alignment* (T≠P≠R), although the latter two are rare.

According to Malchukov et al. (2010), such alignments can be observed in terms of coding properties, as well as behavioural properties. Coding properties include case marking (flagging), word order, and agreement (indexing). Behavioural properties include passivisation, antipassivisation, relativisation, constituent questions, reflexivisation, reciprocalisation, nominalisation, incorporation, and direct-inverse marking (Malchukov et al. 2010:8–18). The Tzutujil example below shows how indirective alignment is realised in case marking.

(6) Tzutujil (Dayley 1985:63, 156, cited in Malchukov et al. 2010:8–9)

a. *x-at-kee-ch'ey*
CPL-2SG.ABS-3PL.ERG-hit
'They hit you.'

b. *N-Ø-kee-ya7* *paq* *cha-qe*
INCPL-3SG.ABS-3PL.ERG-give money to-1PL
'They will give money to us.'

As can be seen, P and T are indexed by absolutive, but R is not indexed at all.

The Tarahumara example below shows how indirective alignment is realised in word order.

(7) Tarahumara (Blansitt 1984:138, cited in Malchukov et al. 2010:9)

a. A-P-V
Siríame *muní* *go'áre.*
chief beans ate
'The chief ate beans.'

b. A-T-V-R
Siríame muní áre mukí.
chief beans gave woman
'The chief gave the woman beans.'

As can be seen, P and T are on the left side of the verb, but R is on the right side.

Languages may show different alignments in coding and in behavioural pattern. Some languages may show different alignments in different kinds of coding properties or different kinds of behavioural properties. So Amharic, for example, shows indirective alignment in case-marking, but secundative alignment in agreement.

(8) Amharic (Amberber 2009:747, cited in Malchukov et al. 2010:10)
ləmma lə-liǰ-u məs'haf sət'-ə-(w).
Lemma to-child-DEF book give.PRF-3M-3M.O
'Lemma gave the book to the child.'

In the example above, case marking shows indirective alignment (the recipient is specially marked by dative), but indexing on the verb shows secundative alignment because R is controlling suffixal person-number indexing on the verb (Malchukov et al. 2010:10).

4 Secundative alignment in Seediq

Let us now turn to Seediq. I will briefly discuss the alignment in coding properties in section 3.1, and then return to behavioural properties in section 3.2. In section 3.3, I will point out some peculiarities concerning secundative alignment. In section 3.4 I will point out some lexical factors. In section 3.5 I will discuss the motivation of secundative alignment.

4.1 Coding properties

In this section I discuss case marking, word order, and agreement.

4.1.1 Case marking

We have to be careful when looking at case marking in languages with 'focus' system, however, that is look at the case marking of P, T, and R when they are not the subject. The case marking of NPs when they are not the subject was already shown in Table 4. The relevant part is summarised in Table 6 for the sake of clarity.

Table 6: Seediq case forms of Patient, Recipient, and Theme

Semantic role	Case
Patient	Oblique case (√-*an* or √-zero)
Recipient	Oblique case (√-*an* or √-zero)
Theme	Direct case (√-zero)

If the oblique case form is realised by √-zero, then Recipient = Patient = Theme, that is, neutral alignment. If oblique case form is realised by √-*an*, then Recipient = Patient ≠

Theme, that is, secundative alignment. Since those NPs higher in animacy hierarchy have √*-an* form, we can say that NPs higher in animacy show secundative alignment and those lower show neutral alignment.

4.1.2 Word order

When both the theme and the recipient appear at the same time, the theme precedes the recipient, as in example (5) in section 2.4. The order cannot be reversed. The theme directly follows the verb, in the same way as the patient in monotransitive verb does (*babuy* in (4a)). So, as for word order, the alignment is Recipient ≠ Patient = Theme, that is, indirective.

4.1.3 Agreement

As for agreement, only the subject NP and the ACTOR NP cause pronoun cliticisation on the predicate, as we saw in section 2.4. The patient, theme, and recipient are irrelevant unless they are chosen as subject. So we can say the alignment is neutral.

4.1.4 Summary

We can summarise the alignment types of Seediq coding properties as not so consistent as follows:

Case marking: Neutral or secundative (depending on animacy)
Word order: Indirective
Agreement: Neutral

4.2 Behavioural properties

In this section I examine behavioural properties: subject/voice assignment, relativisation, nominalisation, constituent question, reflexivisation, and reciprocalisation.

4.2.1 Subject/voice assignment

I have already shown that the subject/voice assignment of monotransitive verbs in section 2.3. Here I look at that of ditransitive verbs.

(9) a. *megay* *bunga* *leqi-'an* *ka* *bubu.*
give.AV sweet.potato child-OBL NOM mother
'The mother gave sweet potato(es) to the/a child.'

b. *biq-an* *bunga* *bubu* *ka* *laqi.*
give-GV2 sweet.potato mother NOM child
'The/A Mother gives sweet potato(es) to the child.'

c. *se-begay* *bubu* *leqi-'an* *ka* *bunga.*
CV-give mother child-OBL NOM sweet.potato
'Mother gave the sweet potato(es) to the/a child.'

Example (9a) is a sentence in ACTOR Voice. The verb form *megay* indicates that the semantic role of the subject *ka bubu* 'mother' is among those which are classified into the ACTOR group. Example (9b) is a sentence in GOAL Voice. The verb form *biq-an* indicates that the semantic role of the subject *ka laqi* 'child' is among those which are classified into the GOAL group. Example (9c) is sentence in CONVEYANCE Voice. The verb form *se-begay* indicates that the semantic role of the subject *ka bunga* 'sweet potato' is among those which are classified into the CONVEYANCE group. The voice of the verb *begay* 'to give' is summarised in Table 7.

Table 7: The voice of *begay* 'to give' and the semantic role of the subject

	Semantic role of the subject	Group which the semantic role of the subject is classified into	Voice	Verb Form
(9a)	Agent	ACTOR	AV	*megay*
(9b)	Recipient	GOAL	GV	*biq-an*
(9c)	Conveyed theme	CONVEYANCE	CV	*se-begay*

Let us now compare Table 2, which shows that the voice assignment of a monotransitive verb and Table 7, which shows the assignment pattern of a ditransitive verb. It is summarised in Table 8.

Table 8: Subject/voice assignment of Seediq monotransitive and ditransitive clauses

Voice/Group	Monotransitive	Ditransitive
AV/ ACTOR	Agent	Agent
GV/ GOAL	Patient, Location	Recipient
CV/ CONVEYANCE	Instrument, Beneficiary	Conveyed theme

The patient of a monotransitive clause and the recipient of a ditransitive clause behave in the same way; they are classified into the GOAL group and become the subjects in GV clauses. The theme of a ditransitive clause behaves differently from them; it is classified into the CONVEYANCE group and becomes subject in a CV clause. It behaves like the instrument or beneficiary of a monotransitive verb. It is clearly a case of secundative alignment.

Patient = Recipient ≠ Theme ► secundative
Verbs such as 'to teach' and 'to tell' show the same pattern.

(10) *t‹em›egesa* 'to teach'

a. *t‹em›egesa* *kari* *teruku* *leqi-'an* *ka bubu*.
<AV>teach language Teruku child-OBL NOM mother
'The mother teaches Teruku language to the/a child.'

b. *tegesa-an* *kari* *teruku* *bubu* *ka laqi*.
teach-GV2 language Teruku mother NOM child
'The mother teaches Teruku language to the child.'

c. *se-tegesa* *bubu* *leqi-'an* *ka kari teruku*.
CV-teach mother child-OBL NOM language Teruku
'The mother teaches Teruku language to the/a child.'

(11) *r‹em›engaw* ‘to tell, to speak’

a. *r‹em›engaw* *kari* *niyi* *leqi-’an* *ka bubu.*
<AV>tell story PROX child-OBL NOM mother
‘The mother tells this story to the/a child.’

b. *rengag-an* *kari* *niyi* *bubu* *ka laqi.*
tell-GV2 story PROX mother NOM child
‘The/A mother tells this story to the child.’

c. *re~rengaw*[5]*=mu* *sediq-un*[6] *ka kari niyi.*
FUT~CV.talk=1SG.GEN person-OBL NOM story PROX
‘I will tell this story to people.’

The person to whom one shows something and the addressee behave the same way as the patient of monotransitive clauses. They become subject in GV clauses. Things shown or messages, on the other hand, become CV subject.

Other Formosan languages and also some Philippine languages also attest of secundative alignment in voice choice (e.g., Tgdaya Seediq (Holmer 1999); Tsou (Chang 2011); Atayal (Huang 1995); Kimaragang Dusun and Tatana’ (Kroeger 2009), among others). The examples (12) and (13) are from Atayal (Formosan). Atayal is a language neighboring Seediq and forms the Atayalic subgroup together with Seediq. Subjects are underlined. Example (12) is a monotransitive sentence where the patient is the subject. The verb form has the suffix *-un*. Example (13a) is a ditransitive sentence where the recipient is the subject (indicated by the bound nominative pronoun *=cu*). The verb form has the suffix *-un*. Example (13b) is a ditransitive sentence where the theme is the subject. The verb form has the prefix *si-*. If the patient of the monotransitive verb or the recipient of the ditransitive verb is the subject, morpheme *-un* marks the verb form, but if the theme of the ditransitive verb is the subject, *si-* marks the verb form. That is, R=P≠T, or a secundative pattern.

(12) Atayal monotransitive (Huang 1995:45)
niq-un *nku’* *nabakis* *ku’ bunga’.*
eat-PF GEN.RF old.man NOM.RF sweet.potato
‘The old man ate the sweet potato.’

(13) Atayal ditransitive (Huang 1995:47)

a. *kal-un=cu* *ni’* *watan* *cku’* *sinubilan* *cu’* *kai’.*
say-PF=1SG.BN GEN Watan ACC.RF legend LIN language
‘Watan told me a legend.’

b. *si-kaal* *ni’* *watan* *’i’* *kuing*
NAF-say GEN Watan ACC 1SG.FN
ku’ sinubilan cu’ kai’.
NOM.RF legend LIN language
‘Watan told the legend to me.’

---

[5] CV future form is the root, with no affix. It is often accompanied by partial reduplication.

[6] *sediq-un* is an oblique case of *se’diq* ‘people’, utilising *-un* suffix, which is irregular.

Example (14) is from Kimaragang Dusun, and example (15) is from Tatana', both Philippine languages (cited in Kroeger 2009). The conveyed theme is the subject in each example, and the verb form is what Kroeger calls its IV form, which is the equivalent of the CV form in this paper. This form is distinct from what is used when the patient of the monotransitive or the recipient of the ditransitive is the subject.

(14) Kimaragang Dusun (Kroeger 2009:3)

i-taak	*ku*	*iti siin*		*sid*	*tanak ku.*
IV-give	1SG.GEN	this[NOM]	money	DAT	child 1SG.GEN

'I will give this money to my child.'

(15) Tatana' (Dillon 1994:48, cited in Kroeger 2009:4)

i-taak	*ku*	*ani'*	*dokou*	*bua' diti.*	
IV-give	1SG.GEN	only	2SG.DAT	fruit	this

'I am just giving this fruit to you (no price).'

4.2.2 Relativisation and nominalisation

Relativisation and nominalisation use the voice system, and the grouping of semantic roles into three groups stays the same with that of subject/voice assignment. In Seediq, only subjects are relativised. One has to make the targeted nominal into the subject by changing the voice of the clause before relativising it. Examples in (16) show the relativisation of monotransitive clauses. When the agent is relativised, the relative clause must be in AV, as in (16a), not in GV, as in (16b). When the patient is relativised, the relative clause must be in GV, as in (16c), not in AV, as in (16d). Compare the following sentences with examples (4a) and (4b).

(16) Relativisation: Monotransitive

a. *senaw* [*k‹em›erut* *babuy*]
man <AV>cut pig
'the man who slaughtered a/the pig'

b. * *senaw* [*keret-an* *ka* *babuy*]
man cut-GV2 NOM pig
Intended meaning: 'the man who slaughtered a/the pig'

c. *babuy* [*keret-an* *masaw*]
pig cut-GV2 Masaw
'the pig which was slaughtered by Masaw'

d. * *babuy* [*k‹em›erut* *ka* *masaw*]
pig <AV>cut NOM Masaw
Intended meaning: 'the pig which was slaughtered by Masaw.'

Examples in (17) show the relativisation of ditransitive clauses. Please compare these examples in (17a), (17b), and (17c).

(17) Relativisation: Ditransitive

a. *kuyuh* [*m‹en›egay* *bunga* *leqi-'an*]
woman AV<PRF>give sweet.potato child-OBL
'a/the woman who gave a sweet potato to a/the child'

b. *laqi* [*biq-an* *bunga* *bubu*]
child give-GV2 sweet.potato mother
'a/the child who was given a sweet potato'

c. *bunga* [*se-begay* *bubu* *leqi-'an*]
sweet.potato CV-give mother child-OBL
'a/the sweet potato which was given to the child by mother'

As for nominalisation, Seediq uses the same morphology as the predicate form of each voice. The predicate of (9a), *megay bunga leqi-'an*, for example, can be used as a nominal, meaning 'the one who gives sweet potato(es) to a/the child', as in (18a). The predicate of (9b), *biq-an bunga bubu*, can be used as a nominal meaning 'the one to whom mother gives a sweet potato', as in (18b). The predicate of (9c), *se-begay bubu leqi-'an*, can be used as a nominal meaning 'what mother gives to the child', as in (18c).

(18) a. *bubu* *ka* [*megay* *bunga* *leqi-'an*].
mother NOM give.AV sweet.potato child-OBL
'The one who gave sweet potato(es) to the/a child is the mother.'

b. *laqi* *ka* [*biq-an* *bunga* *bubu*].
child NOM give-GV2 sweet.potato mother
'The one to whom the/a mother gives sweet potato(es) is the child.'

c. *bunga* *ka* [*se-begay* *bubu* *leqi-'an*].
sweet.potato NOM CV-give mother child-OBL
'What mother gave to the/a child is the sweet potato.'

There are more fixed expressions such as *mpe-tegesa* 'the one who is teaching, teacher', which is 'AV.FUT-teach', and *te~tegesa* 'what will/should be taught', which is 'FUT~CV.teach'.

4.2.3 Constituent questions

There are two types of constituent questions. One is the *wh*-cleft construction and the other is in situ *wh*-construction. In *wh*-cleft construction, the nominalised form is used which was dealt with in section 3.2.2. The nominalised form is made the subject. Since the nominalised form shows a secundative alignment, *wh*-cleft sentences do so as well. Examples:

(19) a. *'ima* *ka* [*megay* *bunga* *leqi-'an*]?
who NOM give.AV sweet.potato child-OBL
'Who gave sweet potato(es) to the/a child?'

b. *'ima* *ka* [*biq-an* *bunga* *bubu*]?
who NOM give-GV2 sweet.potato mother
'To whom did the/a mother give sweet potato(es)?'

c. *manu* *ka* [*se-begay* *bubu* *leqi-'an*]?
what NOM CV-give mother child-OBL
'What did mother give to the/a child?'

In in situ *wh*-construction, the word to be questioned is replaced with an interrogative, in sentences where the word is not the subject. The word cannot be replaced if it is the subject of the sentence. In sentence (20) (=(9a)), neither the recipient nor the conveyed theme is the subject, so one can replace either of them with an interrogative, as in (21b) and (21c). However, (21a) is ungrammatical, because the agent which is the subject of the sentence is being replaced. One can ask who the agent was by using (19a) above.

(20) *megay* *bunga* *leqi-'an* *ka* *bubu.*
give.AV sweet.potato child-OBL NOM mother
'The mother gave sweet potato(es) to the/a child.'

(21) a. * *megay* *bunga* *leqi-'an* *ka* *'ima*?
give.AV sweet.potato child-OBL NOM who
'Who gave sweet potato(es) to the/a child?'

b. *megay* *manu* *leqi-'an* *ka* *bubu*?
give.AV what child-OBL NOM mother
'What did the mother gave to the/a child?'

c. *megay* *bunga* *ma-'an* *ka* *bubu*?
give.AV sweet.potato who-OBL NOM mother
'To whom did the mother give a sweet potato(es)?'

No difference is observed between recipient and conveyed theme in terms of possibility of in situ *wh*-construction. Concerning in situ *wh*-construction, the alignment is neutral.

4.2.4 Reflexivisation and reciprocalisation

In Seediq, reflexivity is expressed by a pronoun and an intensifier *nanak*, as in (22).

(22) *tendahu* *hiya* *nanak* *ka* *rabay.*
praise.AV 3SG by.oneself NOM Rabay
'Rabay praises herself.'

There are no particular restrictions on the use of *nanak* in T or R arguments of ditransitive constructions.[7] One can either say to V something to oneself, as in (23a) and (23b), or to V oneself to somebody, as in (24).

(23) a. *pe'adas* *patas* *hiya* *nanak* *ka* *rabay.*
send.AV letter 3SG by.oneself NOM Rabay
'Rabay sends letter to herself.' (*hiya nanak* as a non-subject argument position)

b. *pedes-un=na* *patas* *ka* *hiya* *nanak.*
send-GV1=3SG.GEN letter NOM 3SG by.oneself
'S/He sends letters to herself/himself.' (*hiya nanak* as the subject)

[7] There is a restriction on *nanak* concerning ACTOR argument. It cannot appear in an ACTOR argument in an AV clause.

(24) *tumun* *'u,* *wada=na* *se-pegekela* *rebiq-an*
Tumun CNJ PST=3SG.GEN CV-introduce Rubiq-OBL
ka *hiya* *nanak.*
NOM 3SG by.oneself
'Tumun introduced herself to Rubiq.' (*hiya nanak* as the subject)

Thus, as far as reflexivisation is concerned, the alignment of P, T, and R in Seediq is neutral.

Let us look at reciprocals then. According to Malchukov et al. (2010:39), it seems to be a general tendency among world's languages that reciprocals derived from ditransitive verbs are likely to mean 'to V something to each other', rather than to mean 'to V each other to somebody/something'. The former is secundative and the latter is indirective.

In Seediq, reciprocals, marked on verbs through partial reduplication and a prefix, mean 'to V something to each other', as (25) shows. Thus, as far as reciprocalisation is concerned, the alignment in Seediq is secundative.

(25) a. *megay* *patas* *mesag-an* *ka* *rubiq* *ni,*
give.AV letter Masaw-OBL NOM Rubiq and
megay *patas* *rebiq-an* *ka* *masaw* *'uri.*
give.AV letter Rubiq-OBL NOM Masaw also
'Rubiq gives letter to Masaw, and Masaw gives letter to Rubiq, too.'

b. *m-be~begay* *patas* *ka* *rubiq* *ni* *masaw.*
AV-RECP~give letter NOM Rubiq and Masaw
'Rubiq and Masaw give letters to each other.'

4.2.5 Summary

We can summarise the alignment types of Seediq behavioural properties as secundative in most cases, but neutral in some cases, as follows:

Subject/voice assignment: Secundative
Relativisation: Secundative
Nominalisation: Secundative
Wh-cleft construction: Secundative
In situ *wh*-construction: Neutral
Reflexivisation: Neutral
Reciprocalisation: Secundative

4.3 Peculiarities concerning secundative alignment in voice system and discussion

There are some interesting differences between Seediq and the ordinary situation of PO/SO (= secundative) languages shown in Dryer (1986). I already mentioned that the most important one, i.e., the difference is not in the way the object is treated but rather in the assignment of subject/voice. Four other differences are discussed below. (See also Tsukida 2012.)

(i) No alternation

Some languages exhibit an alternation between indirective and secundative alignment. In English, for example, there is such an alternation.

(26) a. Mary gave a book to John.
b. Mary gave John a book.

(27) a. John loaded hay onto a truck.
b. John loaded a truck with hay.

In Seediq, there are no such alternations. It is natural since such an alternation is a device to put the recipient/goal argument into grammatical object status, which is the second salient status. In this language, grammatical object status is not so salient, and it is easy to make recipient/goal argument into subject, the most salient status. So, such a device is unnecessary.

(ii) Treatment of beneficiary

The second point to note is the treatment of beneficiary. Dryer (1986:839) argues that beneficiary becomes a PO by advancement, though initially it is not. I quote an English example below. In this sentence *Mary* is the beneficiary and treated as PO.

(28) John baked Mary a pie. (Dryer 1986:838)

Malchukov et al. (2010:2) state that the beneficiary construction, in which a beneficiary argument occurs, is a closely related construction type to ditransitie construction, and in many languages it is expressed like a ditransitive construction.

In Seediq, the beneficiary is never treated as a GV subject, but is treated as a CV subject, as we have already seen in example (4d).

(iii) Treatment of goal

In previous studies on PO/SO languages, sentences with goal argument were not treated in detail, and there are few examples of goals as PO, such as:

(29) Give the tree water (Dryer 1986:816, example (21b))

The recipient is usually human and the goal is non-human. 'The tree' in (29) may be analysed as recipient, but the fact that it is non-human allows us to analyse it as goal.

In Seediq, goals of placement verbs such as 'to paint', 'to cover', 'to tie', 'to write', or 'to stuff', which are usually non-human, are classified into GOAL group and become GV subjects, in the same way as recipients of ditransitive verbs do. Themes of placement verbs, on the other hand, become CV subjects. When one says 'to paint a wall with house paint', 'the wall' is the GV subject and 'the house paint' is the CV subject. Compare the following examples with the Seediq placement verb *risuh* 'to paint'.

(30) a. *ga* *r‹em›isuh* *pingki* *qenabil* *ka* *tama.*
PROG <AV>paint paint wall NOM father
'Father is painting paint on the wall.'

b. *resuh-un=na* *pingki* *niyi* *ka* *qenabil* *hini.*
paint-GV1=3SG.GEN paint PROX NOM wall here
'He will paint this paint on wall here.'

c. *re~risuh=na* *qenabil* *ka* *pingki.*
FUT~CV.paint=3SG.GEN wall NOM paint
'He will paint the paint on the wall.'

The voice/subject assignment of *risuh* 'to paint' is summarised in Table 9.

Table 9: The voice of *risuh* 'to paint' and the semantic role of the subject

	Verb Form	Semantic role of the subject
AV	*r‹em›isuh*	Agent
GV	*resuh-an*	Goal
CV	*se-risuh*	Conveyed theme

This pattern applies to the verbs below.

		CONVEYANCE	GOAL
r‹em›isuh	'to paint'	paint	wall
h‹em›ilaw	'to cover'	blanket	body
q‹em›apang	'to apply'	ointment	body part
mugul	'to tie'	thing tied	place to be tied to
matas	'to write'	things written	papers to be written on
s‹em›puy	'to stuff'	things stuffed	place to where stuffed

So, when one says 'to apply ointment to a body part', 'body part' is the GV subject and 'the ointment' is the CV subject. When one says 'to tie a cow to a stake', 'stake' is the GV subject and 'cow' is the CV subject. Please also refer to Tsukida (2012) for detail.

(iv) Treatment of source

When verbs of separation, such as 'to take away something from somebody/something' are used in perfective form, the source of separation is governed by the GOAL group and becomes GV subject (see (31b)); the separated theme is treated in the same way as the Themes of ordinary ditransitive verbs are and becomes CV subject (31c). Source is not necessarily animate: It can also be inanimate.

(31) a. *g‹em›abal* *sepih* *qempahan=na* *ka* *rubiq.*
<AV>weed weed field=3SG.GEN NOM Rubiq
'Rubiq pulls out weeds in her field.'

b. *g‹en›bal-an* *rubiq* *ka* *qempahan=na.*
<PRF>weed-GV Rubiq NOM field=3SG.GEN
'Rubiq weeded her field.'

c. *g‹en›abal* *rubiq* *ka* *sepih* *niyi.*
<CV.PRF>[8]weed Rubiq NOM weed PROX
'Rubiq pulled out these weeds.'

[8] In some studies on Atayalic languages or dialects, ‹*en*›√ form and ‹*en*›√-an form are analysed differently: ‹*en*›√ form is analysed as patient voice perfect/past and ‹*en*›√-*an* form is analysed as locative voice perfect/past. I analysed these forms of Truku-Seediq according to their use with three-place verbs, which shows clear distribution. With two-place verbs, it is a little messy: Some patients allow ‹*en*›√ form as the predicate, but not ‹*en*›√-*an* form; some patients only allow ‹*en*›√-an form, and still others allow both forms, when they are the subject. See Tsukida (2009, 2012) for details.

Malefactive source also becomes GV subject, as in (32).

(32) *g‹en›guy-an=deha* *qenqaya* *ka* *sapah=mu.*
<PRF>steal-GV=3PL.GEN thing NOM house=1SG.GEN
'They stole things from my house.'

When the verb is not perfective, the voice assignment is the same as that of monotransitive verbs. The CV form is reserved for beneficiary, as in (33a), and both the theme and source are governed by GOAL and become GV subject, as in (33b) and (33c).

(33) a. *se-gabal* *sepih* *rubiq* *ka* *baki.*
CV-weed weed Rubiq NOM old.man
'Rubiq pulls out weed for the old man.'

b. *gebal-un* *rubiq* *ka* *sepih* *qempahan=na.*
weed-GV1 Rubiq NOM weed field=3SG.GEN
'Rubiq will pull out the weed in the field.'

c. *gebal-i* *sepih* *ka* *qempahan=su.*
weed-GV.NFIN weed NOM field=2SG.GEN
'Pull out the weed from your field.'

(v) Discussion

Malchukov et al. (2010:52) presents a semantic map for Recipient and related functions, as follows.

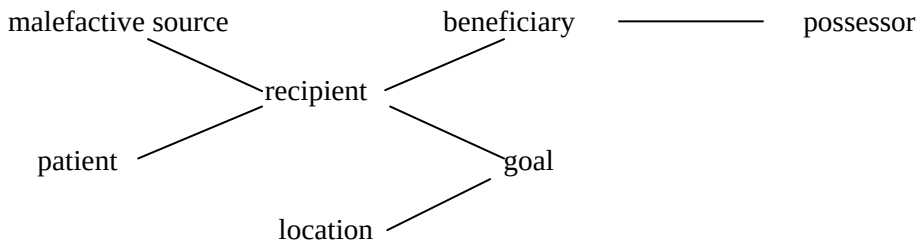


Figure 1: Semantic map for Recipient and related functions
(Malchukov et al. 2010:52)

In Seediq, the GOAL group covers a contiguous area in this map; that is, recipient, malefactive source, patient, goal and location. It does not cover beneficiary or possessor, as shown in Figure 2.

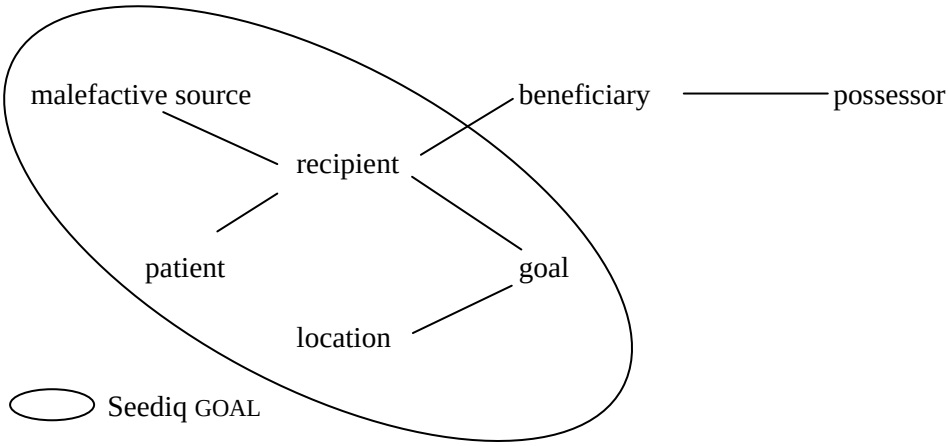


Figure 2: Seediq GOAL group and semantic map for recipient and related functions

4.4 Lexical variation in ditransitive construction

In Seediq, patient of monotransitive verbs is generally treated as GOAL. Some verbs treat its patient as CONVEYANCE. Verbs of ballistic motion, for example, are among such verbs. As for *q‹en›ada* 'to throw', for example, what is thrown becomes the subject in GV sentence, as in (34b), and the goal to which something is thrown becomes GV subject, as in (34c).

(34) a. *q‹en›ada* *sudu* *ngangut* *ka* *lawking.*
<AV>throw trash outside NOM Lawking
'Lawking throws trash outside.'

b. *wada* *se-qada* *tama* *ka* *patas=mu.*
PST CV-throw father NOM book=1SG.GEN
'Father threw my books away.'

c. *qeda-'an=deha* *sudu* *ka* *kulu* *niyi.*
throw-GV2=3PL.GEN trash NOM box PROX
'They throw trash in this box.'

The subject voice assignment is summarised in the Table 10 below.

Table 10: The voice of qada 'to throw' and the semantic role of the subject

Voice/Group	What becomes the subject
AV/ACTOR	Agent
GV/GOAL	Goal
CV/CONVEYANCE	Conveyed theme

The pattern above is that of ditransitive verbs. This means that Seediq *qada* 'to throw' behaves as a ditransitive verb. What is thrown is treated as conveyed theme.

Some verbs hold the middle between verbs like *k‹em›erut* 'to cut' and *q‹em›ada* 'to throw': The patient can be either a CV subject or a GV subject. The verb *mapa* 'to carry on one's back' is such a verb. In both (35a) and (35b), the child, who is carried on someone's

back, is the subject, and the voice of the sentence can be either GV, as in (35a), or CV, as in (35b).

(35) a. *pan-un=na* *bekuy=na* *ka* *laqi.*
carry.on.back-GV1=3SG.GEN back=3SG.GEN NOM child
'S/He carries/will carry a/the child on her/his back.'

b. *'apa*[9] *yawyaw* *ka* *laqi.*
CV.FUT.carry.on.back Yawyaw NOM child
'Yawyaw will carry the child on her back.'

The container for carrying something can become a GV subject, as in (36).

(36) *pan-un=su* *bawyak* *ka* *tawkan* *niyi.*
carry.on.back-GV1=2sg.GEN wild.boar NOM mesh.backpack PROX
'You carry/will carry wild boar in this mesh backpack.'

The voice alternations of *mapa* 'to carry on one's back' and corresponding semantic role of their subject are summarised in Table 11.

Table 11: The correspondence between voice and the semantic role of subject

Voice/Semantic group	What becomes the subject
AV/ACTOR	Agent
GV/GOAL	container (example (36)), Patient (example (35a))
CV/CONVEYANCE	Patient (example (35b))

The patient of ditransitive verbs like *megay* 'to give' are treated as CONVEYANCE, but the patient of some ditransitive verbs is treated either as CONVEYANCE or as GOAL, just like the patient of *mapa* 'to carry on one's back'. The patient of *g‹em›barig* 'to sell' and that of *masug* 'to deal' can be a GV subject or a CV subject. The relation between voice and semantic role of the subject for the verbs *g‹em›barig* 'to sell' and *masug* 'to deal' is summarised in Table 12.

Table 12: *g‹em›arig* 'to sell' and *masug* 'to distribute'

Voice/Semantic group	*gebarig* 'to sell'	*'asug* 'to distribute'
AV/ACTOR	Agent	Agent
GV/GOAL	Recipient (example (39)), Theme (example (37a))	Recipient (example (40)), Theme (example (38a))
CV/CONVEYANCE	Theme (example (37b))	Theme (example (38b))

Examples:

(37) a. *geberig-un=mu* *ka* *padiq* *niyi.*
sell-GV1=1SG.GEN NOM vegetable PROX
'I sell this vegetable.'

[9] The root of the verb 'to carry on one's back' shows somewhat irregular alternation: *pan* (when suffixed) vs *'apa* (otherwise). Root form is used as CV future form, as mentioned in note 4.

b. *g‹en›ebarig=mu* *se'diq* *ka* *padiq.*
<CV.PRF>sell=1SG.GEN person NOM vegetable
'I sold vegetables to people.'

(38) a. *sug-un* *rubiq* *dutut=na* *ka* *siyang.*
distribute-GV1 Rubiq relative=3SG.GEN NOM pork
'Rubiq will distribute the pork among her relatives.'

b. *seg-ani* *dutut=su* *ka* *siyang* *niyi.*
distribute-CV.NFIN relative=2SG.GEN NOM pork PROX
'Distribute this pork among your relatives.'

Recipient also can be a GV subject.

(39) *geberig-un=mu* *padiq* *ka* *rubiq.*
sell-GV1=1SG.GEN vegetable NOM Rubiq
'I sell vegetables to Rubiq.'

(40) *sug-un* *siyang* *rubiq* *ka* *dutut=na.*
distribute-GV1 pork Rubiq NOM relative=3SG.GEN
'Rubiq will distribute the pork among her relatives.'

That the recipient and the conveyed theme can both be a GV subject means that they are both treated as GOAL. We can regard such pattern as the equivalent of "double-object" construction (Haspelmath 2005:426), though they are not actually objects.

There is a semantic map for ditransitive construction proposed by Malchukov et al. (2010:55), shown in Figure 3. It is based on ditransitive construction of various languages. This figure is similar to Figure 1, but at the nodes are verb meaning types, instead of semantic roles. Seediq prototypical ditransitive pattern, represented in Table 7, applies to those verbs that I circled in the figure below. SELL type is near such verbs but outside the circle, since pattern of *gebarig* 'to sell' (Table 12) deviates a little from prototypical pattern.

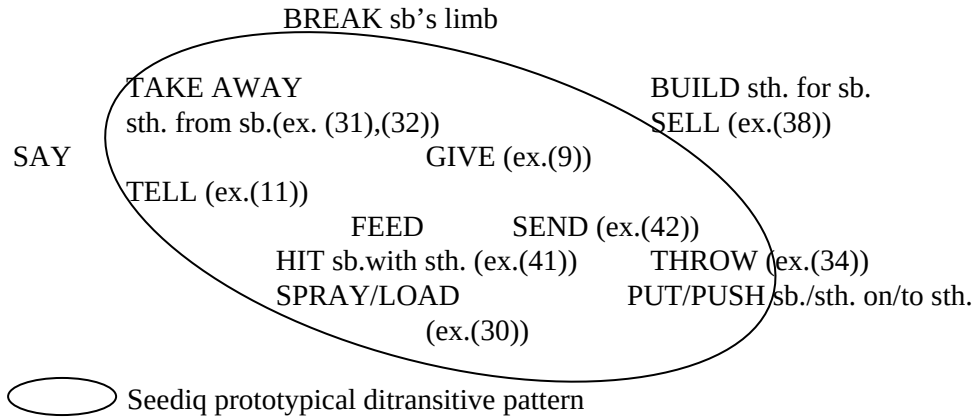


Figure 3: Semantic map for ditransitive constructions

Examples of HIT type verb and SEND type verb are given below. They show prototypical ditransitive patterns.

(41) a. *betaq-an=deha* *senberagan* *ka* *kumay.*
spear-GV2=3PL.GEN spear NOM bear
'They speared a spear at the bear.' / 'They speared the bear with a spear.'

b. *se-baytaq=deha* *kumay* *ka* *senberagan.*
CV-spear=3PL.GEN bear NOM spear
'They speared the spear at a bear.' / 'They speared a bear with the spear.'

(42) a. *pedes-un=na* *patas* *ka* *hiya* *nanak.*
send-GV1=3SG.GEN letter NOM 3SG by.oneself
'S/He sends letters to herself/himself.'

b. *pedes-ani* *nehari* *ka* *laqi=na.*
send-CV.NFIN soon NOM child=3SG.GEN
'Send her/his child right away.'

We can see that the semantic map for ditransitive constructions proposed by Malchukov et al. (2010) is useful for describing Seediq.

4.5 Motivation for secundative alignment

In Seediq, cases of secundative alignment are classified into three categories: (a) case-marking, (b) subject/voice assignment, together with relativisation, nominalisation and *wh*-cleft construction, and (c) reciprocalisation. They are different phenomena, so there may be different motivations.

As for (a) case-marking, secundative alignment is observed when the degree of animacy is higher. That means recipients and patients high in animacy hierarchy are treated alike. Dryer (1986:841) argues that recipients are generally human and more topical, so they are treated as primary objects, whereas themes are treated as secondary objects because they are generally non-human and less topical. This account fits well to the Seediq situation. Recipients that are more topical are likely to be treated as primary objects.

As for (b) subject/voice assignment, this account does not fit well. There are two points.

1. Subject/voice assignment in Seediq treats non-human goals in the same way as recipients.
2. Subject/voice assignment in Seediq treats beneficiary, which is animate, in the same way as conveyed theme.

Dryer says that PO can be considered as secondary topic, next to the subject: "… the PO/SO distinction can be viewed as a grammaticisation of secondary topic vs. non-topic." (Dryer 1986:841)

In the voice system of Seediq, it is easy for patient, recipient, goal, and theme to become the subject, which is the most salient entity of the clause. Since the most salient position is easily available, it seems reasonable to consider that a special slot of secondary topic is not necessary in this language. Grouping must be on semantic basis, but why beneficiary is not treated as GOAL, but as CONVEYANCE, together with Instrument and conveyed theme, remains unclear.

As for (c) reciprocalisation, it seems to follow the general tendency among world's languages. Malchukov et al. (2010:39) account for this by saying that it is related to the fact that a reciprocal relation normally presupposes animacy on the part of the subject and object, which makes the recipient a better choice than theme.

5 Causative

Here I want to show the subject/voice assignment in Seediq causative clauses, because an analogous situation to ditransitive construction is observed in the treatment of arguments of causative construction. I will compare subject/voice assignment in causatives derived from intransitive clauses (IC) and that in causatives derived from transitive clauses (TC). Seediq has only morphological causatives, utilising the prefix *pe-*.

5.1 Correspondence between simple and causative clauses

Kemmer & Verhagen assume a correspondence between simple and causative clauses, and they say that ICs "take the simple transitive clause as a structural and conceptual model" (Kemmer & Verhagen 1994:125) and "…the TC structure is modeled on the structure of a ditransitive clause" (Kemmer & Verhagen 1994:125). The figure below is cited from Kemmer & Verhagen (1994:126).

Simple transitive clauses	Agent	Patient	V_t
	|	|	|
IC clauses	Causer	Causee	$[V_{caus}\ V_i]$

Simple 3-Participant clauses	Agent	Dative (Recipient)	Patient	V_3
	|	|	|	|
TC clauses	Causer	Causee	Affectee	$[V_{caus}\ V_t]$

V_i = Intransitive Verb, V_t = Transitive verb, V_{caus} = Causative Verb, V_3 = 3-Participant Verb

Figure 4: Correspondence between simple and causative clauses

In indirective languages the correspondence between case-marking and semantic roles in transitive clauses and causative clauses would be as follows:

Table 13: Indirective languages: Monotransitive and IC

Case	Monotransitive (Simple transitive clauses)	IC
ACC DAT	Patient	Causee

Table 14: Indirective languages: Ditransitive and TC

Case	Ditransitive (simple three-participant clauses)	TC
ACC	Theme ("Patient" in Kemmer & Verhagen (1994))	Affectee
DAT	Recipient ("Dative" in Kemmer & Verhagen (1994))	Causee

Case marking in causative clauses can be summarised as follows:

Table 15: Case assignment pattern observed in indirective languages

Case	In IC clauses	In TC Clauses
NOM	Causer	Causer
ACC	Causee	Affectee
DAT		Causee

This is what is observed in many indirective languages. Turkish, for example, marks the causee in IC by accusative case, as in (44b). The accusative case is the one used to mark the patient of monotransitive verbs, as in (45a). The causee of TC is marked by dative case, as in (45b). The dative case is the one which marks the recipient in simple ditransitive verbs, as in (43).

(43) *Orhan* *yemeğ-ı* *kopeğ-e* *ver-di.*
Orhan food-ACC dog-DAT give-PST
'Orhan gave the food to the dog.'

(44) a. *Hasan* *öl-dü.*
Hasan die-PST
'Hasan died.'

b. *Ali Hasan-ı* *öl-dür-dü.*
AliHasan-ACC die-CAUS-PST
'Ali made Hasan die. Ali killed Hasan.'

(45) a. *Müdür* *mektub-u* *imzala-dı.*
director letter-ACC sign-PST
'The director signed the letter.'

b. *Dişçi* *mektub-u* *müdür-e* *imzala-t-tı.*
dentist letter-ACC director-DAT sign-CAUS-PST
'The dentist made the director sign the letter.'

Example (43) is from Underhill (1976:67), and (44a), (44b), (45a) and (45b) are from Comrie (1989:175-176).

In secundative languages, on the other hand, the correspondence between case-marking and semantic roles in transitive clauses and causative clauses would be as follows:

Table 16: Secundative languages (Patient=Recipient≠Theme): Monotransitive and IC

Case	Monotransitive (Simple transitive clauses)	IC
ACC[10]	Patient	Causee

Table 17: Secundative languages (Patient=Recipient≠Theme): Ditransitive and TC

Case	Ditransitive (Simple 3-Participant clauses)	TC
ACC	Recipient	Causee
	Theme	Affectee

[10] I use ACC for the case of Patient in languages of secundative alignment, though some scholars may use different set of labels for cases in languages of secundative alignment.

Tarascan (a Mesoamerican isolate), for example, is a secundative language. The patient of a monotransitive verb is marked by *-ni* in (46), and so is the recipient of a ditransitive verb, as in (47) (Maldonado & Nava 2001:161–163).

(46) *Ji* *exe-s-Ø-ka* *wíchu-ni.*
I see-PRF-PRS-IND.1/2 dog-OBL
'I saw the dog.'

(47) *Adrianu* *instïkurhi-s-Ø-ti* *ma* *karakata* *Valeria-ni.*
Adrian give-PRF-PRS-IND.3 a written.thing Valeria-OBL
'Adrian gave Valeria a book.'

Case marking in causative clauses would be as follows:

Table 18: Case assignment pattern observed in secundative languages

Case	IC	TC
NOM	Causer	Causer
ACC	Causee	Causee
		Affectee

A Tarascan example exemplifies this. Example (48) is a Tarascan example of TC clause. The causee *Fernandu* is marked by the object marker *-ni*, but the affectee is not. The causee is "consistently the primary object (Maldonado & Nava 2001:177)" in causatives.

(48) *Ricardu* *itsu-ta-tara-s-ti* *Fernandu-ni*
Ricardo smoke-CAUS-CAUS-PRF-PRS-IND.3 Fernando-OBL
ma *sigarru.*
a cigarette.
'Ricardo made Fernando smoke a cigarette.'

A similar situation is observed in Olutec (Mixe-Zoquean), Cora (Southern Uto-Aztecan), and Sikuani (Guahibo) (Shibatani 2001:17).

5.2 Seediq situation

The subject/voice assignment of Seediq monotransitive and ditransitive clauses was shown in Table 8 in section 4.2. I reproduce the table here, eliminating irrelevant terms.

Table 19: Subject/voice assignment of Seediq monotransitive and ditransitive clauses

Voice	Monotransitive (Simple transitive clauses)	Ditransitive (Simple 3-Participant clauses)
AV	Agent	Agent
GV	Patient	Recipient
CV		Conveyed theme

From the table above, one can expect the following alignment, presuming the above correspondence between simple transitive clauses and causative clauses proposed by Kemmer & Verhagen (1994).

Table 20: Seediq Subject/Voice assignment in IC clauses and in TC clauses

Voice	IC	TC
AV	Causer	Causer
GV	Causee	Causee
CV		Affectee

This correspondence is exactly what is observed in Seediq. Examples are given below. First, examples of Seediq IC clauses are given. The subject is underlined in Seediq sentences and in English translation.

(49) *pe-talang* 'to make somebody run'

a. *pe-talang=ku* *leqi-'an* *ka yaku*.
CAUS[AV]-run=1SG.NOM child-OBL NOM 1SG
'I makes the/a child run.'

b. *pe-telang-un=mu* *ka laqi*.
CAUS-run-GV1=1SG.GEN NOM child
'I will make the child run.'

Then, examples of Seediq TC clauses are given:

(50) *pe-sahug* 'to make somebody serve something'

a. *m-pe-sahug=ku* *'idaw leqi-'an.*
AV.FUT-CAUS-serve=1SG.NOM rice child-OBL
'I will make a/the child/children serve rice.'

b. *pe-sehug-un=mu* *'idaw* *ka laqi*.
CAUS-serve-GV1=1SG.GEN rice NOM child
'I will make the child serve rice.'

c. *pe-sahug=mu* *leqi-'an* *ka 'idaw*.
CV.FUT.CAUS-serve=1SG.GEN child-OBL NOM rice
'I will make a/the child/children eat the rice.'

In Atayal, a similar situation is observed, regarding subject/voice assignment. When the patient is the subject, as in (51a), voice that treats instrument or beneficiary (*si-pa-qaniq* in (51a)) becomes the predicate. When the causee is the subject, as in (51b), voice that treats GOAL as the subject (*pa-pa-qaniq-an* in (51b)) becomes the predicate.

(51) Mayrinax Atayal (Huang 1995:44, glosses are that of Huang 1995)

a. *si-pa-qaniq* *ni'* *yaya'* *cku'* *'ulaqi'*
NAF-CAUS-eat GEN mother DAT.RF child
ku' *bunga'*.
NOM.RF sweet.potato
'Mother fed the sweet potato to the child.'

b. *pa~pa-qaniq-an* *ni'* *yaya'* *cu'* *bunga'*
RED~CAUS-eat-NAF GEN mother ACC.NRF sweet.potato
ku' *'ulaqi'*.
NOM.RF child
'Mother will make the child eat a sweet potato.'

Shibatani (2001:17) discusses causatives in secundative languages. He already observes that a causee nominal (of a transitive-based causative construction) comes to bear a PO relation and is treated like the recipient nominal of a ditransitive clause. Seediq is also a secundative language, and the Seediq situation is consistent with Shibatani's claim.

Let us now look at case marking in causative when it is not chosen as the subject. causee is oblique in the above examples, but it may be direct case, as in (52).

(52) *pe-sipaq* *lawking* *bu-'an=na* *ka* *tama=na.*
CAUS[AV]-hit Lawking mother-OBL=3SG.GEN NOM father=3SG.GEN
'Father makes Lawking hit his mother.'

Table 21 is the summary of case marking in Seediq causative. It is neutral or indirective.

Table 21: Seediq case marking in IC clauses and in TC clauses

	IC	TC
Causer	Genitive (=*mu* '1SG.GEN' in (49b))	Genitive (=*mu* '1SG.GEN' in (50b))
Causee	Oblique (*leqi-'an* 'child-OBL' in (49a))	Oblique (*leqi-'an* 'child-OBL' in (50a), (50c)), Direct (*lawking* 'Lawking' in (52))
Patient		Oblique (*bu-'an=na* 'mother-OBL=3SG.GEN' in (52))

Case marking in causative construction is not as consistent with the alignment in ditransitive construction as subject/voice assignment is. It is neutral or secundative, depending on the degree of animacy.

6 Summary

After having given a brief sketch of the relevant facts of Seediq grammar in section 2 and reviewed concept of indirective and secundative alignment in general in section 3, I showed how the secundative alignment works in Seediq ditransitive clauses in section 3. As for coding properties, the alignment is not consistent. Case marking is secundative or neutral, depending on the animacy. Word order is indirective and as for agreement, it is neutral. As for behavioural properties, secundative alignment is observed more widely (in subject/voice assignment, relativisation, nominalisation, *wh*-cleft construction and reciprocalisation), whereas neutral alignment is observed in few cases (in situ *wh*-construction and in reflexivisation). I have also shown lexical variation in Seediq ditransitive construction in section 4.4.

Then I have had a brief discussion on the motivation of secundative alignment in section 4.5. Dryer (1986) tried to explain secundative alignment in terms of topicality. Since what he tried to explain was mainly the secundative pattern of case-marking, his explanation is appropriate for the situation of Seediq case-marking as well. This, however, does not explain the alignment of the Seediq subject/voice assignment. The situation in Seediq seems to be based on semantics, but why beneficiary is not treated as GOAL, but as CONVEYANCE, together with Instrument and conveyed theme, remains

unclear. As for the secundativeness in reciprocalisation, it is explained as a general tendency among world's languages, which can be explained in terms of animacy.

In section 5, I discussed case-marking of causee and affectee in causative sentences. As Kemmer & Verhagen (1994) show case-marking in intransitive-based causative clauses follows that of monotransitive ones and that in transitive-based causative clauses follows that of ditransitive ones. In Seediq subject/voice assignment, the patient of a monotransitive clause and the recipient of a ditransitive clause become subject in GV, as shown in section 4; it follows from this that the causee becomes subject in GV clauses, whether these are intransitive-based causative clauses or transitive-based causative ones. Case marking in causative construction, however, is not so consistent with the alignment in ditransitive construction. Case marking in ditransitive clause was neutral or secundative, but case marking in causative is neutral or indirective.

References

Blansitt, E. L. Jr. 1984. Dechticaetiative and dative. In *Objects: Towards a theory of grammatical relations*, ed. by Plank, Frans, 127–150. London: Academic Press.

Chang, Henry Y. 2011. Triadic encoding in Tsou. *Language and Linguistics* 12.4:799–843.

Comrie, Bernard. 1989. *Language universals and linguistic typology*. Second edition. Chicago: University of Chicago Press.

Dayley, Jon P. 1985. *Tzutujil grammar.* University of California Publications in Linguistics 107. Berkeley: University of California Press.

Dillon, John Albert. 1994. *A grammatical description of Tatana'*. Ann Arbor: SIL Language and Culture Archives.

Dryer, Matthew S. 1986. Primary objects, secondary objects and antidative. *Language* 62.4:808–845.

Haspelmath, Martin. 2005. Ditransitive construction: The verb 'give'. In *The World Atlas of Language Structures*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie, 426–429. Oxford: Oxford University Press.

Holmer, Arthur. 1999. Structural implications of the function of Instrument Focus in Seediq. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 423–453. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No.1. Taipei: Academia Sinica.

Huang, Lillian M. 1995. *A study of Mayrinax syntax.* Taipei: Crane Publishing Co., Ltd.

Kemmer, Suzanne and Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5.2:115–156.

Kroeger, Paul. 2009. The forms and functions of Instrumental Voice in Northeast Borneo languages. Paper read at 11th International Conference on Austronesian Linguistics. 22–26 June 2009. Aussois, France.

Malchukov Andrej, Martin Haspelmath and Bernard Comrie. 2010. Ditransitive constructions: A typological overview. In *Studies in ditransitive constructions: A comparative handbook*, ed. by Andrej Malchukov, Martin Haspelmath and Bernard Comrie, 1–64. Berlin and New York: Mouton de Gruyter.

Maldonado, Ricardo and E. Fernando Nava L. 2001. Tarascan causatives and event complexity. In *The grammar of causation and interpersonal manipulation*, ed. by Masayoshi Shibatani, 157–195. Typological Studies in Language vol. 48. Amsterdam and Philadelphia: John Benjamins.

Shibatani, Masayoshi. 2001. Introduction: Some basic issues in the grammar of causation. In *The grammar of causation and interpersonal manipulation*, ed. by Masayoshi Shibatani, 1–22. Typological Studies in Language vol. 48. Amsterdam and Philadelphia: John Benjamins.

Tang, Apay Ai-yu. 2011. From diagnosis to remedial plan: A psycholinguistic assessment of language shift, L1 proficiency, and language planning in Truku Seediq. PhD dissertation. Honolulu: University of Hawai‘i at Mānoa.

Tsukida, Naomi. 2000. The CX-*un* and CX-*an* forms in Seediq. In *Modern approaches to transitivity*, ed. by Ritsuko Kikusawa and Kan Sasaki, 53–78. ILCAA study of Language and Cultures of Asia and Africa Monograph series No.35. Tokyo: Institute for the Study of Language and Cultures of Asia and Africa.

Tsukida, Naomi. 2005. Seediq. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 291–325. Routledge Language Family Series. London: Routledge.

Tsukida, Naomi. 2009. Sedekku-go-no bunpōu [A grammar of Seediq]. PhD dissertation. Tokyo: The University of Tokyo [in Japanese]

Tsukida, Naomi. 2012. Goal voice and conveyance voice of Seediq. *Objectivization and subjectivization: A typology of voice systems*, ed. by Wataru Nakamura and Ritsuko Kikusawa, 77–95. Senri Ethnological Studies 77. Osaka: National Museum of Ethnology.

Underhill, Robert. 1976. *Turkish grammar*. Cambridge, MA: MIT Press.

11 *The syntactic derivations of interrogative verbs in Amis and Kavalan*[*]

DONG-YI LIN

1 Introduction

1.1 Interrogative verbs in Amis and Kavalan

Despite the large number of studies on interrogative words and sentences, the possibility that interrogative words can be used as verbs, or interrogative verbs, is still not well known to most linguists. Hagège (2008:3) defines an interrogative verb as "a kind of word which both functions as predicates and questions the semantic content of this predicate". His typological study has revealed the morphological, syntactic, and semantic properties that interrogative verbs share cross-linguistically.

L. Huang et al.'s (1999) study on the interrogative constructions in Formosan languages argues that in addition to nominal and adverbial interrogative words, certain interrogative words in Formosan languages can be used as verbs and exhibit the same morphosyntactic properties as verbal predicates. Lin (2012) shows that interrogative verbs exist indeed in Amis but also in Kavalan—another Formosan language not discussed in the above-mentioned study—in that they have the same morphosyntactic distribution as verbs. Like other verbs, interrogative verbs in the two languages occur in

[*] Fieldwork for this study was sponsored by the research project "The Austronesians: Language, Gene, Culture, and Archaeology" (95R0350-05, 96R0502-06), which was granted to Li-May Sung, National Taiwan University. The dialects of Kavalan and Amis analysed in the paper are Hsinshe Kavalan, which is spoken in Hsinshe village, Hualien County, and Central Amis, which is spoken in Changpin village, Taitung County. The linguistic data for analysis were collected during my fieldwork on these two languages in Taiwan in 2009 and 2010. Most data presented in this paper are elicited data of my fieldwork notes, but secondhand data are also used as a complement to my argumentation. I would like to thank my Amis consultants Ngaday, Ofad, and Panay, and my Kavalan consultants Abas, Buya, Haciang, and Ngengi for sharing their languages with me. I would also like to thank Galia Hatav, Brent Henderson, Felicia Lee, Eric Potsdam, Stacy F. Teng, Joy J. Wu, Elizabeth Zeitoun, and three anonymous reviewers for their comments on earlier versions of this paper. Any errors that remain are my sole responsibility.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 253–289
Asia-Pacific Linguistics, 2015

the sentence-initial position, take tense/aspect markers, attract pronominal clitics, and are affixed with voice markers. The following sentences are for illustration. [1]

(1) Amis (Lin 2012:187)

a. *mi-maan* *ci* *panay*?
AV-do.what NCM PN
'What is Panay doing?'

b. *na* *maan-en* *isu* *k-u-ra* *wacu*?
PST do.what-PV 2SG.ERG ABS-CN-that dog
'What did you do to that dog?'

(2) Kavalan

a. *q‹um›uni=isu* *tangi*?
<AV>do.what=2SG.ABS just.now
'What were you doing just now?' (Lin 2012:186)

b. *quni-an* *na* *wasu* *ya* *saku* *'nay*?
do.what-PV ERG dog ABS cat that
'What does the dog do to that cat?' (Lin 2012:192)

c. *quni-an-su* *m-kala* *ya* *sunis* *a* *yau*?
do.how-PV-2SG.ERG AV-find ABS child LNK that
'How do you find that child?' (Lin 2012:186)

The interrogative words that can be syntactically realised as verbs in the two languages denote 'what', 'how', 'where', and 'how many/much', whereas the interrogative words that denote 'who', 'whose', 'which', 'when', and 'why' cannot serve as verbal predicates as they cannot take voice markers. The affixation of voice markers is unique to verbal predicates, but not non-verbal predicates.

Transitivity of interrogative verbs in Amis and Kavalan is correlated with the voice markers that are affixed to them. Intransitive interrogative verbs are affixed with the agent voice marker, e.g., (1a) and (2a), whereas transitive interrogative verbs are affixed with the patient voice marker, e.g., (1b) and (2b).

It is also found that there are semantic constraints on the use of interrogative words as verbs. For example, the use of *tanian* 'where' as a verb in Kavalan is restricted to questions about the location of the theme argument in a ditransitive event. Questions about the location where an event takes place cannot utilise *tanian* as a verb. Consider the following examples.

(3) Kavalan

a. *tanian-an-su* *ya* *kelisiw-su*?
where-PV-2SG.ERG ABS money-2SG.GEN
'Where do you put your money?' (Not 'Where is your money?')
(Lin 2012:200)

---

[1] Glossing conventions in this paper follow the Leipzig Glossing Rules. Additional glossing conventions are as follows: AV, agent voice; CN, common noun; CP, completive aspect; ENC, enclitic; FAC, factual; HUM, human; IA, instrumental applicative; LA, locative applicative; LNK, linker; NCM, non-common noun marker; NHUM, non-human; PN, proper noun; PREP, preposition; PV, patient voice.

b. * *tanian-an-su* *q‹m›an* *tu/ya* *babuy*?
where-PV-2SG.ERG <AV>eat OBL/ABS pig
(Intended for: 'Where do you eat pork?') (Lin 2012:201)

The intended meaning of the question in (3a) is to inquire about the location of the theme argument, whereas (3b) is intended to ask where the event of eating pork takes place. Only in the former case can *tanian* be used as a verb and be affixed with the patient voice marker. In questions concerning where an event takes place, *tanian* is used as an adverbial expression without taking any voice markers and occurs in the same position as a locative adverbial, as demonstrated below.

(4) Kavalan

a. *tanian* *q‹m›an=isu* *tu* *babuy*?
where <AV>eat=2SG.ABS OBL pig
'Where do you eat pork?'

b. *tanian* *tanuz-an* *na* *tuliq* *ya* *wasu*?
where chase-PV ERG bee ABS dog
'Where do the bees chase the dog?' (Lin 2012:201)

c. *tanuz-an* *na* *tuliq* *ya* *wasu* *tanian*?
chase-PV ERG bee ABS dog where
'Where do the bees chase the dog?' (Lin 2012:201)

The same restriction can be observed for *icuwa* 'where' in Amis.

(5) Amis

a. *icuwa-en* *isu* *k-u* *payci*?
where-PV 2SG.ERG ABS-CN money
'Where do you put the money?'

b. * *icuwa-en* *isu* *mi-saosi* *k-u* *cudad*?
where-PV 2SG.ERG AV-read ABS-CN book
(Intended for: 'Where do you read the book?')

1.2 Goals and organisation

The characteristics and constraints of the interrogative verbs in Amis and Kavalan require a theoretical explanation. In the present paper, I propose a syntactic account for the derivation of the interrogative verbs in the two languages along the lines of Marantz (1997). I will argue that the derivation of interrogative verbs is systematic because whether an interrogative word can be used as a verb can be attributed to universal or language-specific principles or constraints of syntax, the syntactic representations of voice markers, and their corresponding interpretations. This syntactic analysis not only provides a natural explanation for the correlation between voice markers and the transitivity/interpretation of interrogative verbs but also accounts for the semantic restrictions on the use of interrogative verbs in a straightforward and uniform way.

I will also present more empirical evidence for this syntactic analysis by showing that it can explain why certain interrogative verbs must receive a specific interpretation and why some interrogative words cannot be used as verbs. The findings suggest that the

derivations of interrogative verbs are not idiosyncratic, but exhibit a regular pattern and obey syntactic principles and constraints.

I will first clarify the assumptions of the proposed syntactic approach in section 1.3. The main argumentation of this paper is presented in sections 2 and 4. Section 2 discusses the syntactic derivations of interrogative verbs and argues that the derivations obey syntactic principles and constraints. The applicability of the proposed syntactic analysis to other "non-canonical" verbs in Kavalan and Amis is explored in section 3. Based on the analysis formulated in section 2, section 4 explains why certain interrogative words cannot be used as verbs in Kavalan and Amis. Section 5 concludes the study.

1.3 Syntactic assumptions

Adopting the framework of Distributed Morphology (Halle & Marantz 1993, 1994), the present study assumes that roots are not specified for syntactic categories like N and V. What determines the syntactic category of roots are functional heads like v^0, n^0, and a^0. When a root occurs in a verbal environment with the v^0 functional head, it appears as a verb; if instead the root occurs in a nominal environment, it becomes a noun.

Following Starosta (2002[2009]), I analyse voice markers as derivational morphemes. I further suggest that verbal derivations involving voice markers should take place in Syntax. In other words, I reject the assumption that derivational morphology must be implemented in the Lexicon and adopt a syntactic approach to derivational morphology (Harley 2009).

I also assume that the so-called voice markers in Amis and Kavalan are phonological realisations of the category-defining head v^0 due to the following two reasons. First, the affixation of the voice markers is specific to verbal predicates, but not non-verbal predicates. Even though the voice markers also occur in de-verbal nominals, the nominalised words or clauses still possess verbal properties and contain verbal projections (Lin 2010). Nominals with AV/PV exhibit clausal structure higher than VP, unlike English nominalisers *-er*/*-ee*. They should be analysed as headless relative clauses.

Second, the voice markers can derive denominal verbs. In (6a), *nanum* 'water' is an object-denoting noun and appears in a canonical NP position, but when it is affixed with a voice marker as in (6b), it occurs in the predicate position and denotes an activity or action associated with the object denoted by its nominal counterpart.

(6) Amis

a. *mi-sni'* *t-u* *nanum* *i* *takid.*
AV-pour OBL-CN water PREP cup
'(Somebody) pours water into the cup.'

b. *mi-nanum=ho* *kaku.*
AV-drink=IPFV 1SG.ABS
'I am still drinking water.'

In fact, it has been argued that all the lexical roots in Amis are inherently nominal and verbs must be derived via the affixation of voice markers (Wu 2006).

Assuming that the distinction between agent voice and patient voice is correlated with their transitivity (Liao 2002, 2004; Ross & Teng 2005), the present study construes the agent voice marker as an intransitive marker and the patient voice marker a transitive marker. Note that although verbs in the agent voice construction can take a patient argument, this structure is still considered to be syntactically intransitive because the patient argument is demoted and receives oblique case (S. Huang & Tanangkingsing 2011; Liao 2002, 2004). By contrast, the patient voice construction should be analysed as the canonical transitive construction.

2 Syntactic derivations of interrogative verbs

2.1 Syntactic derivations of interrogative verbs based on 'what' and 'how'

Given the assumption that voice markers are verb-defining heads in Syntax, the correlation between the transitivity of interrogative verbs and the voice markers that they take can be attributed to the syntactic nature of *v* that the interrogative roots are merged with. The agent voice marker realises intransitive *v*, whereas the patient voice marker is inserted when *v* is transitive.[2] That is, the transitivity of an interrogative verb is determined by *v* directly. An interrogative root always has at most one argument, and the transitivity of an interrogative verb is derived via the merger of its root with *v* in Syntax.

Consider the following two sets of sentences.

(7) Kavalan

a. *q‹um›uni=isu* *tangi*?
<AV>do.what=2SG.ABS just now
'What were you doing just now?' (Repeated from (2a))

b. *quni-an-su* *ya* *sunis-ku*?
do.what-PV-2SG.ERG ABS child-1SG.GEN
'What did you do to my child?'

(8) Amis

a. *mi-maan* *ci* *panay*?
AV-do.what NCM PN
'What is Panay doing?' (Repeated from (1a))

b. *ma-maan* *cingra*?
AV-what.happen 3SG.ABS
'What happened to him?'

c. *na* *maan-en* *isu* *k-u-ra* *wacu*?
PST do.what-PV 2SG.ERG ABS-CN-that dog
'What did you do to that dog?'

---

[2] I adopt the mechanism of late insertion in Distributed Morphology. A voice marker is analysed as the phonological realisation of a specific type of *v* node, which is defined by morphosyntactic features. Syntax has access to morphosyntactic features or feature bundles, not phonological forms, which are inserted to appropriate nodes after the entire syntactic structure is constructed and sent to the phonological component.

These sentences reveal that the transitivity of an interrogative root like *quni* or *maan* is not lexically specified, but is determined by the voice marker that it takes. When affixed with an agent voice marker, it is morphosyntactically used as an intransitive verb, i.e., 'do what' or 'what happen to' (7a, 8a, 8b); if it takes a patient voice marker instead, it is morphosyntactically used as a transitive verb, i.e., 'do what to' (7b, 8c).[3] In terms of syntactic structure, the interrogative roots in (7a), (8a), and (8b) are merged with an intransitive *v*, which is realised phonologically by an agent voice marker. As for (7b) and (8c), what is merged with the interrogative roots and determines their syntactic category and transitivity is a transitive *v*, which is later realised phonologically by a patient voice marker. I will elaborate on the syntactic structures of these voice markers later in this section after discussing more semantic distinctions among them.

Verbalizing heads exhibit finer semantic distinctions in addition to transitivity. It has been suggested that there are several distinct verb-defining heads with different (combinations of) syntactic/semantic features. One type of *v* that has been extensively discussed is the agent-introducing head, *v*[AG] (Marantz 1997) or Voice (Kratzer 1996). The verbal structure of unaccusative verbs is headed by another type of *v*, which is more like a BECOME-operator (Marantz 1997). Harley (2009) characterises different types of *v* in terms of feature clusters like [±dynamic], [±change of state], and [±cause] as in (9).

(9) The feature specifications of *v* (Harley 2009):
a. v_{CAUSE} : [+dynamic], [+change of state], [+cause]
b. v_{BECOME} : [+dynamic], [+change of state], [-cause]
c. v_{DO} : [+dynamic], [-change of state], [-cause]
d. v_{BE} : [-dynamic], [-change of state], [-cause]

The merger of a root with different types of *v* will thus derive verbs with different Aktionsart properties. The syntactic analysis just presented can account for the interpretation of interrogative verbs if different forms of a particular voice marker are conceived of as phonological realisations of different types of *v* as well.

One clear case in point concerns the contrast between (8a) and (8b). When Amis *maan* is affixed with *mi-*, it is interpreted as an interrogative activity verb; the affixation of *ma-* to this interrogative root derives an interrogative change-of-state verb. This contrast results from the fact that *mi-* and *ma-* realise two distinct *v* heads: v_{DO} and v_{BECOME} respectively. According to Wu's (2006) investigation of the semantics of voice markers in Amis, the affixation of *mi-* to a root, which can inherently denote either an object or an activity, can derive a plain activity verb with an optional motional/purposive/progressive reading. This is illustrated by the following two sentences.[4]

[3] Although the English translation in (7a) and (8a) suggests that 'what' is a direct object of the verb 'do' and that the structure is transitive, the analysis of the English translation cannot apply to the corresponding Kavalan and Amis sentences. The interrogative words in (7a) and (8a) are realised as verbs without taking any direct object syntactically. The structure of the two sentences is intransitive regardless of the structure of their English translation.

[4] All the examples cited from Wu (2006) in this paper have been re-glossed.

(10) Amis

a. *mi-nanum* *ci* *aki* *t-u* *nanum.*
AV-water NCM PN OBL-CN water
'Aki is going to drink water./Aki is drinking water.' (Wu 2006:165)

b. *mi-palu* *ci* *sawmah* *ci* *mayaw-an.*
AV-beat NCM PN NCM PN-OBL
'Sawmah is going to beat Mayaw./Sawmah is beating Mayaw.' (Wu 2006:166)

As for *ma-*, its combination with a root can derive a verb that is interpreted as a result state.[5] The following two sentences demonstrate this meaning of *ma-*.

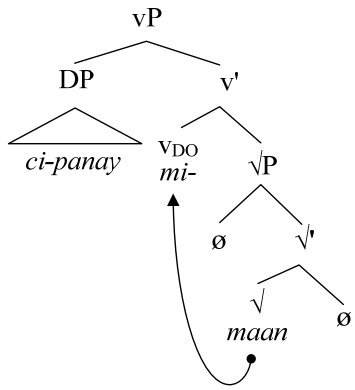
(11) Amis (Wu 2006:183)

a. *ma-adah=tu* *kaku.*
AV-recover=PFV 1SG.ABS
'I have recovered (from illness).'

b. *ma-ruhem=tu* *k-u* *pawli.*
AV-ripe=PFV ABS-CN banana
'The banana is ripe (just now).'

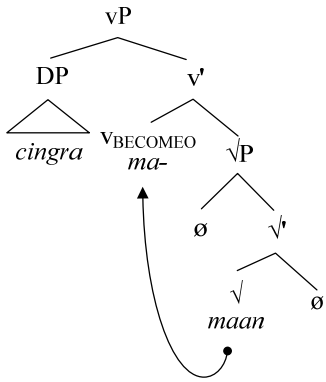
In the theoretical framework of the present paper, *mi-* can be conceived of as an activity-denoting *v*, i.e., v_{DO} and *ma-* can be analysed as v_{BECOME}, which indicates change of state. The different interpretations of (8a) and (8b), i.e., *mi-maan* and *ma-maan*, lie in the feature clusters of *v* that *maan* is merged with. The trees in (12) and (13) represent the derivations of (8a) and (8b), respectively.

(12) (Partial) derivation for (8a)



[5] Wu (2006) classifies *ma-* verbs into four types, each of which is associated with a distinct logical structure. Only the second type, or *ma-*$_2$, is relevant to our discussion here.

(13) (Partial) derivation for (8b)



In (12), *maan* undergoes head movement to v_{DO}, which is the shorthand notation for the feature cluster [+dynamic, -change of state, -cause] and which is phonologically realised as the agent voice marker *mi-*. The resultant *mi-maan* thus denotes a plain activity with an interrogative sense and the DP in the specifier of *v*P is interpreted as the agent of the activity. By contrast, the verbalizing head in (13) consists of the features, [+dynamic], [+change of state], and [-cause] and *ma-* is inserted in this morphosyntactic context. The resultant *ma-maan* is interpreted as a result state and the DP in Spec, *v*P thus refers to a theme argument that undergoes the relevant change of state. The meaning 'what became of him' or 'what happened to him' is thus derived.

Unlike *mi-maan* and *ma-maan*, *maan-en* is interpreted as a transitive interrogative verb 'do what to'. This interpretation is also due to the specific feature cluster of the *v* headed by the patient voice marker *-en*. According to Wu (2006), a verb that is derived via the suffixation of *-en* must have an animate causer/agent and the use of this derived verb emphasises the intention of the agent. This can be demonstrated by the contrast between the following two sentences. The ergative DP in (14a) is an animate causer/agent, but the ergative DP in (14b) is not.

(14) Amis

a.	*tuniq-en*	*aku*	*ku*	*ti'ti'*	*aca.*	
	soft-PV	1SG.ERG	ABS	meat	a.little	
	'I will tenderise the meat a little.'[6]					(Wu 2006:174)

b. *	*tuniq-en*	*nu*	*kuwaq*	*ku*	*ti'ti'*	*aca.*
	soft-PV	ERG	papaya	ABS	meat	a.little

In other words, the verbalizing head that *-en* realises must be [+agentive].

Moreover, the utilisation of a verb suffixed with *-en* always implicates the completion of the action. When *-en* verbs take the imperfective aspect marker *=ho*, they can never receive a progressive interpretation. Compare the following two sentences.

(15) Amis (Wu 2006:176)
a. *ranam-en=ho.*
breakfast-PV=IPFV
'Eat the same thing for the breakfast again!'

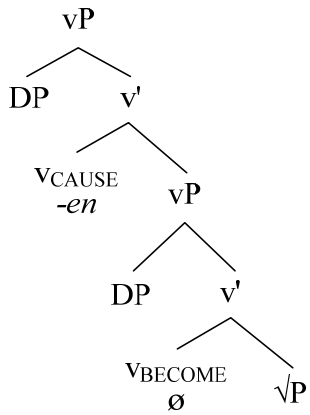
[6] The correct translation should be "I will tenderise only the meat." (Joy Wu, pers. comm.).

b. *mi-nanum=ho ci panay t-u sayta.*
AV-water=IPFV NCM PN OBL-CN soda
'Panay is still drinking soda.'

The verbs in (15a) and (15b) both take the imperfective aspect marker *=ho*. While the verb in (15a), which is suffixed with *-en*, receives an iterative reading, the verb in (15b), which takes the agent voice marker *mi-*, is interpreted as progressive. This suggests that *-en* is inherently [+telic].

In the framework adopted by the present study, the verbalizing head that is realised as *-en* in Amis can be analysed as v_{CAUSE}, which can introduce an agentive causer and implies an endpoint, change of state, or the completion of an action.[7] To capture the inherent semantics of *-en* and its implications, I propose the following verbal structure for verbs that are derived with this suffix.[8]

(16) The verbal structure of *-en*



This structure for *-en* is basically the same as the lexical relational structure assigned to English causative deadjectival verbs by Hale & Keyser (1993). I adopt their conception that the *v*P/VP-shell structure is associated with an asymmetric semantic relation of implication, where a dynamic event encoded in the higher *v*P/VP "implicates" an interrelation or a state encoded in the lower *v*P/VP. The structure in (16) thus aptly reflects the status of *-en* as a causative operator that necessarily implicates an endpoint of the action or change of state. The higher v_{CAUSE} introduces an agent, whereas the lower *v*BECOME [+change of state] ensures that the root merged with it receives a telic interpretation. When this suffix is merged with *maan*, the interpretation of the resultant verb, *maan-en*, follows from the structure in (16).

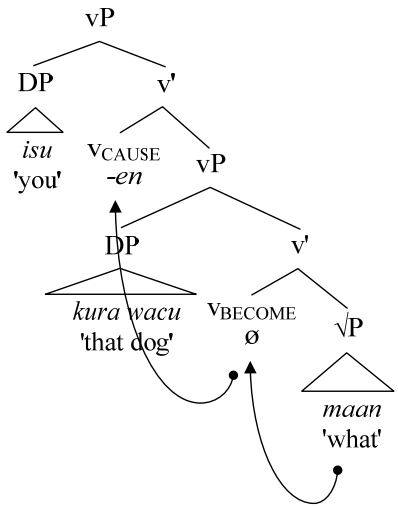
---

[7] As both Amis and Kavalan have a causative prefix *pa-*, the analysis of the patient voice marker as v_{CAUSE} raises a question concerning the differences between the causative prefix and the patient voice marker. The causative prefix is more productive than the patient voice marker in marking a causative event. The two morphemes can co-occur and introduce an agent or causer respectively. How to differentiate the two markers in the syntactic framework we adopt, however, is beyond the scope of the present paper.

[8] An anonymous reviewer points out that v_{CAUSE} does not necessarily introduce an agent, as a non-agentive entity can also be a causer. This suggests that Harley's (2009) classification of *v* needs to make a distinction between two variants of v_{CAUSE}: [+agentive] and [-agentive]. The patient voice marker is inserted under v_{CAUSE} with the feature [+agentive].

Consider the (partial) derivation in (17) for (8c). The higher *v* headed by *-en* introduces an agentive causer, *isu* '2SG.ERG', and implies the existence of an endpoint of the action as indicated by the lower *v*P whose head introduces a theme argument, *kura wacu* 'that dog', which is affected by the action. The derived verb, *maan-en*, is thus construed as a transitive interrogative verb with both an agent argument and a theme argument. The interpretation can be paraphrased as 'X does what such that X causes Y to be in a certain state?' or 'X does something to Y and what happens to Y as a result of this?'.

(17) (Partial) derivation for (8c)



A question arises as to why the lower *v* in the vP-shell structure of (16) or (17) is never realised. It should be noted that there are no verbs that can simultaneously take an agent voice marker *mi-* or *ma-* and a patient voice marker like *-en*. The following verbs are ill-formed.

(18) Amis
a. * *mi-nanum-en*
b. * *ma-ruhem-en*

However, voice markers can co-occur with an instrumental or a locative applicative markers. The following examples are for illustration.[9]

(19) Amis
a. *ka-k‹um›a'en-an* *ni* *ofad* *t-u* *'epah*
KA-<UM>eat-LA ERG PN OBL-CN wine
k-u luma aku.
ABS-CN house 1SG.GEN
'Ofad drinks (wine) at my place. (My place is where Ofad drinks (wine).)'

[9] According to Wu (2006), an AV marker that co-occurs with an applicative affix indicates the conjugation of the verb in the applicative construction, with the conjugated forms determining the semantic role of the applied argument, e.g., location (19a) or purpose (19b). As an AV marker in the applicative construction does not perform the typical function of an AV marker, it is not glossed as AV in this paper, as shown in (19).

b. *mi-cikay-an* *ni* *ofad* *i* *pitilidan* *k-u* *cudad.*
MI-run-LA ERG PN PREP school ABS-CN book
'Ofad runs to school to get the book (for the book). (The book is what Ofad runs to school to get).'

c. *sa-ka-k‹um›a'en* *ni* *ofad* *t-u* *futing* *k-u* *alapit.*
IA-KA-<UM>eat ERG PN OBL-CN fish ABS-CN chopsticks
'Ofad eats fish with the chopsticks. (The chopsticks are what Ofad uses to eat fish.)'

The co-occurrence of voice markers with an applicative marker is one of the reasons why Wu (2006) analyses the so-called locative and circumstantial voice markers in Amis as applicative markers. They perform different functions and should not be classified into the same paradigm. This means that they are governed by different insertion rules and thus are considered separately when insertion takes place. By contrast, the co-occurrence restriction of an agent voice marker and a patient voice marker indicates that they belong to the same set of insertion rules.

I propose that fusion takes place in the *v*P-shell structure of (16) or (17). Fusion is a grammatical process that fuses two terminal nodes that are sisters, e.g., two heads after head-to-head movement, into one single node (Halle & Marantz 1993). As fusion results in one single terminal node, only one vocabulary item can be inserted into this position. In (17), v_{BECOME} and v_{CAUSE} undergo fusion and become one single terminal node, which is a composite of both CAUSE-operator and BECOME-operator. This leads to the semantic implication of the *v*P-shell structure, i.e., 'X does something and causes Y to become Z'. Due to the semantic components of the patient voice marker *-en*, i.e., [+dynamic], [+change of state], [+cause], it is inserted into this position, but not other voice markers. It should be noted that fusion of terminal nodes is subject to cross-linguistic differences (Halle & Marantz 1993). While fusion of v_{BECOME} and v_{CAUSE} takes place in Amis, the two terminal nodes can be realised by independent morphemes in other languages, e.g., Japanese.[10]

Note that Amis *maan* can also be used as a noun as in (20), where it occurs in a case-marked position.

(20) Amis
ma-talaw *ci* *lekal* *t-u* *maan*?
AV-afraid NCM PN OBL-CN what
'What is Lekal afraid of?'

As verbal *maan* is derived in a syntactic context where it can be merged with a verbalizing head via head movement, the use of *maan* as a noun is also contingent on its syntactic environment. In (20), it is moved to *n*, the category-defining head for nouns, so that it can further be case-marked. An equally plausible alternative is to attribute the nominal status of *maan* in (20) to the presence of D, or the case marker *ku*. On this alternative analysis, there is no need to posit the noun-deriving head *n* in Amis. Amis *maan* is an exemplar that shows how the lexical category and interpretation of an

[10] I would like to thank an anonymous reviewer for pointing this out for me.

interrogative root can vary with and be determined by the syntactic context where it occurs.

This syntactic analysis of Amis *maan* can apply to its Kavalan counterpart, *quni* 'do what', the transitivity and interpretation of which is also conditioned by the voice marker that it takes. One prominent difference between Kavalan and Amis concerns the semantics of the different forms of the agent voice. While each form of the Amis agent voice morpheme is associated with a distinct logical structure or interpretation, as shown above for *mi-* and *ma-*, the choice of Kavalan agent voice forms seems to be conditioned by phonology, i.e., phonologically-conditioned allomorphy, and is subject to lexical variation to a great extent. In other words, Kavalan differs from Amis in that it does not utilise distinct lexical items to realise different types of intransitive *v*. However, the overt distinction between the intransitive *v* and the transitive *v* is still preserved in Kavalan. The agent voice construction is an intransitive syntactic structure, whereas the patient voice construction is the canonical transitive structure. For example, *q‹um›uni* 'do what', an intransitive interrogative verb with an agent argument, is derived by moving the root *quni* to v_{DO}, which can assign an agent theta-role.

The function of the patient voice marker *-an* in Kavalan is similar to Amis *-en* in that *-an* also introduces an agent or causer argument and implies an endpoint, a change of state, or the completion of an action. As illustrated below, *-an* is analogous to the causative marker *pa-* in terms of their function to introduce an external argument (21c), 21d). Note that when *sabiqbiq* 'boil' is used in its agent voice form as in (21a), it can only have an unaccusative interpretation, as demonstrated by the ungrammaticality of (21b), where there is an additional external argument.

(21) Kavalan

a. *sabiqbiq=ti ya zanum 'nay.*
boil=PFV ABS water that
'The water has boiled.'

b. * *sabiqbiq=ti=iku tu zanum.*
boil=PFV=1SG.ABS OBL water

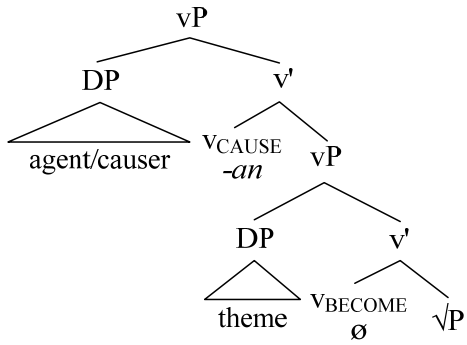
c. *pa-sabiqbiq=ti=iku tu zanum.*
CAUS-boil=PFV=1SG.ABS OBL water
'I boiled water. (I had the water boiled.)'

d. *sabiqbiq-an-ku ya zanum 'nay.*
boil-PV-1SG.ERG ABS water that
'I boiled the water.'

Compare (21a) with (21d). The patient voice marker in (21d) functions as a causative operator that introduces an extra agentive causer, and the action performed by this agentive causer leads to the change of state of the theme argument assigned by the original agent voice predicate. The patient voice marker *-an* should thus be construed as the phonological realisation of v_{CAUSE}. Like Amis *-en*, it also involves a *v*P-shell structure with an implicational causal relation between the higher *v*P and the lower *v*P in accordance with Hale & Keyser's (1993) analysis of de-adjectival verbs in English. Its structure is schematically represented in (22). The merger of *quni* with *-an* leads to the

derivation of a transitive interrogative verb that requires an agentive causer and a theme argument that undergoes the action.

(22) The structure of Kavalan *-an*



It has been found that *-an* can also introduce an additional theme argument (Chang 1997). According to Chang (1997), an intransitive verb is allowed to take an additional argument when it is affixed with the patient voice marker *-an*, but not when it takes the agent voice marker. The contrast between (23a) and (23b) illustrates this function of *-an*.[11] The patient voice marker in (23e) also performs the same function. The addition of an oblique argument that is affected by the event to an agent voice sentence in (23d) is only slightly acceptable. Its patient voice counterpart, (23e), is fully grammatical. The absolutive DP in (23e), 'his mother', is interpreted as an argument that is affected by the action of the agent.

(23) Kavalan

a. ? *maynep=iku tu qaynepan.*
sleep.AV=1SG.ABS OBL bed
(Intended for: 'I am sleeping in a bed.') (Chang 1997:72)

b. *qaynep-an-ku ya qaynepan.*
sleep-PV-1SG.ERG ABS bed
'I slept in the bed.' (Chang 1997:72)

c. *t‹m›alumbi ta-liab-an na takan ya sunis*
<AV>hide LOC-underside-LOC GEN table ABS child
a yau.
LNK that
'The child hides under the table.'

d. ? *t‹m›alumbi ta-liab-an na takan ya sunis a yau*
<AV>hide LOC-underside-LOC GEN table ABS child LNK that
tu tina-na.
OBL mother-3GEN
(Intended for: 'The child hides under the table from his mother.')

[11] The examples in (23a) and (23b) from Chang (1997) have been re-glossed.

e. *talumbi-an* *na* *sunis* *a* *yau* *ta-liab-an* *na*
hide-PV ERG child LNK that LOC-underside-LOC GEN
takan *ya* *tina-na.*
table ABS mother-3GEN
'The child hides under the table from his mother.'

Thus, the argument structure of *-an* includes not only an agent argument, but also a theme argument that is affected by the action of the agent.[12] This provides further justification for the syntactic structure of *-an* in (22).

It is noteworthy that 'what' and 'how' share the same root in both Kavalan and Amis. Moreover, both interrogative words can take the patient voice marker, as illustrated below.

(24) Kavalan

a. (*na*)*quni-an-su* *ya* *sunis* *a* *yau*?
do.what-PV-2SG.ERG ABS child LNK that
'What do you do to that child?'

b. (*na*)*quni-an-su* *m-kala* *ya* *sunis* *a* *yau*?
do.how-PV-2SG.ERG AV-find ABS child LNK that
'How do you find that child?'

(25) Amis

a. *na* *maan-en* *isu* *k-u-ra**wacu*?
PST do.what-PV 2SG.ERG ABS-CN-that dog
'What did you do to that dog?'

b. *na* *maan-en* *ni* *panay* *mi-padang* *kisu*?
PST do.how-PV ERG PN AV-help 2SG.ABS
'How did Panay help you?'

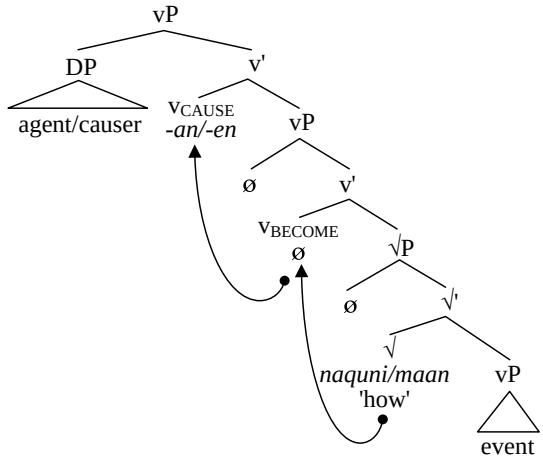
The only difference on the surface lies in the additional verb in the 'how'-question. Nevertheless, 'do what' and 'do how' are conceptually related as a 'how'-question can be easily paraphrased as a 'do.what'-question. For example, 'How did you find the child?' can be paraphrased as 'What did you do to find the child?'. It is thus highly probable that (24a) and (24b) or (25a) and (25b) involve the same verbal derivation with the same category-defining head, v_{CAUSE}.

---

[12] An anonymous reviewer raises a question regarding where the additional affected argument of *-an*, e.g., *tina-na* 'his mother' in (23e), is located in the *v*P-shell structure. The affected argument in (23e) is not really the theme, as the sentence does not mean 'the child caused his mother to hide'. It is likely that the *v* below v_{CAUSE} might actually be an applicative head that licenses an extra argument like an affectee according to Pylkkänen's (2008) analysis of non-core arguments. Another possibility is that the structure of the patient voice marker contains not only v_{CAUSE} and v_{BECOME} but also an applicative head. The questions of whether the patient voice marker is associated with any applicative function and how it interacts with the overt applicative markers cannot be answered without a separate detailed study on the syntax and semantics of applicative constructions in Kavalan and Amis. The evidence presented here, nevertheless, is sufficient to show that the structure of the patient voice marker is more complicated than a single-layered $v\text{P}_{\text{CAUSE}}$.

First, both types of questions require an agent or causer that brings about a certain action or event. Second, they both imply an endpoint. In the case of transitive 'do what', this endpoint interpretation is due to the change of state of the theme argument that is affected by the action. As for 'do how', the endpoint interpretation emanates from the completion of an action. The derivation for 'do how', as represented below in (26), is thus analogous to transitive 'do what', except that there is a *v*P complement to the interrogative root. As with transitive 'do what', 'do how' is also derived via head movement of the interrogative root to v_{CAUSE}, which is realised as the patient voice marker *-en* or *-an*, thus their homogeneity. However, 'do how' requires a *v*P complement and per the implicational causal relation of the *v*P-shell structure, v_{BECOME} indicates that the action/event brought about by the agent/causer induces the completion of another event. In this sense, v_{BECOME} in (26) is slightly different from its counterpart in (17), the structure for transitive 'do what', although both signal the existence of an endpoint.

(26) The structure of the 'do how'-question



2.2 Syntactic derivations of interrogative verbs based on 'where'

The syntactic approach delineated above for the derivation of 'do what' and 'do how' can also provide a natural explanation for the grammatical properties and syntactic distributions of *tanian* and *icuwa* 'where'. The use of Kavalan *tanian* and Amis *icuwa* as verbs is restricted to questions about the location of the theme argument in a ditransitive event. Questions about the location where an event takes place cannot utilise *tanian* or *icuwa* as a verb. Consider the following examples.

(27) Kavalan

a. *tanian-an-su* *ya* *kelisiw-su*?
where-PV-2SG.ERG ABS money-2SG.GEN
'Where do you put your money?'

b. * *tanian-an-su* *q‹m›an* *tu/ya* *babuy*?
where-PV-2SG.ERG <AV>eat OBL/ABS pig
(Intended for: 'Where do you eat pork?')

(28) Amis

a. *icuwa-en* *isu* *k-u* *payci*?
where-PV 2SG.ERG ABS-CN money
'Where do you put the money?'

b. * *icuwa-en* *isu* *mi-saosi* *k-u* *cudad*?
where-PV 2SG.ERG AV-read ABS-CN book
(Intended for: 'Where do you read the book?')

It is argued below that their grammatical properties and restrictions can be derived with reference to the syntactic environment of the interrogatives themselves. Specifically, like other interrogative verbs that have been discussed, *tanian* and *icuwa* serve as verbs when they are selected by a category-defining verbal head, the little *v*.

The adverbial, in situ properties of the adjunct use of *tanian* and *icuwa* as in (29) and (30) follow from their adjunct status. Not being selected by little *v*, *tanian* and *icuwa* cannot be a verb in these constructions and therefore lack verbal properties. Rather, adjunct *tanian* and *icuwa* take scope over the entire verb phrase.

(29) Kavalan

a. *tanuz-an* *na* *tuliq* *tanian* *ya* *wasu*?
chase-PV ERG bee where ABS dog
'Where do the bees chase the dog?'

b. *tanian* *tanuz-an* *na* *tuliq* *ya* *wasu*?
where chase-PV ERG bee ABS dog
'Where do the bees chase the dog?'

(30) Amis

a. *k‹um›a'en* *kisu* *t-u* *hemay* *icuwa*?
<AV>eat 2SG.ABS OBL-CN rice where
'Where do you eat?'

b. *icuwa* *k‹um›a'en* *kisu* *t-u* *hemay*?
where <AV>eat 2SG.ABS OBL-CN rice
'Where do you eat?'

In (29), the question is intended to inquire about the location where the bees chase the dog. Likewise, in (30), the question concerns the location where the addressee eats. Since the scope of *tanian* and *icuwa* in (29) and (30) ranges over an event, it is not unreasonable to assume that they are adjoined to *v*P or TP. The different adjunction positions lead to the word order differences between (29a) and (29b) or between (30a) and (30b).

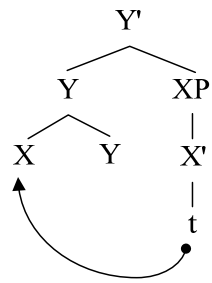
Whether *tanian* or *icuwa* is adjoined to *v*P or TP, there is no way for it to take the voice marker in *v*, which has been merged with the lexical verb. Even if *tanian* is adjoined to the projection of the root phrase before the root moves to *v*, it is still forbidden from moving to *v* because it is inside an adjoined phrase. Head movement out of a specifier or an adjunct is never attested. In the GB framework, this is due to the Head Movement Constraint (Travis 1984) or the Empty Category Principle (Chomsky 1981).

(31) Head Movement Constraint (HMC)
X^0 may only move into Y^0 that properly governs it.

(32) Empty Category Principle (ECP)
An empty category must be properly governed.

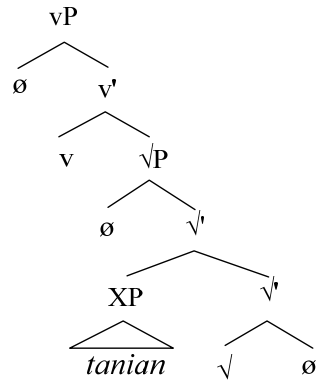
Baker (1988) assumes that the HMC can be derived from the ECP and head movement of X to Y, as represented in (33) below, results in a head-adjunction structure, where the adjunction node does not count as the first branching node for c-command. Under the framework of GB, Baker (1988) proposes that if XP in (33) below is selected by Y, it does not count as a barrier for government from Y. This way, the trace of X in (33) can be antecedent-governed.

(33)



Suppose *tanian* is adjoined to the root phrase instead of *v*P or TP, as represented below.[13]

(34)



As an adjunct, its movement to *v* would violate the ECP because the phrase that it projects is not selected by *v*P and will act as a barrier for government. The illicit movement will lead to a structure where *tanian* cannot antecedent-govern its trace.

The notion of government has been abandoned by the Minimalist Program. However, the empirical fact that a head in a specifier or an adjoined phrase cannot move out of this position still holds. Other theoretical principles or conditions compatible with Minimalist ideas must be sought to explain this syntactic phenomenon. According to Matushansky (2006), the Transparence Condition as formulated in (35) is a potential principle that can generate the same effects as the Head Movement Constraint.

---

[13] I assume with Ernst (2002) that adjuncts can be attached to an intermediate projection instead of a maximal projection.

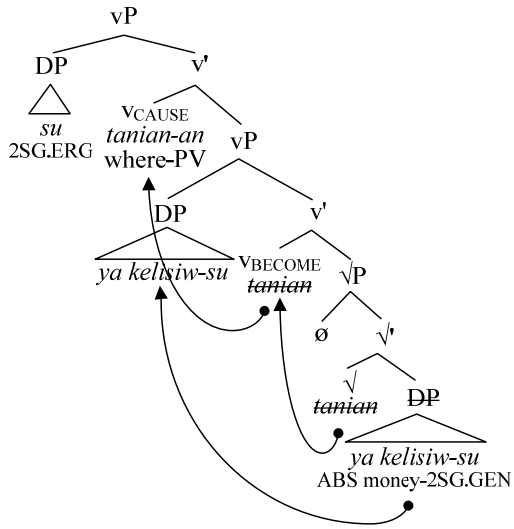
(35) Transparence Condition (Matushansky 2006:78)
A head ceases to be accessible once another head starts to project.

The Transparence Condition functions to ensure that only heads that are still projecting at some point of syntactic derivations are accessible to syntactic operations. When a head X^0 enters the derivation and merges with its complement phrase YP, it is necessary to assess both X^0 and Y^0 in order to determine which head projects. At this point, both heads are likely to project and thus both are still accessible to syntactic operations like movement or Re-merge. Once the selection of X^0 for Y^0 is established, Y^0 is able to move to X^0 at this point of the derivation. This translates into the well-known generalisation of the locality of head movement: A head Z^0 can move to the head W^0 of the phrase WP that takes ZP as the complement, but cannot skip it. The Transparence Condition rules out the configuration where the head in a specifier or an adjoined phrase moves to a c-commanding head that does not select for it. Take (34) as an example. When *v* merges with √P, XP, which has been adjoined to √P, is no longer projecting. Therefore, X^0 is not accessible to syntactic operations and is not allowed to move to *v*. The ban on head movement out of an adjoined phrase can be explained by the Transparence Condition without invoking the notion of government.

Regardless of what theoretical mechanism is adopted, if *tanian* in (34) moves to *v*, this will result in an illicit syntactic configuration. Therefore, when *tanian* is used to question the location where an event takes place, it cannot take a voice marker and be used as a verb. The observation that adjunct *tanian* cannot be used as a verb finds a natural explanation in the proposed syntactic analysis. The analysis assumes that interrogative verbs are derived in Syntax and thus their derivations must conform to established syntactic principles and constraints like the HMC, the ECP, or the Transparence Condition.

By contrast, the verbal derivation for *tanian* or *icuwa* in a question that inquires about the location of a theme argument as in (27a) and (28a) does not incur any violation of syntactic principles and constraints. Take (27a) as an example. The derivation begins with the merger of √TANIAN 'where' with *kelisiw-su* 'money-2SG.GEN'. This is because the DP *kelisiw-su* is the theme argument of √TANIAN 'where'. The interrogative root then moves to v_{BECOME} and v_{CAUSE} in a successive-cyclic fashion. The derivation can be schematically represented in (36). The movement of √TANIAN to v_{BECOME} and v_{CAUSE} obeys the ECP or the Transparence Condition. The higher *v* is the causative operator CAUSE which entails an agent thematic role and defines transitive verbs. This head is spelled out as the patient voice marker *-an* in Kavalan. Together with the inherent locational and interrogative semantics of *tanian*, the result is a transitive construction in which the location of the theme is in question. The movement of the theme DP *ya kelisiw-su* in (36) is motivated by Case checking. Only finite T can check absolutive Case in Kavalan. Before the theme DP can move to Spec, T, it must first move to the edge of vP, which is a phase, or otherwise it would be stranded due to the Phase Impenetrability Condition.

(36) (Partial) derivation for (27a)[14]



Specifically, the *v*P-shell structure with v_{CAUSE} and v_{BECOME} involves an implicational relation where the action performed by the agent introduced by v_{CAUSE} must imply an endpoint. In the case of (36), the endpoint interpretation arises from the change of state of the theme argument, i.e., its ending up being somewhere. The meaning of (36) can thus be paraphrased as 'X (the agent) does something and this causes Y (the theme) to be where?' Without a secondary lexical verb, the details of the action are left under-specified, leading to a meaning of something like 'X put Y where?' When a secondary lexical verb is present, it serves to further specify the action of the transitive event. The secondary lexical verb following *tanian* or *icuwa* must be able to take a location argument, as exemplified below.

(37) Kavalan

tanian-an	*ni*	*abas*	*m-Rupu*	*ya*	*adam*	*'nay*?
where-PV	ERG	PN	AV-shut	ABS	bird	that

'Where does Abas shut the bird?'

(38) Amis

icuwa-en	*isu*	*mi-na'ang*	*k-u*	*riku'*?
where-PV	2SG.ERG	AV-pack	ABS-CN	clothes

'Where do you pack the clothes?'

This restriction on the secondary lexical verb can be ascribed to the structure in (36) and the ditransitive interpretation associated with it. The most natural interpretation of 'X causes Y to be where', the meaning of (36), corresponds to a ditransitive event and is thus compatible with verbs that take a location argument.[15]

---

[14] There is a slight difference between (36) and (22). In (22), the theme is base-generated in the specifier of v_{BECOME}, but the theme in (36) moves to this position. We assume that both configurations are possible structures of *-an*. An alternative analysis is that there are two distinct types of v_{BECOME}.

[15] Due to the limitation of space and the focus of the present study, this paper cannot offer a detailed account for the Interrogative Verb Sequencing Construction (IVSC), which contains

Whether there is a lexical verb following *tanian* or *icuwa* in a verbal 'where'-question, the basic semantic structure of the construction is the same. The interrogative word *tanian* or *icuwa* inherently denotes 'where', while the verbal features follow from its merger with the transitive *v*. The proposed syntactic account can provide a straightforward explanation for the fact that when *tanian* or *icuwa* is used as a verb with both an agent argument and a theme argument, it always takes the patient voice marker *-an* or *-en*, but not the agent voice marker, as illustrated below.

(39) Kavalan
* *tanian=isu* *tu* *kelisiw-su*?
where=2SG.ABS OBL money-2SG.GEN
(Intended for: 'Where do you put your money?')

(40) Amis
* *icuwa* *kisu* *t-u* *payci*?
where 2SG.ABS OBL-CN money
(Intended for: 'Where do you put money?')

This is because only v_{CAUSE}, which is phonologically realised as the patient voice marker *-an* or *-en*, can introduce an agent argument or causer and simultaneously take the projection of v_{BECOME} as its complement to denote a change of state caused by some action. In other words, due to the *v*P-shell structure of the patient voice marker, the ergative argument of *tanian-an* or *icuwa-en* must be interpreted as the agent argument that causes the absolutive argument to be somewhere. This interpretation is compatible with questions about the location of the theme argument in a ditransitive event, but not with questions that concern the location where an event takes place. The semantic restriction on the verbal use of *tanian* thus finds a natural explanation.

2.3 Syntactic derivations of interrogative verbs based on 'how many/much'

In addition to 'do what', 'do how', and 'where', the interrogative words that denote 'how many/much' in Kavalan and Amis can also show up as interrogative verbs.

(41) Kavalan
a. *kin-tani-an-su=pa* *p‹m›ukun* *ya* *sunis*?
HUM-how.many-PV-2SG.ERG=FUT <AV>beat ABS child
'How many children will you beat?'

b. *u-tani-an* *na* *wasu* *q‹m›aRat* *ya* *saku*?
NHUM-how.many-PV ERG dog <AV>take ABS cat
'How many cats does the dog bite?'

---

an interrogative verb followed by a secondary lexical verb. Readers can refer to Lin (2013) for a syntactic analysis of this construction. It is argued that the syntactic relationship between the interrogative verb and the lexical verb in an IVSC is subordination, with the interrogative verb as the main verb.

(42) Amis

a. *pina-en* *ni* *ofad* *k-u* *paysu*?
how.many-PV ERG PN ABS-CN money
'How much money does Ofad want/take?'

b. *pa-pina-en* *isu* *mi-lawup* *k-u* *wawa*?
HUM-how.many-PV 2SG.ERG AV-chase ABS-CN child
'How many children will you chase?'

At first sight, the use of 'how many/much' as a verb in the patient voice construction does not conform to the analysis of *-an* or *-en* as v_{CAUSE} with v_{BECOME} as its complement and thus should constitute a counterexample to my syntactic approach to the derivation of interrogative verbs. A closer examination of the semantics of verbal *tani* or *pina* reveals otherwise.

A question where *tani* or *pina* is employed as a verb and takes the patient voice marker, e.g., (41) and (42), always implies that the quantity of the affected theme argument will or might change from the perspective of the speaker. For example, the utterance of (42a) is appropriate in a scenario where the speaker expected Ofad to take less money, but the contextual evidence s/he had suggested that he might want more money. The utterance of (42b) also has a similar connotation. Suppose that the addressee of this question had chased three children yesterday and he told the speaker that he intended to chase five children today. In this situation, the speaker could utter (42b) to show his suspicion that the addressee might chase even more children. A more appropriate translation of (42b) might be 'HOW MANY MORE children will you chase?' This type of implication is absent in a pseudo-cleft question with *tani* or *pina* as a nonverbal predicate, as illustrated in (43) and (44).

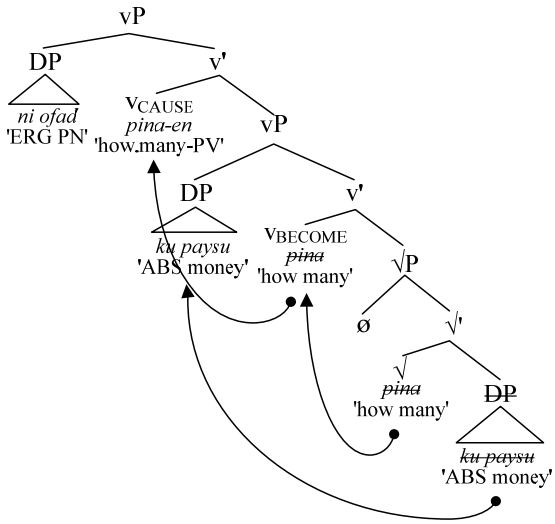
(43) Kavalan
u-tani *ya* *ni-ala-su* *tu* *kelisiw*?
NHUM-how.many ABS PFV-take-2SG.GEN OBL money
'How much money did you take?' (Lit. The money that you took is how much?)

(44) Amis
pina *k-u* *mi-ala-an* *ni* *utay* *a* *payci*?
how.many ABS-CN MI-take-LA ERG PN LNK money
'How much money did Utay take?' (Lit. The money that Utay took is how much?)

Compared with (43) and (44), the questions in (41) and (42), where 'how many' is suffixed with the patient voice marker, emphasise the intention of the agent and simultaneously imply a change of state, specifically the change of the quantity of the theme argument that might be affected.

The semantics of PV-marked *tani* or *pina* is thus compatible with the syntactic structure assigned to the patient voice marker, or v_{CAUSE}. The proposed syntactic analysis correctly predicts that these two interrogative words can show up as a verb and the interpretation of the derived interrogative verb should conform to the semantics of the *v*P-shell structure with v_{CAUSE} and its accompanying v_{BECOME}. The tree in (45) demonstrates the derivation of *pina-en* in (42a).

(45) (Partial) derivation for (42a)



The syntactic structure in (45) reflects three important features of verbal *pina* (or *tani*). First of all, the fact that a question with PV-marked *pina* emphasises the intention of the agent can be ascribed to the agent-introducing function of v_{CAUSE}. This is also the reason why verbal *pina* (or *tani*) must occur in the patient voice construction, but not the agent voice construction, as illustrated by the ungrammaticality of the following sentences.

(46) Kavalan

* *u-tani=isu*	*tu*	*kelisiw*?	
NHUM-how.many=2SG.ABS	OBL	money	

(Intended for: 'How much money do you want/take?')

(47) Amis

* *pina*	*ci*	*ofad*	*t-u*	*payci*?
how.many	NCM	PN	OBL-CN	money

(Intended for: 'How much money does Ofad want/take?')

The verbal meaning of *pina* or *tani* is syntactically derived via the merger with the patient voice marker and the *v*P-shell structure associated with it. As v_{CAUSE}, the patient voice marker can introduce an agent argument or causer and simultaneously take the projection of v_{BECOME} as its complement to denote a change of state caused by some action. The agent voice construction lacks this causative structure.

The second fact that requires an explanation is that a question with *pina-en* or *tani-an* must inquire about the quantity of the theme argument, but not the agent argument. This observation is due to the semantics of the lower *v*P, where *pina* or *tani* is predicated of the theme argument. The agent argument is introduced by v_{CAUSE} and does not belong to the argument structure of *pina* or *tani*. Moreover, there is agreement between *pina* or *tani* and the theme argument in terms of humanness. When the theme argument is human, *pina* takes an agreement prefix *pa-*, which is derived via C*a*-reduplication. The agreement prefixes on Kavalan *tani* are *u-* for non-humans and *kin-* for humans. The syntactic configuration in (45), where *pina* and the noun phrase in question exhibit a local relationship in the root phrase, allows this type of agreement to occur. The agent noun

phrase, which is assigned by v_{CAUSE}, is never part of the argument structure of *pina*. Thus, when *pina* is used as a verb and takes the patient voice marker, it is not the quantity of the agent noun phrase that is in question and it cannot agree with *pina* in terms of humanness.

Finally, the *v*P-shell structure with v_{CAUSE} and v_{BECOME} implicates that there is a causal relation between the two respective sub-events in the upper *v*P and the lower *v*P and further implies a change of state. This implicational relation contributes to the unique interpretation associated with *pina-en* or *tani-an*: The quantity of the affected theme argument will or might change from the perspective of the speaker.

The syntactic derivations of verbal 'what', 'where', and 'how many' with the patient voice marker share two formal properties associated with the *v*P-shell structure of v_{CAUSE} and v_{BECOME}: causal relation and change of state. It should be noted that the labels of *v* are shorthand notations for formal features like [±dynamic] and [±change of state]. The head v_{BECOME} denotes [+change of state] and its presence in the structures of verbal 'what', 'where', and 'how many' should not be directly and literally interpreted as 'becoming what, where, or how many'. In all these cases of interrogative verbs with a patient voice marker, the theme undergoes a certain change of state, and the result state, which can be a general state, location, or quantity, is questioned.

Within the framework of Distributed Morphology, only formal features are visible in the syntactic component, while meanings associated with phonological forms are supplied by a separate component, Encyclopaedia, which regulates appropriate use of expressions instead of their grammatical well-formedness (Harley & Noyer 2000). The syntactic derivations of verbal 'what', 'where', and 'how many' are all grammatically well-formed, as they conform to the Head Movement Constraint or the Transparence Condition. Their interpretations follow from the *v*P-shell structure of v_{CAUSE} and v_{BECOME}, all involving a causal relation and a change of state, even though how the specific interpretations are fine-tuned and conventionalised are still subject to the influence and constraints of Encyclopaedia.

The syntactic mechanisms that are responsible for the derivation of verbal *tani* and *pina* are not peculiar to these two interrogative words, but are shared by the other interrogative verbs. There is no need to resort to lexical stipulation. The grammatical and semantic features of interrogative verbs are the concomitant consequences of the syntactic structure they occur in.

The analysis of *tani* and *pina* as a head that can undergo head movement to *v* has implications for the typology and structure of interrogative and non-interrogative degree expressions like English *how much* and French *combien*. Two anonymous reviewers point out that there is still no consensus on the structural representation of degree expressions. The controversy lies in whether degree expressions should be analysed as an adjunct or as a functional head of DegP and QP projected above AP (Corver 1997; Doetjes 1997). The fact that *tani* and *pina* can take the patient voice marker might lend empirical support to the functional head analysis, as movement to *v* must obey the Head Movement Constraint. However, the adjunct properties of degree expressions regarding extraction and non-selection as discussed in Doetjes (1997) should not be ignored. Cross-linguistic differences in the syntactic representation and derivation of degree expressions are worthy of further detailed investigation. It remains to be seen to what extent *tani* and *pina* are similar to and different from their counterparts in other languages regarding their properties as a head or an adjunct.

3 Extension to non-interrogative words

There are at least two advantages of the syntactic account proposed in the preceding sections for Kavalan and Amis interrogative verbs. First, this syntactic account can be extended to non-interrogative words that share similar morphosyntactic and semantic properties with interrogative verbs. In other words, it can capture the overall grammatical system of Kavalan and Amis. Second, since the derivation of interrogative verbs is constrained by established syntactic principles and operations, either universal or language-specific, this syntactic account can make predictions about what interrogative words can and cannot be used as a verb. I will show that the predictions are borne out. This section deals with the first advantage, and section 4 will elaborate on the second advantage.

3.1 Location verbs

The syntactic analysis proposed in the preceding sections can generalise to non-interrogative cases such as locative deictics, which are also realised as verbs in Kavalan and Amis. In the following examples, the locative deictics occur at the sentence-initial position with the patient voice marker *-an* or *-en*, suggesting that they are used as verbs.

(48) Kavalan

a. *tazian-an-ku* (*pizi*) *ya* *kelisiw-ku.*
here-PV-1SG.ERG put ABS money-1SG.GEN
'I put my money here.'

b. *tawian-an-ku* (*pizi*) *ya* *kelisiw-ku.*
there-PV-1SG.ERG put ABS money-1SG.GEN
'I put my money there.'

(49) Amis

a. *itini-en* *ni* *panay* (*pateli*) *k-u* *payci.*
here-PV ERG PN put ABS-CN money
'Panay put the money here.'

b. *itiraw-en* *ni* *panay* (*pateli*) *k-u* *payci.*
there-PV ERG PN put ABS-CN money
'Panay put the money there.'

Like their interrogative counterparts, *tanian* and *icuwa*, the locative deictics in (48) and (49) are able to serve as the only verb in a sentence without any lexical verb. Moreover, when used as a verb, they must denote the location of the theme argument in a ditransitive event. When they refer to the location where an event takes place, they are not allowed to take the patient voice marker, as illustrated below.

(50) Kavalan

* *tazian-an-ku* *m-Rasa* *tu*/*ya* *sudad.*
here-PV-1SG.ERG AV-buy OBL/ABS book
(Intended for: 'I buy a/the book here.')

(51) Amis

* *itiraw-en*	*ni*	*utay*	*mi-pacu'*	*t-u/k-u*	*fafuy.*
there-PV	ERG	PN	AV-kill	OBL-CN/ABS-CN	pig

(Intended for: 'Utay kills pigs there.')

The locative deictics exhibit the same grammatical properties and observe the same semantic restrictions as *tanian* and *icuwa*. The syntactic analysis that I have elaborated on for the derivation of interrogative verbs can thus be invoked to explain the syntactic distributions of the locative deictics.

3.2 Manner verbs

Given that (*na*)*quni* 'do what/how' in Kavalan and *maan* 'do what/how' in Amis can undergo head-movement to *v* to derive interrogative verbs, it should not be surprising that their non-interrogative counterparts, e.g., manner deictics and manner adverbial expressions, are also syntactically realised as verbs. The following examples are for illustration.[16]

(52) Kavalan

a. *nayau-an-ku.*
that.way-PV-1SG.ERG
'I do (it) in that way.'

b. *nayau-an-na* *ya* *sunis-na.*
that.way-PV-3ERG ABS child-3SG.GEN
'He treats his child in that way.'

c. *paqanas-an-ku* *t‹m›ayta* *ya* *sudad.*
slow-PV-1SG.ERG <AV>see ABS book
'I read the book slowly.'[17] (Chang 2006:46)

(53) Amis

ha'en-en	*k-u*	*kamay.*
this.way-PV	ABS-CN	hand

'Make your hand like this!'

The syntactic analysis of the present study can capture the syntactic similarities between manner interrogatives and manner deictics/adverbials in a straightforward way. Their verbal usage is derived because they can be merged with the verb-defining head via licit head movement. This syntactic analysis predicts that the position of an adverbial in a syntactic tree with respect to the position of *v* can determine whether it can be used as a verb. Only adverbials that occupy the head position of a phrase lower than *v* can undergo

---

[16] The example in (52c) has been re-glossed.

[17] As pointed out by an anonymous reviewer, the embedded event in (24b) and (25b) can be interpreted as the result state, whereas (52c) is not amenable to the same interpretation. Despite the weak semantic connection between (24b)/(25b) and (52c), both exhibit the same structural properties in that the embedded verb occurs in a non-finite complement clause (Chang 2010; Lin 2013).

head movement to *v* and be realised as a verb. This prediction is confirmed by the findings on Formosan adverbial verb syntax in Chang (2006) and Holmer (2010). While low adverbials like manner and frequency expressions are able to take voice markers, high adverbials like epistemic and evaluative expressions cannot be affixed with voice markers.

The two sets of data in (52) and (53), together with the other interrogative sentences discussed so far, suffice to show that there is no absolute underlying distinction between adverbs and verbs in Kavalan and Amis. The notion of adverbs as a distinct syntactic category is also fuzzy in other languages, e.g., Dyirbal, where adverbs modifying verbs show the same inflection as verbs (Dixon 1972). In general, the overlap between adverbial and verbal expressions provides evidence for the proposed syntactic approach, in which roots are not identified with particular lexical categories. The categories of words are defined with respect to the syntactic environments where they occur.

4 Interrogative words that cannot be verbs

The syntactic analysis I have been arguing for can also predict what interrogative words can and cannot be used as verbs in Kavalan and Amis based on the semantics of the voice markers, or verb-defining heads, and established syntactic principles and constraints. Why certain interrogative words in Kavalan and Amis cannot take voice markers and be used as verbs finds a natural explanation in the proposed syntactic framework. The analysis predicts that if an interrogative word must be adjoined to another phrase, it cannot be utilised as a verb as its movement from an adjoined position to *v* would violate the ECP or the Transparence Condition. Also, if the merger of an interrogative word with *v* results in a structure whose interpretation does not correspond to the intended question, that interrogative word should not occur in that verbal environment on the intended interpretation. I have argued that these two considerations rule out the use of adjunct *tanian* or *icuwa* 'where' as a verb. In sections 4.1 and 4.2, I show that 'when', 'why', 'whose', and 'which' cannot be syntactically realised as a verb in Kavalan and Amis for the same reasons. Section 4.3 discusses why 'who' in the two languages cannot be used as a verb.

4.1 'When' and 'why'

The same structural principles that prevent the derivation of adjunct *tanian* or *icuwa* 'where' as a verb can also explain why 'when' (Kavalan *qumni* and Amis *ihakuwa*) and 'why' (Kavalan *mana* and Amis *naw*) cannot be verbs in Kavalan and Amis. As shown in the following examples, the words that denote 'when' and 'why' in the two languages cannot take voice markers and be used as verbs.

(54) Kavalan

a. *qumnitayta-an-su* *ya* *ti-buya*?
when see-PV-2SG.ERG ABS NCM-PN
'When do you see Buya?'

b. * *qumni-an-su* *t‹m›ayta* *ti-buya-an*?
when-PV-2SG.ERG <AV>see NCM-PN-LOC

c. *mana ala-an-su ya kelisiw-ku*?
why take-PV-2SG.ERG ABS money-1SG.GEN
'Why do you take my money?'

d. * *mana-an-su m-ala ya kelisiw-ku*?
why-PV-2SG.ERG AV-take ABS money-1SG.GEN

(55) Amis

a. *ihakuwa ma-alaw isu ci panay*?
when PV-see 2SG.ERG NCM PN
'When do you see Panay?'

b. * *ihakuwa-en ma-alaw isu ci panay*?
when-PV PV-see 2SG.ERG NCM PN

c. *naw ma-ulah ci panay ci lekal-an*?
why AV-like NCM PN NCM PN-OBL
'Why does Panay like Lekal?'

d. * *naw-en ma-ulah ci panay ci lekal-an*?
why-PV AV-like NCM PN NCM PN-OBL

Phrases that denote temporal information are typically adjoined to the phrases that they modify. If it is assumed that 'when' in Kavalan and Amis is also adjoined to VP or IP, the ungrammaticality of (54b) and (55b) then follows from the syntactic structure of adjunction. For Kavalan *qumni* 'when' and Amis *ihakuwa* 'when' to move out of an adjoined position would violate the ECP or the Transparence Condition. On the assumption that 'why' is directly merged in Spec, CP (Ko 2005) or Spec, INT in the fine structure of CP (Rizzi 2001), the ungrammaticality of Kavalan *mana* 'why' and Amis *naw* 'why' as a verb in (54d) and (55d) can be attributed to the Transparence Condition.

4.2 'Which' and 'whose'

Before I explain why 'which' and 'whose' in Kavalan and Amis cannot be syntactically realised as a verb, it is imperative to consider where the two interrogative words are base-generated. While it has become a common assumption that English determiner *the*, demonstratives *this*/*that*/*these*/*those*, and genitive marker *'s*, occupy the head of DP per Abney's (1987) DP hypothesis to account for their complementary distribution, whether the same analysis can apply to other languages is controversial because some languages allow a determiner to co-occur with a demonstrative (Bernstein 1997).

Tang (2006) shows that the syntactic distributions of demonstratives and possessives are quite complicated in Formosan languages as they can occur either in a post-nominal position or in a pre-nominal position. The following Kavalan and Amis examples illustrate the two patterns. Note that Amis demonstratives are bound morphemes that must co-occur with case markers or common noun markers. More importantly, they can only occur in a pre-nominal position, as suggested by the ungrammaticality of (57a).

(56) Kavalan

a. *sudad* *zau*
book this
'this book'

b. *zau=ay* *sudad*
this=REL book
'this book'

c. *sudad* *zaku*
book 1SG.POSS
'my book'

d. *zaku=ay* *sudad*
1SG.POSS=REL book
'my book'

(57) Amis

a. * *c‹m›ikay* *wawa* *k-u-ni.*
<AV>run child ABS-CN-this

b. *c‹m›ikay* *k-u-ni* *(a)* *wawa.*
<AV>run ABS-CN-this LNK child
'This child is running.'

c. *wacu* *nu* *maku*
dog GEN 1SG.POSS
'my dog'

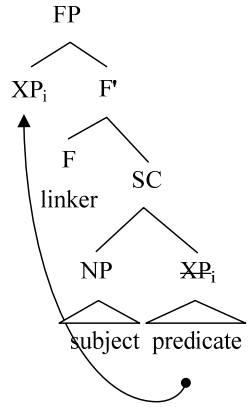
d. *(nu)* *maku* *a* *wacu*
GEN 1SG.POSS LNK dog
'my dog'

It should also be noted that when Kavalan and Amis demonstratives and possessives occur pre-nominally, an additional marker *=ay* or *a* is inserted between them and the noun, as shown in (56b), (56d), (57b), and (57d). The occurrence of the marker *=ay* or *a* is forbidden when demonstratives and possessives follow nouns ((56a), (56c), (57c)). The two markers, *=ay* in Kavalan and *a* in Amis, indicate a modification structure in a noun phrase, occurring between the modifier and the modified noun. The relationship of modification is broadly and loosely defined. They function to introduce diverse kinds of modifiers of a noun, including relative clauses, adjectives, numerals, quantifiers, demonstratives, and possessors.

Due to the parallel functions between Kavalan *=ay* or Amis *a* and linkers connecting a noun and its modifier in other languages, I assume that *=ay* or *a* heads its own functional projection, FP, and triggers DP-internal Predicate Inversion. According to den Dikken & Singhapreecha (2004) and Simpson (2001), a noun phrase where the noun and its modifiers are connected by a linker always involves predication. Moreover, the presence of the linker induces predicate inversion. On den Dikken & Singhapreecha's (2004) account, the noun and its modifier in this construction is base-generated as the subject and predicate of a small clause (SC) respectively. The linker heads its own functional

projection, FP, and prompts the predicate to move to Spec, FP. The derivation is schematically represented by the structure in (58).

(58) DP-internal predicate inversion (den Dikken & Singhapreecha 2004)



Like other modifiers of nouns, *mayni* 'which' and *zanitiana* 'whose' are followed by *=ay* and must occur before a noun. This is also true of Amis *icuwaay* 'which' and *nima* 'whose', which can be followed by *a* and must precede a noun, as illustrated below.

(59) Kavalan

a. [*mayni=ay* *sunis*] *ya* *tayta-an* *ni* *imuy*?
which=REL child ABS see-PV ERG PN
'Which child does Imuy see?' (Lit. 'The person that Imuy sees is which child?')

b. [*zanitiana=ay* *kelisiw*] *ya* *ala-an=ay* *ni* *utay*?
whose=REL money ABS take-PV=REL ERG PN
'Whose money does Utay take?' (Lit. 'The stuff that Utay takes is whose money?')

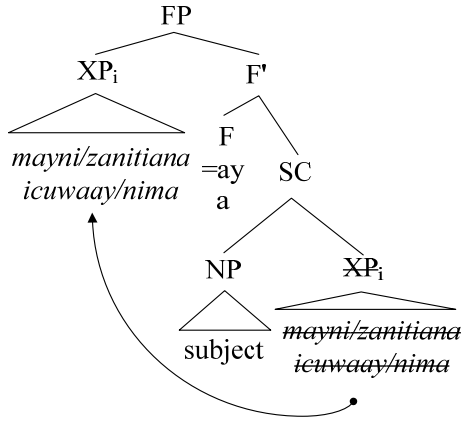
(60) Amis

a. [*icuwaay* *a* *wacu*] *k-u* *ka-ulah-an* *isu*?
which LNK dog ABS-CN KA-like-LA 2SG.ERG
'Which dog do you like?' (Lit. 'What you like is which dog?')

b. [*nima* *a* *wacu*] *k-u* *mi-kalat-ay* *t-u* *pusi* *aku*?
whose LNK dog ABS-CN AV-bite-FAC OBL-CN cat 1SG.GEN
'Whose dog bites my cat?' (Lit. The thing that bites my cat is whose dog?)

I thus assume that these modifier-like interrogative phrases have the structural representation in (61). They must undergo DP-internal Predicate Inversion, triggered by the presence of F, which is headed by *=ay* or *a*. This explains why 'which' and 'whose' in Kavalan and Amis must occur in the pre-nominal position.

(61) Structure of 'which' and 'whose' in Kavalan and Amis



While the fronting of 'which' and 'whose' before the head noun is obligatory, their non-interrogative counterparts can either move across the head noun or remain in its merge position. An anonymous reviewer suggests that the obligatory fronting of 'which' and 'whose' to Spec, FP can be attributed to the [+WH]-Agreement requirement proposed by Aissen (1996) for Tzotzil. While the non-interrogative genitive in Tzotzil follows the head noun, *buch'u* 'who; whose' must precede the head noun when the entire DP is pied-piped to Spec, CP, as illustrated in (62).

(62) Tzotzil (Aissen 1996:456)

a. *I-cham* *x-ch'amal* *li* *Xun-e.*
CP-died A3-child the Xun-ENC
'Xun's child died.' (p. 456)

b. [*Buch'u* *x-ch'amal*] *i-cham*?
who A3-child CP-died
'Whose child died?' (p. 457)

c. * [*X-ch'amal* *buch'u*] *i-cham*?
A3-child who CP-died
'Whose child died?' (p. 457)

Aissen (1996) argues that *buch'u* 'who; whose' in (62b) must move to the specifier of the pied-piped subject DP so that its [+WH] feature can Agree with C[+WH]. On the assumption that Agreement is transitive (Rizzi 1990), after its movement to Spec, DP, *buch'u* 'who; whose' in (62b) first Agrees with this subject DP by Spec-Head Agreement and Head-Projection Agreement and then the entire subject DP Agrees with C[+WH]. Transitivity of Agreement licenses the [+WH]-Agreement between *buch'u* 'who; whose' and C[+WH]. The Wh-Criterion is thus satisfied in this configuration.

Unlike Tzotzil, Kavalan and Amis are wh-in-situ languages (Lin 2013; Wei 2009). Nevertheless, a [+WH] phrase in the two languages still needs to Agree with C[+WH] to be interpreted as an interrogative. This Agreement is crucial for interpretation, as the wh-words in the two languages can function either as interrogatives or indefinites. On the assumption that wh-phrases move to Spec, CP at LF in Kavalan and Amis, the same explanation proposed by Aissen (1996) for Tzotzil can apply to Kavalan and Amis. In

other words, the locality constraint of the Agree operation is what prompts 'which' and 'whose' in (61) to obligatorily move to Spec, FP. After the movement, 'which' and 'whose' Agrees with FP through Spec-Head Agreement and Head-Projection Agreement and then FP Agrees with C[+WH]. Transitivity of Agreement licenses the [+WH]-Agreement between 'which'/'whose' and C[+WH]. Without movement to Spec, FP, the Wh-Criterion would not be satisfied, as 'which' and 'whose', being embedded inside SC, would not be able to Agree with C[+WH] in any direct or indirect way.

If the structure in (61) is correct, the reason why the interrogative words that denote 'which' and 'whose' in Kavalan and Amis cannot be used as a verb can be attributed to their position in a specifier. As adjunct *tanian* 'where' and *icuwa* 'where' are forbidden from moving to *v* due to violations of the ECP or the Transparence Condition, *mayni* 'which', *zanitiana* 'whose, *icuwaay* 'which', and *nima* 'whose' are not allowed to move to *v* either. During the course of the derivation, they must move to the specifier of FP headed by *=ay* or *a* via DP-internal Predicate Inversion. Their movement from the specifier position to a c-commanding head would result in an illegitimate configuration that does not conform to either the ECP or the Transparence Condition. Therefore, 'which' and 'whose' cannot take voice markers and be used as verbs in Kavalan and Amis.

In addition to the syntactic consideration, verbal 'which' and 'whose' in Kavalan and Amis are also ruled out on semantic grounds. As discussed above, *tanian*/*icuwa* 'where' must take the patient voice marker to be used as a verb, and this is because the patient voice marker, as the lexical realisation of v_{CAUSE}, can assign an agent/causer argument and the semantics of the *v*P-shell structure with v_{CAUSE} followed by v_{BECOME} is compatible with a question that inquires about the location of the theme argument in a ditransitive event, which is a typical and canonical type of 'where'-questions. The interrogative words that denote 'how many', i.e., *tani* and *pina*, exhibit the same grammatical patterns when they are used as verbs. Due to the semantics of v_{CAUSE} and v_{BECOME}, a question that is formed with verbal *tani* or *pina* receives a unique interpretation where the speaker suspects that the quantity of the affected theme argument might change. The generalisation is that a question with a PV-marked interrogative verb always implies a change of state of the theme argument with respect to the meaning of the interrogative word. In the case of *tanian*/*icuwa* 'where', the location of the theme argument changes because of some action performed by the agent. As for *tani*/*pina* 'how many', what might change is the quantity of the theme argument that will be affected by the action of the agent.

This type of causal relation and change-of-state implicature is absent in a 'which'-or 'whose'-question. Take (59a) as an example, 'Which child does Imuy see?' The intended meaning of this question does not imply that the theme argument will undergo some change with respect to the meaning of 'which', e.g., from 'this' to 'that'. The same reasoning also applies to a 'whose'-question like (59b), 'Whose money does Utay take?' Its intended meaning does not concern change of possession, e.g., 'the money became whose'. To summarise, the meaning of a 'which'-question or a 'whose'-question is incompatible with the syntactic representation of a PV-marked interrogative verb and its associated semantic interpretation. The reason why a 'which'-question or a 'whose'-question is not associated with the semantics of a PV-marked interrogative verb is elusive, but the empirical generalisation remains intact.

4.3 'Who'

Based on the contention that all the interrogative verbs in Kavalan and Amis are derived in Syntax, their derivation must conform to established syntactic principles and constraints. Conformity to syntactic principles and constraints, however, only leads to grammatical well-formedness, but not appropriate use of expressions. Within the framework of Distributed Morphology, grammatical well-formedness is concerned with the licensing requirements of formal features in the abstract syntactic representation, whereas appropriate use of expressions is regulated by a separate component, Encyclopaedia (Harley & Noyer 2000). This division of labour, as argued by Harley & Noyer (2000), can account for why a nominalised verb can or cannot co-occur with an agent in the possessive form, e.g., the acceptability differences between *the insects' destruction of the crop* and **John's growth of tomatoes*. The unacceptability of **John's growth of tomatoes* is not due to the failure of formal licensing requirements, but results from encyclopaedic anomaly. As *grow* denotes a spontaneous and internally-caused activity, the subject of the nominalised form, *growth*, cannot be construed as an agent. The subject position of *growth* can only be occupied by a theme, e.g., *the tomatoes' growth*.

The examples in (63) also illustrate the importance of encyclopaedic knowledge in addition to grammatical well-formedness. Given the grammaticality of (63a) and (63b), the unacceptability of **Adultery's separation of Jim and Tammy Faye* in (63c) is surprising. However, encyclopaedia knowledge of *separation* can account for the contrast. While *the teacher* in (63b) is a true external causer, *adultery* in (63c) is merely a facilitator, just like *John* in **John's growth of tomatoes*.

(63) Harley & Noyer (2000:365, 366)

a. Jim and Tammy Faye separated.
   Jim and Tammy Faye's separation

b. The teacher separated the children.
   The teacher's separation of the children

c. Adultery separated Jim and Tammy Faye.
   * Adultery's separation of Jim and Tammy Faye

The ungrammatical example occurs in a syntactically well-formed structure just like its grammatical counterparts, but is ruled out by encyclopaedic considerations. In a similar vein, I will argue that the reason why 'who' cannot be an interrogative verb in Kavalan and Amis is due to the contradiction between the canonical interpretation of 'who' in Encyclopaedia and the interpretation imposed on it in a patient voice construction.

Nicolae & Scontras (2010) argue that 'who' in Austronesian languages should be analysed as the interrogative form of a proper noun of the type <e> that denotes individuals based on the following grammatical properties of 'who'. Like a proper noun, 'who' is not able to occur in an existential construction, nor can it be incorporated into a verb. It is used in some languages to question names. It can take a proper noun determiner or a non-common-noun classifier.

A full justification for the analysis of *cima* 'who' in Amis and *tiana* 'who' in Kavalan as the interrogative form of a proper noun is beyond the scope of the present paper, but I am convinced that this analysis is on the right track due to the following grammatical

properties of *cima* and *tiana*. First of all, when the pivot of an existential sentence is a pronoun or a proper noun, the sentence must be interpreted as a locative construction, not an existential construction (Zeitoun et al. 1999), and this is also true of *cima* and *tiana*. Second, the two interrogative words are used to question one's name. Finally, the non-common noun marker *ci* in Amis is inherent in the interrogative word itself; *tiana* in Kavalan can take the non-common noun marker *ti-*, which is also attached to proper names.

These morphosyntactic properties of *tiana* and *cima* indicate that they should be analysed as the interrogative form of a personal proper name. This further suggests that they are of the semantic type <e>, denoting individuals. The encyclopaedic knowledge of a 'who'-question states that 'who' denotes individuals and the function of a 'who'-question is to ask the addressee to pick out a particular individual. The merger of 'who' with *v* would result in a structure whose semantic interpretation is inconsistent with the canonical meaning of a typical 'who'-question in Encyclopaedia. If 'who' is merged with the patient voice marker *-an* or *-en* in the two languages, the resultant interrogative verb should also be construed as a causative verb like PV-marked 'where' and 'how many', with the theme argument undergoing a change of state with respect to the meaning of 'who'. Although this interpretation is not logically impossible, it does not correspond to the canonical meaning of a typical 'who'-question in Encyclopaedia like 'Who did you hit?', which can be paraphrased as 'X did something to some person and that person is who?'. No change of state of the theme argument concerning the status or meaning of 'who' is involved in a typical 'who'-question. Therefore, verbal 'who' in Kavalan and Amis is ruled out due to the encyclopaedic anomaly instead of violations of licensing conditions or syntactic principles.

Verbal 'who' with the patient voice marker is grammatically well formed, but its coerced interpretation in the patient voice construction contradicts the encyclopaedic knowledge of a typical 'who'-question. On this analysis, the lack of PV-marked 'who' in Kavalan and Amis is merely an accidental gap. A typological study on whether 'who' can be a verb in other Formosan languages might shed light on the validity of resorting to Encyclopaedia as an explanation. If we can find a language where 'who' can be utilised as a verb with a specialised meaning, just like the specialised usage of 'how many' in Kavalan and Amis, this will further corroborate the analysis that distinguishes formal licensing conditions from encyclopaedic coercion of interpretations.

5 Conclusion

The possibility or impossibility of using an interrogative word in Kavalan and Amis as a verb is motivated by syntactic principles/constraints, either universal or language-specific. There is no need to stipulate the syntactic categories of interrogative words in the lexicon. Once the assumption that derivational morphology, e.g., the Kavalan and Amis voice system, must operate in the lexicon is abandoned, the syntactic behaviours of interrogative verbs find a uniform explanation in Syntax. Interrogative words are not lexically specified for syntactic categories. Their syntactic categories and the relevant grammatical patterns follow from the interaction of the following factors: The inherent semantics of interrogative words, the available interpretation of the question where they occur, the verbal structures of the voice markers, and the syntactic principles

and constraints that are cross-linguistically valid, e.g., the ECP or the Transparence Condition.

Finally, the syntactic approach can be extended to non-interrogative words as well and makes correct predictions about what interrogative words can and cannot be used as verbs. It is thus able to depict the overall grammatical system of Kavalan and Amis and proves to be a promising way for future typological research. Interrogative verbs are not unconstrained lexical idiosyncrasies. Instead, their derivations are systematically conditioned in Syntax.

References

Abney, Steven Paul. 1987. The English noun phrase in its sentential aspect. PhD dissertation, Cambridge, MA: Massachusetts Institute of Technology.

Aissen, Judith. 1996. Pied-piping, abstract agreement, and functional projections in Tzotzil. *Natural Language and Linguistic Theory* 14.3:447–491.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Bernstein, Judy B. 1997. Demonstratives and reinforcers in Romance and Germanic languages. *Lingua* 102.2-3:87–113.

Chang, Henry Y. 1997. Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu: National Tsing Hua University.

Chang, Henry Y. 2006. The guest playing host: Adverbial modifiers as matrix verbs in Kavalan. In *Clause structure and adjuncts in Austronesian languages*, Hans-Martin Gärtner, Paul Law, and Joachim Sabel, 43–82. Studies in generative grammar 87. Berlin and New York: Mouton de Gruyter.

Chang, Henry Y. 2010. On the syntax of Formosan adverbial verb constructions. In Raphael Mercado, Eric Potsdam, and Lisa deMena Travis (eds.), *Austronesian and theoretical linguistics*, 183–211. Linguistik Aktuell/Linguistics Today 167. Amsterdam and Philadelphia: John Benjamins.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Studies in Generative Grammar 9. Dordrecht: Foris.

Corver, Norbert. 1997. *Much*-support as a last resort. *Linguistic Inquiry* 28.1:119–164.

den Dikken, Marcel, and Pornsiri Singhapreecha. 2004. Complex noun phrases and linkers. *Syntax* 7.1:1–54.

Dixon, R.M.W. 1972. *The Dyirbal language of North Queensland*. Cambridge Studies in Linguistics 9. London: Cambridge University Press.

Doetjes, Jenny Sandra. 1997. Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English. PhD dissertation. The Hague: Leiden University.

Ernst, Thomas. 2002. *The syntax of adjuncts*. Cambridge studies in linguistics 96. New York: Cambridge University Press.

Hagège, Claude. 2008. Towards a typology of interrogative verbs. *Linguistic Typology* 12.1:1–44.

Hale, Kenneth and Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 53–109. Current studies in linguistics 24. Cambridge, MA: MIT Press.

Halle, Morris and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20: Essays in linguistics in honor of Sylvain*

Bromberger, ed. by Kenneth Hale and Samuel Jay Keyser, 111–176. Current Studies in Linguistics 24. Cambridge, MA: MIT Press.

Halle, Morris and Alec Marantz. 1994. Some key features of distributed morphology. In *MIT working papers in linguistics* 21, ed. by A. Carnie, H. Harley, with T. Bures, 275–288. Cambridge, MA: MIT.

Harley, Heidi. 2009. The morphology of nominalizations and the syntax of vP. In *Quantification, definiteness, and nominalization*, ed. by Anastasia Giannakidou and Monika Rathert, 321–343. Oxford studies in theoretical linguistics 24. Oxford: Oxford University Press.

Harley, Heidi and Rolf Noyer. 2000. Formal versus encyclopedic properties of vocabulary: Evidence from nominalisations. In *The lexicon-encyclopedia interface*, ed. by Bert Peeters, 349–374. Current Research in the Semantics/Pragmatics Interface, Vol. 5. Amsterdam: Elsevier.

Holmer, Arthur. 2010. Seediq adverbial verbs: A review of the evidence. In *Austronesian and theoretical linguistics*, ed. by Raphael Mercado, Eric Potsdam, and Lisa deMena Travis, 163–182. Linguistik Aktuell/Linguistics Today 167. Amsterdam: John Benjamins.

Huang, Shuanfan and Michael Tanangkingsing. 2011. A discourse explanation of the transitivity phenomena in Kavalan, Squliq, and Tsou. *Oceanic Linguistics* 50.1:93–119.

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1999. Interrogative constructions in some Formosan languages. In *Chinese languages and linguistics V: Interactions in language*, Yuen-mei Yin, I-li Yang, and Hui-chen Chan, 639–680. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 2. Taipei: Academia Sinica.

Ko, Heejeong. 2005. Syntax of *why*-in-situ: Merge into [Spec, CP] in the overt syntax. *Natural Language and Linguistic Theory* 23.4:867–916.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. by Johan Rooryck and Laurie Zaring, 109–137. Studies in Natural Language and Linguistic Theory 33. Dordrecht and Boston: Kluwer Academic Publisher.

Liao, Hsiu-chuan. 2002. The interpretation of *tu* and Kavalan ergativity. *Oceanic Linguistics* 41.1:140–158.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine Languages. PhD dissertation. Honolulu: University of Hawai‘i at Mānoa.

Lin, Dong-yi. 2010. A syntactic account of nominalization in Kavalan. Paper presented at Syntax of the World’s Languages IV. Institut des Sciences de l’Homme, Lyon, France, 23–26 September.

Lin, Dong-yi. 2012. Interrogative verbs in Kavalan and Amis. *Oceanic Linguistics* 51.1:182–206.

Lin, Dong-yi. 2013. Interrogative constructions in Kavalan and Amis. PhD dissertation. Gainesville: University of Florida.

Marantz, Alec. 1997. No escape from Syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4.2:201–225.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37.1:69–109.

Nicolae, Andrea and Greg Scontras. 2010. What Austronesian tells us about *who*. Paper presented at Austronesian Formal Linguistics Association XVII. Stony Brook University, NY. 7–9 May.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. In *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, ed. by Guglielmo Cinque and Giampaolo Salvi, 287–296. Amsterdam: Elsevier.

Ross, Malcolm and Stacy F. Teng. 2005. Formosan Languages and Linguistic Typology. *Language and Linguistics* 6.4:739–781.

Simpson, Andrew. 2001. Definiteness agreement and the Chinese DP. *Language and Linguistics* 2.1:125–156.

Starosta, Stanley. 2002. Austronesian 'focus' as derivation: Evidence from nominalization. *Language and Linguistics* 3.2:427–479. Reprinted in Elizabeth Zeitoun, ed., *Formosan Linguistics: Stanley Starosta's contributions*, 553–606. Language and Linguistics Monograph Series C6-2. Taipei: Academia Sinica, 2009.

Tang, Chih-Chen Jane. 2006. Demonstratives and DP structure in Formosan languages. *Language and Linguistics* 7.4:935–990.

Travis, Lisa deMena. 1984. Parameters and effects of word order variation. PhD dissertation. Cambridge, MA: Massachusetts Institute of Technology.

Wei, Ting-Chi. 2009. *Wh*-interrogatives in Amis. *Language and Linguistics* 10.2:315–374. [In Chinese]

Wu, Joy J. 2006. Verb classification, case marking, and grammatical relations in Amis. PhD dissertation. Buffalo: Universty at Buffalo, State University of New York.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, and Anna H. Chang. 1999. Existential, possessive, and locative constructions in Formosan languages. *Oceanic Linguistics* 38.1:1–42.

12 *Why exclamatives in Budai Rukai*[*]

LI-MAY SUNG

1 Introduction

Among non-declarative sentence types, exclamatives have received considerably less attention than interrogatives in the literature, even though a strong resemblance appears to exist between exclamative and interrogative constructions. Part of the scarce literature on exclamatives, which can be traced back to the valuable works of Elliott (1971, 1974) and Grimshaw (1979) in the 1970's and later in research concerned with numerous European and Romance languages, has greatly broadened our current understanding of exclamatives, especially in contrast with interrogatives. Canonical exclamatives like (1) in English represent a linguistically coherent clause type. They have roughly the same morphosyntax as embedded interrogatives, as in (2a), an affinity that is well attested crosslinguistically. Exclamatives can, however, be disambiguated from interrogatives with intonational cues, partially reflected in written forms by the exclamation point (!), as well as with the optional employment of a variety of interjections such as *boy*, *my*, and *oh* that are incompatible with interrogatives, as shown in (2b), and finally by the fact that exclamatives appear routinely in the left periphery without displaying subject-verb inversion, unlike interrogatives (3).

(1) Canonical exclamatives in English
   a. What big eyes you have!
   b. What a big car he bought!
   c. How you've grown!
   d. How big a car he bought!
   e. How many books you read!

[*] I wish to express my sincere appreciation to two anonymous reviewers for their very constructive comments. I am also grateful to my two Budai consultants, Salrabu Ka Taikadresengange (Chi-chuan Ke) and Balenge Ka Taikasepelane (Chu-hua Ke), for providing the data on which this paper is based. Errors and shortcomings are of course my own. This research is funded by two projects, "Formation and fluidity of the island world", National Taiwan University (101R3108-103R3108) and "A typological study of Austronesian languages in Taiwan and their revitalisation", National Science Council (NSC 100-2420-H-002-035-MY3).

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, *291–311*
Asia-Pacific Linguistics, 2015

(2) Embedded interrogatives in English
  a. John wonders how Tom has grown.
  b. * John wonders, boy, how Tom has grown.

(3) *Wh*-interrogatives involving subject-verb inversion
  a. What did he buy?
  b. How much did he spend?
  c. How many books did you read?

In addition, not all types of interrogatives correspond to exclamatives. Systematically excluded from exclamative readings are interrogative phrases that do not have a degree-based semantics. For example, in English, *wh*-phrases headed by *who*, *when*, *where*, *which*, and *why* are not grammatical exclamatives, as shown in (4). The constraint exemplified in (4) is not universal, however. As shown in (5), there are many languages that do allow some matrix *where* or *who* exclamatives corresponding to those in (4). Additionally, canonical interrogative-style exclamatives are not the only vehicles for indicating exclamativity. As shown in (6), strings of exclamation points onto the end of declaratives, various adverbials (e.g., *absolutely*, *totally*), degree modifiers (e.g., *so*), or stand-alone nominal phrases (e.g., *The noise!*) all can signal similar exclamative readings as well, in proper context.

(4) *Wh*-phrases not as exclamatives
  a. * Who you've hired!
  b. * When we visited Rome!
  c. * Where she studied!
  d. * Which/What (book) she found!
  e. * Which size eyes you have!
  f. * Why she did what she did!

(5) a. Italian

Dove	*si*	*arrampicano,*	*questi*	*ragazzi*![1]
Where	they	climb.3PL	these	boys

'The places they climb, these boys!' (Michaelis 2001)

b. Turkish

Kimleri	*gördük,*	(*kim*)!
who.PL.OBJ	saw.PST.1PL	who

'The people we saw!' (Michaelis 2001)

c. German

Wen	*der*	*alles*	*eingeladen*	*hat*!
who	he	all	invited	has

'Who all he has invited!' (D'Avis 2002)

[1] The glosses of the data taken from other authors appear as in the original sources. Examples from Sung (2011) are slightly changed to conform to the orthography in this paper. Glossing abbreviations used in this paper follow the Leipzig Glossing Rules. Additional abbreviations that are not covered in the LGR are as follows: ACT, active voice marker; BN, bound nominative; FILL, filler; FIN, finite marker; NFIN, non-finite marker; NFUT, non-future; PASS, passive voice marker; RED, reduplication; STAT, stative marker.

(6) Non-canonical exclamatives
   a. You have big eyes!
   b. It was an absolutely wonderful stay!
   c. It is so hot!
   d. The strange things that he says!
   e. The noise!

All of the examples discussed above illustrate not only diverse syntactic forms but also different types of exclamatives, both within and across languages. Although interest in this clause type has increased in recent years, the overall amount of work on exclamatives is still relatively little in the literature, and research on languages other than English is very limited. Zanuttini & Portner's (2003) study represented a milestone for this increased interest, and a variety of scholarly works on exclamatives have appeared since (e.g., Villalba 2003; Castroviejo Miró 2006; Ono 2006; Rett 2008, 2011; Kaufman 2011; Potsdam 2011, and others). Many of the recent works focus on English, Japanese, Malagasy, Romance, and Austronesian languages, but little attention has been given to exclamatives in Formosan languages. The main goal of the present study is to contribute to the effort to bring Formosan data to the fore and, specifically, to explore two types of *why* exclamatives in Budai Rukai.[2] The two *why* exclamatives manifest distinct syntactic natures; the *a-* type is verbal (7) whereas the *asi* type is nominal (8).

(7) Verbal *a-* exclamative
   a-ni[3] *ka-lragi* *kai* *kaswi-su*!
   do.why-3 STAT.NFIN-long this pants-2SG.GEN
   'How long your pants are!' (Lit: 'How come your pants are (so) long!')

(8) Nominal *asi* exclamative
   asi *amia* *kai* *riladh-ane* *ki* *lasu*!
   ASI as.this this thin-NMLZ GEN person
   'How come the man is so thin!' (Lit: 'How come as so the man's thinness!')

The present study will proceed as follows. Section 2 gives a brief overview of *wh*-interrogatives in Budai Rukai. After that, sections 3 and 4 introduce the two types of *why* exclamatives that this paper is concerned with, from a comparative perspective. Following the research line of Elliott (1971, 1974), Grimshaw (1979), Michaelis & Lambrecht (1996), Michaelis (2001), Zanuttini & Portner (2003), and Rett (2008, 2011), this study shows that both *why* exclamatives in Budai Rukai, despite their syntactic difference, feature the same semantic properties as the better understood canonical degree *what* or *how* exclamatives in other languages. The distinctive properties in question include (i) a degree component, (ii) a speaker-based evaluation, and (iii) a factive nature. Section 5 concludes this work.

---

[2] In this paper, the term "Budai Rukai" or "Budai" is mainly used to refer to the Budai proper dialect of Rukai, spoken in Budai village. This village is located in Wutai Township of Pingtung County in southern Taiwan. The Rukai language includes six major dialects: Tanan, Budai, Labuan, Maga, Tona, and Mantauran. Previous studies on Rukai dialectal subgrouping can be found in Dyen (1963), Ferrell (1969), Li (1973, 1977, 1990, 1996), Tsuchida (1976), Ho (1983), Starosta (1995), and Zeitoun (1995, 2003).

[3] The bound pronoun *-ni* can indicate either singular or plural.

2 Interrogative *wh*-phrases in Budai Rukai

Before delving into a detailed investigation of *why* exclamatives, a brief introduction of some grammatical properties of Budai is in order, with particular reference to those that are relevant to the issues addressed in this paper. Budai is a copula-less and basically predicate-initial language. The word order of arguments is relatively free, being either X(O)S or XS(O), as illustrated in (9).[4] Noun phrases, quantifiers, adjectives, statives, and verbs all may function as matrix predicates. Noun phrases, when used as arguments, are always preceded by demonstratives or nominal case markers. An understanding of these nominal markers is important to the discussion of *asi*'s nominal nature in section 4. Demonstratives are distinguished in terms of visibility and proximity to the speaker: *kai* 'this [+visible, −distance]', *kuini/kui* 'that [+visible, −distance]', *kavai* 'that [+visible, +distance]', and *ku(i)dra* 'that [−visible, +distance]'. As for nominal case markers, Budai makes a three-way case distinction, shown in Table 1, between nominative *ku/ka*, marking a subject argument; genitive *ki*, marking a non-subject agent/actor/possessor; and oblique *ku/ka/ki*, marking a non-subject argument.[5]

(9) a. Nouns as predicates
ngu-dra~drekai [SUBJECT *ka/ku* *Salrabu*]
belong.to-RED~cold.place NOM Salrabu
'Salrabu is a Rukai.'

b. Quantifiers as predicates
takalra [SUBJECT *ka/ku* *lralri* *ki* *Lrangpau*]
many NOM female.friend GEN Lrangpau
'Lrangpau has many female friends.'
(Lit: 'Lrangpau's female friends are many.')

c. Adjectives as predicates
laulapong [SUBJECT *ka* *laimai*]
white NOM clothes
'The clothes are white.'

d. Stative verbs as predicates
ma-lase *ki/*ka/*ku* *Lrailrai* [SUBJECT *ka/ku* *Salrabu*]
STAT.FIN-dislike OBL Lrailrai NOM Salrabu
'Salrabu dislikes Lrailrai.'

---

[4] When two arguments in a sentence are both marked by the same case marker, word order is crucial in distinguishing subject from object. Taking (10a) as an example, the word order within the complement is rather fixed, with *ki Salrabu* 'GEN Salrabu' interpreted as the agent subject. Switching the order between *Salrabu* and *Lrailrai* results in ungrammaticality with the intended reading 'I am surprised at Salrabu's hitting Lrailrai.'

[5] Kuo (1979) uses the terms "article" and "determiner" instead of "case-marker" for *ku*, *ka*, and *ki* in Budai. Based on naturalistic narratives taken from the NTU corpus of Formosan languages (Sung et al. 2008), Shih (2012) investigates the functions of *ku*, *ka*, and *ki* as case markers in Budai Rukai.

e. Action verbs as predicates
wa-lrumai *ki/*ka/*ku* *Lrailrai* [SUBJECT *ka/ku* *Salrabu*]
ACT.FIN-beat OBL Lrailrai NOM Salrabu
'Salrabu beat Lrailrai.'

Table 1: Nominal case system in Budai
(based on Li 1997; Chen 1999, 2008; Zeitoun 2000)

	Nominative	Oblique	Genitive
ku	*ku* (-visible, +distance ± animate)	*ku* (-visible, +distance, -human, ±generic)	
ka	*ka* (+visible, -distance ± animate)	*ka* (+visible, -distance, -human)	
ki		*ki* (+specific, +human) (+generic, -human)	*ki* (± animate)

As noted in Sung (2011), both nominal case markers and demonstratives in Budai have been grammaticalized into new kinds of markers that can indicate nominal complementation (10a), subordination (10b), or modification (10c)–(10d) associated with relative clauses. For instance, the example in (10c) is one instance of an ordinary relative clause, in which the finite clause *wa-lrumai ki Lrailrai* stands in appositive relation to the modified head noun, *lasu* 'man'. The case marker *ku* or *ka* preceding the lexical head *lasu* is reanalyzed as a relativizer. Budai allows either pre-nominal or post-nominal relative clauses. In the post-nominal clauses like the one in (10d), it is then the case marker accompanying the clause that is reanalyzed as a relativizer.

(10) a. *masamali-aku* [*ku/ka/kai/*alaka* *Ø-lrumai-Ø-ini*$_i$
surprise-1SG.NOM OBL/OBL/this/*COMP ACT-hit-NMLZ-3SG.GEN
*ki/*ka/*ku* *Salrabu*$_i$ *ki* *Lrailrai/iniane*]
GEN Salrabu OBL Lrailrai/3SG.OBL
'I am surprised at Salrabu's (frequently) hitting Lrailrai/her.' (Sung 2011 [23a])

b. [*lu* *ka* *ta-ki-tai~tai-ane*], *ngu-a-dalrane madu* *maganaelre*
if/when OBL NFUT-get-RED~taro-NMLZ go-FIN-field have.to all
'At the time of harvesting taro, all (of us) have to go to the field.'
(Sung 2011 [19b])

c. *ngu-a-bere*[6] [SUBJECT *ku* [*Ø-wa-lrumai* *ki* *Lrailrai* *ku* *lasu*]]
go-FIN-escape NOM ACT-FIN-hit OBL Lrailrai REL man
'The man that hit Lrailrai ran away.' (Sung 2011 [11])

[6] My elder language consultant prefers using *mu-a-bere* 'go-FIN-escape'.

d. *ma-lras-aku* *kane* [*ka* *lacenge* [*ka*
STAT.FIN-hate-1SG.NOM NFIN.eat OBL vegetable REL
ta-aga-ane *ki* *Lrailrai*]]
NFUT-cook-NMLZ GEN Lrailrai
'I hate to eat the vegetables that Lrailrai has cooked (because they did not look good in the way that Lrailrai cooked them).' (Sung 2011 [18])

The syntactic properties of Rukai interrogative *wh*-phrases (Li 1973; Chen 1999; Huang et al. 1999; Zeitoun 2000, 2007; Chen & Sung 2005) are not much different from those of Kucapungan Rukai, as elaborated upon in details by Chen (1999). Budai Rukai interrogatives can be grouped into nominal, adverbial, and verbal types based on their categorical distinctions. Consider the examples in (11)–(13), in which (i) nominal interrogatives are argument questions and can be case-marked, featuring either a *wh*-argument in situ or a headless-type pseudocleft/relativized construction; (ii) adverbial interrogatives are freer in position and cannot take case markers; and (iii) verbal interrogatives (including *a-* 'do.why' and a rich array of verbal complexes X-*tumane* 'do.how', literally 'do.what') generally occur clause-initially, and they may attract bound pronouns and inflect for voice/tense/aspect/mood. The verbal interrogative *a-* 'do.why' cannot be used independently; it always appears with bound pronouns: *a-naku/a-nau* 'why-1SG', *a-ta* 'why-1PL', *a-su* 'why-2SG', *a-nu* 'why-2PL', *a-ni* 'why-3'.[7] The resultant portmanteau forms, which do not contain tense and aspect specifications, are followed by an infinitival V2.

(11) Nominal interrogatives: *aneane* 'who; whose', *manemane* 'what', *saininu/kinu* (*ka/ku laimai*) 'which (clothes/kind of clothes)', *pia* 'how many (nonhuman)', *tapia* 'how many (human)', *sainu* 'how much; how (old/heavy/tall)'

a. Active agent subject questioned
pseudocleft
aneane$_i$ [SUBJECT *ka/ku* *wa-lrumai* *ki* *Muni* t_i]
who NOM ACT.FIN-hit OBL Muni
'Who hit Muni?' (Lit: 'The one that hit Muni is who?')

b. Passive patient subject questioned
pseudocleft
aneane$_i$ [SUBJECT *ka/ku* *ki-a-lrumai* *ki* *Salrabu* t_i]
who NOM PASS-FIN-hit OBL Salrabu
'Who was hit by Salrabu?' (Lit: 'The one that was hit by Salrabu is who?')

c. Active patient object questioned
pseudocleft
aneane$_i$ [SUBJECT *ka/ku* *Ø-lrumai-Ø* *ki* *Salrabu* t_i][8]
who NOM ACT-hit-NMLZ GEN Salrabu
'Whom did Salrabu hit?' (Lit: 'The one by Salrabu's hitting is who?')

---

[7] Bound pronouns can be either nominative or genitive, although the difference is only evident in the first person singular: *-(*C*)aku*/*-nau* (nominative) and *-li* (genitive).

[8] Sung (2011) identified two types of nominalizers, *-ane* and *-Ø*. The *-Ø* (zero) nominalizer is used for non-subject arguments, i.e., internal objects, when they are relativized, as in (11c).

d. Active patient object questioned
in situ
wa-lrumai *ki* *aneane* [SUBJECT *ka/ku* *Salrabu*]
ACT.FIN-hit OBL who NOM Salrabu
'Whom did Salrabu hit?' (Lit: 'Salrabu hit whom?')

(12) Adverbial interrogatives: *luigane/kuigane* 'when', *satumatumane* 'when', *inu* 'where'

a. *wa-kane* *kuigane* *ka* *Muni* (kuigane)?
ACT.FIN-eat NFUT.when NOM Muni (NFUT.when)
'When did Muni eat?'

b. *wa-kane* *sa-tuma~tumane* *ka*
ACT.FIN-eat what.time-RED~do.what NOM
Muni (*sa-tuma~tumane*)?
Muni (what.time-RED~do.what)
'When did Muni eat?'

c. *lri-kane* *luigane* *ka* *Muni* (*luigane*)?
FUT-eat FUT.when NOM Muni (FUT.when)
'When will Muni eat?'

(13) Verbal interrogatives: *a-* 'do.why; do.how.come', *ngi-a-tumane* 'do.how; do.what', *wa-(tu-tua-)tumane* 'do.how; do.what', and X-*tumane*, etc.

a. [V1 *a*]-*su* [V2 *Ø-lrumai*/**wa-lrumai*] *ki* *laucu*?
do.why-2SG NFIN-hit/*ACT.FIN-hit OBL laucu
'Why did you hit Laucu?'

b. [VERB *wa-tu-tua~tumane*]-*su* *lu*
ACT.FIN-do-RED~do.what-2SG if/when
Ø-aga-su *kai* *lacenge*?
NFIN-cook-2SG.GEN this vegetable
'How did you cook this vegetable?' (method *how*)
(Lit: 'How did you do if your cooking this vegetable?')

c. [VERB *ngi-a-tumane*]=*nga* *kui* *Ø-aga-su*?
self-FIN-do.what=PFV that NFIN-cook-2SG.GEN
'How is (the result of) your cooking? (Does it come out fine as expected?)'

In addition to the foregoing, there is another, relatively less-discussed *wh*-phrase, *asi*, which roughly translates as 'why; how come', shown in (14).

(14) a. *asi* *pia* *kuini*?
ASI do that
'Why did (you) do that?'

b. *asi* *kai* *kulrung-ane* *ki* *lasu*!
ASI this stupid-NMLZ GEN man
'How come the man is (so) stupid!' (Lit: 'How come the man's stupidness!')

The grammatical status of *asi* 'why' is not straightforwardly evident. Unlike the verbal *a-* 'do.why', *asi* cannot be further decomposed morphologically and does not attract any bound pronouns. Neither does *asi* behave like adverbial interrogatives, which have a freer distribution; *asi* appears strictly in the left periphery. As revealed in (14b), one featuring function of *asi* is to indicate an exclamation, with its distinctive nominality signaled by its embeddedness under a demonstrative *kai* and its co-occurrence with a nominalizing morpheme *-ane* and the presence of a genitive subject. We will come back to *asi* in section 4. The next section explores the verbal *a-* 'do.why', which has the function of expressing an exclamation in addition to that of inquiring for information.

3 Verbal *why* exclamatives in Budai Rukai

As Elliott (1974:232) has notably observed, not all *wh*-phrases are equally able to serve as matrix exclamatives. In English, the only well-formed matrix exclamatives are *wh*-clauses introduced by the degree words *what* or *how* (e.g., *What fun the party was!*; *How enjoyable the party was!*), with the exclamative clause reducible to just a phrase as well (e.g., *What fun!*; *How enjoyable!*). English *wh*-clauses headed by *who*, *which*, *where*, *when*, and *why* cannot stand alone as matrix exclamatives. The counterparts of *what* or *how* in Budai Rukai, however, are not used to indicate exclamations (15).[9] The most intuitive and felicitous way to convey an exclamation is to employ the verbal interrogative *a-* 'do.why; do.how.come' as in (16).[10] To the best of my knowledge, this type of *why* exclamative has rarely been reported cross-linguistically.[11]

---

[9] Some verbal complexes of the form X-*tumane* 'do.what' are used in an exclamative sense, as in (i). They are not construed as asking *what* is to be done, but rather as an exclamation *for what*. Since many oft-tested canonical gradable adjectives cannot appear in this verbal X-*tumane* construction, we will leave it aside in this paper.

(i) *i-tumane-ta* *ka* *ta-ikai-ane* *kai*?
FUT-do.what-1PL.BN NOM NMLZ-exist-NMLZ DEM
'For what are we staying here?' (Chen 1999:61)

[10] Chen (1999:56) remarks in passing on the exclamatory function of *ani* 'why' in addition to its use for inquiring about reasons in Kucapungan Rukai, but provides no detailed discussion. In Mantauran Rukai, Zeitoun (2007:383–384) reports that exclamations are expressed in two ways, one employing the *wh*-phrase *'akosaae-* 'how much', and the other using the prefixation *ka-* 'indeed'.

[11] Much of the literature on *why* centers around its semantic interpretation and its distinction from *how come* or other adjunct *wh*-phrases (see Fitzpatrick 2005; Conroy 2006). Discussion of *why* as an exclamative, in any language, is almost nonexistent. As Villalba (2008:31) acknowledges, "we still lack an answer to the fact that *why* exclamatives are lacking universally." In footnote 38, Zanuttini & Portner (2003) also comment briefly on the inability of *wh*-words like *why* to act as exclamatives. They point out that there seem to be certain occurrences of *why* as exclamatives, for instance, *Why I never heard of such a thing!* This exclamatory use of *why*, however, cannot be embedded under a factive predicate (#*It's amazing why I never heard of such a thing*). In addition, *why* in such constructions may be followed by a pause, distinguishing it from other *wh*-exclamatives. They thus conclude that *why* is not a basic pattern of exclamative.

(15) Question readings only

a. *manemane* *ku* *Ø-kane* *ki* *Muni*?
what NOM NFIN-eat GEN Muni
'What did Muni eat?' (*Intended reading: 'What Muni had eaten!')

b. *sainu* *ku* *lragi-ane* *kai* *kaswi-su*?
how NOM long-NMLZ this pants-2SG.GEN
'How long are your pants?' (*Intended reading: 'How long your pants are!')

(16) a. Question reading
a-ni *ka-lragi*[12] *kai* *kaswi-su*?
do.why-3 STAT.NFIN-long this pants-2SG.GEN
'Why are your pants (so) long?'

b. Exclamative reading
a-ni *ka-lragi* *kai* *kaswi-su*!
do.why-3 STAT.NFIN-long this pants-2SG.GEN
'How long your pants are!' (Lit: 'How come your pants are (so) long!')

In addition to a reason question reading (16a), the same *wh*-phrase *a-* 'do.why', given proper context and emotive cues, can receive an exclamatory interpretation (16b) as well. In the literature, many proposals, though employing different frameworks or taking different perspectives, share a similar intuition about the meaning of exclamatives. For instance, Michaelis & Lambrecht (1996) and Michaelis (2001) present five semantic/pragmatic properties of exclamative constructions: (i) presupposed open proposition, (ii) scalar extent, (iii) affective stance, (iv) identifiability of described referent, and (v) deixis. Also, Zanuttini & Portner (2003) propose that exclamatives feature two central semantic properties, factivity and widening, the latter encompassing (i)–(iii) of the properties suggested by Michaelis & Lambrecht (1996) and Michaelis (2001). Following this line of research, this study will show next that Budai Rukai *a-* 'do.why' constructions are true exclamatives that are characterized by the same semantic/pragmatic properties commonly observed in the literature, here referred to as scalar degree, affective stance, and factivity.

3.1 Scalar degree

For the utterance of an exclamative to be expressively appropriate, its content must be about a degree, and this degree must exceed a relevant standard. That is, exclamatives are scalar and necessitate the use of so-called "gradable" constituents such as adjectives or adverbs.[13] As claimed by many contemporary theories of the semantics of adjectives, in particular in connection with the analysis of comparatives (see Kennedy (1999) and

[12] The morpheme *ka-* in *ka-lragi* marks stativity in non-finite V2 positions.

[13] In prior studies of Rukai, it has been assumed that there are only stative verbs, and that so-called "adjectives" do not exist. The term "adjective" is used in this paper for semantic notational convenience. This paper is not arguing for the existence of an adjective category in Budai Rukai. The present study on exclamations, however, does suggest that adjectival predicates in Budai Rukai form a natural subclass and they all indicate gradability. Not all stative verbs are gradable in nature.

related papers therein), adjectives always denote a gradable property and contain a specification of degree. For instance, saying that *John is tall* amounts to saying "There is a degree to which John is tall that exceeds a particular contextual standard." Thus, for all *a-* 'do.why' sentences to be felicitously interpretable as degree exclamatives, *a-* 'do.why' must be able to occur with any gradable (adjectival) predicate. This is indeed the case. Any so-called adjectival predicate that falls into Dixon's (2004) classification of seven (prototypical) semantic types for adjectives, given in (17), can appear in an *a-* exclamative. The adjective predicate *lragi* 'long' in the example in (16) is one instance of the DIMENSION type. More examples are given in (18), in which all instances receive a gradable degree interpretation.

(17) Prototypical semantic types for adjectives (Dixon 2004)

DIMENSION:	'long', 'short'
AGE:	'young', 'old'
VALUE:	'good', 'bad'
COLOR:	'black', 'white'
PHYSICAL PROPERTY:	'hard', 'soft'
HUMAN PROPENSITY:	'happy', 'kind'
SPEED:	'fast', 'slow'

(18) AGE:

a. *ka* *Salrabu* *lu* *drel-ane,* *a-ni* *kai*
NOM Salrabu if/when look-NMLZ do.why-3 NEG
ka-rudra~rudrange!
STAT.NFIN-RED~old
'How young Salrabu looks!'
(Lit: 'Salrabu when being looked at, how come he doesn't get old.')

VALUE/ HUMAN PROPENSITY:

b. *a-ni* *ka-thariri* *turamuru* *kai* *Salrabu*!
do.why-3 STAT.NFIN-good very this Salrabu
'How nice (handsome) Salrabu is!'

COLOR:

c. *a-ni* *ka-duli* *turamuru* *ka* *laimai-su*!
do.why-3 STAT.NFIN-red very NOM clothes-2SG.GEN
'How red your clothes are!'

PHYSICAL PROPERTY:

d. *a-ni* *ka-telrege* *kai*
do.why-3 STAT.NFIN-heavy this
Ø-udu-su *kai* *becenge*!
NFIN-carry.things.on.head-2SG.GEN this millet
'How heavy the millet that you carry is!'

SPEED:

e. *a-ni* *ka-ridare* *lu* *Ø-lailai* *kai* *vavalake*!
do.why-3 STAT.NFIN-fast if/when NFIN-run this child
'How fast this child runs!'

The *a-* exclamatives can also allow a finite clause or a nominalized clause to appear in the argument subject position (19)–(20). In these cases, *a-* exclamatives range over EVALUATIONS , which involve gradable degrees, as well. In discussing the behavior of how in sentences like *How Buck rode his horse!*, Rett (2008) points out that evaluations such as beautifully, dangerously, or clumsily are instances of available degree interpretations for how. Thus, the exclamative is expressively appropriate in a scenario in which the speaker expected Buck to ride poorly but he instead rode beautifully. As with how in this example, the *a-* exclamatives in (19)–(20) receive the evaluation reading. Even though the exclamatives in (19)–(20) do not contain any overt degree morphology, it can be assumed that they involve a null gradable adverb ADV, which receives its value from context (see Gutiérrez-Rexach 1996; Villalba 2003; Castroviejo Miró 2006; Rett 2008). So in these contexts, the interpretation could include the sense of 'surprisingly' or 'unexpectedly'.

(19) *a-ni* [SUBJECT *ka* *ta-ka-amani-ane* *kai* *lasu*
do.why-3 NOM NFUT-STAT.NFIN-be-NMLZ this man
ka *wa-swa~swape*]!
NOM ACT.FIN-RED~sweep
'How did it come about that the person is always the one who has been sweeping!'
(Lit: 'How did it come by the person being always the one who has been sweeping!')

(20) *a-ni* [SUBJECT *ka* *ma-dalame* *ki* *Muni* *ka* *Salrabu*]!
do.why-3 NOM STAT.FIN-like OBL Muni GEN Salrabu
'(Contrary to my expectation,) How did it come about that Salrabu likes Muni!'

The *a-* exclamatives can also have amount readings, as illustrated in (21). The exclamatory interpretation is due to the assertion of an excessive degree on the scale of quantity.

(21) *a-ni* *pu-madru* *ka-kalra* *kai* *kamadha* *kai* *asane*!
do.why-3 grow-fruit STAT.NFIN-many this mango this now
'How many fruits this mango tree has grown this year!'

3.2 Speaker's affective stance

In addition to the property of degree, there is one central ingredient necessary in the expression of exclamative meaning, in particular in contrast to declaratives: The relevant degree property must exceed the speaker's expectations. In other words, for example, the clothes in (18c) do not need to be objectively red according to a commonly held standard. The degree of their redness only needs to exceed the speaker's expectations, whatever those may be. Such expectations might follow common-ground norms, socially accepted standards, or simply expectations that reflect the speaker's personal evaluation.[14] The purpose of this exclamation for the speaker is to inform the hearer that the degree in

[14] Several authors (Gutiérrez-Rexach 1996; Villalba 2003; Zanuttini & Portner 2003; Castroviejo Miró 2006) have treated such constructions as involving a HIGH or EXTREME degree implicature. Zanuttini & Portner, for example, derive this implicature from the semantic operation of *widening* applied to the domain of the *wh*-expression.

question is extreme in the speaker's conceptualization. The speaker's affective response to a given situation may include surprise (positive or negative), admiration, amazement, or disapproval. This is in fact the case for all the instances of exclamatives given in (18)–(21). They all convey an emotional reaction of some sort. Thus, (18a) *ka Salrabu lu drel-ane, a-ni kai ka-rudra-rudrange!* 'How young Salrabu looks!' can be seen as expressing admiration, and (19) *a-ni ka ta-ka-amani-ane kai lasu ka wa-swa-swape!* 'How did it come about that the person is always the one who has been sweeping!' as conveying amazement.

Notably, *a-* exclamatives co-occur pervasively with two deictic degree adverbs, *mia*/*amia* 'as.this' and *ngithanga* 'as.so', as illustrated in (22)–(24).[15] In addition to expressing that the degree is surprisingly extreme, the deictic degree adverbs *mia* and *ngithanga* in (22)–(24) provide an appropriate reference point in the conversational context mutually known by the speaker and hearer. Moreover, the degree adverb *ngithanga* (23)–(24) often conveys the speaker's strong epistemic reaction like 'why should' and 'there's no need' to the propositional content, given a proper scenario.

(22) *a-ni* ***mia*** *ka-duli* *turamuru* *ka* *laimai-su*!
do.why-3 as.this STAT.NFIN-red very NOM clothes-2SG.GEN
'How come your clothes are so red as this!'

(23) *a-ni* ***ngithanga*** *ka-lisi* *ka* *singsi*!
do.why.3 as.so STAT.NFIN-angry NOM teacher
'How come the teacher is so angry!'

(24) *a-nu* ***ngithanga*** *pa-kidulru* *ki* *lasu*!
do.why.2PL as.so CAUS-work.laboriously OBL man
'Why should you make him work so laboriously!'

3.3 Factivity

The last essential condition of the exclamative speech act is that the speaker emphasizes his/her strong emotional reaction to what s/he takes to be a fact, as first noted in Elliott (1974) and Grimshaw (1979), and later discussed in detail in the papers of Raffaella Zanuttini and Paul Portner (the best known is their 2003 paper). Thus, in contrast to interrogatives all exclamatives are factive, a result of the presuppositional status of exclamations. The contrast in Budai, demonstrated in (25), make this point. Grimshaw argues that exclamatives are restricted to being embedded under emotive factive verbs like *amaze (amazing)*, which requires a presupposed complement as seen in (25a).[16] Therefore, embedding an exclamative under a non-factive predicate like *ask*

---

[15] Michaelis (2001) treats the *so*-class as "anaphoric" degree adverbs with the indexical function of referring to a certain point on a presupposed scalar proposition. Zwicky (1995) distinguishes *so*-class degree adverbs from *very*-class degree adverbs in that the former can have cataphoric reference in the inverted resultative construction (i), while the latter cannot (ii). The distinctive behavior of *ngithanga* 'as.so' and *turamuru* 'very' seems to fit Zwicky's distinction, but this remains to be further investigated.

(i) I almost FAINTED, the sun was so hot.
(ii) * I almost FAINTED, the sun was very hot.

[16] This is not to say that all factives allow exclamative complements.

would be expected to result in ill-formedness, as it does in (25b), in which the co-occurrence of an embedded exclamation and an ask-for-information verb leads to a contradiction.

(25) a. *masamali-aku* *a-ni* *ka-telrege* *kai*
surprise-1SG.NOM do.why-3 STAT.NFIN-heavy this
Ø-udu-su *kai* *becenge*!
NFIN-carry.things.on.head-2SG.GEN this millet
'I am amazed how heavy the millet that you carry is!'

b. * *ki-a-vaga-aku* *ki* *Salrabu* *a-ni*
KI-FIN-ask-1SG.NOM OBL Salrabu do.why-3
ka-lragi *kai* *kaswi* *ki* *Muni*!
STAT.NFIN-long this pants GEN Muni
* 'I asked Salrabu how long Muni's pants are!'

The factivity principle also explains the infelicitousness of the embedded exclamatives in (26). Negating a factive verb like *know* is unacceptable because denying the speaker's knowledge conflicts with the factive presupposition generated by the exclamative.

(26) * *kai-naku* *wa-thingalre* *a-ni* *ngithanga* *ka-telrege*
NEG-1SG.NOM ACT.FIN-know do.why-3 as.so STAT.NFIN-heavy
kai *Ø-udu-su* *kai* *becenge*!
this NFIN-carry.things.on.head-2SG.GEN this millet
* 'I don't know how so heavy the millet that you carry is!'

An additional and connected observation by Grimshaw, Zanuttini, Portner, and many other scholars is the fact that exclamatives cannot play the prototypical role of declaratives as answers to questions like the one in (27a). This follows mainly because the propositional content of an answer is expected to be asserted, as in (27b), not presupposed, as in (27b'). Since exclamatives are inherently factive in the strong sense that they presuppose themselves, they cannot serve as answers.

(27) a. Q: *pingi-a-adrau* *ungulu* *kai* *Salrabu* *kuiya*?
PINGI-FIN-much drink this Salrabu yesterday
'Did Salrabu DRINK a lot yesterday?'

A: *unu*, *pingi-a-adrau* *ungulu*.
yes, PINGI-FIN-much drink
'Yes, (he) DRANK a lot.'

b. A: * *a-ni* *ngithanga* *pingi-adrau* *ungulu* *kai* *Salrabu*!
do.why-3 as.so PINGI-much drink this Salrabu
'How has Salrabu DRUNK so much!'

Summing up, *why a-* exclamatives in Budai Rukai show the same semantic properties as other *wh*-exclamatives with respect to three widely acknowledged constituting criteria. That is, the speaker must find the underlying propositional content surprising in some way, the content must be about a degree that exceeds a relevant standard, and what the speaker exclaims upon is necessarily factive.

4 Nominal *why* exclamatives in Budai Rukai

We now turn to the other type of *why* exclamatives, those involving *asi*, also roughly translated as *why/how come*. The form *asi* is another commonly seen *wh*-phrase that is seldom addressed in the literature on Rukai. Given proper context, *asi* signals an exclamative reading as in (28). (28c) and (28d) are excerpts from two naturalistic narrative stories.

(28) a. ***asi*** (***amia***)[17] [$_{\text{nominalized}}$ *kai* *kulrungu-ane-li*]!
ASI (as.this) this stupid-NMLZ-1SG.GEN
'How come I am (so) stupid!'
(Lit: 'How come (as so) the my stupidness!')

b. ***asi*** (***amia***) [$_{\text{nominalized}}$ *kai* *riladh-ane* *ki* *lasu*]!
ASI (as.this) this thin-NMLZ GEN man
'How come the man is (so) thin!'
(Lit: 'How come (as so) the man's thinness!')

c. 35.. ***asi*** ***amia*** *kaiki* *vavalake* *la* *iya* *si*
ASI as.this this child then say DM
'(They) thought, "How come this child is as such!"'
(NTU Corpus: Pear1)

d. 65... *kuini* *Balenge* *la* *iya* *kai*.\
this PN then say this
66.. ***asi*** ***amia*** *kai* *eh*,_
ASI as.this this FILL
67... *tharir-ane*
pretty-NMLZ
68.. *kaiki* *lasu kikai eh*,_
this man this FILL
69... *la* *kirimu*,_
then fast
70.. *kisamula* *kaungu* *la* *iya* *si*
seriously work then say DM
'Balenge then thought, "How come there is a man that is so handsome and hard-working!"' (NTU Corpus: Princess)

As with *a-* exclamatives, the deictic degree word *amia* 'as.this' often co-occurs with *asi*, and it denotes the speaker's strong epistemic ground. Recall that the verbal *why a-* exclamative construction has the following schematized structure:

---

[17] The use of *ngithanga* after *asi* imposes rather strict restrictions on eventive-type constructions, as in (i). Many of the tests that are acceptable with *amia* are not acceptable with *ngithanga*. Therefore, this section employs only *amia* examples.

(i) *asi* *ngithanga* *tu-kalra*!
ASI as.so do-many
'Why should (you) do so much!'

(29) [$_{V1}$ *a*]-bound.pronoun [$_{V2}$ nonfinite V] [$_{SUBJECT}$ NP]

Unlike the verbal *a-*, *asi* does not attract any bound pronouns and cannot be further decomposed morphologically. The grammatical status of *asi* is not straightforwardly evident. One notable feature distinguishing the *asi* exclamative from the *a-* exclamative is its syntactic nominal status, indicated by its frequent co-occurrence with (i) the nominalizer *-ane*, (ii) a genitive subject (e.g., *-li/ki lasu* '1SG.GEN/GEN man'), and (iii) the definite proximal demonstrative *kai* 'this'. These distinctive features can be represented structurally as in (30).

(30) *ASI* [$_{SUBJECT}$ ***kai*** [ADJ]-*ane* NP$_{GEN}$]

As I have argued elsewhere (Sung 2011), the nominalizer *-ane* occurs pervasively in clausal and lexical nominalizations, and these structures are packaged with certain subjective speaker attitudes toward the resultant status. Take (31) as an example; it presupposes that Lrailrai was hit by Salrabu, and emphasizes the speaker's strong reaction to the serious consequence of the hitting.

(31) *masamali-aku* [*ku*/*ka*/*kai*/**alaka* *ta-lrumadh-ane*
surprise-1SG.NOM OBL/OBL/this/*COMP NFUT-hit-NMLZ
ki/**ka*/**ku* *Salrabu* *ki* *Lrailrai*/*iniane*][18]
GEN Salrabu OBL Lrailrai/3SG.OBL
'I am surprised at (the bad consequence resulting from) Salrabu's hitting Lrailrai/her.' (Sung 2011 [23b])

Furthermore, the use of a demonstrative such as *kai* 'this' in (30) is obligatory. In analyzing NP exclamatives, a primary concern among different proposals pursued in the literature (e.g., ellipsis analysis, Michaelis & Lambrecht (1996); relative clause analysis, Portner & Zanuttini (2005); determiner analysis, Potsdam (2011)) is where the exclamatory sentential force comes from. Drawing on Malagasy data, Potsdam (2011) argues that Malagasy has only nominal exclamatives, and he suggests that the exclamative interpretation is associated with the determiner.[19] This is also the case for *asi* exclamatives in Budai Rukai. Among the various demonstratives of Budai, including *kai* 'this [+visible, –distance]', *kui(ni)* 'that [+visible, –distance]', *kavai* 'that [+visible, +distance]', *ku(i)dra* 'that [-visible, +distance]',[20] the definite proximal *kai* 'this' is predominately preferred in *asi* exclamatives; it can sometimes be optionally replaced by the proximal nominal case marker *ka*. Similar determiner/demonstrative restrictions in nominal exclamatives are observed crosslinguistically (32)–(35). Distal *those* and

[18] The use of a factive complementizer like *alaka* 'that' is unacceptable here. When *alaka* is employed, the embedded verb appears only in a finite form, rather than a nominalized one.

[19] Potsdam (2011) proposes that the illocutionary force of an exclamation is due to an exclamative morpheme E residing in the determiner, rather than in the relative clause. Zanuttini & Portner (2003:n.42) also suggest an alternative analysis for the nominal exclamatives. The definite article *the* in an utterance like *The people she would invite!* plays the crucial role of marking the phrase as factive.

[20] The use of distal *kavai* and *ku(i)dra* is less acceptable and context-dependent. When these two demonstratives are used, they indicate a relatively remote time of the given situation. Such semantic extension to a temporal-referring function is also attested in Abui demonstratives (Kratochvíl 2011).

indefinite *some* in English (32), for instance, are less acceptable in nominal exclamative constructions.

(32) English
a. The/??Those strange people who come from Italy!
(Portner & Zanuttini 2005:5–6, 9)

b. * Oh, some places you'll go! (Rett 2008:612)

(33) Malagasy
Izany *Rasoa!*
DEM Rasoa
'How [pragmatically appropriate adjective] Rasoa is!' (Potsdam 2011:677)

(34) Spanish
la *de cosas* *que* *come* *Juan*!
DET thing.FEM.PL that ate Juan
'How much Juan ate!' (Potsdam 2011:680)

(35) Russian
(*ox už*) *èti* *deti*!
oh yet DEM.PROX.NOM.PL children.NOM.PL
'The children!' (Potsdam 2011:680)

The *asi* exclamative with its distinctive nominality is also characterized by the same semantic/pragmatic properties as the *a-* exclamative, discussed in section 3. The first well-tested fact is that exclamatives necessarily involve gradable degree expressions, including quantity, adjectives, adverbs, evaluations, and so on. Some examples are given in (36).

(36) QUANTITY:
a. *asi* *kai kalra-ane* *kai* *kamadha*!
ASI this many-NMLZ this mango
'How come there are so many mangoes!'
(Lit: 'How come the mangoes' maniness!')

DIMENSION:
b. *asi* *kai* *sua-adrav-ane* *kai* *daane*!
ASI this SUA-big-NMLZ this house
'How big this house is!' (Lit: 'How come the house's bigness!')

HUMAN PROPENSITY:
c. *asi* *kai* *lisi-ane* *ki* *Salrabu*!
ASI this angry-NMLZ GEN Salrabu
'How bad-tempered Salrabu is!' (Lit: 'How come the Salrabu's angriness!')

SPEED:

d. *asi* *kai ridar-ane* *lu* *Ø-lailai* *kai* *vavalake*!
ASI this fast-NMLZ if NFIN-run this child
'How fast this child runs! (Lit: 'How come the fastness when the child runs!')

AGE:

e. *ka* *Salrabu* *lu* *drel-ane,* *asi* *kai* *ngi-vavalak-ane*!
NOM Salrabu if/when look-NMLZ ASI this NGI-child-NMLZ
'How young Salrabu looks!' or 'How childish Salrabu is!'
(Lit: 'Salrabu when being looked at, how come the youngness (or childishness)!')

Given their nominal syntactic form, the exclamatives headed by *asi* in (36) all express something about the degree to which an individual instantiates some gradable property. Also important for the degree restriction is the second ingredient of exclamations, the unexpectedness, an affective feature widely acknowledged by many works on exclamatives. The degree attained by the property has to be above both the relevant standard and the speaker's expectations so that it provokes his/her exclamation. To illustrate, if one exclaims (36b) *asi kai sua-adrav-ane kai daane!* 'How big this house is!', the degree of the house's actual largeness, as observed by the speaker, is higher than the degree the speaker expected. All examples in (36) are used to exclaim at a scenario involving a meaning of surprise, unexpectedness, or annoyance of the speaker.

Finally, *asi* exclamatives are also factive, that is, they presuppose the truth of their complements, and this leads to the contrast shown in (37).

(37) a. *wa-thingalr-aku* *asi* *amia* *kai* *telreg-ane*
ACT.FIN-know-1SG.NOM ASI as.this this heavy-NMLZ
kai *Ø-udu-su* *kai* *becenge*
this NFIN-carry.things.on.head-2SG.GEN this millet
'I know how heavy the millet that you carry is!'

b. * *kai-naku* *wa-thingalre* *asi* *amia* *kai*
NEG-1SG.NOM ACT.FIN-know ASI as.this this
telreg-ane *kai* *Ø-udu-su* *kai* *becenge*
heavy-NMLZ this NFIN-carry.things.on.head-2SG.GEN this millet
'I don't know how heavy the millet that you carry is!'

(37a) is felicitous, with an exclamative embedded under the factive predicate *know*, whereas (37b) is ungrammatical because negating *know* conflicts with the exclamative's nature of presupposing factivity. Factivity is also the reason why interrogative verbs like *ask* cannot embed exclamatives, as illustrated in (38); such embedded utterances could only be interpreted as questions.

(38) * *ki-a-vaga-aku* *ki* *Salrabu* *asi* *amia* *kai*
KI-FIN-ask-1SG.NOM OBL Salrabu ASI as.this this
lragi-ane *ki* *kaswi ki* *Muni*!
long-NMLZ GEN pants GEN Muni
* 'I asked Salrabu how long Muni's pants are!'

Its inability to occur in question-answer pairs again shows *asi* to be an exclamative. Like *a-* exclamatives, *asi* can never serve as an answer to a question:

(39) a. Q: *pingi-a-adrau* *ungulu* *kai* *Salrabu* *kuiya*?
PINGI-FIN-much drink this Salrabu yesterday
'Did Salrabu drink a lot yesterday?'

b. A: * *asi* *amia* *kai* *pingi-adrav-ane* *lu*
ASI as.this this PINGI-much-NMLZ if/when
ungulu *kai* *Salrabu*!
drink this Salrabu
'How Salrabu drank so much!'

5 Concluding remarks

It has long been the goal of syntactic/semantic studies of exclamatives to account for which constructions can form an exclamation, and why. As well-attested crosslinguistically, there is considerable variation in the morphosyntax of exclamatives, which includes subject-verb inversion, subordinate clauses, noun phrases, anaphoric adverbs/adjectives, and *wh*-phrases. Among these, the most prevalent source for constructing exclamations is the set of *wh*-phrases. Yet not all *wh*-phrases correspond to exclamatives. In English, for instance, matrix *wh*-exclamatives can only be formed with *how* and *what*, which are the only *wh*-words that can receive degree readings. In other languages such as Paduan and Italian (Michaelis 2001; Zanuttini & Portner 2003), *who* or *where* also evoke a scalar degree and thus can be used as exclamatives. As for the use of *why* to indicate an exclamation, it seems to be rare crosslinguistically, based on the prior literature (Villalba 2008:31). Crosslinguistic differences seem to hinge on whether or not a particular *wh*-phrase of a language can range over degree, as suggested in Rett (2008). In Budai Rukai, as this paper has shown, *why* is the most intuitive and felicitous way to convey an exclamation. Two types of *why* exclamatives have been distinguished in Budai Rukai, and each manifests a distinct nature. The *a-* type is verbal whereas the *asi* type is nominal. Despite their syntactic difference, both types of *why* in Budai Rukai display the key semantic properties that crucially characterize an exclamative: scalarity, affective stance, and factivity.

References

Castroviejo Miró, Elena. 2006. Wh-exclamatives in Catalan. PhD dissertation. Barcelona: Universitat de Barcelona.

Chen, Cheng-Fu. 1999. Wh-words as interrogatives and indefinites in Rukai. MA thesis, Taipei: National Taiwan University.

Chen, Cheng-Fu. 2008. Aspect and tense in Rukai: Interpretation and interaction. PhD dissertation. Austin: University of Texas at Austin.

Chen, Cheng-Fu and Li-May Sung. 2005. Interrogatives as polarity items in Kucapungan Rukai. *Concentric: Studies in Linguistics* 31.1:95–117

Conroy, Anastasia. 2006. The semantics of *how come*: A look at how factivity does it all. *University of Maryland Working Papers in Linguistics 14*, ed. by N. Kazanina, U. Minai, P. Monahan and H. Taylor, 1–24. College Park, MD: UMWPiL.

D'Avis, Franz-Josef. 2002. On the interpretation of *wh*-clauses in exclamative environments. *Theoretical Linguistics* 28.1:5–31.

Dixon, R.M.W. 2004. Adjective classes in typological perspective. In *Adjective classes: A cross-linguistic typology*, ed. by R.M.W Dixon and Alexandra Y. Aikhenvald, 1–49. Exploration in Linguistic Typology 1. Oxford: Oxford University Press.

Dyen, Isidore. 1963. *The lexicostatistical classification of the Austronesian languages.* London: Yale University Press.

Elliott, Dale Eugene. 1971. The grammar of emotive and exclamatory sentences in English. PhD dissertation. Columbus: Ohio State University.

Elliott, Dale Eugene. 1974. Toward a grammar of exclamations. *Foundations of Language* 11:231–246.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification.* Institute of Ethnology, Academia Sinica Monograph No. 17. Taipei: Academia Sinica.

Fitzpatrick, Justin. 2005. The whys and how comes of presupposition and NPI licensing in questions. *Proceedings of the 24th West Coast Conference on Formal Linguistics 24*, ed. by John Alderete et al., 138–145, Somerville, MA: Cascadilla Proceedings Project.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10.2:279–326.

Gutiérrez-Rexach, Javier. 1996. The semantics of exclamatives. *UCLA Working Papers in Linguistics: Syntax at Sunset*, ed. by Edward Garrett and Felicia Lee, 146–162. Los Angeles: UCLA.

Ho, Dah-an. 1983. lun lukaiyu de qinshu guanxi [The linguistic position of Rukai among Formosan languages]. *Bulletin of the Institute of History and Philology, Academia Sinica* 54.1:121–68.

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1999. Interrogative constructions in some Formosan Languages. In *Chinese languages and linguistics 5: Interactions in language*, ed. by Yuen-mei Yin, I-li Yang and

Hui-chen Chan, 639–680. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 2. Taipei: Academia Sinica.

Kaufman, Daniel. 2011. Exclamatives and temporal nominalizations in Austronesian. *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. by Foong Ha Yap, Karen Grunow-Hårsta and Janick Wrona, 721–754. Typological Studies in Language 96. Amsterdam and Philadelphia: John Benjamins.

Kennedy, Christopher. 1999. *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland Publishing, Inc.

Kratochvíl, František. 2011. Discourse-structuring functions of Abui demonstratives. *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. by Foong Ha Yap, Karen Grunow-Hårsta and Janick Wrona, 757–788. Amsterdam and Philadelphia: John Benjamins.

Kuo, John Ching-hua. 1979. Rukai complementation. MA thesis. Taipei: Fu Jen Catholic University.

Li, Paul Jen-kuei. 1973. *Rukai structure*. Institute of History and Philology, Academia Sinica, Special Publications No. 64. Taipei: Academia Sinica.

Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* 48:1–92.

Li, Paul Jen-kuei. 1990. Classification of Formosan languages: Lexical evidence. *Bulletin of the Institute of History and Philology, Academia Sinica* 61.4:805–835.

Li, Paul Jen-kuei. 1996. The pronominal systems in Rukai. *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer, 209–230. Hamburg: Abera Verlag.

Li, Paul Jen-kuei. 1997. A syntactic typology of Formosan languages: Case markers on nouns and pronouns. *Chinese Languages and Linguistics 4: Typological Studies of Languages in China*, ed. by Chiu-yu Tseng, 343–378. Symposium Series of the Institute of History and Philology, Academia Sinica No. 4. Taipei: Academia Sinica.

Michaelis, Laura A. 2001. Exclamative constructions. *Language typology and language universals: An international handbook*, ed. by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible, 1038–1050. Berlin: Walter de Gruyter.

Michaelis, Laura and Knud Lambrecht. 1996. The exclamative sentence type in English. *Conceptual structure, discourse and language*, ed. by Adele E. Goldberg, 375–389. Stanford: CSLI Publications.

Ono, Hajime. 2006. An investigation of exclamatives in English and Japanese: Syntax and sentence processing. PhD dissertation. College Park: University of Maryland.

Portner, Paul and Raffaella Zanuttini. 2005. The semantics of nominal exclamatives. In *Ellipsis and non-sentential speech*, ed. by Reinaldo Elugardo and Robert J. Stainton, 57–67. Studies in Linguistics and Philosophy Vol. 81. Dordrecht: Kluwer.

Potsdam, Eric. 2011. Expressing exclamatives in Malagasy. In *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. by Foong Ha Yap, Karen Grunow-Hårsta, and Janick Wrona, 659–683. Typological Studies in Language Vol. 96. Amsterdam and Philadelphia: John Benjamins.

Rett, Jessica. 2008. A degree account of exclamatives. In *Proceedings of the 18th Semantics and Linguistic Theory Conference (SALT XVIII)*, ed. by Tova Friedman and Satoshi Ito, 601–618. Ithaca, NY: CLC.

Rett, Jessica. 2011. Exclamatives, degrees and speech acts. *Linguistics and Philosophy* 34.5:411–442.

Shih, Winslow Chia-Lin. 2012. Are there case markers in Budai Rukai? MA thesis, Taipei: National Taiwan University.

Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, Symposium Series of the Institute of History and Philology, Academia Sinica No. 3, 683–726. Taipei: Academia Sinica. Reprinted in Elizabeth Zeitoun, ed., *Formosan Linguistics: Stanley Starosta's contributions*, 737–769. Language and Linguistics Monograph Series C6-2. Taipei: Academia Sinica, 2009.

Sung, Li-May. 2011. Clausal nominalization in Budai Rukai. *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. by Foong Ha Yap, Karen Grunow-Hårsta, and Janick Wrona, 523–559. Typological Studies in Language Vol. 96. Amsterdam and Philadelphia: John Benjamins.

Sung, Li-May, Lily I-wen Su, Fuhui Hsieh and Zhemin Lin. 2008. Developing an online corpus of Formosan languages. *Taiwan Journal of Linguistics* 6.2:79–117.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No. 5. Tokyo: University of Tokyo.

Villalba, Xavier. 2003. An exceptional exclamative sentence type in Romance. *Lingua* 113.8:713–745.

Villalba, Xavier. 2008. Exclamatives: A thematic guide with many questions and few answers. *Catalan Journal of Linguistics: Exclamatives at the Interface* 7:9–40.

Zanuttini, Raffaella and Paul Portner. 2003. Exclamative clauses: At the syntax-semantics interface. *Language* 79.1:39–81.

Zeitoun, Elizabeth. 1995. Problèmes de linguistique dans les langues aborigènes de Taiwan [Issues in Formosan linguistics]. PhD dissertation. Paris, France: Université René Diderot, Paris 7.

Zeitoun, Elizabeth (Qi Li-sha). 2000. *lukaiyu cankao yufa* [*A reference grammar of Rukai*] Formosan Language Series 8. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Keynote address at the *Tenth Annual Meeting of the Austronesian Formal Linguistics Association*, 28–30 March. Honolulu: University of Hawai'i at Mānoa.

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran* (*Rukai*). Language and Linguistics Monograph Series A4-2. Taipei: Academia Sinica.

Zwicky, Arnold M. 1995. Exceptional degree markers: A puzzle in internal and external syntax. *Ohio State University Working Papers in Linguistics* 47:111–123.

13 *The sa construction in Paiwan: A mismatch of syntax and semantics**

FUHUI HSIEH

1 Introduction

In this study, I present an intriguing type of clause combining in Paiwan, an Austronesian language spoken in southern Taiwan. Previous scholars (Tang 1999, 2011; A. Chang 2006) have reported various strategies of clause combining in Paiwan: coordination and subordination, including complementation, serial verb constructions, and adverbial clauses. What is intriguing is the syntactic coding of temporally sequential events in this language. For example:

(1)	*tjara*	*djukul-en=anan*	*a*	*s‹em›u-alap*	*a*	*qipu*	
	surely	beat-PF[1]=CON	LNK	<AF>remove-take	NOM	soil	
	i-zua	*i*	*ta*	*tjapungul-an,*	***sa-na***		
	LOC-that	LOC	OBL	grass.root-NMLZ	and-then		

* This paper is the result of two NSC projects carried out in the past three years. During this period, I have been much honoured to work with the team led by Prof. L. Huang, who has had a profound impact on the field of Formosan languages. I am happy to honour her for the guidance, support, and friendship that she has bestowed upon all the team members. Part of this paper was presented at ICAL-12. I thank the audience at the conference for their precious comments. I am indebted to four anonymous reviewers for their valuable suggestions and comments. Thanks also go to the editors, who read various versions of this paper and gave very helpful comments and suggestions. Research grants from Taiwan National Science Council NSC100-2420-H-036-001-MY3 and NSC 101-2410-H-036-007 are gratefully acknowledged. The usual disclaimer applies.

[1] This paper follows the Leipzig Glossing Rules in glossing examples, with the following additions or exceptions: AF, agent focus; BF, benefactive focus; CAUS, causative; CON, continual aspectual marker; CVB, converb construction; EXIST, existential marker; H, honorific marker; HORT, hortative; IF, instrumental focus; INVOL, involuntary; LF, locative focus; LNK, linker; NAF, non-agent focus; NCM, non-common noun marker; NI, non-indicative form; PF, patient focus; PN, proper noun/personal name; QAL, quality; RED, reduplication; RF, referential focus; STAT, stative.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 313–339
Asia-Pacific Linguistics, 2015

qu<u>ɬ</u>i~qu<u>ɬ</u>its-i,		***sa***	***ɬama-i***	*a*	*q‹in›epu-an,*
RED~clear.cut.bush-PF.NI		and	burn-PF.NI	NOM	<PFV>associate-NMLZ
sa-na	***t‹al›e-talem-i***		*tua*	*vasa.*	
and-then	<QAL>do-plant-PF.NI		OBL	taro	

'We always beat the soil out of the roots, and then clear up and burn the refuse; and then plant the taro.' (*OHPT* 078-004)[2]

The Paiwan construction in (1) translates as "We always beat the soil out of the roots, and then clear up and burn the refuse; and then plant the taro", which looks like an ordinary English coordinate construction. But it is not an ordinary coordinate construction in Paiwan, because only the initial clause contains full finite verb(s), while the verbs in non-initial position connected by the marker *sa* (*sana*) carry the non-indicative morpheme *-i*, which is attested in subjunctive (2a), imperative (2b) and negative (2c) clauses (see Appendix 1 for the verb forms in Paiwan).

(2) a. Subjunctive (*OHPT* 017-019)

ɬakua	*nu*	*tja-sane-qavay-**i***			*tua*	*vaqu,*
but	when	1PL.INCL.GEN-make-dumpling-PF.NI			OBL	millet
sa	*tja-pa-taud-**i***		*tua*	*quvaɬ*	*a*	*qavay,...*
and	1PL.INCL.GEN-CAUS-mix-PF.NI		OBL	hair	NOM	dumpling

'But if we make dumplings of millet and mix hair in the dumplings,...'

b. Imperative (*OHPT* 012-009)

qau	*vaik-en*	*sa*	*tsagtsag-i*	*a*	*si-qeɬev-an*	*sa*
so	leave-PF	and	bump-PF.NI	NOM	IF-block-NMLZ	and
"*su-qeɬev-**i***!"		*aya-in*		*azua*	*ka~kedi-an.*	
remove-block-PF.NI		say-PF		hat	RED~small-NMLZ	

'He went there and knocked on the door, and said to the children: "Open the door!"'

c. Negation (*OHPT* 016-007)

qau	***ini***	*pa-tutu-**i.***
so	NEG	CAUS-breast-PF.NI

'and she wouldn't suckle him.'

Nevertheless, the occurrence of the suffix *-i* in (1) does not color the sentences with a subjunctive, imperative or negative reading. Moreover, the linear order of the linguistic forms (i.e., clauses) is always mirrored with the temporal order of events; Event 1 coded in Clause 1 (S1) happens prior to Event 2 in Clause 2 (S2), which in turn occurs earlier

[2] *One Hundred Paiwan Texts* (*OHPT*) is a collection of Paiwan narratives and conversations edited by Robert Early and John Whitehorn (2003). The first number indicates the text number and the second number is the sentence number; therefore, 76-017 means Text 76 Sentence 17 in *OHPT*. In citing the texts from *OHPT*, I keep the original orthography used and the translations provided by Early & Whitehorn (2003) in order not to confuse readers; yet, I modify glossing abbreviations and symbols to make the data analyses consistent in this paper. I also simplify the three-line interlinear blocks in *OHPT* to two-line interlinear glossing in this study. Some special orthography used in *OHPT* (Early & Whitehorn 2003:14) includes the following: (i) *q* is manifested in some dialects as a glottal stop [ʔ], and in others as a backed-k, (ii) *ng* represents the velar nasal [ŋ], (iii) *l* is a lateral affricate, (iv) *r* is a trilled approximate, (v) *dr* is a retroflexed alveolar plosive, (vi) *ts* is an alveolar affricate.

than Event 3 in Clause 3 (S3), and so on. In other words, events coded in this construction are of equal prominence and/or importance in perception and/or conception. However, syntactic coding of this construction appears not to follow the principle of iconicity in language (Haiman 1985; Dirven & Radden 2004);[3] *sa* clauses always occur in non-initial position and are syntactically realised as dependent clauses as the verbs in those *sa* clauses are all in non-indicative forms. To complicate matters further, the dependent *sa* clauses do not serve canonical functions of subordinate clauses, viz. Ground, as they do not provide any backgrounding information for the figure event, described by the main clause (Radden & Dirven 2007:29; Talmy 2000:327).[4] We can therefore see an intriguing mismatch between the syntax and semantics in this construction, as shown in Table 1.

Table 1: Mismatch between syntax and semantics in the *sa* construction

Form	Function
S1, S2, S3, S4… indicative V non-indicative Vs The clauses are not of equal weight in syntax. unparallel	E1, E2, E3, E4…. Events coded are of perceptually/ conceptually equal prominence or importance. parallel

A. Chang (2006:308–311) is probably the first one who noticed this particular construction in Paiwan and called it "cosubordination". She also claims that there are two types of cosubordinate constructions in Paiwan: the neutral *sa* cosubordinate construction and the sequential *sana* cosubordinate construction, as illustrated in (3a) and (3b) respectively.[5]

(3) A. Chang (2006)

a. *'‹em›au~'aung=aken* ***sa*** *ku=kesa~kesa*.
<AV>RED~cry=1SG.NOM and 1SG.GEN=RED~cook
'I am/was crying and cooking.'
'I am/was crying while cooking.' (p. 309, ex. 12)

b. *kan-u=anan* ***sana*** *su=pavanaw*.
eat-IMP=CON and.then 2SG.GEN=take.bath
'Eat before taking a bath!' (Lit.: Eat and then take a bath!') (p. 311, ex. 16)

A. Chang is correct in pointing out that Paiwan is the mirror image of what Foley & Van Valin (1984) and Foley (2010) described for cosubordination in Papuan languages, where the independent clause follows one or more dependent clauses. In Paiwan

[3] "The principle of iconicity in language means that we conceive a similarity between a form of language and the thing it stands for…Iconicity may manifest itself in three sub-principles, i.e. those of linguistic expressions related to sequential order, distance and quantity." (Dirven & Radden 2004:8)

[4] Sequence principle: "The unmarked (or only possible) linguistic expression for any particular relation between two events in temporal sequence treats the earlier event as a reference point, or Ground, and the later event as requiring referencing—that is, as the Figure. Where the complete syntactic form is a full complex sentence, the two events are in the subordinate and the main clause, respectively." (Talmy 2000:327)

[5] Though A. Chang did use the term "cosubordination", unfortunately, she did not provide any syntactic test to verify this construction (2006:308).

cosubordination, the independent clause precedes the dependent clauses. Unfortunately, she did not explain or explore the semantico-pragmatic functions of this construction and how these particular constructions work in Paiwan, nor did she go further and discuss systematically the morphosyntactic behavior of the claimed two types of cosubordinate constructions.

The purpose of this study is threefold. First, I investigate the morphosyntactic behavior and semantico-pragmatic functions of this particular construction (the *sa* construction henceforth), as in (1). This *sa* construction is of particular interest because the syntactic coding does not conform to the perceptual principle of figure and ground, nor does it observe the principle of iconicity (Haiman 1985; Talmy 2000; Dirven & Radden 2004; Radden & Dirven 2007); in other words, events of conceptually equal prominence or importance are coded unequally in syntax, except for the linear order. The semantico-pragmatic function of this construction is to encode temporally sequential events. Nonetheless, temporally sequential events can also be coded by *saka* 'and' in coordinate construction or be juxtaposed, as in (4).

(4) a. [*ka* *si-tsuay-an* *ka* *ma-iraw* *a* *ka-djunang-an*]
after IF-long.time-NMLZ after AF-melt NOM main-earth-NMLZ
saka [*załum=anga* *pu-linge~lingetj* *a* *gadu*].
and water=COMPL have-RED~entire NOM mountain
'Once upon a time the earth dissolved, and the mountains everywhere became water.' (*OHPT* 023-001)

b. [*qau* *limutseng* *azua* *vaław*];
so AF.angry that spouse
[*qetsi-in* *a* *djurikuku* *nua-zau* *vaław*].
fight-PF NOM fowl GEN-that spouse
'The wife got angry; and she killed the cockerel.' (*OHPT* 002-065)

My second purpose is therefore to investigate if there is any functional difference between the *saka* coordinate construction and the *sa* construction in encoding sequential events, and to explore the motivation underlying the syntactic coding of this particular construction. Finally, I compare the *sa* construction with other Paiwan complex clause constructions along the Interclausal Relations Hierarchy (Van Valin & LaPolla 1997; Lehmann 1988) in terms of event packaging and syntactic integration (Bohnemeyer, Enfield, Essegbey, Ibarretxe-Antuñano, et al. 2007; Bohnemeyer, Enfield, Essegbey & Kita 2011).

All the data under investigation in this study are narratives and conversations taken from *One Hundred Paiwan Texts* (*OHPT*) and my own fieldnotes. Where no appropriate data can be found, elicited examples are used as supplements.

This paper is organised as follows. Section 2 gives a brief sketch of various strategies of clause linkage in Paiwan, with particular focus laid in adverbial subordination and coordination. Section 3 introduces the characteristics of the *sa* construction in Paiwan and shows with various morphosyntactic tests that the *sa* construction is a third clause-combining type in Paiwan, viz., cosubordination. Section 4 investigates event-packaging and syntactic integration of the three clause-combining types in Paiwan, viz., coordination, adverbial subordination and the *sa* construction. Section 5 discusses

the issue whether *sa* and *sana* are two different cosubordinators. The concluding remarks are given in section 6.

2 Clause-combining strategies in Paiwan

Generally speaking, complex clauses in Paiwan can be divided into (i) subordinate constructions, those in which "one of the clauses (the main clause) is modified by one or more subordinate clauses grammatically dependent upon it and generally introduced…by a subordinating conjunction" (Lyons 1968:178); and (ii) coordinate constructions, those in which "the constituent clauses are grammatically co-ordinate, no one being dependent on the others, but all being…added together in sequence, with or without the so-called coordinating conjunctions." (ibid.) To put it in a simple (maybe oversimplified) way, clausal constituents in subordination are of syntactically unequal weight, while those in coordination are of equal status in syntax. Subordinate clauses are of hypotactic structure; they provide such peripheral information as temporal frames, conditions, and causations on the background of the discourse and thus help define the specific meaning of the main clause. On the contrary, coordinate clauses are of conceptually equal prominence, and none is a ground to another. These two types of clause-combining strategies in Paiwan will be discussed in detail in the following subsections.

2.1 Adverbial subordinate constructions in Paiwan

Commonly seen strategies for subordination in Paiwan are complementation, relative clauses, serial verb constructions, and adverbial subordination (Tang 1999; A. Chang 2006). I restrict the discussion to adverbial subordination, viz., constructions which contain adverbial subordinators, in that they are directly related to the current discussion on the syntactic behavior of the *sa* construction.[6]

The commonly seen adverbial subordinators in Paiwan are *nu* 'when' (irrealis temporal marker), and *ka* 'when (in the past); after' (realis temporal marker) (cf. A. Chang 2000, 2006).[7] Paiwan *ka* 'when; after' subordinate clause is used to provide a time frame based on a realis event serving as a grounding for the main clause, as in (5).

---

[6] Complementation and serial verb constructions are excluded from the current discussion due to their very different syntactic behaviors. See Tang (1999) and A. Chang (2006) for detailed discussions on these two constructions in Paiwan.

[7] Note that A.H. Chang (2006) also identifies *uła* 'so that' as a subordinator, as in the example below:

aitsu	*a*	*ini=anan*	*ka*	*se-kełang*	*tua*	*kai*	*nua*	*Tsemas,*
this	LNK	NEG=CON	after	INVOL-know	OBL	speech	GEN	God
tja-tjiak-aw			*a*	*pa-sa*	*tjai-qayaw*	*tjai*	*Yisu-sama,*	
1PL.INCL.GEN-hold.hand-LF			LNK	CAUS-go	OBL-front	OBL	Jesus-H	
[***uła***	*na-ma-pa-se-qełing*			*tiamadju.*]				
so.that	PFV-AF-CAUS-INVOL-save			3PL.NOM				

'Let us lead to Jesus those who do not yet know the word of God, so that they may be saved.' (*OHPT* 096-026)

However, the functions are diverse; moreover, its use as causal links in *OHPT* is not consistent and the examples are few. I therefore exclude this marker in the present discussion.

(5) [***ka*** *mangtjez=anga* *ti* *kaka*]
when AF.come=COMPL NOM elder.sibling
keḍi=anga *a* *ku-sengseng-an.*
AF.small=COMPL NOM 1SG.GEN-work-NMLZ
'When my brother came back, there was less work for me.' (*OHPT* 079-023)

Likewise, *nu* 'when' clause also serves to provide a time frame for the main clause but the event coded in the *nu* clause is irrealis, for example, a habitual event, as in (6). Moreover, Paiwan *nu*, like the English temporal marker *when*, is frequently found to be used as a conditional marker (Traugott 1985), thus providing a conditional situation, as in (7).

(6) "***nu*** *q‹em›au~qaung* *a* *qali~qali,* *q‹em›aung=itjen*
when <AF>RED~cry NOM RED~friend <AF>cry=1PL.INCL.NOM
a *pu-saladj,*" *aya=sun.*
LNK have-companion say=2SG.NOM
'You said, "When other people cry, we join them in crying."' (*OHPT* 003-050)

(7) *qau* *nu* *me-qatsa* *azua* *kuku,* *g‹em›urugur* *a*
so when AF-big that pet <AF>bark LNK
k‹em›ats=anga, [***nu*** *mangtjez* *a* *tsad*] *nu* *qe~zeme~zemetj.*
<AF>bit=COMPL when AF.come NOM bandit when RED~night
'When the puppies get big, they'll bark and bite, if a thief comes at night.'
(*OHPT* 017-022)

As shown by the examples above, clauses with *nu* and *ka* are subordinate and they provide background information showing time frames and conditions for the main clause.

2.2 Coordination

Paiwan makes a distinction between nominal/NP coordination and VP/clausal coordination; the former is connected by markers such as *kati* (8) and *katua* (9), while the latter is connected by markers such as *saka* 'and' (10), *łakua* 'but' (11), and *'aw* 'but; while' (12) (cf. Tang 2011).

(8) *kati* 'and'
i-zua *aya* [*ti* *Quzang* ***kati*** *Tji-nga~ngaḍuy*].
LOC-that say NOM shrimp and KIND-RED~lizard
'They say there was Shrimp and Tree-lizard.' (*OHPT* 060-001)

(9) *katua* 'and'
a. *saka* *nu* *t‹em›ekel* *a~za~zua* [*qaḍis* ***katua*** *ritał*]
and when <AF>drink RED~that hawk and eagle
ma-tsangavu, *ini=anga* *ka* *min-layap* *aya* *tjautsikel.*
AF-swell NEG=COMPL after AF-fly say story
'And when the hawk and eagle drank, they swelled up and couldn't fly, according to the story.' (*OHPT* 024-010)

b. *i-zua* *si-tsuay-an* *i* *pa-sa-zaya~zaya* *tua* *tapaw*
LOC-that IF-long.time-NMLZ LOC CAUS-go-RED~up OBL hut

ni	*Luvutsi*	*a*	*tapaw*	*ni*	*Savali*	*a*	[*Pasusu*
GEN	PN	LNK	hut	GEN	PN	LNK	PN
katua	*kina*].						
and	mother						

'Once upon a time above the Luvutsi house was the house of Savali Pasusu and her mother.' (*OHPT* 018-001)

c. *aza* *zua* *si-paiz-an* *v‹in›etsik-an* *tua* [*tsau~tsau*
that that IF-fan-NMLZ <PFV>mark-LF OBL RED~being
katua *marka* *ḍiḍi*].
and PL pig
'That fan WAS carved with people and pigs.' (*OHPT* 020-002)

The coordinator *kati* is used to connect proper nouns, while *katua* connects common nouns, no matter whether the noun phrases in question are marked as nominative as in (9a), genitive (9b), or oblique (9c).

VP/clausal coordinate constructions can be further divided into (i) syndetic and (ii) asyndetic constructions (Haspelmath 2004). Syndetic coordination is formed by using the coordinators *saka* 'and' (10), *łakua* 'but' (11), and *'aw* 'but; while' (12).[8]

(10) *saka* 'and'

na-c‹em›ulju	*tua*	*kuka*	*timadu*	
PFV-<AF>kill	OBL	chicken	3SG.NOM	
saka *na-c‹em›ulju*		*tua*	*bibi*	(*timadu*).
and PFV-<AF>kill		OBL	duck	3SG.NOM

'He butchered chickens and ducks.' (Lit.: 'He butchered chickens and he butchered ducks.')

(11) *łakua* 'but'

'aca~'aca=aken	***łakua***	*rilja~riljay=aken.*
RED~AF.tall=1SG.NOM	but	RED~AF.thin=1SG.NOM

'I am tall but thin.' (Lit.: 'I am tall but I am thin.')

(12) *'aw* 'but'

na-vaik	*a*	*s‹em›a-taihuku*	*ti*	*paljang*	
PFV-AF.leave	LNK	<AF>go-Taipei	NOM	PN	
'aw	*na-vaik*	*a*	*s‹em›a-tainan*	*ti*	*kaljalju.*
but	PFV-AF.leave	LNK	<AF>go-Tainan	NOM	PN

'Palang went to Taipei, while Kaljalju went to Tainan.'

The other strategy, the more commonly used one, to combine two or more coordinate elements is by means of asyndetic construction, that is, juxtaposition. For example:

(13) a. [*qau* *limutseng* *azua* *vaław*];
so AF.angry that spouse

[8] These Paiwan coordinators, *łakua* and *'aw* in particular, are more like semantic and pragmatic markers than syntactic conjunctors. In other words, they are used mainly for textual cohesion and discourse coherence, indexing interpersonal and textual metafunctions (cf. Halliday 1994) instead of for clause-combining.

[*qetsi-in* *a* *djurikuku* *nua-zau* *vaław*].
fight-PF NOM fowl GEN-that spouse
'The wife got angry; (and) she killed the cockerel.' (*OHPT* 002-065)

b. *nu* *mangtjez* *a* *kasi* *gaku,*
when AF.come LNK from school
[*vaik* *a* *tsałum*]; [*vaik* *a* *ki-tsa~tsevung*].
AF.leave LNK carry.water AF.leave LNK self-RED~meet
'When they came back from school, they went to get water, (or) went to meet people (coming back laden from the fields).' (*OHPT* 079-053)

As shown by these examples, we can see that coordinate constructions, syndetic and asyndetic, are frequently used to connect two or more events of unspecific temporal order, sequential states of events, or simultaneous states of affairs.

2.3 Interim summary

To put in a nutshell, adverbial subordination is of conceptually unequal prominence. It is syntactically dependent; the information encoded in adverbial clauses is backgrounding, providing time frames, conditional situations, or causal links for the main clause. On the other hand, coordination is of conceptually equal prominence, linking events or states of unspecified or sequentially temporal order. I summarise the functions of coordinate and adverbial subordinate constructions in Paiwan as in Table 2.

Table 2: Functions of adverbial subordinate and coordinate constructions in Paiwan

Forms	Functions
Adverbial subordinate constructions	(1) Temporal (2) Conditional
Coordinate constructions	(1) Unspecified temporal order (2) Sequential states or events (3) Simultaneous states of affairs

3 The *sa* construction

Now let us turn our attention back to the *sa* construction in (1), repeated as (14) below.

(14) *tjara* *djukul-en=anan* *a* *s‹em›u-alap* *a* *qipu* *i-zua*
surely beat-PF=CON LNK <AF>remove-take NOM soil LOC-that
i *ta* *tjapungul-an* [***sa-na*** *qulˌi~qulˌits-i*] [***sa***
LOC OBL grass.root-NMLZ and-then RED~clear.cut.bush-PF.NI and
łama-i *a* *q‹in›epu-an,*] [***sa-na***
burn-PF.NI NOM <PFV>associate-NMLZ and-then
t‹al›e-talem-i *tua* *vasa*].
<QAL>do-plant-PF.NI OBL taro
'We always beat the soil out of the roots, and then clear up and burn the refuse; and then plant the taro.' (*OHPT* 078-004)

As shown by the translation in (14), all the events linked by *sa/sana* are temporally sequential with respect to the main clause. Two questions that may naturally arise are: (i)

Is the *sa* construction a coordinate or subordinate construction? (ii) Is there any functional difference between the *saka* coordinate and the *sa* construction in coding sequential events?

3.1 The characteristics of the *sa* construction

The *sa* construction presents two characteristics. First, *sa* clauses always and only occur in non-initial positions. Second, verb forms in the non-initial clauses of the *sa* construction can be in AF, PF, or LF (and IF) non-indicative forms, as in (15). Note that when the verbs in *sa* clauses appear in AF form, they always appear in bare forms, that is, without AF affixes, with or without reduplication. What is intriguing is the appearance of the genitive marked agent in a *sa* AF-clause,[9] as in (15a–b).

(15) a. *k‹em›an=aken* *sa/sana* ***ku**-vai~vaik.*
<AF>eat=1SG.NOM and 1SG.GEN-RED~go
'I ate and then I left.'

b. *kan-u=anan* *sa/sana* ***su**-tekel*!
eat-IMP=CON and 2SG.GEN-drink
'Eat first and then drink!'

c. *qulits-an=anga* *sakamaya* *sa* ***tjugut-i**.*
clear.cut.bush-LF=COMPL only and sow-PF.NI
'We just clear away the vegetation and sow the millet.' (*OHPT* 78-009)

d. *qau* *su-alap-en* *a* *tjakit,*
so remove-take-PF NOM knife
sa ***kelay-an*** *tua* *djaralap* *i* *tsasaw.*
and hang-LF.NI OBL banyan LOC outside
'They took off their swords and hung them on a banyan tree outside.'
(*OHPT* 007-010)

Moreover, though the perfective markers *-in-* and *na-* cannot occur in *sa* clauses, the completive aspectual marker *anga* and the future modality marker *uri* can. For example:

(16) a. *ka* *ika* *maka-laing* *ti* *Tjikunal* *s‹em›u-tjakupun=anga*
after NEG finished-follow NOM PN <AF>remove-hat=COMPL
[*sa* *verits-an=**anga*** *a* *zaman* *a* *qau* *a*
and discard-LF.NI=COMPL LNK torch LNK bamboo LNK
dj‹in›adjas] [*sa* *tju-umaq=**anga***].
<PFV>grasp and there-house=COMPL
'When Tjikunal failed to catch up with the cat, he took off his hat and threw away the grass and bamboo he had been holding, and went indoors.'
(*OHPT* 80-022)

[9] The actor/agent of an Austronesian AF clause is normally marked as a nominative. As shown in (15a), the actor of the first AF-verb 'eat' is marked as a nominative, i.e., *=aken* 'I', while it is marked as a genitive, i.e. *ku-*, in the second AF-verb clause following *sa*.

b. *qau* *azua* *tsaḍ* *ki-rimu-in* *a* *dj‹em›ukul* *timadju*
so that bandit do-quick-PF LNK <AF>beat 3SG.NOM
[*sa* ***uri*** *pa-patsay-i*] [*sa* *su-kava-i* *timadju*] [*sa*
and will CAUS-die-PF.NI and remove-clothing-PF.NI 3SG.NOM and
pa-se-qereng-i] [*sa* *alap-i* *aza*
CAUS-INVOL-lie.down-PF.NI and take-PF.NI that
na-nema~nema=anga] [*sa* *vaik-an*].
PFV-RED~what=COMPL and leave-LF.NI
'And the robbers quickly beat him up, and were going to kill him, and took off his clothes and knocked him over and took everything, and left him.'
(*OHPT* 096-008)

3.2 The *sa* construction is not a subordinate construction

As pointed out by Haspelmath (1995:12ff.), and Van Valin & LaPolla (1997:448), coordination and subordination are not categorically clear-cut but rather form a continuum, and therefore the notion of subordination seems problematic and unworkable for the languages other than Indo-European languages. For a typological study, Haspelmath (ibid.) therefore proposes a set of criteria as sufficient syntactic conditions for subordination. The three criteria adopted here to test Paiwan data are (i) clause-internal word order, (ii) variable position, and (iii) restrictivity and focusability.[10]

3.2.1 Clause-internal word order

While each part of a coordinate construction must be continuous and non-overlapping, a subordinate clause may appear inside its superordinate clause and thus makes the superordinate clause become discontinuous. Consider examples:

(17) Japanese (Kuno 1973:205, cited in Haspelmath 1995:12, ex. 25, 26)
a. *John* *wa* *boosi* *o* *nui-de,* *Mary* *ni* *aisatu* *si-ta.*
John TOP hat ACC take.off-CVB Mary DAT greet do-PST
'John took off his hat and greeted Mary.'

b. *John* *wa* *Mary* *ni* **[*boosi* *o* *nui-de*]** *aisatu* *si-ta.*
John TOP Mary DAT hat ACC take.off-CVB greet do-PST
'John took off his hat and greeted Mary.'

There is no obvious discontinuity in (17a), so word order tells us nothing about the coordinate or subordinate status of this particular construction. However, (17b) shows that *boosi o nuide* comes between two constituents of the superordinate clause *John wa Mary ni aisatu si-ta*, so its subordinate status is beyond doubt.

Let's look at Paiwan examples. Both *nu* and *ka* subordinate clauses can come between the constituents of the superordinate clause and make the main clause discontinuous. For example:

---

[10] The other two are (iv) possibility of backwards pronominal anaphora (i.e., pronominal cataphora) and control, and (v) possibility of extraction (Haspelmath 1995:12ff). Other languages may adopt a different set of criteria.

(18) *aitsu* *a* *malada* [***nu*** *i-zua* *saqetju*]
be.thus NOM shaman when LOC-that painful
djapes-en *niamadju.*
breathe-PF 3PL.GEN
'When someone is in pain the shamans blow on them.' (*OHPT* 85-003)

Occasionally, *sa* clauses may come inside the superordinate clause in a triple-verb serial verb construction.[11] For example:

(19) *qau* *vaik-en* [***sa*** ***ki-qenetj-i*** ***nua*** ***qam***] *a* *s‹em›a* *zua.*
so leave-PF and do-see-PF.NI GEN ant.eater LNK <AF>go that
'The ant-eater went there to have a look.' (*OHPT* 009-057)

In this regard, *sa* clauses behave quite similarly to subordinate clauses.

3.2.2 Variable position

Another positional criterion for subordination is variable position: only subordinate clauses may come after or before the superordinate clause (Haspelmath 1995:13). For example:

(20) a. [***nu*** ***k‹em›an=aken***,] *ku-si-pa-zeka~zekatj=sun*
when <AF>eat=1SG.NOM 1SG.GEN-IF-CAUS-RED~share=2SG.NOM
a *pa-layu~layu.*
LNK CAUS-RED~store
'When I eat I always put some out for you too.'

b. *ku-si-pa-zeka~zekatj=sun* *a* *pa-layu~layu,*
1SG.GEN-IF-CAUS-RED~share=2SG.NOM LNK CAUS-RED~store
[***nu*** ***k‹em›an=aken***].
when <AF>eat=1SG.NOM
'When I eat I always put some out for you too.'

(21) [***nu*** *se-tsevung=itjen* *tua* *se-lapa~lapay*], [***nu***
when INVOL-meet=1PL.INCL.NOM OBL INVOL-RED~severe when
se-tsevung=itjen *tua* *q‹em›au~qaung*], [***nu***
INVOL-meet=1PL.INCL.NOM OBL <AF>RED~cry when
se-tsevung=itjen *tua* *na-s‹em›‹al›imsim*],
INVOL-meet=1PL.INCL.NOM OBL PFV-<AF><QAL>pity
tja-pa-se-qeɬing-aw, *tja-tjiak-aw,*
1PL.INCL.GEN-CAUS-INVOL-save-LF.HORT 1PL.INCL.GEN-hold.hand-LF.HORT
tja-pa-ka-vaɬut-aw.
1PL.INCL.GEN-CAUS-STAT-alive-LF.HORT
'If we meet those who are suffering, if we meet those who are crying, if we meet those who are in trouble, let us save, lead and help them.' (*OHPT* 096-025)

---

[11] This may be due to a very unique syntactic behavior of three-verb Paiwan SVC, which displays intriguing word order (see Hsieh 2012 for a detailed description). For example:

manu ***vaik*** *tiamadju* ***a*** ***tsaɬum*** ***a*** ***s‹em›a*** *tjua* *laku.*
then AF.leave 3PL.NOM NOM carry.water LNK <AF>go there water.well
'And he went to the well to get water.' (*OHPT* 002-057)

The *nu* clause in (20a) precedes the main clause, while it follows the main clause in (20b); in either case, the sentence reading does not differ from each other. In addition, there are four *nu* clauses in the clause-initial position in (21); any positional shift of these four *nu* clauses does not render any change in the sentence meaning, either.

The order of *sa* clauses is rather fixed. To begin with, *sa* clauses can never occur in the sentence initial position. Moreover, when there are more than one *sa* clauses in the sentence, any order reversion of the *sa* clauses results in unacceptability by the native speakers as the events coded may be contrary to the socio-cultural custom and thus sound incomprehensible to the language users. For example:

(22) *tjara* *djukul-en=anan* *a* *s‹em›u-alap*
surely beat-PF=CON LNK <AF>remove-take
a *qipu* *i-zua* *i* *ta* *tjapungul-an*
NOM soil LOC-that LOC OBL grass.root-NMLZ
[*sa-na* ***qulits-i***,]
and-then RED~clear.cut.bush-PF.NI
[*sa* ***łama-i*** *a* *q‹in›epu-an*,]
and burn-PF.NI NOM <PFV>associate-NMLZ
[*sa-na* ***t‹al›e-talem-i*** *tua* *vasa*.]
and-then <QAL>do-plant-PF.NI OBL taro
'We always beat the soil out of the roots, and then clear up and burn the refuse; and then plant the taro.' (*OHPT* 078-004)

There are three *sa* clauses following the main clause in (22), describing the steps that the Paiwan people take in doing taro planting. Any change of the order of these three *sa* clauses will make the order of the planting steps change, and thus will make the procedure of taro planting incorrect.

3.2.3 Restrictiveness and focusability

Only subordinate clauses, but not coordinate clauses, may be modified by focus particles like *also* and *only* and thus can be interpreted restrictively (cf. Tikkanen 1995; Haspelmath 1995), i.e., as modifying the main clause in such a way that its reference is narrowed. For example:

(23) Catalan (Haspelmath 1995:15, ex. 33)
[***Només*** *sortint-nos* *de* *la* *sintaxi* *entesae* *strictament*]
only leaving from the syntax understood strictly
podrem *relacionar* *las* *frases* *de* (6) *amb* *la* *negació.*
we:can relate the sentences of (6) with the negation
'Only by leaving syntax in the strict sense can we relate the sentences in (6) to negation.'

Let's turn back to Paiwan examples:

(24) a. [***amin*** *nu* *mangetjez* *timadju*].
only when return.AF 3SG.NOM

mana *makaya* *pa-ka-leva* *tanuamen*
then able CAUS-STAT-happy 1PL.EXCL.OBL
'Only when he comes will we be happy. (We will be happy only when he comes.)'

b. * *ku-alap-en* *azua* *sunat* [***amin*** *sa* *supu-i*].
1SG.GEN-take-PF that book only and read-PF.NI

c. * *ku-alap-en* *azua* *sunat* [*sa* ***amin*** *supu-i*].
1SG.GEN-take-PF that book and only read-PF.NI

d. * *ku-alap-en* *azua* *sunat* [*sa* *supu-i* ***amin***].
1SG.GEN-take-PF that book and read-PF.NI only

As shown in (24a), the focus marker *amin* 'only' can be used to restrict the adverbial *nu* clause, while the occurrence of *amin* 'only' in *sa* clauses makes (24b–d) ungrammatical.

3.2.4 Interim summary

Based on the data presented, we can see that the *sa* construction fails to meet most of the criteria for subordination discussed previously, as summarised in Table 3.

Table 3: Criteria for subordination

Criteria for subordination	Adverbial subordinate clauses	*sa* clauses
(1) Can occur in clause-internal position?	Yes	Yes (only in triple-verb SVCs)
(2) Can display variable positions?	Yes	No
(3) Can be modified by focus particles?	Yes	No

The asymmetric verb forms in the *sa* construction indicate the subordinate status of the *sa* clauses in the construction; nevertheless, the *sa* construction fails to meet most of the criteria of subordination in that its clausal constituents' positions cannot be moved freely, nor can they be focused or restricted. Therefore, the status of the *sa* construction to be recognised as subordination is doubted.

3.3 The *sa* construction is not a coordinate construction

Then, is the *sa* construction a coordinate construction? Despite the coordinate reading, the answer again is negative. The first piece of evidence against analyzing the *sa* construction as a coordinate construction is the asymmetric verb forms in the construction, because verb forms in the first clause are indicative while those in the *sa* clauses are non-indicative. The second piece of evidence comes from the illocutionary forces of negation and imperative in clauses. In Paiwan, either negation or imperative has its own illocutionary forces restricted within a monoclausal structure. For example:

(25) a. [***sa-u***,] [***kivadaq-u*** *ti* *vuvu* *i* *ḍaḍengeraw*.]"
go-IMP do-ask-IMP NOM.H grandparent LOC PN
'Go! And ask granny Ḍaḍengeraw!' (*OHPT* 074-013)

b. *nu* *mangtjez=sun,* [***kesa-u*** *tu* *kinsa*!]
when AF.come=2SG.NOM cook-IMP OBL food
'When you come home, cook (some) food!'

c. *nu* *masan* *palak=aken,* [***maya*** *ki-dut*!]
when STAT.make sorcery=1SG.NOM don't do-approach
'When I am possessed, don't come near!' (*OHPT* 038-011)

As shown in (25), each operator of negation and imperative has its own illocutionary force within the clause where it occurs. Therefore, in a coordinate construction, both syndetic and asyndetic, the first clause's illocutionary force, be it imperative or negative, does not have its scope over all the clauses to the right. This may suggest that each clausal constituent in the *saka* coordinate construction functions independently. Let's look at imperative examples first.

(26) a. [*kan-u*]! ***saka*** [*vaik-u*]!
eat-IMP and leave-IMP
'Eat! And leave!'

a'. * [*kan-u*]! ***saka*** [*vaik*]!
eat-IMP and leave

b. [*vaik-u*]! [*vecik-u*]!
leave-IMP write-IMP
'Go! Do your homework!'

b'. * [*vaik-u*]! [*vecik*]!
leave-IMP write

As shown in (26), each clause in the *saka* coordinate construction is required to carry an imperative marker, viz. *u* here. Nevertheless, only one imperative marker is allowed in the *sa* construction in that the first clause's illocutionary force has its scope over all the *sa* clauses to the right. Insertion of an extra imperative marker in the *sa* construction will result in ungrammaticality, as illustrated in (27a') and (27b').

(27) a. *kan-u=anan* ***sana***/**sa** *su-vai~vaik*!
eat-IMP=CON and 2SG.GEN-RED~leave
'Eat and then leave.'

a'. * *kan-u=anan* ***sana***/***sa*** *vai~vaik-u*!
eat-IMP=CON and RED~leave-IMP

b. *kan-u=anan* ***sana***/***sa*** *su-tekel*!
eat-IMP=CON and 2SG.GEN-drink
'Eat first and then drink.'

b'. * *kan-u=anan* ***sana***/***sa*** *tekel-u*!
eat-IMP=CON and drink-IMP

In a similar vein, the illocutionary force of negation in the first clause does not have scope over all the clauses to the right in the *saka* coordinate construction. For example:

(28) a. * [*inika* *ku-‹in›alap* *a* *sunat*]
NEG 1SG.GEN-<PF>take NOM book

saka [*ku-s‹in›upu*].
and 1SG.GEN-<PF>read

b. [*inika* *ku-‹in›alap* *a* *sunat*]
NEG 1SG.GEN-<PF>take NOM book
saka [*inika* *ku-s‹in›upu*].
and NEG 1SG.GEN-<PF>read
'I didn't take the booklets and read them.' (Lit.: 'I did not take the booklets and I did not read them.')

As shown in (28), to be negated, each clausal constituent in the *saka* coordinate construction requires one syntactic operator of negation. There is no way to have one negator in the *saka* coordinate construction and have the illocutionary force extend to all the following clauses to the right. To the contrary, only one negator is allowed in the *sa* construction. For example:

(29) a. *inika* [*ku-‹in›alap* *a* *sunat* ***sa*** *ku-supu-i*].
NEG 1SG.GEN-<PF>take NOM book and 1SG.GEN-read-PF.NI
'Do not take the book and read it!' ('Do not take the book and do not read it!')

b. * [*inika ku-‹in›alap* *a* *sunat*] ***sa*** [*inika ku-supu-i*].
NEG 1SG.GEN-<PF>take NOM book and NEG 1SG.GEN-read-PF.NI

Based on the data presented, it is suggested that the interclausal relation of the *sa* construction is very different from that of the *saka* coordinate construction. Just like the *saka* coordinate construction, the *sa* construction encodes sequential events of conceptually equal prominent; however, the *sa* construction behaves totally differently from the *saka* coordinate construction in that the scope of illocutionary force of negation and imperative in the initial clause of the *sa* construction can be extended over to the clauses to the right, while those in the *saka* construction cannot. In other words, the clausal units in the *sa* construction do not function independently. I summarise these findings in Table 4.

Table 4: The scope of illocutionary force in the *saka* and the *sa* constructions

Can the scope of illocutionary force be extended over to the clauses to the right?	*saka* coordinate construction	*sa* construction
Negation	No	Yes
Imperative	No	Yes

3.4 The *sa* construction is a cosubordinate construction

As stated above, the *sa* construction appears not to be a subordinate construction, nor is it a coordination. Nonetheless, it exhibits some characteristics of both subordination and coordination at the same time, and can thus be called cosubordination (Foley & Van Valin 1984: chap 6; Foley 2010).[12] On the one hand, the *sa* construction is like coordination in that it is used to link sequential events of conceptually equal semantics. On the other hand, the *sa* construction behaves like subordination in that it is structurally asymmetric and its clausal constituents do not function independently. Moreover, *sa* clauses can come between the constituents of the main clause. However, unlike

[12] Literature has different names for such a construction: clause-chaining and cosubordination.

flexibility of the clause constituents in other adverbial subordinate clauses in Paiwan, any positional switch of *sa* clauses will result in unacceptability or ungrammaticality.

In addition to the above-mentioned negative evidence, positive evidence showing the *sa* construction is a cosubordination construction include the fact that only one syntactic operator of negation and imperative is allowed in the *sa* construction, while each clausal constituent needs one operator in both coordinate and subordinate constructions in making negation and imperative, as described earlier in detail. Moreover, event packaging and syntactic integration in the *sa* construction appear to be tighter than those in the other two constructions (see the following section for a detailed discussion on this point).

4 Event packaging and syntactic integration in three clause-combining types

Now, we may go back to our second question: Are there any functional differences in coding sequential events between coordinate constructions and the *sa* construction? We may repeat the functions of these three clause-combining types in Paiwan as in Table 5.

Table 5: Functions of the three major clause-combining types in Paiwan

Forms	Functions
Subordinate constructions	(1) Temporal
	(2) Conditional
Coordinate constructions	(1) Unspecified temporal order
	(2) Sequential states or events
	(3) Simultaneous states of affairs
Cosubordinate constructions	(1) Sequential states or events
	(2) Simultaneous states of affairs

As shown in Table 5, we can see that coordinate and cosubordinate constructions in Paiwan overlap quite a lot with each other in functions, as both code sequential and simultaneous events and states. This may bring in our second question: What is the functional difference characterizing the events coded in these two constructions? Scrupulously probing into the narratives in *OHPT*, I have found that there is a strong tendency showing that event packaging in the *sa* construction is tighter than those in coordinate and subordinate constructions. In other words, those events coded in coordinate constructions are two or more distinct events that happen in sequence, but those coded in the *sa* construction can be viewed as different phases of a single macro-event (as proposed by Bohnemeyer, Enfield, Essegbey, Ibarretxe-Antuñano, et al. 2007; Bohnemeyer, Enfield, Essegbey & Kita 2011). We may use the following examples to illustrate this point. For example:

(30) a. [*qau* *limutseng* *azua* *vaław*];
so AF.angry that spouse
[*qetsi-in* *a* *djurikuku* *nua-zua* *valaw*].
fight-PF NOM fowl GEN-that spouse
'The wife got angry; and she killed the cockerel.' (*OHPT* 002-065)

b. *qau* *me-qatsa* *azua;* *ki-vadaq* *tua* *kina:* "*tima* *ku-kama?*"
so AF-big that do-ask OBL mother who 1SG.GEN-father
'He grew up, and asked his mother: "Mother, who is my father?"'
(*OHPT* 005-010)

As shown in (30), the events coded in each sentence are individual events, one happening after another and loosely bound. Moreover, it is often the case that the sentence meaning remains intact, even if the order of the clause constituents of a coordinate construction is reversed. For example:

(31) a. *nu* *mangtjez* *a* *kasi* *gaku,*
when AF.come LNK from school
[*vaik* *a* *tsałum,*] [*vaik* *a* *ki-tsa~tsevung*].
AF.leave LNK carry.water AF.leave LNK self-RED~meet
'When they came back from school, they went to get water or went to meet people (coming back laden from the fields). (*OHPT* 079-053)

b. *nu* *mangtjez* *a* *kasi* *gaku,*
when AF.come LNK from school
[*vaik* *a* *ki-tsa~tsevung*], [*vaik* *a* *tsalum*].
AF.leave LNK self-RED~meet AF.leave LNK carry.water
'When they came back from school, they went to get water or went to meet people (coming back laden from the fields).

However, it is impossible to switch the order of any clausal constituents in the *sa* construction. For example:

(32) *a* *si-sane* *ka~kan-en-an* *tja-alap-en* *azua* *seti*
NOM IF-make RED~eat-PF-NMLZ 1PL.INCL.GEN-take-PF that meat
sakamaya [***sa*** ***tja-senaw-i*** *tua* *qavu* *a*
only and 1PL.INCL.GEN-wash-PF.NI OBL ash NOM
s‹em›ane *lia~liaw,* *tua* *sika* *su-alap-an* *nua-zua*
<AF>make RED~many OBL reason remove-take-LF GEN-that
dj‹ar›aya~djayadjay,] [***sa-na*** ***tja-qirqir-i*** *tua* *ławław,*]
<QAL>RED~sticky and-then 1PL.INCL.GEN-fry-PF.NI OBL oil
[***sa*** ***tja-tau'iyu-i,***] [***sa*** ***tja-djamay-an***].
and 1PL.INCL.GEN-mix.with.soy-PF.NI and 1PL.INCL.GEN-side.dish-LF.NI
'When we are going to prepare them to eat, we take the meat only and clean it many times over with ash/lime to get rid of the stickiness; then we fry them in oil, add soya sauce, and eat them as a side dish.' (*OHPT* 086-005)

In this particular excerpt, the speaker was talking about how they prepared, cooked, and ate snails. Each clause in the *sa* construction can be viewed as a step of this particular food-preparation process: step 1, take out the snail meat only; step 2, clean it many times over with ash to get rid of the stickiness; step 3, fry the meat in oil; step 4, add soya sauce; and finally, step 5, serve it as a side dish. All these five events are chained or bound tightly together as different phases of a single event or action. Excerpt (33) shows a ritual procedure, (34) tells the fieldwork procedure of taro planting, and (35) describes the procedure of caring for a child.

(33) Ritual procedure (*OHPT* 085-025)
nu uri ma-lada tiamadju
when will STAT-Ritual 3PL.NOM
me-suaw a g‹em›edged
AF-yawn LNK <AF>repeat
sa-na *pa-ngetje~ngetjez-i*
and-then CAUS-RED~come-PF.NI
sa-na *sane djala~djalan.*
and-then make RED~road
'When shamans are going to perform a trance rite, they yawn repeatedly and then the spirits come on them, and they make their way (to heaven).'

(34) Procedure of taro planting (*OHPT* 078-001)
nu s‹em›an pu-vasa-an dangdang-an=anan,
when <AF>make have-taro-NMLZ cut.bush-LF=CON
sa-na *łama-i* ***sa-na*** *qusul-i.*
and-then burn-PF.NI and-then dig.roots-PF.NI
'When we make a taro field, we clear the ground and burn the vegetation and dig out the roots.'

(35) Procedure of caring for a child (*OHPT* 027-003)
qau alap-en azua ałak, sa pa-ki-ramata-i
so take-PF that child and CAUS-self-wash-PF.NI
sa pa-itung-i sa pa-kan-i.
and CAUS-cloth-PF.NI and CAUS-eat-PF.NI
'She took the child, washed, clothed and fed it.'

Are there any syntactic characteristics that support such a functional difference? The answer appears to be positive. Three parameters, viz., co-temporality, syntactic operators, and event participant sharing (Bohnemeyer, Enfield, Essegbey, Ibarretxe-Antuñano et al. 2007; Bohnemeyer, Enfield, Essegbey & Kita 2011; Ameka 2006), are employed here to determine event packaging and syntactic integration of three clause-linkage types, i.e., coordination, subordination, and cosubordination.

To begin with, according to Bohnemeyer et al.'s (Bohnemeyer, Enfield, Essegbey, Ibarretxe-Antuñano et al. 2007; Bohnemeyer, Enfield, Essegbey & Kita 2011) concept of Macro-Event Property (MEP), an "event-denoting construction has the MEP iff it combines only with those time-positional or durational operators that have scope over all sub-events it entails" (Bohnemeyer, Enfield, Essegbey & Kita 2011:48). In other words, events that are in temporal contiguity are more "tightly integrated" syntactically—i.e., more similar to simple sentences—than others (Bohnemeyer, Enfield, Essegbey & Kita 2011:46–47). The syntactic evidence I am looking for is aspectual operators indicating co-temporality of events coded in each construction.

As mentioned in section 3.3, of the four TAM markers, *na-* perfective, *‹in›* perfective, *uri* future modality, and *anga* completive marker, only the completive aspectual marker *anga* and the future modality marker *uri* are allowed in non-initial *sa* clauses. Nevertheless, the four aspectual markers can occur freely in each clausal constituent in both coordination and subordination. Table 6 summarises the distribution of the occurrence of different aspectual markers in the three clause-linkage types.

Table 6: The distribution of the occurrence of aspectual markers in each clause-combining type

co-temporality		*saka* coordinate	subordinate	*sa* cosubordinate
aspectual markers	*na-* perfective	✓	✓	×
	‹*in*› perfective	✓	✓	×
	uri future modality	✓	✓	✓
	anga completive	✓	✓	✓

Next, as mentioned earlier, only one syntactic operator of negation and imperative is allowed in the *sa* cosubordination, while each clausal constituent needs one operator in both coordinate and subordinate constructions in making negation and imperative. In other words, the fact that the occurrence of one single operator in the first clause can have its illocutionary force of negation or imperative over all the clauses to the right suggests that each clausal unit in the *sa* construction does not function independently and that the interclausal dependency within the *sa* construction is so strong that all the clausal constituents are integrated and behave as a simple sentence.

Finally, each clause in the *sa* construction can have its own argument structure; in other words, all the arguments in each clause can be, though not necessarily, explicitly realised. Nevertheless, there is always one event participant shared by all the clauses in the *sa* construction. In most cases, the shared event participant is the agent, especially when the agent is in the first personal plural inclusive form, that is, *tja*. For example:

(36) a. [***tja****-alap-en* *a* *ałak;*] [*qau* ***tja****-katsu-in*
1PL.INCL.GEN-take-PF NOM child so 1PL.INCL.GEN-carry-PF
a *tju-umaq* *zua* *kuku,*] [*sa* ***tja****-pa-kan-i*].
LNK there-house that pet and 1PL.INCL.GEN-CAUS-eat-PF.NI
'We'll take the puppies, and bring them home, and feed them.' (*OHPT* 017-021)

b. *manu* [*mangtjez* ***a*** ***tsema*****s**], [*sa* *su-alap-i* *ti*
then AF.come NOM god and remove-take-PF.NI NOM
aḍing]; [*sa* *katsu-i=aken* *a* *sa* *va~vaw*
boy and carry-PF.NI=1SG.NOM LNK <AF>go RED~above
i-tjua *ka-levelev-an*].
LOC-there STAT-dazzle-NMLZ
'And a god came and took the boy off me; and I was taken up to heaven (by the god).' (*OHPT* 016-038)

The patient can sometimes serve as the shared event participant in the *sa* construction, too. For example:

(37) *qau* [*tsingtsing-an*] [*sa* *su-alap-i* ***a*** ***ałis***].
so cut-LF and remove-take-PF.NI NOM tooth
'So the teeth were cut and removed.' (*OHPT* 021-008)

As shown in the examples, there is always one shared event participant serving as a pivot connecting all the sub-events in the *sa* construction. On the contrary, the *saka* coordinate construction and adverbial subordinate construction can have totally different event participants in each clause, which is not allowed in the *sa* construction. For example:

(38) a. *katiaw* [*na-d‹em›ukul=aken* *tai* *kaljalju;*] **(*saka*)**
yesterday PFV-<AF>hit=1SG.NOM OBL PN and
[*na-d‹em›ukul* *ti* *paljang* *tai* *muakay*].
PFV-<AF>hit NOM PN OBL PN
'Yesterday I hit Kaljalju; (and) Paljang hit Muakay.'

b. *katiaw* [***ka*** *d‹em›ukudukul=aken* *tai* *kaljalju*],
yesterday when <AF>hit=1SG.NOM OBL PN
[*d‹em›ukudukul* *ti* *paljang* *tai* *muakay*].
<AF>hit NOM PN OBL PN
'Yesterday when I hit Kaljalju, Paljang was hitting Muakay.'

Based on the data presented, I find that the events coded in the *sa* construction have a tighter event packaging than those coded in coordinate or subordinate constructions, and that the syntactic integration among the clausal constituents in the *sa* construction is also tighter than those in coordinate and subordinate. Table 7 summarises these findings.

Table 7: Event packaging and syntactic integration in the three clause-combining types in Paiwan

Parameters	Features	*saka* construction	subordinate constructions	*sa* construction
Co-temporality	Each verb can have its own aspectual marking	Yes	Yes	Yes (limited)
Each clausal constituent can have its own syntactic operator in making negative or imperative forms	Negation	Yes	Yes	No
	Imperative	Yes	Yes	No
Event-participant-sharing	Each verb can have its own argument structure	Yes	Yes	Yes
	At least one shared event participant is required	No	No	Yes

5 Are *sa* and *sana* different cosubordinators?

Before I wrap up this paper, there is one more issue I would like to tackle. A. Chang (2006:308–311) proposes that there are two different cosubordination constructions in Paiwan: the one introduced by the neutral *sa* 'and; while', and the other by the sequential *sana* 'and then'; the former denotes a simultaneous action/event, whereas the *sana* construction denotes a sequential event (A. Chang 2006:310). For example:

(39) A. Chang (2006:309, ex. 12 & 15)

a. *'‹em›au~'aung=aken* ***sa*** *ku-kesa~kesa.*
<AF>RED~cry=1SG.NOM and 1SG.GEN-RED~cook
'I am/was crying and cooking. I am/was crying while cooking.'

b. *na-ki'ivu=anan=aken* *tjay* *kama* ***sa-na*** *ku-vai~vaik*
PFV-AF.speak=CON=1SG.NOM OBL father and-then 1SG.GEN-RED~go

katiaw.
yesterday
'I had spoken to my father and then I left yesterday.'

However, as revealed by the data in *OHPT*, *sa* and *sana* are both used to mark temporally sequential events. For example:

(40) *a* *si-sane* *ka~kan-en-an* *tja-alap-en* *azua*
NOM IF-make RED~eat-PF-NMLZ 1PL.INCL.GEN-take-PF that
seti *sakamaya* ***sa*** *tja-senaw-i* *tua* *qavu* *a*
meat only and 1PL.INCL.GEN-wash-PF.NI OBL ash LNK
s‹em›ane *lia~liaw,* *tua* *sika* *su-alap-an* *nua-zua*
<AF>make RED~many OBL reason remove-take-LF GEN-that
dj‹ar›aya~djayadjay, ***sa-na*** *tja-qirqir-i* *tua*
<QAL>RED~sticky and-then 1PL.INCL.GEN-fry-PF.NI OBL
ławław, ***sa*** *tja-tau'iyu-i,*
oil and 1PL.INCL.GEN-mix.with.soy-PF.NI
sa *tja-djamay-an.*
and 1PL.INCL.GEN-side.dish-LF.NI
'When we are going to prepare them to eat, we take the meat only and clean it many times over with ash/lime to get rid of the stickiness; then we fry them in oil, add soya sauce, and eat them as a side dish.' (*OHPT* 086-005)

Moreover, numerous examples from *OHPT* indicate that the *sa* construction is not used to mark simultaneous events, as illustrated by the following examples, just to name a few.

(41) a. *qau* *mi-gatsał* ***sa*** *vaik* *a* *s‹em›a* *pa-teku~teku*
so AF-stand and leave LNK <AF>go CAUS-RED~down
tua *taqtaq.*
OBL platform
'He got up and went under the sleeping platform.' (*OHPT* 011-029)

b. *qau* *azua* *kaka* *a* *si-sekam* *nua* *tsemas*
so that elder.sibling LNK IF-spread GEN god
sa *pa-qereng-i.*
and CAUS-lie.down-PF.NI
'And the god made a bed for the younger brother and laid him down to sleep.' (*OHPT* 016-018)

c. *qułits-an=anga* *sakamaya* ***sa*** *tjugut-i.*
clear.cut.bush-LF=COMPL only and sow-PF.NI
'We just clear away the vegetation and sow the millet.' (*OHPT* 078-009)

d. *ka* *i-zua=anga* *ku-itung,* *vaik=anga=ken* *a*
after LOC-that=COMPL 1SG.GEN-cloth leave=COMPL=1SG.NOM LNK
s‹em›a *umaq;* ***sa*** *ku-pa-itung-i=anga* *tia*
<AF>go house and 1SG.GEN-CAUS-cloth-PF.NI=COMPL NOM.PL
kina *a* *mareka* *ku-kaka.*
mother LNK many 1SG.GEN-elder.sibling
'When I had got some clothes I went back home; and I gave clothes to my mother and my siblings.' (*OHPT* 079-028)

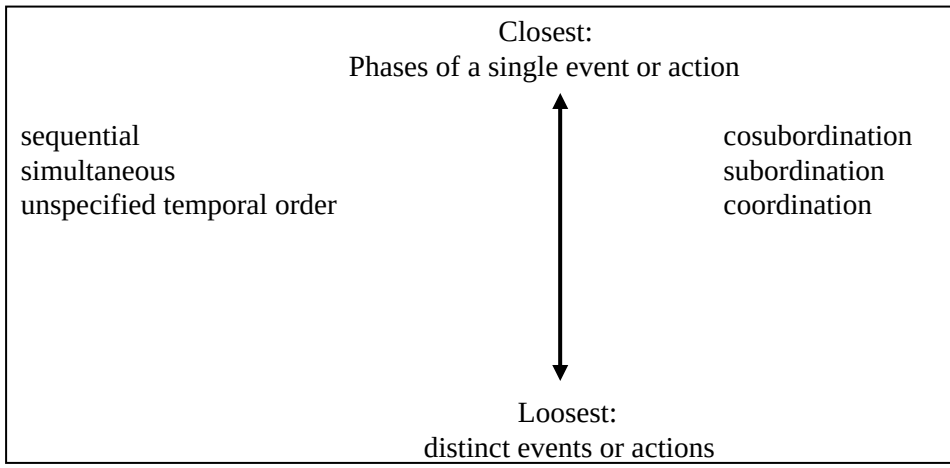
These sentences are perfectly grammatical and acceptable to our informants. All evidence suggests that the proposal that *sa* and *sana* are used to mark different cosubordination types may not be feasible nor plausible. Then, what is the difference between *sa* and *sana* if they do not differ in function?

I investigated the data in *OHPT* attempting to look for some historical clues to answering the question. There are 407 instances of *sa* and 156 instances of *saka* used to connect sequential or simultaneous events in Texts 1–74; however, *sana* can be found only of totally 5 instances occurring in only Texts 72 and 73,[13] where *sana* is glossed as 'and-?' indicating *-na* is a later attachment, whose functions were not yet clear and whose usage was not yet popularised then. It is not until Texts 75–100 where I can find more instances of *sana* and where it is clearly glossed as 'and-then'. According to the introduction chapter of *OHPT* (Early & Whitehorn 2003:3–7), Texts 1–74 of *OHPT* were collected by Ogawa & Asai in 1935, while Texts 75–100 were collected by Early & Whitehorn during 1952–1954. The nearly twenty-year gap may suggest that the use of *sana* as a cosubordinator is a later emergence. At present, no positive evidence showing that *sa* and *sana* are two different cosubordinators has been found. More studies are needed to tease this issue out.

6 Conclusion

This study may be of some theoretical implications. It has shown that Paiwan has a very intriguing clause-combining type, viz., the *sa* cosubordinate construction, in addition to coordinate and subordinate constructions. It has also shown that the event packaging in the *sa* construction is tighter than those in coordinate and subordinate constructions. Moreover, evidence has also revealed that the syntactic integration is stronger in the *sa* construction than those in coordination and subordination. We may borrow Van Valin & LaPolla's (1997:481) concept of Interclausal Relations Hierarchy and arrange the different clausal relations in Paiwan as in (42).

(42) Interclausal Relations Hierarchy in Paiwan (adapted from Van Valin & LaPolla 1997:481)



[13] Two instances are in 072-043; two more instances occur in Text 73-057 and one more in 073-058.

This finding may suggest that there exists isomorphism between event packaging and syntactic integration of these three types of clause-combining in Paiwan (Givón 2001; Radden & Dirven 2007). It is hoped that such a study may shed some light on crosslinguistic studies of clause-combining types in Formosan languages.

Appendix

Verb forms in Paiwan

Focus	Indicative forms	Non-indicative forms (Imperative/negation)
Agentive Focus	*ma-; ø; ; m-; <en>*	*Ø*
Patient Focus	*<in> ; in-; -in; -en*	*-i*
Locative Focus	*-an*	*-an*
Referential / Instrumental / Benefactive Focus	*si-*	*-an*

References

Ameka, Felix K. 2006. Ewe serial verb constructions in their grammatical context. In *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 124–143. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.

Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke and Felix K. Ameka. 2007. Principles of event segmentation in language: The case of motion events. *Language* 83.3:495–532.

Bohnemeyer, Jürgen, N. J. Enfield, James Essegbey and Sotaro Kita. 2011. The macro-event property: The segmentation of causal chains. In *Event representation in language and cognition*, ed. by Jürgen Bohnemeyer and Eric Pederson, 43–67. Language, culture, and cognition 11. New York: Cambridge University Press.

Chang, Anna Hsiou-chuan. 2000. *paiwanyu cankao yufa* [*A reference grammar of Paiwan*] Formosan Languages Series 9. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Chang, Anna Hsiou-chuan. 2006. A reference grammar of Paiwan. PhD dissertation. Canberra: Australian National University.

Chang, Henry Yungli. 2006. The guest playing host: Adverbial modifiers as matrix verbs in Kavalan. In *Clause structure and adjuncts in Austronesian languages*, ed. by Hans-Martin Gärtner, Paul Law, and Joachim Sabel, 43–82. Berlin: Mouton de Gruyter.

Dirven, René and Günter Radden. 2004. The cognitive basis of language: Language and thought. In *Cognitive exploration of language and linguistics*, ed. by René Dirven and Marjolijn Verspoor, 1–23. Cognitive Linguistics in Practice 1. Amsterdam and Philadelphia: John Benjamins. 2nd edition

Early, Robert and John Whitehorn. (ed.) 2003. *One hundred Paiwan texts*. Pacific Linguistics 542. Canberra: The Australian National University.

Foley, William A. 2010. Clause linkage and nexus in Papuan languages. In *Clause linking and clause hierarchy: Syntax and pragmatics*, ed. by Isabelle Bril, 27–50. Studies in Language Companion Series 121. Amsterdam and Philadelphia: John Benjamins.

Foley, William A. and Robert D. Van Valin, Jr. 1984. *Functional syntax and universal grammar*. Cambridge Studies in Linguistics 38. Cambridge: Cambridge University Press.

Givón, T. 2001. *Syntax: An Introduction*. Amsterdam and Philadelphia: John Benjamins.

Haiman, John. 1985. Symmetry. In *Iconicity in syntax: Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983*, ed. by John Haiman, 73–95. Typological Studies in Language 6. Amsterdam and Philadelphia: John Benjamins.

Halliday, M. A. K. 1994. *An introduction to functional grammar*. London: Arnold. 2nd edition

Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms—Adverbial participles, gerunds*, ed. by Martin Haspelmath and Ekkehard König, 1–55. Empirical Approaches to Language Typology 13. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2004. Coordinating constructions: An overview. In *Coordinating constructions*, ed. by Martin Haspelmath, 3–39. Typological Studies in Language 58. Amsterdam: John Benjamins.

Hsieh, Fuhui. 2012. Multi-verb Constructions in Paiwan. Paper presented at the 12th International Conference on Austronesian Linguistics (12-ICAL), 2–6 July. Bali, Indonesia: Udayana University.

Kuno, Susumu. 1973. *The strucutre of the Japanese language*. Cambridge, MA: MIT Press.

Lehmann, Christian. 1988. Towards a typology of clause linkage. In *Clause combining in grammar and discourse*, ed. by John Haiman and Sandra A. Thompson, 181–225. Typological studies in language 18. Amsterdam and Philadelphia: John Benjamins.

Lyons, John. 1968. *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press.

Ogawa, Naoyoshi and Erin Asai. 1935. *Gengo ni yoru Taiwan takasago-zoku densetsu-shu* [*The myths and traditions of the Formosan native tribes (texts and notes*)]. Taihoku [Taipei]: Taihoku Imperial University. [In Japanese]

Radden, Günter and René Dirven. 2007. *Cognitive English grammar*. Cognitive Linguistics in Practice 2. Amsterdam and Philadelphia: John Benjamins.

Talmy, Leonard. 2000. *Toward a cognitive semantics, Vol. 1: Concept structuring systems*. Cambridge, MA: MIT Press.

Tang, Chih-Chen Jane. 1999. On clausal complements in Paiwan. In *Selected papers from the eighth international conference on Austronesian linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 529–578. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

Tang, Chih-Chen Jane. 2011. Comitative vs. conjunctive constructions in Paiwan. *Language and Linguistics* 12.1:141–170.

Tikkanen, Bertil. 1995. Burushaski converbs in their South and Central Asian areal context. In *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms—adverbial participles, gerunds*, ed. by Martin Haspelmath and Ekkehard König, 487–528. Empirical Approaches to Language Typology 13. Berlin: Mouton de Gruyter.

Traugott, Elizabeth Closs. 1985. Conditional markers. In *Iconicity in syntax: Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983*, ed.

by John Haiman, 289–307. Typological Studies in Language 6. Amsterdam and Philadelphia: John Benjamins.

Van Valin, Robert D. Jr. and Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

14 *Reported evidentials in Saaroa, Kanakanavu and Tsou**

CHIA-JUNG PAN

1 Introduction

1.1 Evidentiality

In Aikhenvald's (2004:17) monograph *Evidentiality* (with examples drawn from over 500 languages), evidentiality is found in about a quarter of the languages of the world. Evidential systems are widespread in South American and North American Indian languages, in the languages of the Caucasus, and in the languages of the Tibeto-Burman family. This can be achieved by **evidentials** or by other means (i.e., **evidentiality strategies**), such as reported speech, evidential-like markers and lexical expressions of perception and cognition.[1] This study will only discuss evidentiality as a grammatical category, and will not mention evidential strategies.

There are some semantic parameters in languages with grammatical evidentiality. Aikhenvald (2004:63–64, 2014:9) states that the recurrent semantic parameters are:

I. VISUAL: covers information acquired through seeing.
II. NON-VISUAL SENSORY: covers information acquired through hearing, and is typically extended to smell and taste, and sometimes also to touch.

* First of all, this paper is dedicated to Prof. Huang for her achievements on Formosan languages. I am grateful to Alexandra Aikhenvald for bringing 'evidentiality' to my attention while she was supervising my PhD thesis, and for providing me with a number of valuable comments on the Saaroa language. I would like to thank R.M.W. Dixon, Jackson Sun, Elizabeth Zeitoun, Henry Y. Chang, Stacy F. Teng, and Hsiu-chuan Liao for their discussions and comments. I am grateful to my Saaroa-, Kanakanavu- and Tsou-language consultants for helping me learn their languages, and thankful to the participants of the workshop on "A typological study of Austronesian languages in Taiwan and their revitalisation" held at the Institute of Linguistics, Academia Sinica, Taiwan, 1–2 December 2012, for their suggestions. I would like to thank the Formosan Language Archive Project, Academia Sinica, directed by Elizabeth Zeitoun, for providing me with a research grant to conduct fieldwork. I am indebted to four anonymous reviewers and the editors of this volume for their valuable suggestions and comments. All remaining errors are mine.

[1] Having lexical means for optional specification of the source of knowledge is probably universal (Aikhenvald 2003:1).

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 341–362
Asia-Pacific Linguistics, 2015

III. INFERENCE: based on visible or tangible evidence, or result.
IV. ASSUMPTION: based on evidence other than visible results: This may include logical reasoning, assumption, or simply general knowledge.
V. HEARSAY: for reported information with no reference to those it was reported by.
VI. QUOTATIVE: for reported information with an overt reference to the quoted source.

Aikhenvald (2004:25) classifies languages consisting of evidentials into numerous subtypes based on how the information source acquires distinct grammatical marking. Some languages distinguish just two choices, while others three, four, or five choices. Systems with just two choices cover:

A1. Firsthand and non-firsthand
A2. Non-firsthand versus 'everything else'
A3. Reported (or 'hearsay') versus 'everything else'
A4. Sensory evidence and reported (or 'hearsay')
A5. Auditory (acquired through hearing) versus everything else

1.2 Motivation of the present study

This study is motivated by the following. First, evidentiality in Formosan languages has not been thoroughly and systematically investigated yet. This is because typologically, evidentiality is not a salient grammatical feature in the grammar of Formosan languages. Previous works dealing with this (or a similar) topic include for Tsou, Yang (2000a, 2000b); for Paiwan, A. Chang (2012); for Kavalan, Hsieh (2012); and for Saaroa, Pan (2014). Second, while there is increasing tendency to treat evidentiality as a grammatical category, as in Aikhenvald (2004, 2014), previous works in Formosan languages, except for Pan (2014) on Saaroa, have focused on evidentiality strategies. Pan's (2014) study, for instance, investigates the expression of knowledge through evidentials and other means via evidentiality strategies. It is found that Saaroa has a closed grammatical category, evidentiality. In addition, evidentiality in Saaroa can be expressed through other means, including reported speech, evidential-like meanings through special markers, and verbs of perception.

In this study, 'evidential' is treated as a term principally relating to 'information source' as a closed grammatical system, so that all the other means through evidentiality strategies will be excluded. In addition, two research issues which are typologically important in evidentiality studies will be included and discussed. First, in Pan (2014), when discussing evidentials as a grammatical category, the author only pays attention to grammatical expressions and semantics. However, according to Aikhenvald (2004), a grammatical category of evidentiality interacts with different clause types, because evidentials used in questions and commands differ in their semantics and pragmatic connotations from their counterparts in declarative clauses. Second, in languages with evidentiality as part of grammar, evidentials are linked with different sorts of knowledge and experience, together with conventionalised ways of talking about them. Thus, evidentials have their own stereotyped evidential uses, e.g., the way of talking about dreams and new cultural innovations.

At the moment, there are sixteen officially recognised indigenous languages in Taiwan. In order to manage the scope of this paper, three languages are chosen—Saaroa, Kanakanavu, and Tsou—as the sample of evidentiality as a grammatical category. These three languages are selected not only because I am familiar with them, but also for two other important reasons. First, these three languages are on the verge of extinction, as shown in Table 1. Data collection can contribute to other linguistic studies and language revitalisation. Second, in these three languages, expression of 'reported' knowledge is often connected with evidentiality which is attested as a grammatical category in its own right. Sentences may contain an indication of how the information was acquired by the speaker: whether they heard it, or know about it because somebody else told them, etc.

Table 1 presents some general background of Saaroa, Kanakanavu, and Tsou.

Table 1: General background of Saaroa, Kanakanavu, and Tsou

	Saaroa	Kanakanavu	Tsou
Location	Taoyuan District and Namasia District, Kaohsiung City, Taiwan	Namasia District, Kaohsiung City, Taiwan	Alishan Township, Chiayi County and Xinyi Township, Nantou County, Taiwan
Population	Approximately 400	Approximately 500	Approximately 4000
Fluent speakers	less than 10	less than 10	Unknown
Language proficiency	The older they are, the better language proficiency they have	The older they are, the better language proficiency they have	The older they are, the better language proficiency they have
Language use	Old and native speakers: Saaroa/Bunun > Japanese/Mandarin	Old and native speakers: Kanakanavu/Bunun > Japanese/Mandarin	Old and native speakers: Tsou > Japanese/Mandarin
	Young and semi-speakers: Mandarin > Bunun > Saaroa	Young and semi-speakers: Mandarin > Bunun > Kanakanavu	Young and semi-speakers: Mandarin > Tsou
Dialects	No	No	Iimucu (extinct), Tfuya, Tapang~~u~~, Luhtu

2 Morphosyntactic characteristics of reported evidentials

Saaroa, Kanakanavu and Tsou have a system with two terms, i.e., two choices (A3), distinguishing reported (or 'hearsay') evidentials from 'everything else' by grammatical marking.[2] The reported term is marked, and the non-reported ('everything else') term is

[2] The meaning of the term 'reported' is verbal report; alternative terms are 'hearsay' or 'quotative'. However, 'quotative' should be reserved for a reported evidential involving exact indication of the person who provided the information. In this study, 'reported' will be used, since in Saaroa, Kanakanavu, and Tsou reported information typically does not involve an overt reference to those it was reported by.

unmarked. An analytic problem that arises in Saaroa, Kanakanavu and Tsou is whether the unmarked form should be considered a term in the system. Aikhenvald (2004:73) takes the position that "an informed decision between these alternatives can be made only on language-internal grounds. If zero marking has a specific semantic connotation, then it should be looked upon as part of the system." In the three languages, the absence of an evidential marker only tends to represent knowledge that the speaker is sure of, or to imply that the speaker was an eyewitness of the event. Based on Aikhenvald's (2004:73) typological generalisation, zero-marking (i.e., non-reported 'everything else') is not considered as a special term in the evidential system of Saaroa, Kanakanavu and Tsou.

Reported evidentials in Saaroa, Kanakanavu, and Tsou signal the way how the information is obtained, typically through someone else's narration. When reported evidentials are used, they imply that the speaker's contribution to the precise information source has been made. The use of the reported evidential is regarded as a way of shunning possible miscommunications.

(1) Saaroa[3] (Pan 2014:97)

akuisa	*lh‹um›ivuru*	*civuka-isa,*	*m-utu-pulhu=ami*
when	<AV>stab	belly-3.GEN	AV-move.toward-come.out=EVID
a	*ma~maini-isa,*	*rian~~u~~=ami*	*al~~u~~m~~u~~lh~~u~~.*[4]
CORE	RED~small-3.GEN	all=EVID	wild.boar

'When (he) stabbed her belly, they said that her children came out, and all (children turned out to be) wild boars.'

(2) Kanakanavu (Tsuchida 2003:16)

para-tea-nené	***=kani***	*=kiái*	*sua=*	*sanapisápi.*
scoop(PF)	it-is-said	her	NOM	driftwood

'She scooped a driftwood.'

---

[3] Examples of Saaroa all come from my fieldwork, mostly based on texts from narratives with a few examples based on limited corroborative grammatical and lexical elicitation. Examples drawn from Kanakanavu and Tsou texts are based on Tsuchida (2003) and Tung (1964), respectively. Examples of Kanakanavu and Tsou without due acknowledgment are from my own fieldwork. Tsou examples based on Tung (1964) are re-analysed by the author. Interlinear glossing of Kanakanavu examples based on Tsuchida (2003) is retained. Though Tsuchida's Kanakanavu texts were published in 2003, they were collected in 1969. Unlike Tsou, forty five years after the data was collected, Kanakanavu texts cannot be fully understood by any Kanakanavu speaker. In order to avoid revising the data, interlinear glossing of Kanakanavu examples is retained.

[4] Abbreviations in this study follow the Leipzig Glossing Rules. Those which do not appear there are as follows: ACHI: achievement marker, AF, Actor focus; ASP, aspect; AV, Actor voice; CORE, core case; COS, change of state; EVID, evidential; FILL, pause filler; INDEP, independent; INFER, inference; LINK, linker; LOC, location; LV, locative voice; MIR, mirative; NAV, non-Actor voice; PF, patient focus; PN, pronoun; PR, polite request; PV, patient voice; REAL, realis; RED, reduplication; TEMP, temporal.

Orthography follows IPA except where *'* = glottal stop, *c* = unaspirated voiceless alveolar affricate, *e* = schwa (in Kanakanavu only), *ng* = velar nasal, *l* = alveolar flap, *lh* = alveolar lateral fricative, *r* = alveolar trill, ~~*u*~~ = high central unrounded vowel, and *:* = long vowel.

(3) Tsou (Tung 1964:260)
moso ***nana*** *eoh~~u~~* *na* *eaazuon~~u~~.*
AV.REAL EVID hunt NOM Eaazuon~~u~~.people
'The Eaazuon~~u~~ people went hunting.'

Reported evidentials are expressed by *=ami* in Saaroa, *=kani* in Kanakanavu, and *nana* in Tsou. Their morphosyntactic properties of expression, multiple occurrences in the same clause, and scope will be discussed in the following subsections.

2.1 Expressing the reported evidential

The reported evidential in Saaroa is *=ami*. It is not selective to its host, and encliticises to the first constituent of the clause. As shown in examples (4a)–(4f), the reported evidential *=ami* attaches to different types of host, e.g., the dynamic verb *m-u-sala* 'go/walk', the stative verb *ma-aru* 'exist', the conditional subordinator *maaci* 'if/when', the negator *uka'a*, the 'concerning' clause *maacu*, and the quantifier *rian~~u~~* 'all'. Another piece of evidence in favour of the enclitic analysis is that if *=ami* were an independent word, both the constituent preceding *=ami* and *=ami* itself should bear their own stress respectively. However, this is not the case.

(4) Saaroa

a. *m-u-sala=**ami*** *a* *kana* *cucu-isa=na*
AV-motion.on.foot-road=EVID CORE FILL person-3.GEN=DEF
puaili *k‹um›ita* *ta'alhara-isa.*
return <AV>look chicken.coop-3.GEN
'It is said that their people returned to guard their chicken coop.'

b. *maaci* *ka* *kiariari* *a* *ucani* *ka* *uka'a=mana*
if/when LINK past LINK one LINK NEG=IPFV
ka *kiariari* *lhi-pu'a* *na* *kana* *sikam~~u~~* *ia*,
LINK past PFV-buy OBL FILL mat TOP
ma-aru=ami *a* *tapa~~u~~=na* *m~~u~~u~~u~~m~~u~~a* *tu-sikam~~u~~.*
AV-exist=EVID CORE shell.flower=DEF also make-mat
'When (people) had no money to buy a mat in the past, it is said that the Shell flower could be used to make a mat.'

c. *maaci=**ami*** *utulu=cu* *vulalh~~u~~* *ia,* *um-a-urapi=cu.*
if/when=EVID three=COS moon/month TOP AV-IRR-sow.seed=COS
'It is said that when (it is) March, (we) sow seeds.'

d. *uka'a=cu=**ami*** *ka* *vutukulhu* *m-aa* *isana.*
NEG=COS=EVID CORE fish AV-be:LOC/TEMP there
'It is said that there is no fish over there.'

e. *maacu=**ami*** *alham~~u~~* *a* *kani'i* *ia,*
concerning=EVID bird LINK this TOP

m-aa *n* *kani'i* *mapulhar~~u~~.*
AV-be:LOC/TEMP OBL this a.flat.land.of.low.altitude
'It is said that concerning this (type of) bird, (it perches) at a flat land of low altitude (tableland).'

f. *rian~~u~~=**ami*** *tam* *sa'au* *ka* *viaru=na* *maaci*
all=EVID very tasty CORE corn=DEF if/when
avava *alha* *capa*
boil(PV) DISJ.COOR broil(PV)
'It is said that the corns are all very tasty if (they are) boiled or broiled.'

In Kanakanavu, the reported evidential is expressed via encliticising *=kani*. Similar to Saaroa *=ami*, Kanakanavu *=kani* usually attaches to the right of the first clausal element. As exemplified in (5a–i), the reported evidential *=kani* attaches to different types of host, e.g., the dynamic verb *ala-ene* 'take', the irrealis marker *tia*, the negator *ka'ane* 'not', the existential word *'aisi*, the locational expression *'aravange* 'inside', the adverbial temporal expression *maka-asua* 'then', the expression of temporal duration *si-pitu-unu* 'seven nights', the numeral word *neeme* 'six', and the case marker *sua*.

(5) Kanakanavu (Tsuchida 2003)

a. *ala-ené* *=kani* *=kiái* *t‹um›áini,* *valúvalu* *canúmu.*
take(PF) it-is-said her throw-away rapids water
'She picked it up and threw it away in the rapids of water.' (p. 16)

b. *tia=* *=kaní* *=cu* *mu-á-ca* *mata* *valí* *=ini.*
fut. it-is-said already leave with friend his
'He was going to go with his friend.' (p. 22)

c. *ka'áne* *=kaní* *=pa* *ma-aránge.*
not it-is-said still old.
'It was not ripe yet.' (p. 43)

d. *'áisi* ***=kani*** *ukula-tumúlu* *sua=* *nungúnungu.*
be it-is-said full NOM creek
'The creeks were full of water.' (p. 63)

e. *sua=* *ta-u-kusa-á* *=ini* *ta-'áisi-a* *taru'án* *=ini* *=ia,*
NOM place-go-toward their place-stay mountain-hut their TOP,
'arávange ***=káni*** *váatu.*
inside it-is-said stone
'Their temporary dwelling in the mountains in that area was the inside of stones (i.e., a stone cave).' (p. 81)

f. *maka-ásua* ***=kaní*** *=cu* *sua=* *tamú* *=ini,* *"náai,*
then it-is-said already NOM grandfather his well

ki-tanam-ái *ki-suanái* *maánu* *íisi,"* *mi-sá* *=kani.*[5]
ask-try ask(AF) child this say it-is-said
'Then his grandfather said, "Well, try to ask this child."' (p. 18)

g. *si-pitu-unu* ***=káni*** *ara-pite'é* *=kani* *sua=* *ka-kaángeca.*
seven-nights it-is-said become-dark it-is-said NOM skies
'On the seventh day the skies became dark.' (p. 88)

h. *néeme* ***=kani*** *nganái* *=ini.*
six it-is-said name his
'His name was Neme (six).' (p. 92)

i. *sua=* ***=kani*** *naa-néeme* *=ia,*
NOM it-is-said ex-Neme TOP,
ni-m-ú-usa *=kani* *napálanga* *m-u-cúmuku* *m-u-'aravánge.*
went-toward it-is-said Napalanga go-ripened enter-house
'Neme went to Napalanga to enter a family as an adopted son.' (p. 92)

In Tsuchida's Kanakanavu texts, the reported evidential may also attach to the second element in a clause.

(6) Kanakanavu (Tsuchida 2003:25)
'áisi *paati-'uri-'uríngi* ***=kani*** *arupa-a-'uri-'uríngi.*
be carry-fire it-is-said each-other-carry-fire
'People were carrying fire to each other.'

The factor in determining whether the reported evidential attaches to the first or second clause element is not clear in Tsuchida's Kanakanavu texts (2003). At the moment of writing, the Kanakanavu language consultant rejects the attachment to the second clausal element.

(7) Kanakanavu
* *tavar~~u~~'~~u~~=ku* *m-asip~~u~~c~~u~~=**kan(i)*** *tikuru.*
able=1SG.NOM AV-wash=EVID clothes
'It is said that I am able to wash clothes.'

In Tsou, the reported evidential is expressed through *nana*. Every verbal clause of Tsou begins with an auxiliary, which denotes the voice and reality status of the clause.

(8) Tsou (Pan 2010:200)
te-c'u *auesiesi* *eon* *to* *pnguu* *'o* *voyu.*
IRR-ASP forever(AV) live(AV) OBL place.name NOM male.name
'Voyu will live in Pnguu forever.'

The reported evidential typically occurs between the preverbal auxiliary and the verb.[6] In example (9a), the reported evidential *nana* appears between the auxiliary *moso* and the verb *teel~~u~~* 'on time'. In example (9b), it occurs between the auxiliary *moh* and the

[5] *=kani* may occur more than once in some sentences. The multiple occurrences develop additional semantic overtones (section 2.2).

[6] Though some optional elements may occur next to *nana*, e.g., a negator and an aspectual marker, *nana* still occurs between the preverbal auxiliary and the verb.

verb *cm~~u~~'ho* 'return' in the subordinate clause, and appears between the auxiliary *mi* and the verb *mitungucu* 'sacrifice' in the main clause.

(9) Tsou (Tung 1964)

a. *moso* ***nana*** *o'te* *teel~~u~~* *ho* *moso* *mitungucu.*
AV.REAL EVID NEG on.time when AV.REAL sacrifice
'(They) were not on time when (people at home) sacrificed.' (p. 260)

b. *ho* *moh* *cu* ***nana*** *cm~~u~~'ho* *ho* *mo* *eoh~~u~~,*
when AV.REAL ASP EVID return when AV.REAL hunting
mi *cu* ***nana*** *so* *mitungucu.*
AV.REAL ASP EVID already sacrifice
'When (they) returned when (they) went hunting, (people at home) already sacrificed.' (p. 260)

The reported evidential *nana* is not an enclitic or suffix, in that it always bears its own stress. Example (10) shows that both the auxiliary *moso* and the evidential *nana* bear their own stress, respectively. If *nana* were an enclitic or suffix, *moso* would not have its own stress.

(10) Repeated from (9a)
móso ***nána*** *ó'te* *teél~~u~~* *ho* *móso* *mitungúcu.*
AV.REAL EVID NEG on.time when AV.REAL sacrifice
'(They) were not on time when (people at home) sacrificed.'

2.2 Occurring more than once in the same clause

In Saaroa, Kanakanavu and Tsou, the reported evidential indicates the source of information, when it occurs once in a clause. Examples consisting of the reported evidential once per clause are shown below.

(11) Saaroa (Pan 2014:97)
*um-a-c~~u~~ka=**ami*** *ka* *kana* *tamu-isa* *lhi-calhum-a*
AV-IRR-get.up=EVID CORE FILL grandparent-3.GEN PFV-bury-PV
n *kana* *kalalhapa* *taar~~u~~* *rumalha~~u~~,* *k‹um›a~kita=**ami***
OBL FILL under bed when <AV>IRR~look=EVID
ka *kana* *lhamu-isa* *kana* *tamu-isa.*
CORE FILL grandchild-3.GEN FILL grandparent-3.GEN
'It is said that when their grandparent who have been buried under a bed get up, their grandchildren will guard their grandparent.'

(12) Kanakanavu (Tsuchida 2003:18)
mata-vecekái *=**kani*** *si-puu-mua-muár* *=ini* *'ecép* *=ini,*
go-middle it-is-said things-talked his dream his
pa-tiul-ái *=**kani*** *sa=* *tamú* *=ini* *sua=* *ivíci* *maánu.*
make-stop-talk it-is-said OBL grandfather his NOM mouth child
'When he talked half of his dream, his grandfather covered the mouth of the child with his hand to stop him talking.'

(13) Tsou (Tung 1964:260)

ho	*moh*	*cu*	***nana***	*fengna*			
when	AV.REAL	ASP	EVID	evening			
pan	*no*	*moso*	***nana***	*aoko*	*mongsi*	*ci*	*oko.*
exist	OBL	AV.REAL	EVID	incessantly	cry	REL	child

'When (it) was evening, there was an incessantly crying child.'

However, in the three languages the reported evidential may occur more than once in the same clause. The multiple occurrences of reported evidentials acquire additional semantic overtones. For example, in (14), the reported evidential occurs more than once: first on the first constituent of the clause *kiariari* 'past' and then on the verbs *k‹um›a-kalii* 'dig' and *a-calhumu* 'bury'. The reported evidential on the first constituent of the clause indicates the source of information, and those on the verbs acquire additional emphatic overtones.

(14) Saaroa (Pan 2014:95)

maaci	*m-iungu=cu*	*aari-isa*	*ia,*	*kiariari=**ami***	*ka*	
if/when	AV-arrive=COS	day-3.GEN	TOP	past=EVID	LINK	
*k‹um›a~kalii=**ami***	*n*	*kalalhapa*	*taare*	*a-calhumu-isa=**ami***		
<AV>IRR~dig=EVID	OBL	under	bed	IRR-bury-3.GEN=EVID		
kalalhapa	*taar~~u~~.*					
under	bed					

'If his time had come, it is said that in the past they would dig (the ground) under a bed and bury (the body) under the bed.'

In Kanakanavu, the multiple occurrences in the same clause can be found on verbal constituents occurring in succession, as shown in (15).

(15) Kanakanavu (Tsuchida 2003:59)

cu-cúlu	*=**kani***	*m-aka-cangí*	*=**kani***	*sua=*	*canúmu*
true	it-is-said	do-soon	it-is-said	NOM	water
m-ata-nakáre	*=**kani,***	*sua=*	*'aapacú='ai*	*na=*	
get-dry	it-is-said,	NOM	even	at	
ni-ala-túmulu	*canúmu.*				
much	water				

'Water soon got dried really even at the deep place.'

When there are two reported evidentials in a clause, the reported evidential attaching to the first element of the clause indicates that the information is acquired from someone else, and the one occurring in the modifying phrase develops an additional emphatic connotation.[7]

[7] An anonymous reviewer suggests that the multiple evidentials *=kan(i)* in Kanakanavu could be a case of evidential semantic harmony. The emphatic or reinforced hedging interpretation arises from the repeated marking of evidential *=kan(i)*. Due to the limitations of time and space, I leave this research question open for further investigation.

(16) Kanakanavu (Tsuchida 2003)

a. *'áisi* ***=kani*** *t‹in›ú-puru* *='inía* *sua=* *cáau*
be it-is-said sat there NOM man
[*ni-maru-mánenge* ***=kani***]$_{VP}$.
became-big-good it-is-said
'A **fine-built** man was sitting there.' (p. 44)

b. *'áisi* ***=kani*** *ni-m-i-ta-tíini* *='inía* *tavenevéne*
be it-is-said hooked there banana
[*u-rúcini* ***=kaní*** *=cu*]$_{NUMP}$.
two it-is-said already
'There were **two** bananas hooked.' (p. 51)

Unlike Saaroa and Kanakanavu (i.e., reported evidentials occurring on verbal constituents in succession), Tsou allows one reported evidential on the verb and the other on the NP. The multiple occurrences of the Tsou reported evidential enhance the speaker's objectivity and unwillingness to vouch for the information.

(17) Tsou (Yang 2000a:83)
sua *ho* *mi-'o* ***nana*** *bon~~u~~* *no* ***nana*** *eosk~~u~~-mu.*
dream that Aux:AF-1SG EVI eat OBL EVI fish-your
'I dreamed that I ate your fish.'

2.3 Scope

The reported evidential itself cannot fall within the scope of negation. In a sentence consisting of a reported evidential and a negator, the negator negates the whole proposition, but it does not negate the information source.

In (18), the evidential enclitic is added to a negator, meaning '(They) went to that place where goats gathered but couldn't find it, I heard'.

(18) Saaroa (Pan 2014:94)
ku=ami *pai-ta~t~~u~~alh~~u~~* *isana* *m-u-sala*
NEG=EVID find-RED~ACHI 3.INDEP.PN AV-motion.on.foot-road
m-aki‹ka›kua *n* *kana* *'u~'ukui-a.*
AV-action.toward.location<RED> OBL that RED~goat-LOC.NMLZ
'It is said that (they) went to that place where goats gathered but couldn't find it.'

In (19), the evidential enclitic attaches to a negator, meaning 'That woman didn't sleep, I heard'.

(19) Kanakanavu
*kuu=****kan(i)*** *si-'ic~~u~~p~~u~~* *nanak~~u~~* *i:sa.*
NEG=EVID sleep-dream woman that
'It is said that that woman didn't sleep.'

In (20), the reported evidential appears in a negative sentence, meaning 'They didn't ever see, I heard'.

(20) Tsou (Tung 1964:260)

o'a	*i-he*	***nana***	*aht-a*	*teol~~u~~-i.*
NEG	NAV.REAL-3PL	EVID	ever-PV	see-LV

'They did not ever see.'

When negating the information source, one has to adopt an evidentiality strategy through lexical means, e.g., verbs of auditory perception 'hear'.

In Saaroa, for example, to say 'Ancestors had a love affair with a wild boar, I didn't hear', one has to use the verb *timalha* 'hear'.

(21) Saaroa (Pan 2014:96)

ku	*lhi-**timalha**-ku*	*na*	*alhaama*	*kiariari*	*n*
NEG	PFV-hear(PV)-1SG.GEN	OBL	ancestor	past	LINK
kana	*m-uritalhiva~~u~~*	*n*	*al~~u~~m~~u~~lh~~u~~.*		
FILL	AV-have.a.love.affair	OBL	wild.boar		

'I didn't hear that ancestors had a love affair with a wild boar.'

In Kanakanavu, for example, to say 'This woman was netting fish, I didn't hear', one has to use the verb *timana* 'hear'.

(22) Kanakanavu

kuu=ku	*t‹um›a~**timana***	*'e:si*	*l‹um›a~liu'u*
NEG=1SG.NOM	<AV>RED~hear	PROG	<AV>RED~fish.by.scoopnet
sua	*nanak~~u~~*	*i:si.*	
NOM	woman	this	

'I didn't hear that this woman was netting fish.'

In Tsou, to say 'They said Pasuya is sick, I didn't hear', one has to employ the verb *tac'~~u~~h-i* 'hear'.

(23) Tsou

o'a	*os-'o*	***tac'~~u~~h-i***	*ho*	*i-he*	*yainca*
NEG	NAV.REAL-1SG	hear-LV	that	NAV.REAL-3PL	say
mo	*tma'congo*	*'o*	*pasuya.*		
AV.REAL	sick	NOM	male.name		

'I didn't hear that they said Pasuya is sick.'

3 Mirative overtones

Mirativity covers speaker's unprepared mind, unexpected new information and concomitant surprise, and is often connected with evidentiality (Aikhenvald 2004:195; DeLancey 1997). In Saaroa, the reported evidential *=ami* usually attaches to the right of the first constituent in a clause, as mentioned in section 2.1. When the reported evidential *=ami* is not attaching to the right of the first constituent in a clause, it may acquire an overtone of mirativity.[8] Example (24) shows that the reported evidential is encliticised to

[8] Tsou and Kanakanavu are not discussed here, because similar phenomenon is not attested in these two languages. Mirativity in these two languages are expressed via other particles, rather

the locational noun *vuvulungaa* 'mountain' which occurs in the clause-final position. Those people who are not living in the mountains stun the speaker, because they don't know kiwi fruit exists in the mountains.

(24) Saaroa (Pan 2014:95)
ku *tararian~~u~~* *kana* *cucu=lika'a* *kaaiu* *ma-aru*
NEG hear that person=outside far.there AV-exist
luuvi-ta *vuvulungaa=**ami.***
kiwi.fruit-1PL.INCL.GEN mountain=EVID
'Those outsiders didn't hear that we have kiwi fruit in the mountain!' (uttered with amazement by the speaker).

4 The reported evidential and person

4.1 Nature of perceiver

Reported evidentials in Saaroa, Kanakanavu and Tsou usually occur in sentences containing a 'third' person participant in S or A function.[9] The 'third' person participant can be explicitly or implicitly specified. However, the 'third' person participant is not the perceiver or transmitter of information. In default use, reported evidentials always presuppose that the speaker is the perceiver or transmitter of information.

In examples (25)–(27), reported evidentials occur in sentences with an explicitly specified 'third' person participant. In examples (28)–(30), reported evidentials occur in sentences with an implicitly specified participant (which can be retrieved from the context), shown in boldface in English translation.

(25) Saaroa (Pan 2012:76)
[***saa-***]$_A$*pala-va~vililh-a=**ami***
3.GEN-stealthily.follow-RED~stealthily.follow-PV=EVID
ka *kana* *cucu* *salia-isa* *k‹um›ita* *aunaana=iau*
CORE FIL person house-3.GEN <AV>look/see like.that=MIR
rumalha~~u~~ *m-uritalhiva~~u~~* *na* *al~~u~~m~~u~~lh~~u~~**.***
when AV-have.a.love.affair OBL wild.boar
'It is said that **he** stealthily followed the person to her house and had a look. Like that, (he saw her) have a love affair with a wild boar.'

(26) Kanakanavu (Tsuchida 2003:23)
tara-kari-ái *=**kani*** [*=**'inía***]$_A$ *sua=* *tamú* *=ini*.
ask it-is-said NOM grandfather his
'**He** asked his grandfather.'

---

than evidentiality. Due to the limitations of space and time, mirativity in these three languages deserve another research paper.

[9] In Basic Linguistic Theory, S refers to intransitive subject, and A transitive subject (Dixon 2010:116).

(27) Tsou (Tung 1964:262)
mo ***nana*** *mieboc~~u~~* [*na* ***meefucu***]$_S$.
AV.REAL EVID break.wind NOM Dwarf
'The **Dwarf** broke wind.'

(28) Saaroa (Pan 2014:99)
*puaili=cu ka lhalhusa, m-ia‹laa›lang~~u~~ m-ian~~u~~=**ami.***
return=COS CORE man AV-prepare<RED> AV-pound=EVID
'After the men returned (home), it is said that (**they**) prepared to pound rice.'

(29) Kanakanavu (Tsuchida 2003:22)
*ala-la'únusu sua= íisua.'áisi =**kani** =cu ki-a-talísi.*
finish NOM that. be it-is-said already make-rope
'That ended. (**He)** was making a rope.'

(30) Tsou (Tung 1964:262)
i-si ***nana*** *cohiv-i no mamespingi*
NAV.REAL-3SG.GEN EVID know-LV GEN woman
ho te-si t~~u~~tp~~u~~t-a no meefucu. mi cu
that IRR-3SG.GEN seize-PV GEN Dwarf AV.REAL ASP
nana *aoko eoho~~u~~.*
EVID incessantly yell
'The woman knew that (**she**) would be seized by the Dwarf. (**She**) incessantly yelled.'

4.2 'First person' effects

Evidentials may acquire additional semantic overtones when they co-occur with a first person participant (Aikhenvald 2004:219). The additional semantic expansions are **'first person' effects**. Aikhenvald (2004:225–226) exemplifies some of the semantic expansions. A reported evidential with first person implies (i) overtones of lack of control and subsequent surprise on behalf of the speaker,[10] (ii) overtones of irony, surprise, and disagreement with what was said about the speaker,[11] and (iii) the speaker does not remember what s/he said.[12]

In Saaroa, when the speaker uses the reported evidential in talking about himself, the utterance acquires additional semantic overtones consistent with first person effects such as are mentioned above. Example (31) shows that the speaker was **stunned** because he was required to tell a story about traditional Saaroa marriage.

[10] It has been reported in Estonian (Aet Lees, pers. comm. with Aikhenvald).

[11] It has been reported in Shilluk (Miller & Gilley 2007:14–15).

[12] It has been reported in Tucano (Ramirez 1997, vol. I:142), Shipibo-Konibo (Valenzuela 2003:41), Eastern Pomo (Sally McLendon, pers. comm. with Aikhenvald), Wintu (Schlichter 1986:49), and Wichita (Rood 1976:93).

(31) Saaroa (Pan 2014:99)
*ki-a-lha~lhamu=**aku**=**ami*** *kani'i* *lhalhusa* *maaci*
tell-IRR-RED~tell=1SG.NOM=EVID this man if/when
um-aala *n* *alhaina.*
AV-take OBL woman
'I am going to tell this (story) about men marrying women.'

In addition, the reported evidential in Saaroa can have strong overtones of **irony and disagreement** with what was said about the speaker. Example (32) from Saaroa implies the speaker's vehement disagreement with what was said about him.

(32) Saaroa
*ku=**aku**=**ami*** *palhu-salhi.*
NEG=1SG.NOM=EVID sing-song
'It is said that I did not sing. (at the gathering when we practiced singing our ritual songs—this is absolutely not true)'

In Kanakanavu, the reported evidential co-occurs with the 'first' person (I) when the speaker 'I' (i) was drunk, (ii) had to learn the facts about himself from someone else's report, or (iii) has the disease of sleepwalking or dementia. In other words, the reported evidential may be used with first person if **the speaker does not remember what s/he did**.

(33) Kanakanavu: fact about the speaker from someone else's report
*una=**ku**=**kan(i)*** *taciin* *kanarua* *nanak~~u~~.*
exist=1SG.NOM=EVID one younger.sibling woman
'It is said that I have one younger sister.'

(34) Kanakanavu: sleepwalking
*paira=**ku**=**kan(i)*** *mocaaca.*
often=1SG.NOM=EVID walk
'It is said that I often walk.'

(35) Kanakanavu: dementia
*tavar~~u~~'~~u~~=**ku**= **kan(i)*** *m-asip~~u~~c~~u~~* *tikuru.*
able=1SG.NOM=EVID AV-wash clothes
'It is said that I am able to wash clothes.'

5 Correlations of reported evidentials and clause types

More evidential choices can be attested in statements than in any other clause type in a majority of languages (Aikhenvald 2004:242). In the case of questions and commands, the use of the reported evidential in Saaroa, Kanakanavu and Tsou may reflect the information source of the speaker or a third party, and carry different additional overtones.

5.1 Questions

Both Saaroa and Tsou can use the reported evidential in questions. When this occurs, the evidential can have additional semantic overtones. In Saaroa, examples (36a) and (36b) show that the speaker is speaking **on behalf of someone else** to direct the question to the addressee (second person), but its origin is not within the speech act situation; that is, the question's author comes from a third party.

(36) Saaroa

a. *ngasa=**ami*** *ngalha-u*?
what=EVID name-2SG.GEN
'(He or she wants to know) what's your name?'

b. *um-a-usalh~~u~~=i=**ami*** *maataata*?
AV-IRR-rain=Q=EVID tomorrow
'(He or she wants to know) will it rain tomorrow?'

In Saaroa, the reported evidential in questions may carry other additional overtones. In example (37), the reported evidential implies that the speaker is **puzzled** by something overheard or unsure of the information, and expects **further discussion or clarification** of the matter.[13]

(37) Saaroa

*cu-lhaumangi=u=**ami*** *m-u-sa* *taipak~~u~~*?
IRR-when=2SG.NOM=EVID AV-motion.on.foot-road Taipei
'When will you go to Taipei?'

In Tsou, the reported evidential in questions implies that the speaker is questioning some information that s/he has been given. Using the reported evidential just enhances the speaker's uncertainty, so an answer is not expected to be provided by the hearer. This can be treated as a kind of **rhetorical question** in Tsou. (There are other types of rhetorical questions as well)

(38) Tsou (Yang 2000a:79)

te ***nana*** *mcoi* *'o* *ihe* *tiuna*?
FUT EVID die NOM 3.SG beat
'Will the person who was beaten die?'

In Tsou, the reported evidential in questions may develop an additional overtone of **politeness**. Example (39a) is a plain question. The speaker inquires the addressee whether it is possible for him to eat the tangerine. To compare with example (39a), example (39b) shows that the inclusion of the reported evidential ameliorates what would otherwise (i.e., without the reported evidential) sound like a mild command. The speaker wants to be respectful or polite in directing the question to the addressee.

(39) Tsou

a. *te-'o* *peel-a* *an-a* *'e* *huv'o* *eni*?
IRR-1SG able-PV eat-PV NOM tangerine this
'Can I eat this tangerine?'

---

[13] This is confirmation-seeking in Korean (Ahn & Yap 2014).

b. *te-'o* ***nana*** *peel-a* *an-a* *'e* *huv'o* *eni*?
IRR-1SG EVID able-PV eat-PV NOM tangerine this
'May I eat this tangerine?'

5.2 Commands

In Saaroa and Kanakanavu, reported evidentials can be used in commands and have a **secondhand imperative** meaning 'do something on someone else's order'.[14]

In Saaroa and Kanakanavu, a command containing the reported evidential comes from a third party. In other words, the reported evidential is used in commands **marking an order on behalf of someone else**. In addition, because the addressee does not know who is the exact author of the command, the use of the reported evidential functions as a deflection of blame. Examples are demonstrated in (40) and (41).

(40) Saaroa
a. *t‹um›a~timalha=kia=**ami***!
<AV>RED~listen=PR=EVID
'Please listen (on someone else's order)!'

b. *kuu=kia=**ami*** *alusap~~u~~*!
NEG.IMP=PR=EVID sleep
'Please do not sleep (on someone else's order)!'

(41) Kanakanavu (Tsuchida 2003:112)
"*m-u-ciri-**á*** *=**kani***!" *ki-sa-ené* *=kani* *=kiái* *sua=* *iná* *=ini.*
stand it-is-said say it-is-said NOM mother his
'"Stand up!" he said to his mother.'

6 Conventions and knowledge of reported evidentials

In Saaroa, Kanakanavu, and Tsou, a traditional story is typically acquired from someone's verbal report. The kind of knowledge encoded in such stories is marked with a reported evidential, which, in turn, becomes a token of the narrative genre. Thus, kinds of knowledge heavily cling to conventions and traditions in each speech community.

The focal point of this section is to explore (i) what one sees in dreams and (ii) what new cultural conventions—writing, radio, television, Internet, telephone, and the like—are introduced, shedding light onto evidential meanings.

6.1 Reported evidentials in dreams

Dreams are subject to different choices of evidentiality. For example, dreams which are supposed to be seen are couched in firsthand evidential in Jarawara (Dixon 2004). In Saaroa, Kanakanavu and Tsou, dreams are cast in reported evidentials. They signal that what one experienced in a dream is unconsciously acquired information, outside the real

[14] Evidentials cannot be used in imperative clauses in an overwhelming majority of languages (Aikhenvald 2004:250).

world. Therefore, using reported evidentials in these three languages is not 'seen' in the same way one sees objects and activities in the real world.

(42) Saaroa
pausasalili=aku *m-ilakupu=aku=**ami.***
dream=1SG.NOM AV-fall=1SG.NOM=EVID
'I dreamed that I fell.'

(43) Kanakanavu
s‹in›i-'~~u~~cu~~p~~u=ku *mi~~u~~ra* *ni-musutupuku=ku=**kan(i)***.
<PFV>sleep-dream=1SG.NOM yesterday PFV-fall=1SG.NOM=EVID
'Yesterday, I dreamed that I fell.'

(44) Tsou
mi-'o *yacei* *ho* *mi-'o* ***nana*** *mah~mahafo.*
AV.REAL-1SG dream that AV.REAL-1SG EVID RED~bring
'I dreamed that I was pregnant.'

6.2 Reported evidentials and cultural innovations

Boas (1942:183) states that 'when changes of culture demand new ways of expression, languages are sufficiently pliable to follow new needs'. As a result of cultural innovations, further insights on the Saaroa reported evidential can be acquired from the way in which the reported evidential is employed to describe existing cultural practices.

For example, the reported evidential is used in Saaroa to talk about something one heard on the PHONE or RADIO.

(45) Saaroa
lhi-makari=aku *'arivungua* *m-ilakupu=**ami*** *kana'ana.*
PFV-talk=1SG.NOM telephone AV-fall=EVID 3.INDEP.PN
'When I talked on the phone, (I heard on the phone) he fell.'

Saaroa speakers consider televised images on a par with visual information, and consequently, do not use the reported evidential to retell what they have just seen on TELEVISION. However, if one hears the news on television without watching images, the reported evidential is appropriate.

7 Origin of the reported evidential

The development of quotative markers from generic speech verbs like 'say' has been discussed in grammaticalisation research, e.g., Frajzyngier (1996), Kachru (1979), Lord (1976, 1993), Saxena (1995), and Subbarao et al. (1983).[15] Along the same line, in

[15] Such claim has been challenged by authors such as Güldermann (2008). He argues against the view that non-predicative uses of speech verbs like 'say' are inevitably the result of 'grammaticalisation'. Rather, he proposes another view that the non-predicative uses may exist earlier than the predicative uses. This view is the result of what he labels 'lexicalisation'. Since there is no sufficient evidence to argue for the alternative view in Saaroa, I leave it for further investigation.

Aikhenvald's (2004:17) monograph *Evidentiality* (examples drawn from over 500 languages), reported evidentiality and quotative evidentials often come from a grammaticalised verb of speech.[16] In Saaroa, the reported evidential =*ami* also comes from the verb of saying *ami*, and has some morphosyntactic peculiarities (Pan 2014:100–101). Unlike Saaroa, there is no sufficient evidence to track the origin of reported evidentials in Kanakanavu and Tsou.

8 Concluding remarks

This study has shown that in Saaroa, Kanakanavu and Tsou, evidentiality is a grammatical category in its own right, and a closed grammatical system, reported evidentiality, is attested.

Table 2 summarises the findings of reported evidentials in Saaroa, Kanakanavu and Tsou.

Table 2: Summary of the reported evidential in Saaroa, Kanakanavu, and Tsou

Parameter	Saaroa	Kanakanavu	Tsou
Grammatical means—§2.1	enclitic =*ami*	enclitic =*kani*	*nana*
More than once in the same clause—§2.2	Yes: one the first constituent of the clause and the other on the following verb(s) (*contrastive focus*)	Yes: on verbal constituents in succession (*emphatic connotations*)	Yes: one on the verb and the other on the noun (*enhance speaker's objectivity and unwillingness to vouch for the information*)
Scope—§2.3	Not within the scope of negation		
Mirative overtones—§3	Yes: on the constituent triggering mirativity	—	
Nature of perceiver—§4.1	'I' (the speaker)		
First person effects—§4.2	(i) Overtones of lack of control and subsequent surprise on behalf of the speaker; (ii) irony and disagreement	If the speaker does not remember what s/he did	—
Correlations with questions—§5.1	(i) On behalf of someone else; (ii) puzzled and expects further clarification	—	(i) Enhance the speaker's uncertainty (rhetorical question); (ii) politeness
Correlations with commands—§5.2	On behalf of someone else (secondhand imperative)		—
In dreams—§6.1	Unconsciously acquired information, outside the real world		
In cultural innovations—§6.2	Telephone, radio	—	
Origin—§7	A grammaticalised verb of speech *ami*	—	

[16] This has been reported in Lezgian (Haspelmath 1993), Tauya (MacDonald 1990), Tibetan (Tournadre 1994:152), Maricopa (Gordon 1986:86), Akha (Thurgood 1986:221), Cora (Casad 1992:154–156), Korean (Ahn & Yap 2014), and many others.

Owing to the limitations of space and time, I leave some residual issues open for further investigation, including (i) the actual size of the three languages' evidential system, (ii) the association between case markers and evidential markers, (iii) the possible origin of reported evidentials in Kanakanavu and Tsou, and (iv) the alternative analysis (i.e., lexicalisation) of the origin of Saaroa =*ami*. These remaining puzzles go beyond the scope of the present paper, and deserve at least another research paper.

References

Aikhenvald, Alexandra Y. 2003. Evidentiality in typological perspective. In *Studies in evidentiality*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 1–31. Typological Studies in Language 54. Amsterdam: John Benjamins.

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2014. The grammar of knowledge: a cross-linguistic view of evidentials and the expression of information source. In *The grammar of knowledge: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 1–51. Explorations in Linguistic Typology 7. Oxford: Oxford University Press.

Ahn, Mikyung and Foong Ha Yap. 2014. On the diachronic development of Korean 'say' evidentials and their extended pragmatic functions. *Diachronica* 31.3:299–336.

Boas, F. 1942. Language and culture. In *Studies in the history of culture: The disciplines of the humanities*, 178–84. Menasha: The George Banta Publishing Co.

Casad, Eugene H. 1992. Cognition, history and Cora *yee*. *Cognitive Linguistics* 3.2:151–86.

Chafe, Wallace and Johanna Nichols. 1986 (eds). *Evidentiality: The linguistic coding of epistemology*. Advances in Discourse Processes 20. Norwood, NJ: Ablex Publishing Co.

Chang, Anna Hsiou-chuan. 2012. Evidentials in Paiwan. *Tajen Journal* 40:115–129.

Chang, Henry Y. 2006. Rethinking the Tsouic subgroup hypothesis: A morphosyntactic perspective. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 565–583. Language and Linguistics Monograph Series W-5. Taipei: Academia Sinica.

DeLancey, Scott. 1997. Mirativity: the grammatical marking of unexpected information. *Linguistic Typology* 1.1:33–52.

Dixon, R.M.W. 2004. *The Jarawara language of Southern Amazonia*. Oxford: Oxford University Press.

Dixon, R.M.W. 2010. *Basic Linguistic Theory, Vol. 2: Grammatical topics*. Oxford: Oxford University Press.

Frajzyngier, Zygmunt. 1996. *Grammaticalization of the complex sentence: A case study in Chadic.* Studies in Language Companion Series 32. Amsterdam: John Benjamins.

Gordon, Lynn. 1986. The development of evidentials in Maricopa. *Evidentiality: The linguistic coding of epistemology*, edited by Wallace Chafe and Johanna Nichols, 75–88. Advances in Discourse Processes 20. Norwood, NJ: Ablex Publishing Co.

Güldemann, Tom. 2008. *Quotative indexes in African languages: A synchronic and diachronic survey*. Empirical Approaches to Language Typology 34. Leipzig: Leipzig University.

Haspelmath, Martin. 1993. *A grammar of Lezgian*. Mouton grammar library 9. Berlin: Mouton de Gruyter.

Hsieh, Fuhui. 2012. On the grammaticalization of the Kavalan SAY verb *zin*. *Oceanic Linguistics* 51.2:464–489.

Kachru, Yamuna. 1979. The quotative in South Asian languages. *South Asian Languages Analysis* 1:63–77.

Lord, Carol. 1976. Evidence for syntactic reanalysis: from verb to complementizer in Kwa. In *Chicago Linguistic Society* 12:179–191 (*Papers from the parasession on diachronic syntax, April 22, 1976*, ed. by Sanford B. Steever, Carol A. Walker and Salikoko S. Mufwene).

Lord, Carol. 1993. *Historical change in serial verb constructions*. Typological Studies in Language 26. Amsterdam and Philadelphia: John Benjamins.

MacDonald, Lorna. 1990. Evidentiality in Tauya. *Language and Linguistics in Melanesia* 21:31–46.

Miller, Cynthia L. and Leoma G.Gilley. 2007. Evidentiality and mirativity in Shilluk. In *Advances in Nilo-Saharan Linguistics: Proceedings of the Eighth Nilo-Saharan Linguistics Colloquium, University of Hamburg, August 22–25, 2001*, ed. by Doris L. Payne and Mechthild Reh, 191–206. Nilo-Saharan Volume 22. Cologne: Rüdiger Köppe.

Pan, Chia-jung. 2010. *The grammatical realization of temporal expressions in Tsou*. LINCOM Studies in Austronesian Linguistics 7. Muenchen: Lincom Europa.

Pan, Chia-jung. 2012a. 'Adjectives' in Lha'alua and Tsou: a comparative study. Paper read at the Thirteenth National Conference on Linguistics. 23–24 May. Tainan: Southern Taiwan University of Science and Technology.

Pan, Chia-jung. 2012b. A grammar of Lha'alua, an Austronesian language of Taiwan. PhD dissertation. Cairns: James Cook University.

Pan, Chia-jung. 2014. The grammar of knowledge in Saaroa. In *The grammar of knowledge: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 89–107. Oxford: Oxford University Press.

Ramirez, H. 1997. *A Fala Tukano dos Yepâ-masa*. Tomo I. *Gramática*. Tomo II. *Dicionário*. Tomo III. *Método de aprendizagem*. Manaus: Inspetoria Salesiana.

Rood, David and Jean Ormsbee Charney. 1976. *Wichita grammar*. New York: Garland Publications.

Saxena, Anju. 1995. Unidirectional grammaticalization: diachronic and cross-linguistic evidence. *Sprachtypologie und universalienforschung* 48.4:350–372.

Schlichter, Alice. 1986. The origins and deictic nature of Wintu evidentials. In *Evidentiality: The linguistic coding of epistemology,* ed. by Wallace Chafe and Johanna Nichols, 46–59. Advances in Discourse Processes 20. Norwood, NJ: Ablex Publishing Co.

Subbarao, K. V. et al. 1983. The verb 'say' in South Asian languages: A study in linguistic convergence. In *Language change and language variation*, ed. by Aditi Mukerjee, 89–103. Hyderabad: Osmania University Publications.

Thurgood, Graham. 1986. The nature and origins of the Akha evidentials system. In *Evidentiality: The linguistic coding of epistemology*, ed. by Wallace Chafe and Johanna Nichols, 214–222. Advances in Discourse Processes 20. Norwood, NJ: Ablex Publishing Co.

Tournadre, Nicholas. 1994. Personne et médiatifs en tibétain. *Faits de langues* 3:149–158.

Tsuchida, Shigeru (ed.). 2003. *Kanakanavu Texts* [*Austronesian Formosan*]. Endangered Languages of the Pacific Rim. ELPR Publications Series A3-014. Osaka: Osaka Gakuin University

Tung, T'ung-ho, with the assistance of S.H. Wang, T.K. Kuan, T.F. Cheng, and Margaret M. Yan. 1964. *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica, Special Publications 48. Taipei: Academia Sinica. Reprinted in 1991 with notes and errata prepared by Paul Jen-kuei Li.

Valenzuela, P. 2003. Evidentiality in Shipibo-Konibo, with a comparative overview of the category in Panoan. In *Studies in evidentiality*, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 33–61. Amsterdam: John Benjamins.

Yang, Gloria Fan-pei. 2000a. The semantics-pragmatics of the hearsay evidential *nana* in Tsou. *National Taiwan University Working Papers in Linguistics* 3:69–86.

Yang, Gloria Fan-pei. 2000b. Tsou case markers as evidentials. *National Taiwan University Working Papers in Linguistics* 3:41–67.

15 *Functions of ma' 'isa:a' in Saisiyat: A discourse analysis**

MARIE MEILI YEH

1 Introduction

Many corpus-based studies (Thompson & Mulac 1991; Biq 2001, 2003, 2007, among others) have unveiled the significance of frequency and collocation in grammaticalisation. For example, Biq (2003:92–93) points out that "frequent collocations create syntagmatic association, which often triggers semantic change of the lexical items involved in the collocation." Paying attention to the interaction between collocation and semantic change, this paper discusses the functions of *ma' 'isa:a'* (lit.) 'also that' in Saisiyat. Before proceeding to the distribution of this idiomatic expression, an introduction to the source of data is necessary.

This study is based on the data drawn from the NTU Corpus of Formosan Languages established by National Taiwan University under the supervision of Shuan-fan Huang, Lily I-wen Su, and Li-May Sung (Su et al. 2008; Sung et al. 2008). It contains data on Saisiyat, Kavalan, Tsou, and Amis. The data are transcribed largely according to Du Bois et al. (1993), with the Intonation Unit (IU) serving as the basic unit for a detailed recording of linguistic phenomena, including pauses, repetitions, repairs, and intonations. The Saisiyat corpus contains twenty-two texts (with sound files), with nineteen of them being natural spoken narrations and the other three, conversations.[1] For comparison or support, data from the *Workshop on Saisiyat Ritual Songs, Dances and Folklores:*

* This study is supported by the NSC project "Demonstratives in Saisiyat" (100-2410-H-134-013). I would like to thank all the Saisiyat informants for sharing with me their knowledge of this language. An earlier version of this paper was presented at the NTU Workshop of Formosan Languages on 8 January 2011, where I benefited a lot from discussions with the participants. Thanks are also due to the two anonymous reviewers for their helpful comments. I am also grateful to the editors Elizabeth Zeitoun, Stacy F. Teng, and Joy J. Wu for their insightful and constructive comments. I would like to dedicate this paper to my PhD advisor, Prof. Lillian M. Huang, who has always been a mentor in my research career.

[1] Another corpus displaying Saisiyat data is the Formosan Language Archive developed by Academia Sinica under the direction of Elizabeth Zeitoun (http://formosan.sinica.edu.tw), closed at this moment. For details please refer to Zeitoun et al. (2003) and Zeitoun & Yu (2005). My analysis is based on the data drawn from the NTU Corpus of Formosan Languages since the collocation of *ma'* and *'isa:a'* occurs more frequently in this source than in the other one.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 363–384
Asia-Pacific Linguistics, 2015

Project Report[2] and the *Scripts for Saisiyat Folklores*[3] will also be cited. It is found that *ma' 'isa:a'* occurs more often in conversations than in narrations. In the NTU Corpus of Formosan Languages, around 62% (i.e., twenty-one out of the thirty-four tokens) of *ma' 'isa:a'* are from conversations, which constitute only 14% of the corpus. By contrast, it does not occur as frequently in narratives: among the nineteen narrations in the NTU Corpus of Formosan Languages, only three (about 16%) are found to contain *ma' 'isa:a'*, and the tokens constitute 38%.

Biq (1994), examining how interactionality influences the use of Mandarin *jiushi, jiushishuo, jiushuo,*[4] *ranhou* 'then', and *qishi* 'in fact', concludes that although interactionality can be seen in all text types, it is most salient in conversations. This calls into question whether the unequal distribution of *ma' 'isa:a'* can be attributed to the difference of interactionality inherent in narrations and conversations. To pursue this issue, this paper examines the functions of *ma' 'isa:a'* in the two genres.

2 Background and previous studies

Saisiyat is a Formosan language spoken in the mountain areas of northwestern Taiwan. According to Li (1985) and Blust (1999), it belongs to the Northwestern subgroup, which includes Saisiyat, Pazeh, and Kulon. Compared to other Formosan languages, Saisiyat displays a rich case marking system with six sets of case forms—nominative, accusative, genitive, possessive, locative, and dative (Yeh 2000). Like most of the other Formosan languages, Saisiyat verbs display a Philippine-type focus system, where four foci—agent focus, patient focus, locative focus, and referentialfocus (or instrumental focus)[5]—are morphologically distinguished, though syntactically locative focus is active only in nominalised constructions (see Yeh 2003 and Yeh 2011a for details). With respect to word order, a discrepancy is found between agent focus and non-agent focus constructions—while the former follows SVO order in general, the constituent order of the latter is basically free (Yeh 1991).

Previous studies on *ma' 'isa:a'* include Huang (2008) and Yeh (2011b). Huang (2008) is a dissertation on polysemy and its application to corpus documentation. In a chapter discussing the multiple senses of *ma'*, it is pointed out that *ma'* conveys four senses—COMPARISON, INCLUSION, CONJUNCTION, and CONNECTION. Both COMPARISON and INCLUSION emphasise the sharing of an identical element; while the former views sharing as forming a chaining relation, as in (1), the latter takes a collective view, as shown in (2).

---

[2] The report contains ten stories from a storytelling contest held in Tungho, the village where Saisiyat is spoken by a majority of inhabitants. They were transcribed from Chinese into Saisiyat by the native speaker Lalo' a tahesh kaybaybaw. Seven tokens of *ma' 'isa:a'* are found in the ten stories; the collocation of *ma' 'isa:a'* constitutes only around 5% of the total tokens of *'isa:a'*.

[3] It contains five plays translated by native speakers from Chinese scripts. Only one token of *ma' 'isa:a'* was found.

[4] The expression *jiushi* is composed of the copula *shi* 'be' and the adverb *jiu* 'just' with versatile meanings, and *jiushishuo* with *shuo* 'say' adding to it.

[5] As this marker *Si-* can mark peripheral arguments other than instrument as subject, besides instrumental focus, it is also subsumed under the term referential focus (in the NTU Corpus) or circumstantial focus (in Ross 1995).

(1) COMPARISON (Narration: Kathethel 2, NTU, cited from Huang 2008:44)[6]

156. ...(1.9) *in'alay* *ri'sa* *bazae'* *ka* *intot* *hiza*[7]
be.from there hear ACC fart that

157. ... *kabinao*
lady

158. ...(1.0) *isa*
FILL

159. ...(0.8) *baza'* *na'ora==*
hear suddenly

160. ... *sawa'* *ila* *hiza* *ka==* *kamo'alay* *ma'* *sawa'* *ila*
laugh PFV that NOM young.man also laugh PFV

'Upon hearing the dog fart, the lady laughed. The young man also laughed.'

(2) INCLUSION (Conversation: Holiday, NTU, cited from Huang 2008:156)

60. C:...(1.5) *sa'sa'eh* *ka* *mae'iyaeh* *ma==*'
every LNK person DM

61. C:.. *'iska* *nonak* *ba:iw*
each self buy

'Everybody buys his own things to offer.'

CONJUNCTION and CONNECTION both denote the relation between two clauses, with the former referring to textual connection, usually a causal or a chronological relation as in (3) and the latter to an idiomatic sense designated by *ma'* in collocation with *'isa:a'* to manage discourse cohesion in spontaneous speech, as in (4).

---

[6] In the heading to the example sentences, NTU is used to refer to the NTU Corpus of Formosan Languages, preceded by the title of the narration or conversation. Abbreviations for the other sources include Story for the *Workshop on Saisiyat Ritual Songs, Dances and Folklores: Project Report* and Scripts for *Scripts for Saisiyat Folklores*. Note that the transcription of *ma' 'isa:a'* varies even in the same source. We will stick to *ma' 'isa:a'* in the main text while following the transcription of the original sources in the examples cited. One reason for sticking to the transcription is because phonological reduction may be an indicator of grammaticalisation. The glossing in the NTU Corpus has been revised slightly according to the Leipzig Glossing Rules. Especially, to avoid confusion, *ma' 'isa:a'* will consistently be glossed as 'also that'. What follows are abbreviations added: AF, agent focus; E, exclusive; EXIST, existential; INT, interjection; LF, locative focus; LNK, linker; PF, patient focus; PN, personal name; RED, reduplication; RF, referential focus.

For typographic convenience and following the orthography of Saisiyat, *ae*, *oe*, *e*, *:*, *'*, *S*, and *ng* are used to represent the low front unrounded vowel /æ/, the mid front rounded vowel /œ/, the mid central vowel /ə/, the long vowel, the glottal stop /ʔ/, the voiceless alveolar fricative /s/ (with *s* representing voiceless dental fricative) and the velar nasal /ŋ/ respectively. For details on transcription conventions in discourse analysis, please refer to Du Bois et al. (1993). Abbreviations on discourse in this paper include: BC, backchannel; DN, discourse marker; FILL, filler; FS, false starter.

[7] A discrepancy is found between the data presented in Huang (2008) and the NTU Corpus searched on 28 March 2014. In the Corpus, there is a *ka*, glossed as the nominative case marker, after *hiza*. For ease of presentation, we delete the Chinese glosses and translations (which are presented by Roman characters in Huang (2008)).

(3) CONJUNCTION (Conversation: Life, NTU, cited from Huang 2008:157)

221. F: .. *so:* *la'oz* *ila* *ma'*
COND enough PFV DM
'If (the ducks have grown) enough…'

222. F: ...(1.2) *ka-hi-hiwa'-en* *ila* *ka-si'ael-en* *ka-powa'-en* *so:* *tatini'-in*
KA-RED-kill-PF PFV KA-eat-PF KA-why-PF COND old-PF
'…the ducks can be killed as soon as they grow old. Why do you raise them until old age?'
(Lit: When (the ducks have grown) enough, they will be killed and be eaten, when they grow old …)[8]

(4) CONNECTION (Conversation: Life, NTU, cited from Huang 2008:159)

147. F: *m==*
FIL

148. F: *'okik* *wa'isan* *ka-ba:iw-an* *ma'* *isaa*
NEG expensive KA-buy-NMLZ also that
...'am *nak'ino'*
FUT so.what
'The market price is not good. What can (we) do? '

According to Huang (2008), the exact meaning of *ma'* in fact depends on the word with which it collocates or the construction it occurs in. For example, as mentioned above, CONNECTION sense is expressed by *ma*' in collocation with '*isa:a*'; besides, INCLUSION is rendered when it co-occurs with *sa'sa'ih* 'each' or *'ana* 'whether', as in (2). The constructional patterns for different senses are summarised as below.[9]

Table 1: Meanings and constructional patterns of *ma'* (Huang 2008:171)

Meanings	Structure
COMPARISON	
[similarity]	NP_1 VP_1, NP_2 *ma'* VP_1
[listing]	NP_1 ma' VP_1, NP_2 *ma'* VP_1
[periphery]	(NP_1 VP_1), NP_2 *ma'* VP_1
[counter-expect]	(NP_1 VP_1), NP_1 *ma'* VP_1
INCLUSION	
[indefiniteness]	*'ana* indefinite_pro *ma'* VP
[extremity]	*'ana* NP/VP *ma'* VP
[collectiveness]	*sasa'ih/saboeh ma'* VP
CONJUNCTION	CL_1 *ma'* CL_2
(CONNECTION)	CL_1 *ma' isaa* CL_2
	CL_1 *isaa ma'* CL_2

[8] Literal translation will be added when the original translation does not seem to be clear enough.

[9] Note that Huang (2008:54–56) makes a distinction between meanings independent of and dependent on constructional patterns. The former is listed in brackets and the later (i.e., idiomatic sense), in parentheses.

Yeh (2011b), comparing the uses of *ma' 'isa:a'* in conversations and narrations, points out that in conversations it often occurs at the end of an IU. In other words, it occurs at transition relevance places (TRP, Sacks et al. 1974) where another speaker may take over the floor. Based on Yeh (2011b), *ma' 'isa:a'* will be examined with an aim to explain the relationship between its different functions. In the following two sections, the distribution and functions of *ma' 'isa:a'* will be introduced. Section 5 brings under discussion the relationship between its different functions, and section 6 presents concluding remarks.

3 Distribution of *ma' 'isa:a'*

Table 1 may leave us with the impression that, as argued by Huang (2008:40), *ma'* typically occurs between the subject and the predicate with right or left modifying scope, whereas *ma' 'isa:a'* occurs only between clauses. However, according to my observation, *ma' 'isa:a'* also occurs between the subject and the predicate, as illustrated in (5) and (6).

(5) Subject_predicate (Conversation: Holiday, NTU)

196. C: [*ehe*]					
BC					
197. K: (0)*baba:i'*	*ma' isaa*		*kala-baba:i'*	*nonak*	*ma==*
PN	also	that	LOC-PN	self	FILL
198. C:... *k‹om›si'ael*			*risa==*		
<AF>have.lunch			here		

'The Fengs have their lunch at the Feng homes.'

(6) Subject_predicate (Narration: Kathethel 2, NTU)

140....*kabinao'*	*ki*	*kamo'alay*	*ma' isaa*	
lady	and	youth	also	that
141....*'i'ini*	*paehraehraeng*			
NEG	speak			

'Both of them have not spoken to each other.'

Besides, the clause preceding or following *ma' 'isa:a'* can be left unspecified, resulting in the occurrence of *ma' 'isa:a'* at boundary positions. Take extract (7), for example. Speaker K ends his turn by *ma' 'isa:a'* without providing the consequence of the precondition. Such implicitness leaves the space for the hearer to infer the consequence on his own.

(7) Clause1_ (Conversation: Holiday, NTU)

42. C: ... *a=*					
BC					
43. K: ... *nak*	*kakhayza'an*		*<L2shiL2>:*		*nakhara'*
like	before		[Man]		like
44. K: ...(1.2) *wa'-wa'isan*		*r‹oem›a'oe*		*ka*	*pinobae:aeh*
RED-good.at		<AF>drink		ACC	wine
'akoy ka	*pinobae:aeh*	*tabin*			
AF.many	NOM wine	until			

45. K:.. <L2s=*antian*L2>..*san*--<L2*sanji*L2>*o:* *aw'itol*
[Man] [Jap] DM around
ma' isaa
also that
'People used to be good drinkers and used to drink a lot until around three in the afternoon.'
46. C: [@]
47. C: *haysia* [XX@@]
still xx
'(They were) still (drinking).'

In the following extract (8) drawn from a narration, unlike the above-mentioned examples, no clear relationship between IU 16 and IU 17 can be established, and therefore the *ma' 'isa:a'* in IU 17 will be counted as occurring in initial position. Note that both the long pauses (indicated in 0.9 by …) and the lengthening (as indicated by =) surrounding *ma' 'isa:a'* seem to indicate that it constitutes a self-contained unit.[10]

(8) _(Clause 2) (Narration: 'anhi, NTU)
16. ...(2.3) *haysani==' 'alikaeh* *sowiti'*
now quick a.bit
17. ...(0.9) *ma' isaa=* *... So: ...* *t‹in›alek-en* *ila* *kayzaeh* *ila* *mari'-in*
also that COND <PFV>cook-PF PFV good PFV take-PF
18. ...(1.1) *taboe-on* *ray==* *hobos*
put-PF LOC bag
'Now it is a bit faster. That is…The bamboo shoots are put in a container after being heated.'

The fact that *ma' 'isa:a'* may constitute an IU on its own is revealed more clearly in (9), where it is uttered by a different speaker.

(9) One IU (Conversation: Life, NTU)
F and M, talking about M's kids, mentioned one kid, who is not yet married.
266. M: ...(1.4) *nisia* *nonak ...* *k‹in›ita'* *o:* *tatini'* *o:*
3SG.GEN self <PFV>see DM old.man DM
hayza' *ay* *ka* *howaw* *haysani*
EXIST Q ACC work now
'Let him see for himself. (For us) old people, is it our business nowadays?'
267. F: ...(1.1) [*ma'* *isaa*]
also that
'It's like that.'
268. M: [*niSo*] *korkoring* *hayza'* *ka* *'ima* *min-owa'*
2SG.GEN child EXIST ACC IMA MIN-like
269. M: ...(1.3) *isaa* *k‹in›ita'* *ila*
there <PFV>see PFV
'Your child (Ataw) has somebody he likes.'

[10] Note that there exists a mismatch between sentential boundaries and IU boundaries. A sentence may be composed of many IUs.

The distribution of *ma' 'isa:a'* in the NTU Corpus is summarised in Table 2 below. As shown in Table 2, whether in conversation or in narration, *ma' 'isa:a'* mostly occurs between two constituents, be they intra-clausal constituents such as the subject and the predicate or inter-clausal ones. This distribution seems to reveal that its prototypical function is basically connective. In the following section, the functions of *ma' 'isa:a'* will be elaborated.[11]

Table 2: Distribution of *ma' 'isa:a'*

Genre	Conversation				Narration			
Title	H	L	E	Total	K2	A1	F6	Total
Function								
S _ P	2	1	1	4 (19%)	2	0	1	3 (23%)
(CL1)_(CL2)	14	3	0	10 (81%)	9	1	0	10 (77%)
Total	16	4	1	21	11	1	1	13

4 Functions of *ma' 'isa:a'*

In this section, the functions of *ma' 'isa:a'* will be examined according to its distribution.[12] Huang (2008) holds that *ma' 'isa:a'* appears between two clauses and conveys an idiomatic sense CONNECTION, which is related to the function of managing discourse cohesion in spontaneous speech. However, as mentioned in the previous section, *ma' 'isa:a'* not only appears between the subject and the predicate but as shown in (10), it also points to the similarity between the subject (in this case *'aehoe'*) and the subject in the previous text (*korkoring*). In other words, it also manifests the sense of COMPARISON.[13]

(10) Comparison (Narration: Frog 6, NTU)

129....(1.5)	*hiza'*	*korkoring*	*ki*	*'aehoe'*	*hiza*	*...'okik*	*ra:am*	
	there	child	and	dog	there	NEG	know	
130....(1.1)	*ma*	*h‹om›ses==*	*si-'intani'*	*hiza*	*sahae'*	*ila*	*ilahini*	*ray*
	also	<AF>frightened	RF-stop	there	fall	PFV	go.here	LOC
131....(1.6)	*ray*	*'ataS*						
	LOC	cliff						

'The child and the dog did not know (where they were) and were frightened; the deer stopped and they fell off the cliff.'

132. ... (0.9)	*ra*	*'aehoe'*	*ma isaa*	*losalaz*	*a*	*malobaz*	*ila*	
	FILL	dog	also that	first	FILL	heavily	PFV	
133. ... sahae'	ila	'oyae'						
fall	PFV	together						

'The dog fell heavily and fell together with the child.'

[11] The notation S _ P stands for Subject _ Predicate and (CL1) _ (CL2) for the position between the first and the second clauses, with parentheses indicating optionality. The abbreviations for the title include: H = Holiday, L = Life, E = Election, K = Kathethel, A = *'anhi*, F = Frog.

[12] I am indebted to one of the editors for directing my attention to the correlation between functions and distribution.

[13] For ease of comparison, the terminology in Huang (2008) is followed.

For the *ma' 'isa:a'* occurring between two clauses, the function basically pertains to Huang's CONJUNCTION—to indicate such logical relationships as causal or chronological relations. For example, in (11) and (12) *ma' 'isa:a'*, or the pair *So: ~ ma' 'isa:a'*, marks a temporal or a causal relationship between two clauses.

(11) Temporal (Conversation: Holiday, NTU)

147. K: *ai*==						
BC						
148. C:... *a*--***So:***	*ba:iw-in*	*ma' isaa*				
FS COND	buy-PF	also that				
149. C:...(1.0) *ba:iw-in*	*m-wa:i'*	*aras-en*	*ila*	*rini*	*raremewan*	
buy-PF	AF-come	take-PF	PFV	here	PN	

'After buying the things, they take them here to Xiangtianhu.'

(12) Temporal (Narration: Kathethel 2, NTU)

543. … *hiza*	*tabin*						
there	until						
544. …(1.3) *haysani*	*ma' isaa*	*'oka'*	*ila*	*ka-kita'*	*ki*==	*saysiyat*	
today	also there	NEG	PFV	RED-see	with	PN	

'From then on, she didn't meet with the SaiSiyat people again.'

Though there is the temporal marker *So:* in (11), the exact meaning (in this case temporal sequential) is in fact inferred from both the word order (on the basis of the iconic principle of sequential order) and real-world knowledge. In other words, pragmatic inference is involved. As for (12), the temporal frame is established by the clause *hiza tabin haysani* 'from then till now.' The narration is about a traditional story on how the goddess Kathethel came from the water to teach the Saisiyat the skill of weaving but ended up being treated with hostility by the Saisiyat women. Disappointed with the Saisiyat, she decided to leave them and never to come back again. In view of the whole context (how the goddess came to help the Saisiyat and ended up to being mistreated by them), *ma' 'isa:a'* in fact serves to introduce the result or outcome of the story (the goddess left and from then on never met again with the Saisiyat). In this case, a sense of consequential relationship is also rendered. The following example serves as another example to illustrate the point that *ma' 'isa:a'* can also preface a consequential clause.

(13) Consequence (Narration: Kathethel, NTU)

528....*'am*	*ma*==*..nom'i'iya*		*mita'*	*ti-roSa'-en*					
FUT FILL	??	AF.live	1IPL.GEN	cut-two-PF					
529....*s-‹in›roSa'*	*ma' isaa*	*masay*	*ila*	*bo:ok*	*ila.. 'oka'*	*ila*			
RF-<PFV>two	also that	die.AF	PFV	rot	PFV	NEG	PFV		

'We cut it in two, and it died, rotten, and gone.'

In the above examples, where *ma' 'isa:a'* follows an antecedent clause and prefaces another clause, *ma' 'isa:a'* seems to function as a connective marker, of which the exact logical relationship is pragmatically inferred from the context or the real world situation.

The above-mentioned temporal and consequential relationships belong to textual functions and therefore are subsumed under CONJUNCTION. In Su (1998:171) they are categorised as sequential uses, in contrast with interaction uses such as conditionals, concessives, topic successions, and fillers. Though not clearly stated, the idiomatic sense CONNECTION ascribed to *ma'* in collocation with *'isa:a'* according to Huang (2008)

functions at the discourse level and therefore should belong to the interaction use.[14] With regard to the CONNECTION functions, *ma' 'isa:a'* quite frequently marks a conditional relationship between two clausal elements as shown in (14).

(14) Conditional (Conversation: Life, NTU)
F talking about his land, being at the halfway of a hill …

190. F: ...	*so==*	*rikrika-en*	*ila*	*ma'*	*isaa*			
	COND	hot-PF	PFV	also	that			
191. F:	*'ana*	*p‹in›amowa'-an*	*tatimae'*	*ma*	*ayayo'*	*ila*	*saboeh*	
	regardless	<PFV>plant-NMLZ	vegetable	also	wilt	PFV	all	

'If the weather is hot,
the plants will all wilt.'

Note that in (14) the conditional relationship between the two clauses is marked by *ma' 'isa:a'* pairing with the temporal or conditional marker *So:*. Nevertheless, a conditional reading can also be inferred even without the presence of *So:*, as shown in (15) below.[15]

(15) Conditional (Conversation: Holiday, NTU)

179. C:...(1.1)<L2*jiushishuo*L2>			*nakhini*	*'am*	*patabir*	*ma*	*isaa*	
[Man]			like.this	FUT	do.worship	also	that	
180. C:...*hini*	*ray*	*asang*	*ma'*	*rima'*	*ka*	*saboeh*		
here	LOC	neighborhood	FIL	go	NOM	all		
ka	*korkoring*	*tatini'*						
LNK	children	old						
181. C:...*matna'amet*	*ila*	*isaa*	*o:*					
all.go.AF	PFV	there	DM					

'In rituals like this, everybody in the neighborhood goes.'

In the above example, the conditional reading is rendered because the antecedent clause is marked by the future tense marker *'am*. According to Schiffrin (1987:246), the difference between temporal and conditional lies in the fact that for conditional relationships the sequence is potential rather than actual. Besides, for conditionals, occurrence of the antecedent is a prerequisite for assurance of the temporal succession of the consequent.

In the following example, *ma' 'isa:a'* also functions to connect the given condition 'the price of mushrooms not being good' and the consequence or conclusion 'what can we do.' However, a concessive reading can also be rendered through pragmatic inference.

---

[14] Note that Huang (2008) does not go deep into the discussion on the forces of this sense.

[15] As will be discussed in the next section, this example is in fact ambiguous between a temporal and a conditional reading. Following Traugott's (1985:295) notion of frame, if we take *'am patabir* as providing the temporal frame for the event everybody goes to, a temporal reading is obtained. On the other hand, if *'am patabir* is interpreted as 'whenever the ritual comes' then a conditional reading is inferred. In my opinion, it pertains more to the conditional reading since it refers to general situations instead of a single case of event. Thanks go to one of the editors for bringing up the possibility of the temporal reading.

(16) Concessive (Conversation: Life, NTU)
F and M talking about cultivating mushrooms to make a living…
146. F: [XX]
XX
147. F: *m*==
FILL
148. F: *...'okik wa'isan ka-ba:iw-an ma' isaa ... 'am nak'ino'*
NEG expensive KA-buy-NMLZ also that FUT so.what
'The market price is not good. What can we do?'

The concessive reading is stronger when *ma' 'isa:a'* occurs in constructions containing a universal quantifier, such as *hiya'* as in (17).

(17) Concessive (Conversation: Holiday, NTU)
103. C:..*uhN*
BC
104. K:...***hiya'*** *m-alibih ma' isaa itayso'-on*
who AF-take also that ignore-PF
'We don't care if somebody takes them home.'

As shown in the foregoing discussion, ambiguity may exist between temporal and conditional or temporal and consequential relationship. In the following example (18), such ambiguity is also attested. From the viewpoint of semantic analysis, it is tempting to assign the sense COMPARISON to this example since it denotes the similarity between the *Sayna'ase:* clan and the other clans. However, if we take into account the characteristics of spontaneous speech, the repetition (*lasia* in IU 117 and IU 118) together with the long pause (indicated by 0.8) after *ma' 'isa:a'* points to the alternative that IU 118 is a repair and thus *ma' 'isa:a'* here functions more like a pause filler showing the speaker's hesitation.[16]

(18) Pause (Conversation: Holiday, NTU)
117. C: ... *noka sayna'ase:* ***lasia*** *ma' isaa*
GEN PN 3PL.NOM also that
118. C: ... **(0.8)** ***lasia*** *nonak mak-sasome ba:iw*
3 PL.NOM self AF-contribute buy
'The Sayna'ases contribute money to buy the things they need.'

In the same vein, Example (19) is also ambiguous. Looking at the logical relationship between the two clauses, we may categorise it under the CONJUNCTION sense denoting a temporal relationship. However, taking into account the repetition (*ba:iw-in* 'buy (PF)' in IU 148 and IU 149) as well as the long pause (indicated by 1.0), we may regard it as a pause filler showing the speaker's hesitation.

[16] Such fillers are called placeholders (Wouk 2005; Hayashi & Yoon 2006; Podlesskaya 2010, among many others) and are held to be different from interjective fillers like English *uh* and *um* in that they perform the function of a preparatory substitute for a yet-to-be-produced lexical item and holds the syntactic slot for it (Hayashi & Yoon 2006:490).

(19) Pause (Conversation: Holiday, NTU)
147. K: *ai*==
BC
148. C:...*a--So:* ***ba:iw-in*** *ma' isaa*
FS COND buy-PF also that
149. C:...(1.0) ***ba:iw-in*** *m-wa:i' aras-en ila rini raremewan*
buy-PF AF-come take-PF PFV here PN
'After buying the things, they take them here to Xiangtianhu.'

In (20), there are two instances of *ma' 'isa:a'*—the first one in IU 161 connects the precondition of 'not finish' and the consequence 'take home' whereas the evaluation part 'good' of the second *ma' 'isa:a'* is left unspecified.[17]

(20) Pause (Conversation: Holiday, NTU)
159. C: ... (0.8) *e*== *pa-'apol ri'saza*
FILL CAUS-distribute there
160. C: ... (0.8) *so: hiya k‹om›oSa:*
COND who <AF>say
'The food is distributed to everyone.'
161. C: ..*'okay 'amet-i ma' isaa*
NEG finish-PF also that
162. C: ... (0.9) *alibih-in isaa so:*
take.home-PF there ??
'Leftover food is taken home.'
163. C: ... (1.3) *hiya' nihiya' 'amet-en ma' isaa*
who who.GEN finish-PF also that
'(It is good) if one's food is entirely consumed.'

Note that the underspecified part concerns the speaker's evaluation. In other words, in this example, *ma' 'isa:a'* expresses the speaker's stance. In the following example from Story, the speaker uses *ma' 'isa:a'* to achieve a down-toning effect to show that the story he is going to tell is nothing special, 'just' what everyone often hears.

(21) Down-toning (Kowaa a taain, Story)
kaysa'an yao 'oka' ila ma'an ka papnabih,
today 1S.NOM NEG PFV 1S.GEN ACC say
ma' isaa ima rengreng ka bazae'-en saeboeh, …
also that REL often LNK hear-PF all
'Today I have nothing to say…it's just what we all often hear…'

Note that it is mentioned in section 3 that *ma' 'isa:a'* can constitute a self-contained unit on its own. In regard of the function, *ma' 'isa:a'* is here exploited by the speaker as a

[17] The conversation entitled 'Holiday' involves two elders talking about the Saisiyat tomb sweeping ritual (*pasankinolol*). At tomb sweeping, people of the same *sinrahoe'* 'clan' get together to weed the tomb and make sacrifices. After the ceremony, they distribute the offerings on-site.

backchannel.[18] As shown in (22), Speaker F uptakes by using *ma' 'isa:a'* to show her agreement to M's comment on marriage of the younger generation.

(22) Backchannel (Conversation: Life, NTU)
F and M, talking about M's kids, mentioned one kid, who is not yet married.
266. M: ...(1.4) *nisia* *nonak ... k‹in›ita'* *o:* *tatini'* *o:*
3SG.GEN self <PFV>see DM old.man DM
hayza' *ay* *ka* *howaw* *haysani*
EXIST Q ACC work now
'Let him see for himself. (For us) old people, is it our business nowadays?'
267. F: ...(1.1) [*ma' isaa*]
also that
'It's like that.'
268. M: [*niSo*] *korkoring* *hayza'* *ka* *'ima* *min-owa'*
2SG.GEN child EXIST ACC IMA MIN-like
269. M: ...(1.3) *isaa* *k‹in›ita'* *ila*
there <PFV>see PFV
'Your child (Ataw) has somebody he likes.'

An interactional function as a backchannel is also evident in (23), where Speaker K and C talked about how they distributed the offerings after tomb sweeping. After C mentions that as for the Chu family, they sliced pork and rice cake on their own. Speaker K's uptake comes with *'a'apol* 'distributed' followed by *ma' 'isa:a'*, functioning to provide feedback rather than to take the floor from C, as evidenced by the fact that Speaker C continues his floor right after K's uptake. Also note that C repeats K's *'a'apol ma' 'isa:a'*.

(23) Backchannel (Conversation: Holiday, NTU)
237. C: *niya'om* *isaa* *hiza*
1EPL.GEN that there
238. C: ...(0.9) *talo--* *sa-saboeh-en* *noka==*
FS RED-all-PF GEN
239. C: ...(1.0) *aniyami'a* *inoka* *titiyon* *no* *ma* *isa*
1EPL.POSS GEN Chu.clan ?? also that
yami *nonak* *h‹om›iwa'* *ka==='aeyam*
1EPL.NOM self <AF>cut ACC meat
240. C: ... *tinawbon*
rice.cake
'We slice our own meat and rice-cakes.'
241. K: ... *'a'apol* *ma' isaa*
distribute also that
242. K: (0) *'a'apol* *ma'*/
distribute DM
'And then distribute, and then?'
243. C: (0) *ta--*
FS

[18] Backchannels are vocal indications of attentiveness such as *uh-uh*, *yeah*, *mmm* uttered by addressees to provide feedback to the speakers. (Alba-Juez 2009:113).

244. C: ... *'a'apol* *ma'* *isaa* *pa-'apol* *ka* *saboeh* *ila*
distribute DM ?? CAUS-distribute NOM all PFV
hiza *si-paysiri* *ila* *hiza* *ray* *sinrahoe*
that RF-move.around PFV there LOC clan
'We distribute them to all the clans.'

The only instance of *ma' 'isa:a'* found in the Script source also functions as a backchannel.

(24) Backchannel (Sako:o', Script)
Iban: *'oka'* *o:* *'ita'* *'okay* *'araS* *ka* *ta~ta:oS*!
NEG EXCL 1IPL.NOM NEG take ACC RED~net
'But, we do not take a fishing net with us.'
Ba'ay: *'a:i'!* *ma'* *'isa:a'* *sa'owaz* *siya.*
yes also that real DM
'Yes, indeed!'

As revealed in the above discussion, ambiguity is found in many cases. In addition, pragmatic inference is involved in deciding the exact relationship between the two clauses connected by *ma' 'isa:a'*. Hence, it is difficult to assign an exact function to a given token. However, in order to get a picture on the functions of *ma' 'isa:a'* in narrations and conversations, efforts have been made to assign to each token a single function that is considered the most appropriate. Table 3 summarises the functions of *ma' 'isa:a'*. As shown in Table 3, more frequent uses at textual level (CONJUNCTION) are attested in narration, whereas interaction uses (CONNECTION) at the interpersonal level appear to be more prevalent in conversation. Besides, backchannel uses at the interpersonal level are attested only in conversation.

Table 3: Functions of *ma' 'isa:a'*

Function		Genre / Title	conversation H	L	E	Total	narration K2	A1	F6	Total
S_P	COMPARISON		2	1	1	4 (19%)	2	0	1	3 (23.1%)
(CL1)_ (CL2)	CONJUNCTION	temporal	0	0	0	0	1	0	0	1 (7.7%)
		sequential	1	0	0	1 (5%)	2	0	0	2 (15.4%)
		consequential	1	0	0	1 (5%)	1	0	0	1 (7.7%)
		conditional	6	2	0	8 (38.1%)	4	0	0	4 (30.8%)
	CONNECTION	concessive	2	0	0	2 (9.5%)	1	0	0	1 (7.7%)
		pause	2	0	0	2 (9.5%)	0	1	0	1 (7.7%)
		backchannel	2	1	0	3 (14.3%)	0	0	0	0
Total			16	4	1	21	11	1	1	13

5 Discussion

As shown in the above section, *ma' 'isa:a'* displays three major functions according to its distribution—COMPARISON between the subject and the predicate, CONJUNCTION and CONNECTION between two clauses. Table 4 summarises the senses of *ma'* in Huang (2008) and the functions of *ma' 'isa:a'* discussed in this paper.

Table 4: Comparison on the functions of *ma'* and *ma' 'isa:a'*

ma' (Huang 2008)			***ma' 'isa:a'*** (this study)	
Force	Structure	Sense	Distribution	Sub-Function
similarity	NP_1 VP_1, NP_2 *ma'* VP_2			
listing	NP_1 *ma'* VP_1, NP_2 *ma'* VP_2	COMPARISON	S *ma' ' isa:a'* P	
periphery	(NP_1 VP_1), NP_2 *ma'* VP_2			
counter-expect	(NP_1 VP_1), NP_1 *ma'* VP_1			
indefiniteness	*ana'* INDEFINITE PRO *ma'* VP			
extremity	ana' NP/VP *ma'* VP	INCLUSION		
collectiveness	*sasa'ih/saboeh ma'* VP			
	CL_1 *ma'* CL_2	CONJUNCTION	(CL_1) *ma' ' isa:a'* (CL_2)	temporal sequential consequential
	CL_1 *ma' isa:a* CL_2 CL_1 *isaa ma'* CL_2	(CONNECTION)	(CL_1) *ma' ' isa:a'* (CL_2)	conditional concessive pause
			ma' ' isa:a'	backchannel

As demonstrated in Table 4, both the functions and distributions (or structures in Huang's (2008) terminology) of *ma' 'isa:a'* are in fact wider than those reported in Huang (2008). In terms of functions, *ma' 'isa:a'* generally follows from *ma'* with the exception of the INCLUSION sense. The INCLUSION sense as shown in Table 4 is triggered under the context of collocation with *ana'* 'even' or quantifiers such as *sasa'ih* 'each' and *saboeh* 'all' (Huang 2008). As *ma' 'isa:a'* is absent from such collocation, this sense is thus not attested. The underlining reason, following Huang's (2008:165—170) study, may be ascribed to the fact that INCLUSION and CONNECTION both involve a change in viewpoint. The following discussion will focus on the relationship between the sub-functions of CONJUNCTION and CONNECTION.

For the textual connective function subsumed under CONJUNCTION, both Huang (2008:160) and Yeh (2010, 2013) point out that *'isa:a'* functions as a connective marker even without collocating with *ma'*.[19] However, it is found that compared to *'isa:a'*, *ma' 'isa:a'* displays a stronger tendency to mark conditional relationships (see Table 3). Besides conditional, the logical relationship can also be temporal, causal, or concessive. In light of grammaticalisation (Hopper & Traugott 1993:186; Traugott 1985:292), temporal markers are the source of conditional or causal connectives.[20] Traugott (1985:295) resorts to the notion "frame" to account for the motivation of development from temporal to conditional. She argues that sentence (25) has two readings: a temporal reading where 'Bill came home' provides the temporal frame for John's leaving as well as a conditional reading in which *when* is treated as *whenever*, framing the set of occasions of John's leaving.

(25) When Bill came home, John left. (Traugott 1985:295)

[19] Yeh (2013) points out that because of its indexing or pointing nature, *'isa:a'* begins to assume the function of a connective in discourse.

[20] According to Traugott (1985:292), temporal is the commonest lexical source for conditionals.

As for the relationship between conditionals and causals, the former PRESUPPOSES a condition, whereas the latter ASSERTS a condition (Traugott 1985:297). According to Hopper and Traugott (1993:187), concessive, being more abstract and complex logically, is a late development from conditional. The following example from König (1986:239) illustrates *if* being used as a concessive.

(26) This is an interesting, <u>if</u> complicated, solution. (König 1986:239)

König (1985:267–272) remarks that concessive markers in some languages are composed of a marker conditional or temporal in origin plus an additive focus particle or emphatic particle like English *even*, *also*, or *too*.[21] Besides, due to the semantic similarity, in Old English, Middle English and Old High German, the distinction between conditional and concessive is not clear. In Old High German, both are introduced by *wenn* or *ob*, with concessive meaning being marked by extra markers such as pro-adverbials (*doch*) in the main clause or focus particles (*auch, gleich, schon* …) in the subordinate clause. These markers are combined to form Modern German concessive connectives like *wenngleich*, *wenn auch*, *obschon*, *obgleich*. Given this, it is not strange that *ma' 'isa:a'* composed of an emphatic marker *ma'* and a demonstrative *'isa:a'*, which has developed into a connective marker, can designate concessive relationship. Note that the grammaticalisation process to a concessive marker may just have begun since the concessive reading is rendered through contexts where a universal quantifier or a negator occurs, as shown in (27) and (28) respectively.

(27) Concessive (Conversation: Holiday, NTU)
103. C:..*uhN*
BC
104. K:...***hiya'*** *m-alibih* <u>*ma' isaa*</u> *itayso'-on*
who AF-take also that ignore-PF
'We don't care if somebody takes them home.'

(28) Concessive (Conversation: Life, NTU)
F and M talking about cultivating mushrooms for life …
146. F: [XX]
XX
147. F: *m*==
FILL
148. F: ...***'okik*** *wa'isan* *ka-ba:iw-an* <u>*ma' isaa*</u> ... *'am* *nak'ino'*
NEG expensive KA-buy-NMLZ also that FUT so.what
'The market price is not good. What can we do?'

Before moving on to the CONNECTION function of *ma' 'isa:a'* at interpersonal level, a fact worthy of mentioning is that 71% of the tokens (twenty out of the thirty-four) appear in utterance final position. Even if syntactically it functions to link the subject and the predicate, it is also possible for *ma' 'isa:a'* to occur in utterance final position, as illustrated in (29).

---

[21] In Malayalam the form is *enkil-um* 'if-even/also', and in Turkish, it is composed of conditional marker plus *de* 'also.'

(29) Utterance final (Narration: Kathethel 2, NTU)
30. ...(1.2) *lasia* *ma'* *isaa==*
3PL.NOM also that
31. ... *rima'* *ila* *'‹oem›alep*
go PFV <AF>hunt
32. ... *lobih*
return
'They went back from hunting.'

And even in cases where *ma' 'isa:a'* functions as a connective marker prefacing the consequence, it is common for it to appear in the final position of the antecedent clause, as in (30).

(30) Utterance final (Holiday, NTU)
131.C:...*ma==*
then
132. C:...*So:*
COND
133. C:..*hiya'*
who
134. C:...(0.9) *pinabilis* *ka* *rayhil* *ma'* *isaa*
take.charge ACC money also that
135. C:...*isaa* *rima'* *ba:iw* *ka==*
there go buy ACC
'The person in charge of the money has to buy the things.'

Note that the above extract is characterised by pauses and repetitions, indicating that the speaker is encountering difficulty and therefore *ma' 'isa:a'* here is exploited as a filler. In light of this, the distribution in final position is motivated by conceptual planning.

Another motivation for *ma' 'isa:a'* to occur in utterance final position is interaction-based. Take for example the following extract from a conversation.

(31) Utterance final (Conversation: Holiday, NTU)
48. K: [XX @ @]
XX
49. K: ... (1.0) *isaa* *isahini* *kayzaeh* *so:* *'aehae'* *taew'an* *'aehae'*
DM now good COND one family one
50. K: ... *kinae'hae'an* *ma'* *isaa*
one.bottle also that
'Now the situation is much better. So if each household brings one bottle…'
51. K: ... *'al'alak* *hara* *r‹oem›a'oe*
youth like <AF>drink
52. K: .. *o:* *wa'-wa'isan* *saboeh*
DM RED-good.at all
'The young people were very good drinkers. (There might not be enough wine.)'

As the consequence of the pre-condition 'if each household brings one bottle' is left implicit, it produces the effect of inviting the hearer to imagine the consequence. Due to this effect of getting the hearer involved, it can have the effect of both turn-holding and turn-yielding when occurring in turn-final position (cf. Koivisto 2012).[22] As shown in the following example, *ma' 'isa:a'* in IU 45 appears in the turn final position, following by the addressee's uptake.

(32) Turn-final (Conversation: Holiday, NTU)

42. C: ... *a=*
BC

43. K: .. *nak* *kakhayza'an* *<L2shiL2>:* *nakhara'*
like before [Man] like

44. K: ...(1.2) *wa'-wa'isan* *r‹oem›a'oe* *ka* *pinobae:aeh*
RED-good.at <AF>drink ACC wine
'akoy *ka* *pinobae:aeh* *tabin*
many NOM wine until

45. K:..<L2s=*antian*L2>..*san*--<L2*sanji*L2> *o:* *aw'itol* *ma' isaa*
[Man] [Jap] DM around also that
'People used to be good drinkers and used to drink a lot until around three in the afternoon.'

46. C: [@]

47. C: *haysia* [XX@@]
still xx
'(They were) still (drinking).'

As mentioned in section 4, the interlocutor can make use of *ma' 'isa:a'* to show his/her agreement with the speaker's comment. Such a function at the interpersonal level is manifested more clearly in the following extract from a conversation, where after Speaker K complained about their situation of distributing the tomb sweeping offering last year, Speaker C uptakes by talking about what they (the Chu family) did. After Speaker C ends his turn at IU 240, Speaker K takes up the turn, and his turn again contains *ma' 'isa:a'* in the final position, showing the intention to yield the turn.

(33) Backchannel (Conversation: Holiday, NTU)

235. C: .. *eN*
BC

236. K: ... <P XX P>
XX

237. C: *niya'om* *isaa* *hiza*
1EPL.GEN that there

238. C: ...(0.9) *talo*-- *sa-saboeh-en* *noka==*
FS RED-all-PF GEN

[22] Koivisto (2012) points out that pausing after a conjunction make it possible to become a point for turn-transition.

239. C: ...(1.0) *aniyami'ai* *noka* *titiyon* *no* *ma* *isa*
1EPL.POSS GEN Chu.clan ?? also that
yami *nonak* *h‹om›iwa'* *ka== 'aeyam*
1EPL.NOM self <AF>cut ACC meat
240. C: ... *tinawbon*
rice.cake
'We slice our own meat and rice-cakes.'
241. K: ... *'a'apol* *ma'* *isaa*
distribute also that
242. K: (0) *'a'apol* *ma'*/
distribute DM
'And then distribute, and then?'
243. C: (0) *ta--*
FS
244. C: ... *'a'apol* *ma'* *isaa* *pa-'apol* *ka* *saboeh* *ila*
distribute also that CAUS-distribute NOM all PFV
hiza *si-paysiri* *ila* *hiza* *ray* *sinrahoe*
that RF-move.around PFV there LOC clan
'We distribute them to all the clans.'

Around seventy percent of the *ma' 'isa:a'* in conversation are found to operate at the interpersonal level. It is speculated that *ma' 'isa:a'*, being composed of *ma'* and a demonstrative *'isa:a'*, is able to refer back and draw attention to a specific elements in the prior discourse and therefore is able to maintain cohesion in interaction.

6 Concluding remarks

In Huang (2008), the polysemy *ma'* is examined in a cross-linguistic context, and *ma' 'isa:a'* is treated as an idiomatic expression with the meaning CONNECTION. In this paper, the functions of *ma' 'isa:a'* in Saisiyat are examined in detail from a discourse analytic perspective. It is found that besides CONNECTION, *ma' 'isa:a'* displays two other functions — CONJUNCTION and COMPARISON. As is pointed out in Huang (2008), CONJUNCTION and CONNECTION operate at the textual level and the interaction level respectively. As for COMPARISON, as it applies in contexts where *ma' 'isa:a'* occurs between the subject and the predicate, it should correspond to the propositional level. In terms of grammaticalisation, the functions displayed by *ma' 'isa:a'* at the three different levels correspond to the propositional, textual and expressive meanings in Traugott (1982). Therefore, in light of the path of development, it follows the cline from propositional to textual and to expressive meanings or scope within the proposition to scope over the proposition and scope over the discourse units in Traugott & Dasher (2009:40), as represented in Figure 1.

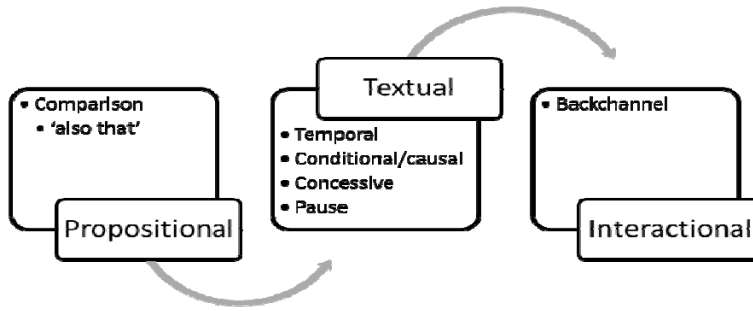


Figure 1: Development of *ma' 'isa:a'*

As the functions of *ma' 'isa:a'* mostly concern COMPARISON and CONNECTION, the mechanism of subjectivisation is also involved. Triggered by this mechanism, *ma' 'isa:a'* developed into a connective marker indexing "the speaker's view of the sequential relationship between units of discourse" (Traugott & Dasher 2009:152). It also involves intersubjectivisation, especially at the interaction level where the interaction between speaker and addressee is brought into play. In sum, while discussing the functions of *ma' 'isa:a'*, it is found that pragmatic factors come into play and must be taken into consideration. In other words, through the examination of *ma' 'isa:a'*, we are investigating the Saisiyat language in use. An obstacle encountered in this study is the lack of conversation data. Although there are two online corpora available and there are stories and play scripts transcribed and translated by native speakers, only three conversations are available in the NTU Corpus of Formosan Languages. Therefore, this study is based on forty two tokens of *ma' 'isa:a'*, with thirty four of them from the NTU Corpus of Formosan Languages. As such, an implication that can be obtained from this study is that both the documentation of conversation data and the exploration of Saisiyat on the basis of conversation analysis are areas in need of further research investigation.

References

Alba-Juez, Laura. 2009. *Perspectives on discourse analysis: Theory and practice.* Newcastle Upon Tyne: Cambridge Scholars Publishing.

Biq, Yung-O. 1989. *Ye* as manifested on three discourse planes: polysemy or abstraction? In *Functionalism and Chinese Grammar*, ed. by James H-Y. Tai and Frank F. S. Hsueh, 1–18. Chinese Language Teachers Association Monograph Series No. 1. South Orange, NJ: Chinese Language Teachers Association.

Biq, Yung-O. 1994. huihua hudongxing han yuyan shiyong [Conversational interactionality and language use]. *Proceedings of the 4th Annual Meeting of the World Chinese Language*, 227–236. [In Chinese]

Biq, Yung-O. 2001. The Grammaticalization of *Jiushi* and *Jiushishuo* in Mandarin Chinese. *Concentric: Studies in English Literature and Linguistics* 27.2:53–74.

Biq, Yung-O. 2004. From collocation to idiomatic expression: The grammaticalization of *hao* phrases/constructions in Mandarin Chinese. *Journal of Chinese Language and Computing* 14.2:73–95.

Biq, Yung-O. 2007. Lexicalization of phrases involving the distal demonstrative *na* in spoken Mandarin. In *Proceedings of the 18th Annual Meeting of the North American Conference on Chinese Linguistics* (*NACCL 18*), ed. by Janet Z. Xing, 24-41. 23–25 June 2006. Los Angeles, CA: GSIL, USC.

Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica No. 1. Taipei: Academia Sinica.

Du Bois, John, Stephan Schuetze-Coburn, Susanna Cumming and Danae Paolino. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, ed. by Jane A. Edwards and Martin D. Lampert, 45–89. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hayashi, Makoto and Kyung-eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30.3: 485–540.

Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Huang, Shuping. 2008. Analysis of polysemy and its applications to corpus documentation: A study based on Saisiyat. PhD dissertation. Taipei: National Taiwan University.

Koivisto, Aino. 2012. Discourse patterns for turn-final conjunctions. *Journal of Pragmatics* 44.10:1254–1272.

König, Ekkehard. 1985. Where do concessives come from? On the development of concessive connectives. In *Historical semantics: Historical word-formation*, ed. by Jacek Fisiak, 263–282. Berlin: Walter de Gruyter.

König, Ekkehard. 1986. Conditionals, concessive conditionals and concessive: Areas of contrast, overlap and neutralization. In *On conditionals*, ed. by Elizabeth Closs Traugott, Alice T. Meulen, Judy Snitzer Reilly and Charles. A. Ferguson, 229–246. Cambridge: Cambridge University Press.

Li, Paul Jen-kuei. 1985. The position of Atayal in the Austronesian family. In *Austronesian linguistics at the 15th Pacific Science Congress*, ed. By Andrew Pawley and Lois Carrington (eds), 257–280. Pacific Linguistics C-88. Canberra: The Australian National University.

Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In *Fillers, pauses and placeholders*, ed. by Nino Amiridze, Boyd H. Davis and Margaret Maclagan, 11–32. Typological Studies in Language 93. Amsterdam and Philadelphia: John Benjamins.

Ross, Malcolm. 1995. Reconstructing Proto-Austronesian verbal morphology: Evidence from Taiwan. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng, 727–791. Symposium Series of the Institute of History and Philology, Academia Sinica No. 3. Taipei: Academia Sinica.

Sacks, Harvey, Emanuel Schegloff and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.

Schiffrin, Deborah. 1987. *Discourse markers.* Cambridge: Cambridge University Press.

Su, Lily I-wen. 1998. Conversational coherence: The use of *ranhou* in Chinese spoken discourse. In *Selected papers of the Second International Symposium on Languages in Taiwan*, ed. by Shuanfan Huang, 167–181. Taipei: Crane Publishing Co., Ltd.

Su, Lily I-wen, Li-May Sung, Shuping Huang, Fuhui Hsieh and Zhemin Lin. 2008. NTU corpus of Formosan languages: A state-of-the-art report. *Corpus Linguistics and Linguistic Theory* 4.2:291–294.

Sung, Li-May, Lily I-wen Su, Fuhui Hsieh and Zhemin Lin. 2008. Developing an online corpus of Formosan languages. *Taiwan Journal of Linguistics* 6.2:79–117.

Thompson, Sandra A., and Anthony Mulac. 1991. A quantitative perspective on the grammaticization of epistemic parentheticals in English. In *Grammaticalization II*, ed. by Elizabeth Closs Traugott and Bernd Heine, 313–339. Current Issues in Linguistic Theory 24. Amsterdam: John Benjamins.

Traugott, Closs Elizabeth. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In *Perspectives on historical linguistics*, ed. by Winfred P. Lehmann and Yakov Malkiel, 245–271. Amsterdam: John Benjamins.

Traugott, Closs Elizabeth. 1985. On conditionals. In *Iconicity in syntax: Proceedings of a symposium on iconicity in syntax, Stanford, 24–26 June 1983*, ed. by John Haiman, 289–307. Amsterdam: John Benjamins.

Traugott, Closs Elizabeth and Richard B. Dasher. 2009. *Regularity in semantic change*. Cambridge Studies in Linguistics 97. Cambridge: Cambridge University Press.

Wouk, Fay. 2005. The syntax of repair in Indonesian. *Discourse Studies* 7.2:237–258.

Yeh, Marie M. 1991. Saisiyat structure. MA thesis. Hsinchu: National Tsing Hua University.

Yeh, Marie M. 2000. *saixiayu cankao yufa* [*A reference grammar of Saisiyat*]. Series on Formosan languages 2. Taipei: Yuan-Liou Publishing Co., Ltd. [In Chinese]

Yeh, Marie M. 2003. A syntactic and semantic study of Saisiyat verbs. PhD dissertation. Taipei: National Taiwan Normal University.

Yeh, Marie M. 2010. Grammaticalization of the demonstrative *'isa:a'* in Saisiyat . Paper read at the Workshop on Pragmatic Markers in Asian Languages: A pre-conference workshop of The 4th Conference on Language, Discourse and Cognition. 30 April 2010. Taipei: National Taiwan University.

Yeh, Marie M. 2011a. Nominalization in Saisiyat. In *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. by Foong Ha Yap, Karen Grunow-Hårsta and Janick Wrona, 561–588. Typological Studies in Language 96. Amsterdam and Philadelphia: John Benjamins.

Yeh, Marie M. 2011b. Discourse functions of *ma' 'isa:a'* in Saisiyat. Paper read at the NTU Workshop on Discourse and Grammar in Formosan languages. 8 January 2011. Taipei: National Taiwan University.

Yeh, Marie M. 2013. From referential to non-referential: Grammaticalization of the demonstrative *'isa:a'* in Saisiyat. Paper under review.

Zeitoun, Elizabeth and Ching-hua Yu. 2005. The Formosan Language Archive: Linguistic analysis and language processing. *Computational Linguistics and Chinese Language Processing* 10.2:167–200.

Zeitoun, Elizabeth, Ching-hua Yu and Cui-xia Weng. 2003. The Formosan Language Archive: Development of a multimedia tool to salvage the languages and oral traditions of the indigenous tribes of Taiwan. *Oceanic Linguistics* 42.1:218–232.

16 *Amis personal pronouns revisited*[*]

JOY J. WU

1 Introduction

This paper studies the personal pronouns in Amis, focusing on revisiting two issues discussed in Huang (1995):[1] the case marking of personal pronouns, and the morphosyntactic status of genitive pronouns. Huang has made two important claims regarding these two issues. First, she reports that there is a correlation between the case marking of pronouns and that of different classes of nouns in Amis. Second, she argues that the genitive pronouns in Amis are free forms, which is very different from the claim made in quite a few studies (e.g. He et al. 1986; Chen 1987; Yan 1992). Both of Huang's claims have been very influential and have been partially or fully followed by many subsequent works on Amis morphosyntax (e.g. D. Liu 1999; Wu 1995; E. Liu 2003; D. Lin 2013).

---

[*] I dedicate this paper to Prof. Lillian M. Huang. Ever since I first met her in my freshman year in college, Prof. Huang has played many roles in my life and career, and each role is equally important and influential. She has been a teacher, a supervisor, a mentor, and a friend to me. Her painstaking efforts in the research and revitalisation of Formosan languages have greatly inspired me and motivated me to work in the same field. Without her guidance, I would have never accomplished what I have done today. My gratitude to her can never be fully expressed in words.

I would also like to thank the following Amis informants for generously providing data and sharing their knowledge with me: Mr. Jin-long Chen (Ofad in Amis) for the data for Kakacawan Amis, Mr. Ming-yi Wu (Namoh in Amis) for the data for Paheko Amis, and Ms. Ruy-yu Chu (Lamlu in Amis), Ms. Mei-hua Chung (Tengu in Amis), and Ms. Ku-lin Liu (Kolas in Amis) for the data for Farangaw Amis. Special thanks also go to the anonymous reviewers, the other two editors of this volume and Hsiu-Chuan Liao for their valuable comments and suggestions, and to Chih-hsien Lin for the map that he drew for this paper. All the remaining errors are my own responsibility.

[1] The major findings of the two issues were first reported in Huang (1995), who investigates Amis only. Those findings were later cited in Huang et al. (1999), where pronouns in some Formosan languages are discussed from a typological perspective. Although the later collaborative work is more familiar and more frequently cited in relevant studies, it is the arguments of Huang (1995), not entirely shown in Huang et al. (1999), that will be reexamined in this work. Hence, unless necessary, I will only cite Huang (1995) in my discussion.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 385–406
Asia-Pacific Linguistics, 2015

The goals of this study are twofold. First, based on cross-dialectal data, I hope to offer a more complete picture regarding the correlations between pronominal case marking and nominal case marking mentioned in Huang (1995). Second, after reexamining the data in Huang (1995), I will argue for a bound form analysis for the genitive pronouns, following Chen (1987).

This paper is organised as follows. Section 1 offers a brief introduction to the Amis language, in particular the distribution of the sub-dialects discussed in this paper and the aspects of the grammar that are pertinent to this study. Section 2 presents Amis personal pronouns in different sub-dialects. In addition to a further investigation of Huang's (1995) observation on the correlations of case markers between nouns and pronouns, this section also presents a cross-dialectal view of the different distribution of deictic elements that occur in third person pronouns. Section 3 reviews the dispute regarding the morphosyntactic status of the genitive pronouns and proposes an analysis that supports Chen's (1987) claim. Section 4 concludes this paper and offers issues for further research.

1.1 The dialects of Amis

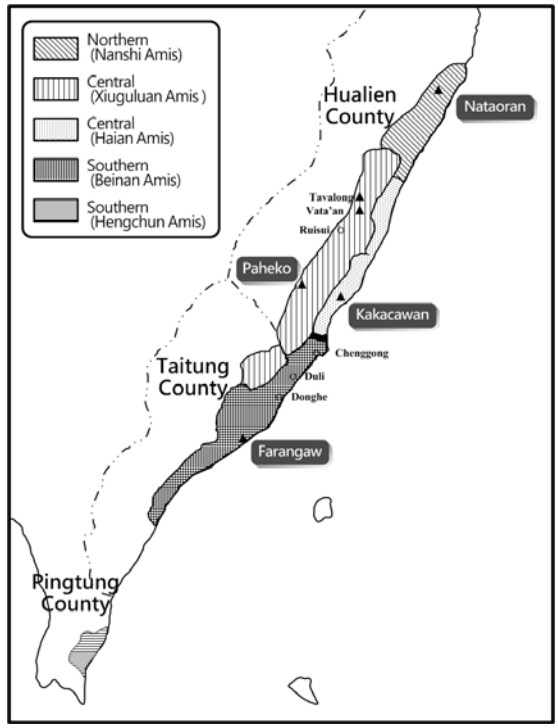
Amis is an Austronesian language mainly spoken on the east coast of Taiwan, stretching along three counties (mostly Hualien and Taitung and part of Pingtung). According to Huang (1995:218, cited from Tsuchida (1982) and P. Li (1992:3)), there are five major dialects of this language: Sakizaya (or Sakiraya),[2] Northern (often referred to as Nan-shi Amis), Tavalong-Vata'an, Central (Haian Amis and Xiuguluan Amis excluding Tavalong-Vata'an), and Southern (Beinan Amis and Hengchun Amis).[3] See Map 1 for their distribution.[4]

---

[2] Following Tsuchida's (1982, 1988) classification, Sakizaya is treated as one of the Amis dialects even though it has been officially recognised as an independent indigenous tribe and language by the government of Taiwan in 2007.

[3] According to P. Li (1997), the Central dialect is the one that is the most commonly used, whereas Sakizaya is the most conservative dialect, retaining more archaic phonological and lexical features.

[4] Notice that Sakizaya is not shown on the map. The communities of Sakizaya speakers are all located in Hualien County and are mixed with the communities of Northern, Tavalong-Vata'an, and Central Amis (especially Xiuguluan Amis; see the map in S. Lin (2007:15)). The Sakizaya examples are cited from Shen (2008). Her data were collected from Maibor, Wuhe village, Ruisui, which is located in the communities of Xiuguluan Amis as shown on Map 1. She also collected data from speakers having immigrated to Daxi, Taoyuan County, which is in the northwest of Taiwan (Shen 2008:1). Also notice that Tavalong-Vata'an is not shown as an independent dialectal area on the map; instead, it is shown as a part of Xiuguluan Amis.



Map 1: The distribution of Amis dialects[5]

Table 1 below shows the information of the sub-dialects discussed in this paper, including their Amis names (referred to in this paper), their Chinese names, the dialectal group each sub-dialect belongs to, and sources of the data.

Table 1: Information on the sub-dialects discussed in the paper

Amis	Chinese	Dialectal group	Sources of the data
Sakizaya	Saqilaiya	Sakizaya	Shen (2008)
Sakiraya		Sakiraya	
Nataoran	Nanshi	Northern	Chen (1987)
Paheko	Yuli	Central	My own field notes
Kakacawan	Changbin	Central	Wu (2006)
			My own field notes
Farangaw	Malan	Southern	My own field notes

1.2 Basic morphosyntax

This subsection offers a brief introduction to the features of Amis morphosyntax that are relevant to this research. To begin with, like most Formosan languages, Amis is verb-initial (or predicate-initial). The order of the arguments following the predicate is

[5] This map is adapted from Figure 1.2 in Wu (2006:5), which is based on a map from the following website, but the link is not working anymore: http://tcemap.gcc.ntu.edu.tw/sub_2/ethno_theme.htm#

determined by their semantic roles. For two-place actor-voice (AV) sentences, the order can be either actor-undergoer or vice versa, but the former is preferred. However, for dyadic undergoer-voice (UV) sentences, the actor has to precede the undergoer (Huang 1995:223; Huang et al. 1999:185).[6]

In addition, Amis makes a tripartite case distinction: nominative, genitive, and dative (Wu 2006).[7] The so-called case markers are actually composed of two parts: a case marker and a noun class marker. Table 2 shows the composites of case and noun class (data from Kakacawan).[8] The noun class markers are bold-faced.

Table 2: The Composites of case and noun class markers (Kakacawan)[9]

Case / Nouns		Nominative	Genitive	Dative
Common		*k-**u***	*n-**u***	*t-**u***
Personal	Singular	***Ø-ci***	*n-**i***	***ci** …-an*
Proper[10]	Plural	***Ø-ca***	*n-**a***	***ca** …-an*

As we will see later in the discussion, the case marking is also shown in the pronominal system, including personal pronouns, demonstrative pronouns (displayed in Table 3), and interrogative pronouns.

---

6 Huang et al. (1999:185) only mentions the ungrammaticality of an undergoer preceding an actor in undergoer-voice (or her patient-focus) sentences for pronominal participants. But, this observation is also true for non-pronominal participants according to my survey. However, Chen (1987) offers examples in Nataoran in which the undergoer can precede the actor in UV sentences. There might be dialectal differences regarding word order restrictions.

7 The dative case has been analysed as locative (e.g., Huang 1995, for the personal proper noun set), accusative (e.g., Huang 1995, for the common noun set; Wu 2000) and oblique (e.g., Kuo & Sung 2010). All of these works are based on data from Kakacawan.

8 Unlike personal pronouns, case markers do not exhibit significant differences among the sub-dialects except for their pronunciation (e.g., *ci* in Kakacawan, but *ci* or *si* in Farangaw). Hence, I only list those collected from Kakacawan.

9 The phonemic symbols used in the transcription follow the orthography published by the Ministry of Education and the Council of Indigenous Peoples in 2005 (http://www.edu.tw/userfiles/url/20121127164752/aboriginal.pdf). This system generally follows IPA, with the following exceptions: /c/ stands for the voiceless palato-alveolar affricate [ts], /e/ for schwa [ə], /d/ for the voiceless lateral [ɬ], /ng/ for [ŋ], /y/ for the palatal glide [j], /'/ for the epiglottis stop [ʔ]. Notice that in Amis, /o/ and /u/ are in free variation and can be used in orthography interchangeably though there exists dialectal preference in using one instead of the other. The glosses generally follow Wu (2006), who adopts some of D. Liu's (1999) morphemic analyses (e.g., the separation of noun classifiers and case markers) with some slight modifications. Glossing conventions follow the Leipzig Glossing Rules. Additional glossing abbreviations are as follows: ASP, aspect, AV, actor voice, CONJ, conjunction; CN, common noun; LNK, linker; NV, neutral voice; PPN, personal proper noun; PREP, preposition; RED, reduplication; UV, undergoer voice.

10 One reviewer suggests that the word 'proper' should be removed, as every personal noun is also a proper noun. However, in Amis orthography, both personal and non-personal proper nouns (e.g., the personal name Sawmah and the place name Taipei) are capitalised initially although the latter are case-marked as common nouns. In addition, non-proper personal nouns such as *teacher* are also marked as common noun. Hence, for better clarification, I keep the feature 'proper' in the tables and the glosses.

Table 3: Demonstrative pronouns (Kakacawan)

		Nominative	Genitive	Dative	Gloss
Proximal		*ku-ni*	*nu-ni*	*tu-ni*	'this'
Distal	Visible	*ku-ra*	*nu-ra*	*tu-ra*	'that'
	Invisible	*ku-ya; ya*	*nu-ya*	*tu-ya*	'that'

As shown in Table 3, the demonstrative pronouns are distinguished in terms of case, proximity between the speaker and the referent, and visibility. Comparing the forms, one can easily see that the proximal meaning is designated by the *-ni* part while the distal meaning is expressed by the *-ra* or (-)*ya* parts in the pronouns. Ross (2006) mentions that third person pronouns in Formosan languages "either are derived from or still are demonstrative pronouns" (Ross 2006:536). This observation is borne out in Amis as the deictic elements *-ni* and *-ra* also appear in the third person pronouns. However, as we will see later, the distribution of these elements varies cross-dialectally.

Amis makes a two-voice distinction: actor voice (AV, i.e., actor as subject) and undergoer voice (UV, i.e., undergoer as subject), the latter of which includes a plain undergoer subject, an instrumental applicative subject, and a locative applicative subject (Wu 2006).[11] Amis has been claimed to be a split ergative language (Chen 1987; Wu 2006), but displays syntactic ergativity in the formation of relative clauses and cleft sentences (Wu 2006). After this brief introduction to the Amis language, the main discussion of this paper is presented below.

2 Personal pronouns in different sub-dialects

This section compares the forms of the personal pronouns in five sub-dialects that belong to the four of the five main dialect groups mentioned in section 1.1. In particular, I will discuss the forms of the case markers in pronouns, and the deictic elements present in third person pronouns. As this discussion is inspired by the observation made in Huang (1995), I will first review her findings in the following section.

2.1 Huang's (1995) observations

Huang (1995) is the first work that offered a thorough description of the case marking and pronominal systems in Amis (with data collected from Kakacawan). One of her important findings is that Amis seems to treat first and second person pronouns, both singular and plural, as common nouns, and third person pronouns as proper nouns based on the observation that the case markers for common nouns (e.g., *k-*) in the case system also appears in first and second person pronouns while those for personal proper nouns (i.e., *ci-* and *ca-*) show up in third person set. This can be illustrated from the nominative set in Table 4 below.

[11] These instrumental and locative applicative forms have been treated as voice forms in earlier studies of Amis (e.g., Yan 1992 (based on Dulik, which belongs to the Southern dialect); Wu 2000; E. Liu 2003 (based on Kakacawan)). That is, Amis was claimed to display a four-voice (or a four-focus) distinction: actor voice, undergoer/patient voice, instrumental voice, and locative voice. The two-voice analysis follows Wu (2006).

Table 4: Nominative personal pronouns (Kakacawan)

Number	Person		Nominative
Singular	1st		***k**aku*
	2nd		***k**isu*
	3rd		***ci**ngra*
Plural	1st	Inclusive	***k**ita*
		Exclusive	***k**ami*
	2nd		***k**amu*
	3rd		***ca**ngra*

After investigating more sub-dialects, I found that the case marking correlations between nouns and pronouns actually vary among them. In addition, the dialectal variation is also found in the presence of the deictic elements (i.e., *-ni* and *-ra*) in third person pronouns. These findings are reported below.

2.2 The data

In this section, personal pronouns from different Amis sub-dialects will be presented. Let us first take a look at the data from Nataoran (which belongs to the Northern dialect) in Table 5.

Table 5: Nataoran personal pronouns (based on Chen 1987:135)

Number	Person		Top	Nominative	Genitive	Locative
Singular	1st		*ako*	*kako*	*nako* *nomako* *ako*	*itakoan*
	2nd		*iso*	*kiso*	*niso* *nomiso* *iso*	*itisoan*
	3rd		*cira*	*cira*	*nira*	*iciraan*
Plural	1st	Inclusive	*ita*	*kita*	*mita* *nomita* *ita*	*itamian*[12]
		Exclusive	*ami*	*kami*	*niam* *niniam*[13]	*ititaan*
	2nd		*amo*	*kamo*	*namo* *amo*	*itamoan*
	3rd		*ohni*	*kohni*	*nohni* *ohni*	*itohnian*

[12] I suspect that *itamian* and *ititaan* are reversed in Table 5 based on paradigmatic comparisons.

[13] I suspect that this is a typo. I think the correct form should be *noniam* based on the paradigm, and this is actually the form that appears in Chen's (1987) examples (e.g., Chen 1987:61, 64, 85, 91 and 105).

As we can see, first and second person pronouns in Nataoran are treated as common nouns as the case marking elements for common nouns are also found in the two sets, but there is a difference among third person pronouns. third person singular pronouns are treated as personal proper nouns (marked by *c-*) while the plural ones are treated like common nouns (marked by *k-*). A difference is also found in the distribution of the deictic elements; the distal part (*-ra*) shows up in the third person singular set while the proximal part (*-ni*) is found in the plural set.

Next, let us examine the pronouns from Sakizaya based on Shen (2008).

Table 6: Sakizaya personal pronouns (based on Shen 2008:41)

Number	Person		Nominative	Genitive	Oblique/Locative	Possessive
Singular	1st		*kaku*	*aku*	*taku-an* *taku-anan*	*nu maku*
	2nd		*kisu*	*isu*	*tisu-an* *tisu-anan*	*nu misu*
	3rd		*ciniza* *cingza*	*niza*	*ciniza-an*	*niza*
Plural	1st	Inclusive	*kita*	*ita*	*kita-nan*	*nu mita*
		Exclusive	*kami*	*niyam*	*tami-anan*	*nu niyam*
	2nd		*kamo*	*namu*	*tamu-an* *tamu-anan*	*nu namu*
	3rd		*kuheni* *kunuheni*	*heni*	*tu*(*nu*)*heni-an*	*nu heni*

As Table 6 shows, Sakizaya displays the same phenomena as those in Nataoran. That is, first and second person pronouns as well as third person plural pronouns are treated as common nouns. third person singular pronouns are treated as personal proper nouns. The correlations of the presence of the deictic markers in third person pronouns are also the same in the two sub-dialects. That is, the distal part is found in the singular set, but the proximal marker is observed in the plural pronouns in Sakizaya.

Next, let us move on to the sub-dialect spoken in Paheko, Hualien County.

Table 7: Paheko personal pronouns

Number	Person		Nominative	Genitive	Dative	Possessive
Singular	1st		*kako*	*ako*	*takowanan*	*mako*
	2nd		*kiso*	*iso*	*tisowanan*	*miso*
	3rd		*cira* *ciira* *cingra*	*nira* *niira* *ningra*	*ciraan* *ciiraan* *cingraan* *cingranan*	*nira* *mira* *ningra*
Plural	1st	Inclusive	*kita*	*ita*	*kitanan*	*mita*
		Exclusive	*kami*	*niyam*	*kamiyanan*	*niyam*
	2nd		*kamo*	*namo*	*tamoanan*	*namo*
	3rd		*caira* *cangra*	*naira* *nangra*	*cairaan* *cangraan*	*naira* *nangra*

The noun class correlation between the nominal and pronominal case marking in Paheko is slightly different from that in Nataoran and Sakizaya. While first and second person

pronouns have the same case marking elements for common noun, third person pronouns all have the case markers for personal proper nouns regardless of number. Moreover, only the distal deictic marker (i.e., *-ra*) shows up in third person pronouns in this sub-dialect. The description for Paheko can also be applied to the pronouns in Kakacawan, given in Table 8 below.

Table 8: Kakacawan personal pronouns (Wu 2006)

Number	Person		Nominative	Genitive	Dative	Possessive[14]
Singular	1st		*kaku*	*aku*	*takuwanan*	*maku*
	2nd		*kisu*	*isu*	*tisuwanan*	*misu*
	3rd		*cingra*	*nira*	*cingranan*	*nira*
				ningra		*ningra*
Plural	1st	Inclusive	*kita*	*ita*	*kitanan*	*mita*
		Exclusive	*kami*	*niyam*	*kamiyanan*	*niyam*
	2nd		*kamu*	*namu*	*tamuanan*	*namu*
	3rd		*cangra*	*nangra*	*cangraan*	*nangra*

Finally, let us take a look at the pronouns in Farangaw, spoken in Taitung City.

Table 9: Farangaw personal pronouns

Number	Person		Nominative	Genitive	Dative	Possessive
Singular	1st		*kaku*	*aku*	*takuwanan*	*maku*
	2nd		*kisu*	*isu*	*tisuwanan*	*misu*
	3rd		*ku mira*	*ira*	*tu mira*	*mira*
				nira		
				nu mira		
Plural	1st	Inclusive	*kita*	*ita*	*kitanan*	*mita*
		Exclusive	*kami*	*niyam*	*kamiyanan*	*niyam*
	2nd		*kamu*	*namu*	*tamuanan*	*namu*
	3rd		*ku mira*	*nu mira*	*tu mira*	*mira*

The pronouns in Farangaw exhibit different phenomena compared to the sub-dialects that we have seen so far. Almost all the pronouns have the same case-marking element as common nouns. The only exception is one of the third person genitive pronouns, *nira*. Moreover, in spite of sharing a common feature with Paheko and Kakacawan in having only the distal marker found in third person pronouns, these pronouns are composed of a case marker and the possessive pronominal noun (e.g. *ku mira*).

2.3 Interim summary

Table 10 below summarises the case marking information in personal pronouns, featuring the noun class resemblance between nouns and pronouns.

[14] The possessive pronouns are called possessive pronominal nouns in Wu (2006:85) as they behave like nouns. However, for the convenience of comparison, I will simply label them as "possessive pronouns" in this paper.

Table 10: Noun class marking resemblance between nouns and personal pronouns

Sub-dialect	Person			
	1st	2nd	3rd	
			singular	plural
Nataoran	common	common	personal proper	common
Sakizaya	common	common	personal proper	common
Paheko	common	common	personal proper	
Kakacawan	common	common	personal proper	
Farangaw	common	common	common	

Third person pronouns in Nataoran and Sakizaya show a split in terms of the noun class marking elements; the singular ones are treated like personal proper nouns while the plural ones are marked like common nouns. In Paheko and Kakacawan, all third person pronouns are treated like personal proper nouns while in Farangaw, the same set of pronouns are regarded as common nouns.

How about the presence of the deictic elements in third person sets? This summary is given below:

Table 11: Presence of the deictic elements in third person pronouns

Sub-dialect	Third person pronouns	
	Singular	Plural
Nataoran	distal (*-ra*)	proximal (*-ni*)
Sakizaya	distal (*-za*)	proximal (*-ni*)
Paheko	distal (*-ra*)	
Kakacawan	distal (*-ra*)	
Farangaw	distal (*-ra*)	

If we combine the information of Table 10 and Table 11 above, we find that there seems to be an increase of neutralisation of forms in the personal pronouns from north to south in terms of the geographical location of the sub-dialects. This neutralisation further affects the number distinctions of third person pronouns in Farangaw, especially for the nominative set and the dative set. This is illustrated in Table 12, with examples from the nominative set only. We can see that the third person singular pronoun and the plural one share the same form. As for the genitive pronouns, the same ambiguity arises in Farangaw when *nu mira* is used instead of *ira* or *nira*.

Table 12: Formal distinctions of third person nominative pronoun in different sub-dialects

Sub-dialect	Singular	Plural
Nataoran	*cira*	*kuhni*
Sakizaya	*ciniza*	*kuheni*
Paheko	*ciira*	*caira*
Kakacawan	*cingra*	*cangra*
Farangaw	*ku mira*	*ku mira*

3 Morphosyntactic status of genitive pronouns

The second issue I would like to revisit is the analysis of the genitive pronouns mentioned in Huang (1995) as opposed to the possessive pronouns. Her paradigm is given in Table 13:[15]

Table 13: Genitive and possessive pronouns (Kakacawan, from Huang 1995:234)

	Genitive	Possessive
1SG	*aku*	*maku*
2SG	*isu*	*misu*
3SG	*nira; ningra*	*nira; ningra*
1PL.INCL	*ita*	*mita*
1PL.EXCL	*niyam*	*niyam*
2PL	*namu*	*namu*
3PL	*nangra*	*nangra*

There are two proposals regarding the morphosyntactic status of the genitive set of pronouns. One treats them as bound forms. The representative work is Chen (1987), which explicitly terms them as clitics (Chen 1987:136).[16] The other is a free form analysis, proposed by Huang (1995). These two proposals will be reviewed respectively in sections 3.2 and 3.3. Although Huang (1995) has given reasons for why she thinks these pronouns should be treated as free morphemes, one cannot ignore the fact that the distribution of the genitive pronouns is indeed more constrained than the possessive set, as I will describe in section 3.1. Hence, in this section, I will re-examine Huang's evidence and argue for the clitic analysis made in Chen (1987). My reassessment is given in section 3.4.

3.1 Features of the genitive pronouns and the possessive pronouns

Although both genitive and possessive pronouns can designate a possessor and an actor in the UV constructions, the genitive pronouns behave differently from the possessive set in at least three aspects.

First, only the possessive pronouns can be preceded by case markers to show different arguments in a sentence; the genitive pronouns cannot be preceded by any case markers.

(1)	a.	*Ma-nengneng*	*aku/n-u*	*maku*	*kisu.*
		UV-watch	1SG.GEN/GEN-CN	1SG.POSS	2SG.NOM
		'I saw you.'			

[15] As one can see in Table 13, although Huang (1995) lists two sets of pronouns, there seems to be syncretism (the shaded cells) between third person singular, first person plural exclusive, second person plural, and third person plural sets. I will only focus on those sets that do not show syncretism in my discussion.

[16] In addition to Chen (1987), there are other studies which adopt the bound form analysis, such as He et al. (1986) and Yan (1992), but these two studies do not mention what kind of bound morpheme these pronouns serve. Notice that none of the above-mentioned works is based on Kakacawan. He et al. (1986) collected their examples from Madawdaw (Chenggong) and Fafokod (Donghe) and Yan's (1992) data were from Dulik (Duli); these places all belong to Beinan Amis (the Southern dialect) as one can see in Map 1.

a'. * *Ma-nengneng* ***n-u*** ***aku*** *kisu.*
UV-watch GEN-CN 1SG.GEN 2SG.NOM
'I saw you.'

b. *Mi-ala* *Ø-ci* *aki* ***t-u*** ***maku***
AV-take NOM-PPN Aki DAT-CN 1SG.POSS
atu ***misu.***
CONJ 2SG.POSS
'Aki is going to take mine and yours.'

b'. * *Mi-ala* *Ø-ci* *aki* ***t-u*** ***aku*** ***atu*** ***isu.***
AV-take NOM-PPN Aki DAT-CN 1SG.GEN CONJ 2SG.GEN
'Aki is going to take mine and yours.'

c. *Ma-ra~ramod* *k-u* *wawa* *isu*
NV-RED~get.married NOM-CN child 2SG.GEN
atu *n-u* *maku,* *ma-nga'ay* *tu?*
CONJ GEN-CN 1SG.POSS NV-good ASP
'Is that OK if your child and mine get married?'

c'. * *Ma-ra~ramod* *k-u* *wawa* *isu*
NV-RED~get.married NOM-CN child 2SG.GEN
atu *aku* *ma-nga'ay* *tu?*
CONJ 1SG.GEN NV-good ASP

The examples in (1) indicate that genitive pronouns cannot be preceded by case markers (1a'), nor can they be conjoined (1b') and (1c').

Second, in terms of word order, genitive pronouns can only follow the possessum while possessive pronouns can appear before or after it. In other words, a genitive pronoun has to be adjacent to the head noun. Compare their difference in (2):

(2) a. *wawa* ***aku***
child 1SG.GEN
'my child'

a'. * ***aku*** *a* *wawa*
1SG.GEN LNK child

b. *wawa* *n-u* ***maku***
child GEN-CN 1SG.POSS
'my child'

b'. (*n-u*) ***maku*** *a* *wawa*
GEN-CN 1SG.POSS LNK child
'my child'

Third, possessive pronouns can appear independently as an answer to a question while genitive pronouns cannot. Consider:

(3) a. Q: *Nima* *wawa* *k-u-ni?*
who.GEN child NOM-CN-this
'Whose child is this?'

b. A: (*N-u*) *maku.*
GEN-CN 1SG.POSS
'Mine.'

b' A: * *Aku.*
1SG.GEN

The above-displayed differences, especially the second and third ones, suggest that genitive pronouns are bound forms while possessive pronouns are free forms. This is indeed the proposal made in some previous studies such as He et al. (1986), Chen (1987), and Yan (1992). Among them, Chen (1987) is the one that explicitly mentions that the genitive pronouns are clitics. However, Huang (1995) argues for the free form status of the genitive forms, and her analysis has been adopted by many studies later (e.g. Wu 2006 and E. Liu 2003). The two analyses will be reviewed respectively in the following two sections.

3.2 Genitive pronouns as bound forms: Chen's (1987) analysis

We have seen the pronominal system provided by Chen (1987) in Table 5. In fact, according to Chen, genitive pronouns in Nataoran show three realisation patterns. To facilitate the discussion, I label them as Gen_1, Gen_2, and Gen_3 in Table 14 below:

Table 14: Nataoran genitive personal pronouns (based on Chen 1987:135)

	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
TOP	*ako*	*iso*	*cira*	*ita*	*ami*	*amo*	*ohni*
GEN_1	*nako*	*niso*	*nira*	*mita*	*niam*	*namo*	*nohni*
GEN_2	*nomako*	*nomiso*		*nomita*	*niniam*		
GEN_3	*ako*	*iso*		*ita*		*amo*	*ohni*

Chen (1987:136) regards the forms in Gen_2 as the fullest realisation, those in Gen_1 as the intermediate form, and Gen_3 as the short form. She makes two special remarks about the short forms. First, she proposes that these short forms are clitic pronouns as "…even though their attachment to the N does not result in a shift in stress position, nothing is allowed to intervene between the head N and the short form of a possessive pronoun" (Chen 1987:136). Second, she claims that a genitive agent is never realised as a short form. In other words, a short form can only denote a possessor. Chen's clitic analysis has also been adopted in Ross's (2006) reconstruction of the pronouns in Proto Amis, as seen in his Gen_1 set in Table 15:

Table 15: Pronominal paradigm of Proto Amis (based on Ross 2006:555)

	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
NEUT	*aku*	*isu*	*c-ira*	[*k*]*ita*	*ami*	*amu*	*uhni*
NOM	*k-aku*	*k-isu*	*c-ira*	*k-ita*	*k-ami*	*k-amu*	*k-uhni*
GEN_1	*=aku*	*=isu*	--	*=ita*	--	*=amu*	*=uhni*
GEN_2	*m-aku* (*n-aku*)	*m-isu* (*n-isu*)	*n-ira*	*m-ita*	*niam*	*n-amu*	*n-uhni*
OBL	[*i*]*t-aku-an*	[*i*]*t-isu-an*	[*i*]*c-ira-an*	[*i*]*t-ita-an* [*i*]*kita-an*	[*i*]*t-aman*	[*i*]*t-amu-an*	[*i*]*t-uhnian*

Chen's short forms are analogous to Huang's genitive pronouns. However, unlike Nataoran, the Kakacawan genitive pronouns can denote an agent (or actor) (as in (1a)). Moreover, other elements can intervene between the genitive pronouns and their heads. It is the latter feature that has led Huang (1995) to conclude that the genitive pronouns in Kakacawan should be regarded as free forms. Her proposal is summarised in the next section.

3.3 Genitive pronouns as free forms: Huang's (1995) analysis

Huang's (1995) argument lies on the status of the aspectual marker *tu*, which can appear between the verb and the genitive pronoun as seen in (4) (Huang 1995:233).

(4) *Ma-nanum* ***tu*** ***aku*** *k-u* *sayta.*
UV-drink ASP 1SG.GEN NOM-CN cider
'I already drank the cider.'

My further investigation on Kakacawan has found that *tu* can also appear between a genitive pronoun that manifests a possessor and its head. Consider the nominal predicate in (5):

(5) *O* *singsi* ***tu*** ***aku*** *t-u* *ro~mi'a~mi'ad*
CN teacher ASP 1SG.GEN DAT-CN day~RED~day
Ø-ci *sawmah.*
NOM-PPN Sawmah
'Sawmah is my instructor for every day (for sure now).'

The intervention of *tu* between the genitive pronoun and its preceding element challenges the bound-form analysis of the genitive pronoun proposed by Chen (1987), unless this aspectual marker is also a bound element. However, Huang argues that *tu* is a free form (instead of a bound morpheme, particularly, an affix) based on the following examples, especially (6a) and (6c) (Huang 1995:233, original gloss, emphasis mine).

(6) a. *Uya* *ci* *rapih* ***tu***.[17]
that NOM Rapih ASP
'Rapih has become a different person.' (She was too happy.)

b. *Han* *ni* *puda'* *kuya* *wawa* *nira* *ci* *rapih.*
HAN GEN Puda' that.NOM child 3SG.GEN NOM Rapih
'Puda' told his child Rapih so.' (This sentence appears after a direct quote.)

c. *Suwal* ***han*** ***tu*** ***aku*** *ku* *wawa* *aku.*
tell HAN ASP 1SG.GEN NOM child 1SG.GEN
'I will tell my child.'

Huang analyses (6a) as an equational sentence and *tu* as an aspectual marker functionally similar to *le* in Mandarin Chinese. In her opinion, "it is not legitimate to regard it as being attached to the preceding noun and as a bound form" (Huang 1995:233). Her further

[17] Both (6a) and (6b) are cited from a text. A small part of the text is provided in the appendix of the paper. See the appendix of Huang (1995) for the whole text.

evidence for treating *tu* as a free form and consequently *aku* (the genitive form) also as a free form is from (6c), in which there is an element *han* preceding *tu* and *aku*. She analyses *han* as a free morpheme based on its position in (6b), though *han* is left unglossed in both of the examples.

3.4 My reassessment of Huang's analysis

Huang's study was carried out at least twenty years ago,[18] at a time when the grammar of Amis was not as fully investigated as now. My further investigation of Amis shows that the functions of *tu* are actually more complicated than we knew about twenty years ago, and there are two more possible explanations of (6a). Consider the sentences in (7).

(7) a. *Palu-en* *n-i* *dongi*$_i$ *cingra*$_i$ ***tu***.
beat-UV UV-PPN Dongi 3SG.NOM REFL
'Dongi beat herself.'

b. *Ma-sti'* ***tu*** *aku* *kaku* ***tu***.
UV-beat ASP 1SG.GEN 1SG.NOM REFL
'I have beaten myself.'

c. *Kisu* ***tu***.
2SG.NOM TU
'It is your turn (to do something).' (aspectual marker reading)
'You (do something) yourself.' or 'How about you yourself?' (reflexive marker reading)

In addition to being an aspectual marker, it may also be a form with a referring function to the preceding noun. This later function is frequently found in a reflexive construction such as (7a) and (7b) (Wu 2006; Shen 2008). The two different functions of *tu* can co-occur in a sentence, as illustrated in (7b).

The *tu* in (7c), structurally similar to (6a), is ambiguous as it can have an aspectual reading or a reflexive reading. Nevertheless, if *tu* in (7c) functions as an aspectual marker, it is possible to receive an emphatic stress, but emphatic stress never falls on the reflexive *tu*. Based on this further understanding of *tu*, the translations of (6a) may either be 'Now it is Rapih's turn to say something' or 'How about Rapih herself?' besides the one provided by Huang (1995). These translations all fit the following discourse after (6a) in the original text as there is a direct quote following this sentence that describes Rapih's comments. See (8b) (Huang 1995:233, original gloss).

(8) a. *Uya* *ci* *rapih* ***tu***.
that NOM Rapih ASP
'Rapih has become a different person.' (She was too happy.) (Huang's translation)[19]
'Now it is Rapih's turn to say something.' (My translation 1)
'How about Rapih herself?' (My translation 2)

---

[18] Huang's data were collected from 1992 to 1994 (Huang 1995:218, n. 1).

[19] The translation provided in the appendix is 'Rapih then has become a different person', which is slightly different from the one given in the main text. See the cited excerpt in the appendix.

b. *U ma-tini-ay ku suwal nu mama aku*
U MA-this-AY NOM word GEN father 1SG.GEN
ha:i saan.
approve so.said
'Since my father said so, I agreed (Rapih said).'

No matter which function *tu* serves in the sentence, it is impossible to perfectly guarantee its morphosyntactic status as a free form. If it is a reflexive marker, it has to be bound to an NP that it refers to. If it is an aspectual marker, it has to follow a predicate. Hence, we still cannot say that a genitive pronoun is a free form solely based on the intervention of *tu*.

Now, let us examine another piece of evidence employed by Huang (1995): the morphosyntactic status of *han*. The morpheme *han* originally is a quotative verb, which always follows or precedes a direct quote (Wu 1995). This function is also indicated in (6b).[20] However, the functions of *han* seem to have evolved, and it has become a marker in constructions such as the ideophone-forming construction that describes a manner in (9a)[21] (Wu 2006) or the disposal construction[22] in (9b) (E. Liu 2003; Wu 2006).

(9) a. *Sa-rayaray han ita pa-sasuluy k-u-ni*
form-row like.this[23] 1PL.INCL.GEN CAUS-pass NOM-CN-this
a anengang itiya kalamkam.
LNK chair ITIYA fast
'Let's line up (and in this way) and we can pass this chair quickly.'

b. *Tanam han n-i kacaw k-u*
taste like.this GEN-PPN Kacaw NOM-CN
dateng, anger.
vegetable bitter
'Kacaw (then) tasted the vegetable and (found that it was) bitter.'

In constructions exemplified in (9), *han* is preceded by another predicate that appears in its root form. Moreover, the entire predicate (i.e., the root predicate + *han*), carries undergoer voice marking function (E. Liu 2003; Wu 2006; Shen 2008) as one can see from the case-marking pattern of the examples. Note that *han* cannot be stressed in these constructions; it can only be stressed in sentences like (6b), when it serves as a quotative verb. These features all indicate that the *han* in (9) is not the same as the quotative verb in

---

[20] Its preceding direct quote is given in the appendix (Sentences #1 and #2) of the paper.

[21] The two examples in (9) are cited from Wu (2006:440, 401). The glosses are mine.

[22] The disposal construction is a construction headed by V + *han*, in which the whole predicate V + *han* carries a meaning of treating someone or handling something with force (Tsai & Zeng 1997; E. Liu 2003). According to E. Liu (2003:88–89), this construction has two functions. First, it may emphasise the way of handling something. Second, it may emphasise the degree/result of handling. Examples in (9) are related to its second function.

[23] Wu (2006) originally glosses *han* as 'so.said'. One reviewer suggests that *han* should be glossed as HEDGE. However, I do not think this morpheme serves as a hedge, at least in the disposal construction. Another reviewer has pointed out that it is very common cross-linguistically for a quotative verb to have a corresponding 'similitude' meaning, and could be glossed as 'like.this'. I adopt the latter suggestion and regloss *han* in (9).

(6b). The unit *suwal han* in (6c) is structurally more like the constructions in (9). It is thus possible to treat *han* as a bound form or as part of a complex verb[24] and *tu* in (6c) as an enclitic that has to attach to the complex verb. The enclitic analysis of *tu* is also adopted in NTU Corpus of Formosan Languages and D. Lin (2013).[25]

The reinvestigation of *tu* and *han* has cast some doubts over Huang's proposal of treating the genitive pronouns as free forms, but it goes well with the enclitic analysis proposed by Chen (1987:136) even though the absence of intervening forms between the genitive pronoun and its head that is mentioned by Chen is not observed in Kakacawan. In fact, in addition to *tu*, other morphemes such as the aspectual marker *ho* 'still; yet' and the quantifier *aca* 'only; a little' can also appear between the genitive pronouns and the predicate or the nominal head in Kakacawan. Examples follow:

(10) a. *Nanum-en* ***ho*** *aku.*
drink-UV ASP 1SG.GEN
'I will drink it first.'

b. *Tenger-en* ***aca*** ***aku*** *k-u* *titi.*[26]
cook.till.tender-UV only 1SG.GEN NOM-CN meat
'I will cook the meat a little longer.'

c. *Tenger-en* *aku* *k-u* *titi* ***aca.***
cook.till.tender-UV 1SG.GEN NOM-CN meat only
'I will cook only the meat.'

It is highly possible that the intervening elements between the predicates and the genitive pronouns in (10) are all bound elements. For example, the aspectual marker *ho*, analogous to *tu*, has also been treated as a clitic in D. Lin (2013). The quantifier *aca* has to follow the elements that it quantifies (cf. (10b) and (10c)). If we view all those intervening elements as clitics, then Amis seems to display clitic stacking in examples like (10a), (10b), and (10c)).[27] Although we still need more direct evidence to support this proposal, the bound-form (in particular, the clitic) analysis better accounts for the distribution and the functions of the genitive pronouns, and it also fits better with most of the Formosan languages in which clitic pronouns are also found.

---

[24] Shen (2008:44) analyses *han* in this construction as a suffix that carries voice-marking function. However, this analysis does not seem appropriate, at least for Kakacawan, as *han* in this positions cannot receive stress like other suffixes (e.g., the voice marker *-en*) in this sub-dialect. It seems to behave more like an unstressed enclitic.

[25] None of the two sources provide justifications for why *tu* is treated as an enclitic in Amis.

[26] I would like to thank Cheng-chuen Kuo for sharing his finding on the position of *aca* with me.

[27] Similar clitic stacking is found in many Malayo-Polynesian languages such as Tagalog (Schachter 1973), and in Formosan languages, such as Isbukun Bunun (L. Li 2010). For example, in Tagalog, there are two types of clitics: clitic personal pronouns and clitic particles. When they appear together (i.e., stack), they follow certain order constraints; for instance, clitic particles always precede di-syllabic clitic pronouns (Schachter 1973:218). The clitic stacking in Amis seems to follow this constraint as well.

4 Conclusion and suggestions for future research

This paper revisits two important claims of the personal pronouns in Amis proposed in Huang (1995) based on data collected from other sub-dialects and further investigation of the Amis grammar. First, the cross-dialectal data have given a more complete picture regarding the case marking correlation between nouns and pronouns, which was initially reported in Huang (1995). It is found that in Nataoran and Sakizaya, first and second person pronouns are case-marked like common nouns. As for third person pronouns, the singular pronouns are treated like personal proper nouns whereas the plural ones are marked like common nouns. Both Paheko and Kakacawan have their first and second person pronouns case-marked like common nouns, but their third person ones are all marked like personal proper nouns. Finally, in Farangaw, almost all the personal pronouns are case-marked like common nouns with only one exception (i.e., *nira*). Besides the above-mentioned differences, these sub-dialects also show variations in terms of the presence of the deictic element in third person pronouns. For example, the distal element *-ra* is found in the singular pronouns in Nataoran and Sakizaya, but it is the proximal element *-ni* that is found in the plural ones. In the other three sub-dialects investigated in this paper, namely, Paheko, Kakacawan, and Farangaw, only the distal element is found in third person pronouns. The second issue that is revisited in this study is the morphosyntactic status of the genitive pronouns. After reassessing the argument in Huang (1995), I have suggested that the clitic analysis proposed by Chen (1987) is better supported as it not only accounts for the data but also fits the overall observation in most Formosan languages, and many Malayo-Polynesian languages as well.

Nevertheless, this research can never be brought to an end as there are still quite a few issues that wait for answers. For example, during the investigation of the pronominal forms, we can see an increase of formal neutralisation in the personal pronouns along from the north to the south in terms of the geographical location of the sub-dialects. Interestingly enough, such structural neutralisation in Farangaw is also found in the other two structures: the case marking patterns of the *sa-…-aw*/*sa-…-an* optative constructions and the use of the conjunctors. Table 16 below shows the comparison of the two structures in Kakacawan and Farangaw (Wu 2009, 2011):[28]

Table 16: Comparison of two structures in Kakacawan and Farangaw

Sub-dialect	Case patterns of optative constructions		Use of conjunctors	
	sa-V-*aw*	*sa*-V-*an*	Common nouns	Personal proper nouns
Kakacawan	UV pattern	AV pattern	*atu*	*aci*
Farangaw	AV pattern		*atu*	

As displayed in the table, while Kakacawan shows two case marking patterns (i.e., AV and UV) for the optative constructions, Farangaw only has the AV pattern for the two different optative verbs. Moreover, two conjunctors, *atu* and *aci*, are employed in Kakacawan to connect different types of NPs (*atu* for both common and personal proper nouns but *aci* for personal proper nouns only) in the NP conjunction constructions, only *atu* is found in Farangaw. In other words, the structural distinctions found in Kakacawan

[28] The data of the two structures in the other three sub-dialects is not available at this moment.

are neutralised in Farangaw. Will more cases of neutralization be found in Farangaw or other Southern sub-dialects? Further investigation is needed.

Another point that deserves further study is the deictic forms in third person pronouns. Although it is not typologically unique that third person pronouns are related partially or completely to the demonstrative pronouns (Ross 2006; Bhat 2007), the correlation between grammatical number and distance in Nataoran and Sakizaya (i.e., singular/distal and plural/proximal) seems to be an interesting issue left unexplored. Related to these questions are the historical development of the forms of Amis pronouns, and the grammaticalisation patterns and analogical processes by which they were formed, issues that are discussed in detail in Reid (2007).

Finally, if the aspectual marker *tu* and the genitive pronouns are both clitics, does Amis also possess different kinds of clitics (e.g. phrasal clitics and head-adjacent clitics)[29] as shown in other Formosan languages such as Isbukun Bunun (L. Li 2010)? Such issues require more investigation.

[29] One reviewer points out that there are two homophonous *le* in Mandarin Chinese; one is an enclitic to the verb and the other is positioned after the object. The former encodes aspect and is a head-adjacent clitic whereas the latter denotes a currently relevant state and seems to be a phrasal clitic. e/she further suggests the verb-final *tu* in Amis is like the verb-final *le* while the *tu* that appears after an NP is like the second *le* in Mandarin. I would like to thank him/her for offering such a direction for future research. Furthermore, I have found that the co-occurrence of the aspectual markers such as *tu* and *ho* and the quantifier *aca* is not allowed in Kakacawan. Does this suggest that some clitics are not allowed to be stacked in Amis? This is also an issue that requires more investigation.

Appendix

The following excerpt is taken from the appendix (sentences #41 to #50) of Huang (1995). I have retained the original glosses but rewritten them with the glossing conventions adopted in this paper.

(1) *mangaay kuni alu-fa'inay isu.*
may this.NOM act-husband 2SG.GEN
'That man can be your husband.'

(2) *unini ku fa'inay isu, tata'ak ku ka-hmek*
this NOM husband 2SG.GEN greatly NOM -happy
niyam, a ma'ulip tu kami tu ina isu.
3NOM.GEN[30] LNK life ASP 1PL.EXCL LOC mother 2SG.GEN
'If he is your husband, we will be very happy and your mother and I will live very well.'

(3) *han ni puda' kuya wawa nira ci rapih.*
HAN GEN Puda' that.NOM child 3SG.NOM NOM Rapih
'Puda' told their child Rapih so.'

(4) *ha:i saan.*
approve so.said
'(Rapih) agreed'

(5) *ya ka-matiya ma-hahuding tulu a rumi'ad a*
that -alike MA-hold.contest three LNK day LNK
ma-hahuding, pa-ta-maku~maku han tu nira ci sawmah.
MA-hold.contest -smoke~RED HAN ASP 3SG.GEN NOM Sawmah
'During the three-day contest, she (Rapih) invited Sawmah to smoke'

(6) *ulah han tu ni sawmah.*
love HAN ASP GEN Sawmah
'Sawmah (then) loved Rapih'

(7) *ha:i satu.*
approve
'(Sawmah) approved (then)'

(8) *ma-hemek tu.*
MA-happy ASP
'(They two) were very happy'

[30] I think this is wrongly glossed. It should be a first-person plural exclusive genitive pronoun.

(9) *uya ci rapih tu.*
that NOM Rapih ASP
'Rapih then became a different person'

(10) *u ma-tini-ay ku suwal nu mama aku ha:i saan.*
U MA-this-AY NOM word GEN father 1SG.GEN approve so.said
'Since my father said so, I agreed (Rapih said).'

References

Bhat, D.N.S. 2007. *Pronouns*. Oxford: Oxford University Press.

Chen, Teresa M. 1987. *Verbal constructions and verbal classification in Nataoran-Amis*. Pacific Linguistics C-85. Canberra: The Australian National University

He, Rufen, Zeng Siqi, Tian Zhongshan and Lin Dengxian. 1986. *gaoshanzu yuyan jianzhi: ameisi yu* [*The sketch of the languages of mountain tribes* (*Amis*)]. Beijing: Min-Zu Publishing Co. [In Chinese]

Huang, Lillian M. 1995. The case markers and pronominal system in Amis. *The Journal of National Chengchi University* 70.1:217–58.

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1999. A typological overview of pronominal systems of some Formosan languages. In *Selected papers from the Fifth International Conference on Chinese Linguistics*, ed. by H. Samuel Wang, Feng-fu Tsao, and Chin-fa Lien, 165–198. Taipei: Crane Publishing Co., Ltd.

Kuo, Cheng-chuen and Li-May Sung. 2010. On the syntax of Amis comparative constructions. *Concentric: Studies in Linguistics* 36.1:25–57.

Li, Lilian Li-ying. 2010. Clitics in Nantou Isbukun Bunun (Austronesian). MA Thesis. Nantou: National Chi Nan University.

Li, Paul Jen-kuei. 1992. *Orthographic systems for Formosan languages*. Taipei: Ministry of Education, Republic of China.

Li, Paul Jen-kuei. 1997. A syntactic typology of Formosan languages -- case markers on nouns and pronouns. In *Chinese Languages and Linguistics: Typological studies of languages in China IV*, ed. by Tseng Chiu-yu, 343–378. Symposium Series of the Institute of History and Philology, Academia Sinica No. 2. Taipei: Academia Sinica.

Lin, Dong-yi. 2013. Interrogative constructions in Kavalan and Amis. PhD dissertation. Gainesville: University of Florida.

Lin, Shih-hui. 2007. Some investigation in Sakizaya: Language description and reference grammar. Project Report for National Science Council (NSC 95-2411-H-259-006)

Liu, Dorinda Tsai-hsiu. 1999. Cleft constructions in Amis. MA Thesis. Taipei: National Taiwan University.

Liu, Emma En-hsing. 2003. Conjunction and modification in Amis. MA Thesis. Hsinchu: National Tsing Hua University.

Reid, Lawrence A. 2007. Another look at the marking of plural personal noun constructions in Austronesian languages. *Oceanic Linguistics* 46.1:232–252

Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto-Austronesian. In *Streams converging into an ocean: Festschrift in honor of professor Paul Jen-Kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M.

Huang and Dah-an Ho, 521–563. Language and Linguistics Monograph Series Number W-5. 2006. Taipei: Institute of Linguistics, Academia Sinica.

Schachter, Paul. 1973. Constraints on clitic order in Tagalog. *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, ed. by Andrew B. Gonzalez, F.S.C., 214–231. Quezon City: Linguistic Society of the Philippines.

Shen, Wen-chi. 2008. Sakizaya syntax: With special reference to negative, interrogative, and causative constructions. MA Thesis. Taipei: National Taiwan University.

Tsai, Zhuenghan. and Siqi Zeng. 1997. *ameizu muyu yufa jiegou fenxi* [*The structural analysis of Amis grammar*]. Taipei: Taiwan Aboriginal Foundation. [In Chinese]

Tsuchida, Shigeru. 1982. Subclassification of Amis dialects. Unpublished Manuscript.

Tsuchida, Shigeru. 1988. Amis. In *The Sanseido encyclopaedia of linguistics Vol. 1: Languages of the world, Part one*, ed. by Takashi Kamei, Rokuro Kono and Eiichi Chino, 447–449. Tokyo: Sanseido Press.

Tsukida, Naomi. 1993. The use of *-en* form in Fataan-Amis. *Asian and African Linguistics* 22:123–140.

Wu, Jing-lan Joy. 2000. *ameiyu cankao yufa* [*A reference grammar of Amis*]. Formosan Language Series 11. Taipei: Yuan-Liou Publishing Company. [In Chinese]

Wu, Jing-lan Joy. 1995. Complex sentences in Amis. MA Thesis. Taipei: National Taiwan Normal University.

Wu, Jing-lan Joy. 2006. Verb classification, case marking, and grammatical relations in Amis. PhD dissertation. Buffalo: State University of New York at Buffalo.

Wu, Jing-lan Joy. 2009. On the optative mood constructions *Sa-….-aw* and *Sa-….-an* in Amis. Paper Presented at The 11th International Conference on Austronesian Linguistics, 22-25 June, Aussois, France.

Wu, Joy. 2011. Noun phrase conjunction in Amis. *Language and Linguistics* 12.1:109–139.

Yan, Zhi-kuang. 1992. ameiyu de yufa jiegou: canyuzhe yu shijian de yanjiu [Syntactic structure of Amis: A study of participants and events]. MA Thesis. Taipei: National Chengchi University. [In Chinese]

Zeng, Siqi. 1991. *taiwan ameisiyu yufa* [*Amis grammar*]. Beijing: Central Min-zu College Press. [In Chinese]

17 *A comparative study of the personal pronominal systems in three Puyuma dialects**

STACY F. TENG

1 Introduction

1.1 Some background information

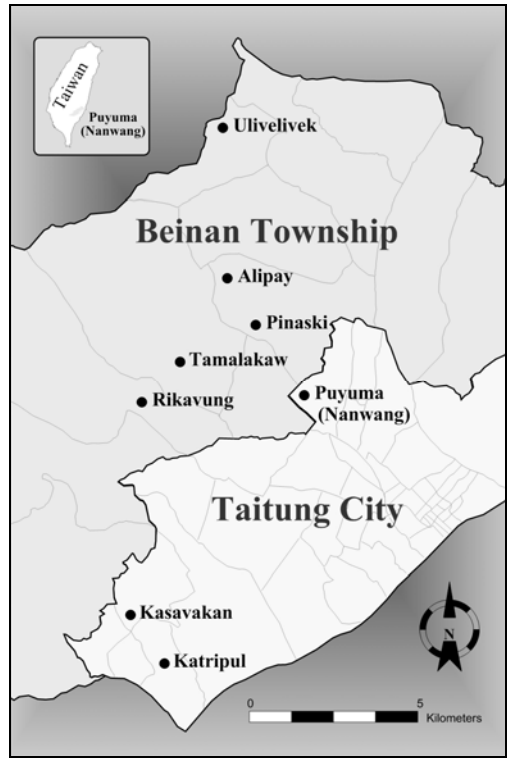
In Huang et al.'s (1999) pioneer survey of the pronominal systems of Formosan languages (including Amis, Atayal, Saisiyat, Bunun, Tsou, Rukai, Paiwan, Puyuma and Kavalan), Puyuma exhibits the most complex systems among the languages under study. Among the different dialects of Puyuma,[1] it is found that there is a great degree of diversity in terms of both forms and functions. Part of the pronominal system of

* It is my great pleasure to dedicate this paper to my MA supervisor, Prof. Lillian M. Huang, who introduced me to the field of Austronesian linguistics. She has been very supportive and encouraged me all the way in my research. This paper presents results of a research project supported by the National Science Council (NSC 99-2410-H-001-096). Parts of this paper were presented at the Workshop on "Discourse and grammar in Formosan languages", National Taiwan University, 8 January 2011, and at the Workshop on a "Typological study of Austronesian languages in Taiwan and their revitalisation", Academia Sinica, 1–2 December 2012. I wish to thank the participants for the discussions and remarks. I would also like to show my appreciation to Malcolm Ross, Hsiu-chuan Liao, Amy Pei-jung Lee, Joy Jing-lan Wu, Rui-wen Wu, Rik De Busser, Loren Billings, Maya Yu-ting Yeh, and two anonymous reviewers for their constructive and insightful comments on an earlier draft version. A special word of thanks must go to Elizabeth Zeitoun, who has read and commented on the paper many times. They are, however, not responsible for any remaining errors and omissions.

[1] These three dialects are chosen for comparison due to the following reasons. In many previous studies (e.g., Ting 1978; Li 1991, among others), Puyuma is usually regarded as consisting of two dialects from a phonological point of view, including the one spoken in Nanwang and the one spoken elsewhere (Katripul is usually treated as the representative). Teng (2009, 2011a), on the other hand, proposes that at least one more dialect should be distinguished based on morphosyntactic evidence. Thus, Puyuma may tentatively be said to consist of three dialects: Nanwang, Katripul (including Katripul and Kasavakan), and the one spoken in the other five villages (including Ulivelivek, Tamalakaw, Rikavung, Alipay, and Pinaski). Tamalakaw is chosen as the representative of the third dialect because it is relatively well studied and there are two important pieces of work published (see Tsuchida 1980, 1995).

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 407–429
Asia-Pacific Linguistics, 2015

Proto-Puyuma has been reconstructed in Ross (2006:554–555) based on the dialects spoken in Nanwang and Tamalakaw (see appendix I). However, the paradigms are not complete and the functions of these forms are not touched upon. In drawing data from one more dialect spoken in Katripul, the present paper seeks to provide a fuller picture of the forms and functions of the pronominal systems of Puyuma dialects,[2] and to determine what the diachronic changes are and how they may have caused dialectal differences. The geographical setting of these villages is given in Map 1.



Map 1: Geographical setting of the Puyuma dialects

The present study compares the personal pronominal systems spoken in three Puyuma villages: Nanwang, Katripul and Tamalakaw.[3]

---

[2] Puyuma pronouns display some peculiar features that are seldom mentioned in previous studies on Formosan languages, including their predicative uses vs non-predicative uses, the marking of irrealis in pronouns, pronominal derivations (as verbs), and the functional extension of suffixal pronouns, etc. However, due to the scope of the present paper, these will not be included in the discussion here.

[3] The sources of data include: (i) Nanwang: Ogawa & Asai (1935), Teng (2008), and fieldnotes and texts that I have collected over the past ten years; (ii) Katripul: Ogawa & Asai (1935), Quack (1981), and fieldnotes and texts that I have collected over the past four years; (iii) Tamalakaw: Tsuchida (1980, 1995) and fieldnotes and texts that I have collected over the past four years. I would like to express my gratitude to the following Puyuma consultants: Pu'ay Kiringan (A-Sin Chan), Risem Kadadepan (Er-Lang Yiu), Ukesang Mavaliw (Chin-Mei Chen), and Awkayan Pawkayan (Chin-Mei Lin).

1.2 An overview

Four categories of pronouns are distinguished in terms of their morphological properties and the environments in which they occur. Table 1 gives an overview of the four categories of pronouns in Puyuma.

Table 1: An overview of the four categories of pronouns in Puyuma

	Clitic pronouns	Free pronouns	Suffixal pronouns	Combined forms
Morphological status	clitics	portmanteau words	suffixes	prefixes
Complete paradigm or not	YES	YES	NO	NO
Functions	denoting core arguments	denoting topics, possessors, and oblique arguments	mainly used in kin terms	denoting core arguments

Examples from the dialect spoken in Katripul are given in (1).[4]

(1) Katripul[5]

a. Clitic pronoun
tu*=*pa*-*ka*-**vi’as-*aw*=*ku *zananku* *kualrengan*.
3.GEN=CAUS-STAT-hot-UV:P=1SG.NOM 1SG.PSR.OBL illness
‘My illness made me have a fever.’

b. Free pronoun
ta=pa-ka-vulray-aw ***nanu*** *d‹in›away*.
1PL.EXCL.GEN=CAUS-STAT-beautiful-UV:P 2SG.PSR.NOM <PFV.NMLZ>produce
‘Let us make your product beautiful.’

c. Suffixal pronoun
atras *i* *wadi-**li***.
tall SG.NOM younger.sibling-1SG.PSR
‘My younger brother is tall.’

d. Combined form
kunu*-*kedreng*-*aw
1SG.GEN+2SG.NOM-pull.out-UV:P
‘I pulled you out.’

The paper is organised in the following way. In sections 2–5, forms and functions of the four categories of pronouns in the three named dialects are compared. The diachronic changes that may have caused the variations will also be discussed. Section 6 deals

[4] Abbreviations used in this paper that are not included in the Leipzig Glossing Rules are: AV, actor voice; NEUT, neutral; PSR, possessor. The abbreviations UV:P, UV:L, and UV:C indicate that the named verb is in undergoer voice form, with a patient [P], location [L] or a conveyance theme [C] as subject respectively.

[5] Unless specified otherwise, the pronouns in the three dialects exhibit the same properties when examples are given in one dialect.

specifically with the change of position of the clitic pronouns in Katripul and Tamalakaw. Section 7 presents a summary of findings.

2 Clitic pronouns

2.1 Forms

A paradigmatic overview of the clitic pronouns in the three dialects under study is given in Table 2. Differences among the dialects are marked in color.

Table 2: Puyuma clitic pronouns

			Singular			Plural			
			1	2	3	1 INCL	1 EXCL	2	3
NOM	Subj	Nanwang	*=ku*	*=yu*	–	*=ta*	*=mi*	*=mu*	–
		Katripul	*=ku*	*=u*	–	*=ta*	*=mi*	*=mu*	–
		Tamalakaw	*=ku*	*=u*	–	*=ta*	*=mi*	*=mu*	–
	PSR	Nanwang	*ku=*	*nu=*	*tu=*	*ta=*	*niam=*	*mu=*	*tu=*
		Katripul	*ku=*	*nu=*	*tu=/taw=*	*ta=*	*niam=*	*mu=*	*tu=/taw=*
		Tamalakaw	*ku=*	*nu=*	*tu=/taw=*	*ta=*	*niam=*	*mu=*	*tu=/taw=*
GEN		Nanwang	*ku=*	*nu=*	*tu=*	*ta=*	*niam=*	*mu=*	*tu=*
		Katripul	*ku=*	*nu=*	*tu=/taw=*	*ta=*	*mi=/niam=*	*mu=*	*tu=/taw=*
		Tamalakaw	*ku=*	*nu=*	*tu=/taw=*	*ta=*	*mi=/niam=*	*mu=*	*tu=/taw=*

The second singular nominative subject in Nanwang is *=yu* while it is *=u* in the other two dialects. The third person nominative PSR and genitive is *tu=* in Nanwang whereas in the other dialects there is another form *taw=* serving the same function. The first exclusive plural nominative PSR and genitive is *niam=* in Nanwang, whereas in Katripul and Tamalakaw there is one more form *mi=* serving as genitive. The diachronic reasons for causing these variations will be given in section 6.1.

2.2 Functions

There is a three-way distinction among the clitic pronouns: the NOM manifests either the subject (e.g., *=yu* 'you (sing.)' in (2a)), or the possessor (PSR) of the subject (e.g., the second *ku=* 'my' in (2b)). The GEN denotes the non-subject actor (e.g., the first *ku=* 'I' in (2b)). Note also that PSR and GEN are syncretic, with the exception of the third person and first person plural exclusive forms.[6] In cases where the GEN clitics and the PSR clitics have the same form, a GEN clitic cannot be replaced with a free pronoun, while a PSR clitic can, as shown in (2b') and (2b'').

(2) Nanwang

a. *s‹em›anan=yu* *i* *dalran.*
<AV>stray=2SG.NOM LOC road
'You lost your way.'

[6] For more detailed discussion on case syncretism, interested readers are referred to Teng (2009).

b. ***ku****=belriyas-aw* ***ku****=tilril.*
1SG.GEN=turn.over-UV:P 1SG.PSR.NOM=book
'I turned my book over.'

b'. *ku=belriyas-aw* *nanku* *tilril.*
1SG.GEN=turn.over-UV:P 1SG.PSR.NOM book
'I turned my book over.'

b''. * *nanku* *belriyas-aw* *ku=tilril.*
1SG.PSR.NOM turn.over-UV:P 1SG.PSR.NOM=book

2.3 Diachronic changes

In section 2.1 three groups of variations among the clitic pronouns are given (i.e., *=yu* vs *=u* '2SG.NOM', *tu=* vs *taw=* '3.GEN/PSR', *niam=* vs *mi=* '1PL.EXCL.GEN/PSR'). The possible diachronic explanations for these variations will be discussed in this section. Before doing that, let us first have a look at Ross's (2006) reconstructed PAn personal pronouns, which are given in Table 3.

Table 3: PAn personal pronouns, based on Ross (2006:532)

	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL
NEUT	i-aku	iSu[qu]	s-ia	ita	i-ami	i-mu[qu] i-amu
NOM1	aku	Su[qu]	ia	(i)ta	ami	mu[qu], (amu)
NOM2	=ku =[S]aku	=Su	(-ya)	=(i)ta	=mi[a] =[S]ami	=mu
ACC	i-ak-en	iSu[qu]-n	---	ita-en	i-ami-n	i-mu[qu]-n
GEN1	=[a]ku	=Su	(-ya)	=(i)ta	=mi[a]	=mu
GEN2	(=)m-aku	(=)m-iSu	---	(=)m-ita	((=)m-ami)	(=)m-amu
GEN3	n-aku	n-iSu	n-ia	ni-ta	**ni-am** **n-ami**	ni-mu n-amu

When we compare Table 2 with Table 3, we see that the Puyuma clitic pronouns generally reflect Ross's NOM2 and GEN1, with the following exceptions, which also correspond to the three groups of variations observed in section 2.1.

(i) While Tamalakaw and Katripul's second personal singular nominative form *=u* reflects the PAn NOM2 *=Su unambiguously (i.e., PAn *S became zero in Puyuma), the corresponding form in Nanwang is *=yu*.

(ii) The third person forms *tu=/taw=* are innovative; they are not in Ross's reconstructed forms.

(iii) The Puyuma PSR/GEN form *niam=* in all three dialects reflects PAn GEN3, not the expected GEN1.

In addition, the forms in PAn GEN1 are enclitics, while synchronically, GEN and PSR are proclitics in modern Puyuma dialects. The issue regarding the change of position of these clitics will be further addressed in section 6.

In what follows, I will discuss the three sets of variations (i.e., *=yu* vs *=u* '2SG.NOM', *tu=* vs *taw=* '3.GEN/PSR', *niam=* vs *mi=* '1PL.EXCL.GEN/PSR') in turn.

First, regarding the form *=yu*, the appearance of an extra glide /j/ may be a result of reanalysis.[7] Two phonological rules are involved here, namely glide insertion and resyllabification. Let us first assume that the second singular nominative clitic was *=u* before reanalysis took place in Nanwang. In Nanwang, a homorganic glide is inserted when a high vowel is followed by an unlike vowel. So, when a verb ended with /i/ and there was a second singular nominative enclitic, /j/ was inserted. For instance:

(3) Glide insertion

a. ***madreki=u*** *dra* *trau.*
/maɖəki**j**u/
scold=2SG.NOM INDF.OBL person
'You scolded people.'

b. ***tawasi=u****=diya* *dra* *kiruan*?
/tawasi**j**u/
brush=2SG.NOM=yet INDF.OBL clothes
'Have you brushed clothes?'

Resyllabification happened when there was a second singular nominative pronoun following the voice markers *-ay* or *-anay*, as shown in (4a) and (4b).

(4) Resyllabification

a. *pasisi-**ay=u*** /pa.si.si.a**j.u**/ → /pa.si.si.a.**ju**/
(*pasisi* 'force sb. to do something)

b. *lukluk-**anay=u*** /luk.luk.a.na**j.u**/ → /luk.luk.a.na.**ju**/
(*lukluk* 'wrestle')

The pressure for reanalysis was reinforced because in the paradigm, only the second singular nominative enclitic lacked an onset. Thus, to make the paradigm look more regular, an onset was added to the pronoun *=u*, and the most available candidate was, of course, /j/.

With regard to *tu=/taw=*, while Nanwang only has *tu=* in the clitic pronoun paradigm, the other two dialects have both forms. There are two possibilities to explain the variation. We may either assume that they were free variations in Proto-Puyuma, and in the modern Nanwang dialect, *taw=* is lost, while Katripul and Tamalakaw still preserve both forms. Or, we may assume that there was only *tu=* in Proto-Puyuma, and *taw=* is innovated in Katripul and Tamalakaw. The first hypothesis is more probable because also the form /*taw*/ occurs in free neutral pronouns in the Nanwang dialect (e.g., *taytaw* '3.NEUT'; *nantaw* '3.PSR.NOM'; *kantaw* '3.OBL').

Last, I turn to the discussion of the form *niam=*. Unlike *=yu*, for which I discard the possibility of its coming from the NEUT set, there is a good reason for me to assume that *niam=* comes from the GEN3 set in Ross's reconstruction and replaced the original form *mi=* due to the effect of homonymy avoidance (Hopper & Traugott 2003:102–103; De Vogelaer & Coussé 2011).

---

[7] The form *=yu* may superficially look like a reflex of *iSu '2SG.NEUT', but this possibility is rejected because there seems to be no reason for the speakers to abandon the original form *=Su '2SG.NOM' and replace it with a form from another set.

In all dialects of Puyuma, there is a prefix *mi-* denoting existential, instrumental and possessive meanings (cf. Teng 2014), and this morpheme can, at least, be reconstructed to Proto-Puyuma, if not in a higher node. To avoid possible ambiguity, speakers of Proto-Puyuma used *niam=* instead of *mi=* when they wanted to express a possessive meaning. For example:

mi-paisu	'have money'	*niam=paisu*	'our money'
mi-walak	'have children'	*niam=walak*	'our children'

The replacement is extended to the GEN category in the Nanwang dialect, but is incomplete in the other two dialects, where *niam=* and *mi=* both occur, as summarised below.

	PSR	GEN
Tamalakaw	*niam=*	*mi=/niam=*
Katripul	*niam=*	*mi=/niam=*
Nanwang	*niam=*	*niam=*

The effect of homonymy avoidance has caused another change in the Nanwang verbal construction. As pointed out earlier (see also Teng 2008 and Teng & Ross 2010), a replacement of a bound pronoun with a free pronoun is a crucial test of whether the following head is verbal or nominal. That is, while a clitic PSR can be replaced by a free PSR, a genitive is always bound, as shown in (2) and (5).

(5) Nanwang

a. ***ku=****belriyas-aw* ***ku=****tilril*.
1SG.GEN=turn.over-UV:P 1SG.PSR=book
'I turned my book over.'

b. *ku=belriyas-aw* ***nanku*** *tilril*.
1SG.GEN=turn.over-UV:P 1SG.PSR book
'I turned my book over.'

c. * ***nanku*** *belriyas-aw* *ku=tilril*.
1SG.PSR turn.over-UV:P 1SG.PSR=book

However, the test mentioned above fails to serve as a diagnostic when the pronoun is 1PL.EXCL, as in (6).

(6) Nanwang

a. ***niam=****belriyas-aw* ***niam=****tilril*.[8]
1PL.EXCL.GEN=turn.over-UV:P 1PL.EXCL.PSR.NOM=book
'We turned our books over.'

b. *niam=belriyas-aw* ***naniam*** *tilril*.
1PL.EXCL.GEN=turn.over-UV:P 1PL.EXCL.PSR.NOM book
'We turned our books over.'

[8] As mentioned earlier, in Katripul and Tamalakaw, both *mi=* and *niam=* are acceptable before the verb, and thus the replacement is partial. Whether *naniam* is acceptable to occur before the verb needs further research.

c. ***naniam*** *belriyas-aw* *niam=tilril.*
1PL.EXCL.PSR.NOM turn.over-UV:P 1PL.EXCL.PSR.NOM=book
'We turned our books over.'

In this section I have (i) compared the forms/functions of the clitic pronouns, (ii) identified three groups of variations among the dialects, and (iii) given diachronic explanations for the variations. It is found that when a variation occurs, it is always between Nanwang on the one hand and Katripul and Tamalakaw on the other, and Nanwang is always the most innovative one morphosyntactically.

3 Free pronouns

Four sub-categories of free pronouns can be distinguished in terms of the functions they play in sentences, including manifesting a topic or being used in reflexive situations (Neutral), denoting an oblique argument (Oblique), indicating the possessor of a subject (Nominative PSR), or being the possessor of a non-subject (Oblique PSR). While nominative PSRs are always definite, a further distinction between definite vs indefinite is made among the oblique PSRs. In the following discussion, I will first compare the forms (section 3.1) and then discuss their functions individually (section 3.2). Table 4 gives a summary of free pronouns in the three dialects.

Table 4: Free pronouns in Nanwang, Katripul and Tamalakaw

			Singular			Plural			
			1	2	3	1 INCL	1 EXCL	2	3
NEUT	Nanwang		*kuiku*	*yuyu*	*taitaw*	*taita*	*mimi*	*muimu*	—
	Katripul		*inku*	*inu*	*intaw*	*inta*	*iniam*	*inmu*	—
	Tamalakaw		*inku*	*inu*	*intaw*	*inta*	*iniam*	*inmu*	—
OBL	Nanwang		*kanku*	*kanu*	*kantaw*	*kanta*	*kaniam*	*kanemu*	—
	Katripul		*kaninku*	*kaninu*	*kanintaw*	*kaninta*	*kaniniam* *kaniam*	*kanimu*	—
	Tamalakaw		*kananku* *kaninku*	*kanu* *kaninu*	*kantaw* *kanintaw*	*kanta* *kaninta*	*kaniniam* *kaniam*	*kanmu* *kanimu*	—
NOM PSR	Nanwang		*nanku*	*nanu*	*nantu*	*nanta*	*naniam*	*nanemu*	*nantu*
	Katripul		*nanku* *ninku*	*nanu* *ninu*	*nantu* *nantaw* *nintu* *nintaw*	*nanta* *ninta*	*naniam* *niniam*	*nanemu* *ninemu*	*nantu* *nantaw* *nintu* *nintaw*
	Tamalakaw		*nanku* *ninku*	*nanu* *ninu*	*nantu* *nantaw* *nintu* *nintaw*	*nanta* *ninta*	*naniam* *niniam*	*nanmu* *ninmu*	*nantu* *nantaw* *nintu* *nintaw*
OBL PSR	Nanwang	DEF	*kanku* *kananku*	*kanu* *kananu*	*kantu* *kanantu*	*kanta* *kananta*	*kaniam* *kananiam*	*kanemu* *kananemu*	*kantu* *kanantu*
		INDF	*draku* *drananku*	*dranu* *drananu*	*dratu* *dranantu*	*drata* *drananta*	*draniam* *drananiam*	*dramu* *drananemu*	*dratu* *dranantu*

			Singular			Plural			
			1	2	3	1 INCL	1 EXCL	2	3
OBL	Katripul	DEF	*kanku*	*kanu*	*kantu*	*kanta*	*kaniam*	*kanemu*	*kantu*
PSR			*kananku*	*kananu*	*kanantu*	*kananta*	*kananiam*	*kananemu*	*kanantu*
			kaninku	*kaninu*	*kanintu*	*kaninta*	*kaniniam*	*kaninemu*	*kanintu*
					kanintaw				*kanintaw*
		INDF	*zaku*	*zanu*	*zatu*	*zata*	*zaniam*	*zamu*	*zatu*
			zananku	*zananu*	*zanatu*	*zananta*	*zananiam*	*zananemu*	*zanantu*
			zaninku	*zaninu*	*zanintu*	*zaninta*	*zaniniam*	*zaninemu*	*zanintu*
	Tamalakaw	DEF	*kaninku*	*kaninu*	*kanantu*	*kaninta*	*kaniam*	*kanemu*	*kanantu*
			kananku	*kananu*	*kanintu*	*kananta*	*kananiam*	*kananemau*	*kanintu*
					kanintaw		*kaniniam*	*kaninemu*	*kanintaw*
		INDF	*zaninku*	*zaninu*	*zanantu*	*zaninta*	*zaniam*	*zanemu*	*zanantu*
			zananku	*zananu*	*zanintu*	*zananta*	*zananiam*	*zaninemu*	*zanintu*

3.1 Forms

Free pronouns are portmanteau words; they consist of noun phrase markers and the clitic pronouns discussed in section 2. In the following discussion, I will take Katripul as an example to demonstrate how free pronouns are formed. Before doing that, let us have a look at the noun phrase markers in this dialect.

Table 5: Noun phrase markers in Katripul Puyuma

	Personal nouns		Common nouns	
	Singular	Plural	Definite	Indefinite
NOM	*i*	*na*	*na*	*a*
GEN	*ni*	*nina*	*nina/na*	*za*
OBL	*kani*	*kana*	*kana*	*za*

The formation of free pronouns can be formulated as (7):

(7) Neutral: i + /N/ + clitic pronoun[9]
Oblique: *ka*+ /N/ + neutral pronoun;[10]
Nominative possessor: *na* or *ni* + /N/ +clitic pronoun;
Definite oblique possessor: *ka* + /N/ + clitic pronoun; *ka* + /N/ +free nominative possessor
Indefinite oblique possessor: *za* + clitic pronoun; *za* + free nominative possessor; note that the element /N/ after *za* is usually omitted.

Examples for first person singular and first person inclusive plural are given below:

[9] It is suggested that in forming free pronouns, there is usually a linking element /N/ between the noun phrase marker and the clitic pronoun.

[10] Alternatively, we may say that it is *kan*, rather than *ka* that is involved in the formation process. That is, it is always a noun phrase marker (*i*, *na*, *ni*, *za* and *kan*) plus /N/, and because the final nasal in *kan* and the linking element /N/ are the same, only one /n/ is kept. This alternative analysis may better capture the unity of the formation process.

	clitic form	NEUT	OBL	NOM.PSR	OBL.PSR
1SG	*=ku*	*i-n-ku*	*ka-n-inku*	*na-n-ku* *ni-n-ku*	*ka-n-ku* *ka-n-nanku* *ka-n-ninku* *za-(n)-ku* *za-(n)-nanku* *za-(n)-ninku*
1PL	*=ta*	*i-n-ta*	*ka-n-inta*	*na-n-ta* *ni-n-ta*	*ka-n-ta* *ka-n-nanta* *ka-n-ninta* *za-(n)-ta* *za-(n)-nanta* *za-(n)-ninta*

As we can observe from Table 4, Nanwang again displays some differences from Katripul and Tamalakaw. These variations fall into two groups: The first is about the formation of neutral pronouns, and the second has to do with the genitive noun phrase marker *ni*. The diachronic changes that cause the variations will be discussed in section 3.3.

3.2 Functions

Unlike the clitic pronouns, where the functions the pronouns play in each dialect are in general the same, some subcategories of free pronouns have distinct functions in different dialects. In the following discussion, I will first present their similarities and then their differences (if any) will also be compared.

3.2.1 Neutral free pronouns

In Nanwang, neutral pronouns have four major functions: (i) to affirm identities (reply to an interrogative sentence starting with *i manay* 'who'), as in (8a); (ii) to encode an emphatic meaning, as in the first free translation given in (8b); (iii) to indicate a reflexive meaning, as in the second free translation given in (8b); and (iv) to appear in a topic position, as in (8c).

(8) Nanwang

a. Q: *i* *manay* *na* *s‹em›a~senay*?
SG.NOM who DEF.NOM <AV>IPFV~sing
'Who's singing?'
A: *kuiku*.
1SG.NEUT
'Me.'

b. *ku=na'u-ay* *kuiku*.
1SG.GEN=see-UV:L 1SG.NEUT
'I myself looked after it.' / 'I looked at/looked after myself.'

c. *kuiku* *i,* *a* *Puyuma*.
1SG.NEUT TOP INDF.NOM Puyuma
'As for me, I am a Puyuma.'

In addition to the functions mentioned above, Katripul and Tamalakaw neutral forms may be treated as personal proper nouns, i.e., they can be case-marked. But all forms are treated as singular. Note that in those cases, the clitic pronoun still needs to occur obligatorily, as shown in (9b).

(9) Katripul

a. *ma-rengay* *i* *intaw*.
AV-say SG.NOM 3.NEUT
'He said, ...'

b. *ma-zalram=**mu*** *i* *inmu*.
AV-know=2PL.NOM SG.NOM 2PL.NEUT
'You all know.'

In Tamalakaw, the neutral forms are also used in inclusive constructions (cf. Teng 2011b) to convey the comitative/conjunctive expression.

(10) Tamalakaw

a. *m-ukua=**mu*** *i* *zenan* ***inmu*** *i* *Senayan*.
AV-go=2PL.NOM LOC mountain 2PL.NEUT SG.NOM Senayan
'You and Senayan went to the mountains.'

b. *m-ukua=**mi*** *i* *zenan* ***iniam*** *i* *Senayan*.
AV-go=1PL.EXCL.NOM LOC mountain 1PL.EXCL.NEUT SG.NOM Senayan
'Senayan and I went to the mountains.'

In Nanwang, only neutral forms can be relativised, but there is no such restriction in Katripul and Tamalakaw, as shown in the following sentences.

(11) Nanwang

an *ma-ruwa=ta* *ki-ma-drayar* ***taita***
when AV-can=1PL.INCL.NOM get-AV-discuss 1PL.INCL.NEUT
[*na* *t‹em›ara-puyuma* *na* *paseket*] *i,* *ala*
REL <AV>speak-Puyuma DEF.NOM clear TOP maybe
ma-ladram *kilengaw*
AV-know listen
'If we, those who speak Puyuma clearly, can converse, maybe they can understand (Puyuma).'

(12) Katripul

*mawmaw=**ku*** [*na* *mi-puran* *za* *mar-sa'az*].
only=1SG.NOM REL have-betelnut INDF.OBL RECP-branch
'Only I who have the kind of betelnut with crotched branches.'

A comparison of the functions that neutral pronouns display in each dialect is given in Table 6.

Table 6: Functions of neutral pronouns

	To affirm identities	To encode an emphatic meaning	To indicate a reflexive meaning	To appear in a topic position	To be used as personal proper nouns	To be used in inclusive constructions
Nanwang	✓	✓	✓	✓	×	×
Katripul	✓	✓	✓	✓	✓	×
Tamalakaw	✓	✓	✓	✓	✓	✓

3.2.2 Oblique free pronouns

Oblique free pronouns are used to denote a non-core argument, i.e., arguments excluding subject or agent. The functions of oblique free pronouns are the same in the three dialects.

(13) Nanwang
m-uai=yu *mi-kataguin* ***kanku****?*
AV-willing=2SG.NOM have-spouse 1SG.OBL
'Will you marry me?'

(14) Katripul
ku=i-pa-veray-anay *nanku* *piniduwa* ***kaninu.***
1SG.GEN=MOD-CAUS-give-UV:C 1SG.NOM.PSR treasure 2SG.OBL
'I will give my treasure to you.'

(15) Tamalakaw
'inava=u ***kaninku****.*
good=2SG.NOM 1SG.OBL
'You are good to me.'

3.2.3 Free pronouns denoting the possessors

Free pronouns that denote possessors may be nominative or oblique. In Nanwang, the possessor of a subject is manifested as nominative, while the possessor of a non-subject, including an actor or a non-core argument, is manifested as oblique. Note that there is an additional genitive proclitic agrees with the oblique NP manifesting the actor. Examples are given below.

(16) Nanwang
a. *bangsar* ***nantu*** *wadi*.
handsome 3.NOM.PSR younger.sibling
'His younger brother is handsome.' (possessor of subject)

b. ***tu$_i$****=pinatray-aw* *idru* *na* *trau* ***kantu*** ***wadi***.
3.GEN=kill-UV:P that.NOM LK person 3.OBL.PSR younger.sibling
'His younger brother/sister killed that person.' (possessor of non-subject actor)

c. ***tu****=veray-anay* *nantu* *ruma'* ***kanantu*** ***walak***.
3.GEN-give-UV:C 3.NOM.PSR house 3.OBL.PSR child
'He gave his house to his child.' (possessor of non-core)

In Katripul, the possessor of a subject or that of a non-subject actor is manifested as nominative, whereas the possessor of non-core arguments is manifested as oblique. Likewise, there is an additional genitive proclitic agreeing with the nominative NP which manifests the actor.

(17) Katripul

a. *alruzun* ***ninku*** *tuaktuk.*
heavy 1SG.NOM.PSR hammer
'My hammer is heavy.' (possessor of subject)

b. ***tu$_i$***=*kerutr-aw* *na* *vurasi* ***ninku*** ***wadi$_i$***.
3.GEN=dig.out-UV:P DEF.NOM sweet.potato 1SG.NOM.PSR younger.sibling
'My younger brother dug out the sweet potatoes.' (possessor of non-subject actor)

c. *ma-redek=ta* ***kananta*** *'alup*.
AV-arrive=1PL.INCL.NOM 1PL.INCL.OBL.PSR hunting.ground
'We arrived at our hunting ground.' (possessor of non-core)

In Tamalakaw, the possessor of a subject is manifested as nominative, whereas the possessor of non-core arguments is manifested as oblique. The possessor of a non-subject actor may be manifested as either nominative or oblique. In any case, there is an additional genitive proclitic agreeing with the nominative NP which manifests the actor.

(18) Tamalakaw

a. *kemay* *kanizu,* *z‹em›iyar* ***nantu*** *tezek.*
from that <AV>red 3.NOM.PSR buttock
'From then on, its buttock was red.' (possessor of subject)

b. *taw$_i$*=*kan-aw* *na* *i-salrem* ***nintu/kanintaw*** *hung$_i$*.
3.SG=eat-UV:P DEF.NOM NMLZ-plant 3.NOM.PSR/3.OBL.PSR cattle
'Their cattle ate what was planted.' (possessor of non-subject actor)

c. *taw=pu-rebu-ay* *na* *lrapung* ***kanantaw*** *nganan*.
3.GEN=CAUS-fire-UV:L DEF.NOM straw 3.OBL.PSR hole
'He burned the straw in the hole (of the tree).' (possessor of non-core)

To avoid possible ambiguity, in Katripul and Tamalakaw the actor is normally placed in the initial position, as shown below.

(19) Katripul

a. *tu=ada-aw=la* ****kananta/nanta*** *lalak.*
3.GEN=take-UV:P=PFV 1PL.INCL.OBL.PSR/1PL.INCL.NOM.PSR child
'Our children took it away.'
'Our children were taken away.'

b. *nanta* *lalak* *tu=ada-aw* *i* *temuwantaw*.
1PL.INCL.NOM.PSR child 3.GEN=take-UV:P SG.NOM grandparent
'Our children took their grandfather away.'

c. *i* *temuwantaw* *tu=ada-aw* *nanta* *lalak.*
SG.NOM grandparent 3.GEN=take-UV:P 1PL.INCL.NOM.PSR child
'Their grandfather took our children away.'

The coding of possessors of different syntactic roles (subject, non-subject actor, and non-core) in the three dialects is summarised below. It is found that while all three dialects under study make a distinction between nominative vs oblique possessor, there exists a variation as to whether the possessor of a non-subject is coded as nominative or oblique.

Table 7: Coding of the possessor of different syntactic roles

	PSR of subject	PSR of non-subject actor	PSR of non-core
Nanwang	Nominative	Oblique	Oblique
Katripul	Nominative	Nominative	Oblique
Tamalakaw	Nominative	Nominative/Oblique	Oblique

3.3 Diachronic changes

In section 3.1 two types of variation in forms are identified, i.e., the formation of neutral pronouns and the occurrence of the genitive noun phrase marker *ni* in Katripul and Tamalakaw.

First, regarding the formation of neutral pronouns, let us have a look at Ross's reconstructed forms.

Table 8: Proto-Puyuma neutral pronouns and their reflexes in Nanwang, Katripul and Tamalakaw

	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
Ross's reconstruction	*i-(ŋ)ku	*i-(n)u	*(i-ʐiw) *(i-taw)	*i-(n)ta	*i-niam	*i-(n)mu	*naḍiw
base forms	*ku=	*u=	*taw=	*ta=	*niam=/*mi=	*mu=	–
Nanwang	*kuiku*	*yuyu*	*taitaw*	*taita*	*mimi*	*muimu*	–
Katripul	*inku*	*inu*	*intaw*	*inta*	*iniam*	*inmu*	
Tamalakaw	*inku*	*inu*	*intaw*	*inta*	*iniam*	*inmu*	

Ross (2006:555) suggests that the forms in Nanwang represent a restructuring with reduplication of the base in front of the *i- prefix, e.g., *ku-i-ku* < *ku+*i-(n)ku. While his observation about the formation of these forms is correct, he does not explain the origin of the prefix *i-, and the occurrence of the nasal in forms like *i-(n)ku. The explanations for their occurrences have been given in section 3.1. The only form that needs a word of explanation is Nanwang *yuyu* '2SG.NEUT', because the expected form would be **u-i-u. I suggest that Nanwang neutral pronouns represent a later development after the change of *=yu* from *=u in Nanwang. In section 2.3 the diachronic reasons for the formation of Nanwang *=yu* '2SG.NOM' has been given. The form *yuyu* indicates that the base which serves in the formation of *yuyu* is not *=u but *=yu* (i.e. *yuyu* < *yu+* *i-u).

The second major variation among forms in the formation of free pronouns is the occurrence of /ni/ in Katripul and Tamalakaw PSR pronouns. The explanation that can be given is straightforward; *ni* is a genitive noun phrase marker in Katripul and Tamalakaw, whereas in Nanwang genitive and oblique are syncretic (cf. Teng 2009). This fact also

indicates that the formation of free pronouns occurred after syncretism had taken place in Nanwang.

4 Suffixal pronouns

This section describes the suffixal pronouns. A summary of the forms are given below:

Table 9: Puyuma suffixal pronouns

		Grandparent	Father	Mother	Older sibling	Younger sibling	Grandchild
Nanwang	1st	*mu-li*	*nama-li*	*nana-li*	*bae-li*	*ku=wadi*	*ku=temuwan*
	2nd	*temu-u*	*temama*	*taina*	*bae-u*	*nu=wadi*	*nu=temuwan*
	3rd	*temu-taw*	*temama-taw*	*tina-taw*	*bae-taw*	*tu=wadi*	*tu=temuwan*
Katripul	1st	*mu-li*	*ma-li*	*na-ni*	*va-li*	*wadi-li*	*mu-li*
	2nd	*temu-u*	*temama*	*taina*	*va-u*	*wadi-u*	*temu-u*
	3rd	*temuwan-taw*	*temama-taw*	*taina-taw*	*vau-taw*	*wadi-taw*	*temuwan-taw*
Tamalakaw	1st	*mu-li*	*ama-li*	*ina-ni*	*va-li*	*wadi-li*	*mu-li*
	2nd	*temu-u*	*temama*	*taina*	*va-u*	*wadi-u*	*temu-u*
	3rd	*temu-taw*	*tema-taw*	*taina-taw*	*vau-taw*	*wadi-taw*	*temu-taw*

Unlike the pronouns discuss in previous sections, suffixal pronouns do not form a complete paradigm; there are only singular forms, i.e., *-li* 'first singular', *-u* 'second singular' and *-taw* 'third singular'.[11] Possessor pronouns (clitics or free pronouns) are used if the speaker wants to express 'our/your(pl.)/their'. Compare:

(20) Tamalakaw

wadi-li	'my younger sibling'
mu-li	'my grandparent/grandchild'
naniam wadi	'our younger sibling'
naniam temuwan	'our grandparent/grandchild'

Suffixal pronouns do not carry case themselves; there is always a noun phrase marker preceding the noun to indicate the case of the NP, and the NP behaves like a personal noun. For instance:

(21) Katripul

a. *m-ukua* *i* *zenan* *i* *ma-li.*
AV-go LOC mountain SG.NOM father-1SG.PSR
'My father went to the mountains.'

b. *m-ukua* *i* *zenan* *nanku* *alak.*
AV-go LOC mountain 1SG.NOM.PSR child
'My child went to the mountains.'

[11] In Tamalakaw, there appears to have a plural form *-ta* 'our (inclusive)' used in very limited situations. There are only two examples in my corpus: *wadi-ta* 'our (inclusive) younger sibling', *alak-ta* 'our child'.

While the forms of the suffixal pronouns that appear in the three dialects are similar (especially those forms that appear in Katripul and Tamalakaw), the distribution of these forms varies across the dialects. In Nanwang, these forms only and obligatorily co-occur with elder kinship terms. In Katripul, most kinship terms (elder or younger) preferably co-occur with suffixal pronouns, but a replacement of it with a possessor pronoun seems to be more and more acceptable nowadays. Thus, while in Nanwang and Tamalakaw it is not acceptable to say *nanku temama* 'my father' or *nanku taina* 'my mother', it is well accepted in Katripul. In Tamalakaw, the distribution of this category of pronouns is not limited to kinship terms, but also to nouns denoting social relationships (such as 'friends', 'spouse' and 'age group') and nouns designating body parts.[12] For instance *turuma'an-li* 'my spouse', *alri-u* 'your male friend', *matra-li* 'my eyes'.[13]

5 Combined forms

This category of pronouns was first reported by Tsuchida (1980).[14] Among Formosan languages, Atayal and Seediq[15] are another two that exhibit combined forms.[16]

Table 10: Puyuma pronominal combined forms

	Nanwang	Katripul	Tamalakaw[17]
1SG.GEN+2SG.NOM	*kanu-*	*kunu-*	*kanu-/kunu-*
1SG.GEN+2PL.NOM	×	×	*kan(e)mu-*
2SG.GEN+1SG.NOM	×	×	*takanu-/takunu-*
2PL.GEN+1SG.NOM	×	×	*takan(e)mu-*
2PL.GEN+1PL.EXCL.NOM	×	×	*takan(e)mu-*

First, these forms are most likely affixes, i.e., they cannot be separated from verbs by clitics. Consider:

(22) Tamalakaw
a. *kanu-vavetra'-aw=la*.
1SG.GEN+2SG.NOM-lie-UV:P=PFV
'I have lied to you.'

---

[12] By analogy this category of pronouns can even attach to almost all kinds of common nouns. For example: *takil-li* 'my cup'. Because the use of these pronouns in Tamalakaw is very peculiar, they are treated in more detailed in Teng (2011b).

[13] Tsuchida (1995) attributes this category of pronouns to mark inalienable possession in Tamalakaw. However, the pair of examples, *ku=alak* 'my child' vs *alak-li* 'my adopted child', shows that this may not be totally correct.

[14] The forms given in Tsuchida (1980:200) are different from the forms I collected. For instance, Tsuchida reported both forms *kanu-* and *kunu-* in Tamalakaw, but gave them different glosses (*kanu-* '1SG.GEN + 2SG.NOM' and *kunu-* '2SG.GEN +1SG.NOM').

[15] A list of the Atayal and Seediq combined forms is given in appendix II. Interested readers are referred to Huang (1993, 1995, 2006) and Holmer (1996) for more detailed discussions.

[16] Readers may wonder whether these Puyuma forms are innovations or retentions from PAn. Because the combined forms in Puyuma and in Atayal and Seediq vary and cannot be reconstructed, it is more likely that in each language the combined forms are triggered by different sources.

[17] There are some more forms beginning with *tu/taw* that are reported in Tsuchida (1980:200). They are not considered as combined forms but are analyzed as clitic series in this paper. See section 6 for more discussion.

b. * *kanu=la* *vavetra'-aw*.
1SG.GEN+2SG.NOM=PFV lie-UV:P

Second, Nanwang and Katripul only preserve one form, whereas Tamalakaw has a fuller paradigm. It is worth mentioning that the difference between *kanu-* (*takanu-*) and *kunu-* (*takunu-*) in Tamalakaw is subject to affirmative/negative polarity. Compare (22a) with (23).

(23) Tamalakaw
'azi *kunu-vavetra'-i=la*.
NEG 1SG.GEN +2SG.NOM-lie-UV:P=PFV
'I didn't lie to you.'

Constructions with combined forms can always be replaced by one with a genitive proclitic and a nominative enclitic. In some villages, i.e., Nanwang and Katripul, constructions with combined forms are no longer productive; in others, i.e., Tamalakaw, (as well as Ulivelivek and Rikavung), constructions with combined forms are still often used.

(24) Nanwang
a. ***kanu****-beray-ay=la* *dra* *paisu*.
1SG.GEN+2SG.NOM-give-UV:L=PFV INDF.OBL money
'I have given you money.'

b. ***ku=****beray-ay****=yu****=la*** *dra* *paisu*.
1SG.GEN=give-UV:L=2SG.NOM=PFV INDF.OBL money
'I have given you money.'

(25) Katripul
a. ***kunu****-kedreng-aw*
1SG.GEN+2SG.NOM-pull-UV:P
'I pulled you out.'

b. ***ku=****kedreng-aw****=u***
1SG.GEN=pull-UV:L=2SG.NOM
'I pulled you out.'

Because Nanwang and Katripul have only one form and do not actively use them, more data from other villages, including Ulivelivek, Rikavung, Alipay and Pinaski, are needed for us to have a fuller picture about these forms and their functions.

6 Clitic sequence

In section 2 we saw that genitive clitic pronouns occur before the verb while nominative pronouns occur after the verb. In Katripul and Tamalakaw, we see that the nominative clitic pronouns sometimes occur right after the genitive pronouns to form a clitic sequence. For instance:

(26) Katripul

a. ***tu=ta**=zua-aw* *kanini*.
3.GEN=1PL.INCL.NOM=come-UV:P here
'They came to us.'

b. ***tu**=zua-aw=**ta*** *kanini*.
3.GEN=come-UV:P=1PL.INCL.NOM here
'They came to us.'

The occurrence of clitic sequence is restricted. From the corpus, only examples with third person actor are found.

In Tsuchida (1980) clitic sequences are analyzed as combined forms, but they are treated differently in the present study for the following reasons. First, unlike the forms discussed in section 5, where we are not able to break down the combined forms into two elements, it is self-evident that clitic sequences, such as *tuta* in (26), consist of two clitic pronouns *tu* and *ta*. Second, and more importantly, an aspectual marker may occur between the two clitics. Examples in (27) show some possible positions of the pronominal clitics in sentences.

(27) Katripul

a. *'azi=tu=ta=la* *i-piyamau-i*.[18]
NEG=3.GEN=1PL.INCL.NOM=PFV MOD-take.care-UV:P
'They will no longer take care of us.'

b. *'azi=tu=la=ta* *i-piyamau-i*.
NEG=3.GEN=PFV=1PL.INCL.NOM MOD-take.care-UV:P
'They will no longer take care of us.'

c. *'azi=tu* *i-piyamau-i=ta=la*.
NEG=3.GEN MOD-take.care-UV:P=1PL.INCL.NOM=PFV
'They will no longer take care of us.'

d. *'azi=tu=ta* *i-piyamau-i=la*.
NEG=3.GEN=1PL.INCL.NOM MOD-take.care-UV:P=PFV
'They will no longer take care of us.'

e. * *'azi=la=tu* *i-piyamau-i=ta*.
NEG=PFV=3.GEN MOD-take.care-UV:P=1PL.INCL.NOM

f. * *'azi=tu* *i-piyamau-i=la=ta*.
NEG=3.GEN MOD-take.care-UV:P=PFV=1PL.INCL.NOM

The above examples show that genitive pronouns were once enclitics; when there is a negator, they may attach to the negator. This seems to prove that the arguments given in Starosta et al. (1981, 1982) about the position of genitive pronouns in Puyuma and Paiwan is correct. They propose that PAn had a set of auxiliary verbs which included "not only words marking tense or aspect…but also logical and existential negators and certain kinds of adverbs denoting manner and instrumentality…" (Starosta et al. 1981:15) These auxiliary verbs were the head of the predication, to which clitics were attached. Then

[18] The prefix *i-* is a marker that relates to the willingness of the actor of carrying out the action denoted by the verb.

"when the auxiliary is lost, the clitic-initial structure results" (Starosta et al. 1981:157–158). That is, in languages where auxiliary use declined (e.g., Puyuma and Paiwan), some auxiliaries disappeared, but the pronoun formerly cliticised to the auxiliary has remainded "stranded" and has instead become a proclitic to the verb.

Although Ross (2008) agrees that PAn had enclitic sequences and their ordering was genitive-nominative, he infers that in modern and pre-Puyuma, no enclitic sequence occurred based on the Nanwang dialect. The Katripul data show that enclitic sequence does exist in Puyuma negative transitive clauses, and with the loss of auxiliary/negator, the sequence changed the host and became procliticised to the verb, as shown in (25a). What I have not yet found are clauses with sequence of clitics encliticised to verb.

7 Conclusion

In this paper, I have described the forms and functions of four categories of pronouns in Puyuma, including clitic pronouns, free pronouns, suffixal pronouns and combined forms. In the first two categories, I also offer diachronic explanations wherever there are variations. In the latter two categories, I have found that Tamalakaw (and possibly other close dialects, such as Rikavung and Ulivelivek) has fuller paradigms and these forms are more productive than those in Nanwang and Katripul. The development of the latter two categories in Tamalakaw grammar deserves a fuller description in another paper. In addition to the description of the four categories of pronouns, I have also dealt with clitic sequence and its relevance in the grammar of PAn in section 6.

Appendix

Appendix I. The pronominal system in Proto-Puyuma (based on Ross 2006:554–555)

	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
NEUT:TPC, FRFOC, DISJ							
P-Puyuma	*i-(ŋ)ku	*i-(n)u	*(i-ʐiw) *(i-taw)	*i-(n)ta	*i-niam	*i-(n)mu	*naɖiw
Tamalakaw	*i-ŋku*	*i-nu*	*i-ʐiw* *ta-y-taw*	*i-nta*	*i-niam*	*i-nmu*	*naʐiw*
Nanwang[1]	*ku-i-ku*	*yu-yu*		*ta-i-ta*	*mi-mi*	*mu-i-mu*	*Ø*
NOM:SBJ							
P-Puyuma	*=ku	*=(y)u	*Ø	*=ta	*=mi	*=mu	*Ø
Tamalakaw	*=ku*	*=u*	*Ø*	*=ta*	*=mi*	*=mu*	*Ø*
Nanwang	*=ku*	*=yu*	*Ø*	*=ta*	*=mi*	*=mu*	*Ø*
NOM:AGT[2]							
P-Puyuma	*ku=	*nu=	*taw=	*ta=	*mi=	*mu=	*taw=
Tamalakaw	*ku=*	*nu=*	*taw=*	*ta=*	*mi=*	*mu=*	*taw=*
Nanwang	*ku=*	*nu=*	*tu=*	*ta=*	*mi=*	*mu=*	*tu=*
PSR1:PSRA							
P-Puyuma	*=li	*=u	*=taw	*(=ta)	*(=mi)	*(=mu)	*(ni-naɖiw)
Tamalakaw[3]	*=li*	*=u*	*=taw*	*=ta*	*=mi*	*=mu*	*ni-naʐiw*
Nanwang	*=li*	*=u*	*=taw*	–	–	–	–
PSR2:[4] PSRA							
P-Puyuma	*ku=	*nu=	*taw=	*ta=	*niam=	*mu=	*taw=
Tamalakaw	*=[ŋ]ku*	*=[ŋ]nu*	*=ntaw*	*=nta*	*=niam*	*=nmu*	*=ntaw*
Nanwang	*ku*	*nu*	*tu*	*ta*	*niam*	*mu*	*tu*
GEN-OBL:AGT, PAT							
P-Puyuma	*kan-(iŋ)ku	*kan-nu	*kan-taw	*kan-ta	*ka-niam	*kan-mu	*kan-(an-)taw
Tamalakaw	*kan-iŋku*	*kan-nu*	*kan-taw*	*kan-ta*	*ka-niam*	*kan-mu*	*kan-an-taw*
Nanwang	*kan-ku*	*ka-nu*	*kan-taw*	*kan-ta*	*ka-niam*	*kan-emu*	*kan-taw*

Ross's notes (2006:555):

1. Nanwang NEUT forms seem to represent a restructuring with reduplication of the base in front of the *i- prefix, e.g., *ku-i-ku* NEUT:1SG < *ku+*i-(ŋ)ku.
2. GEN acts as an agreement marker in both dialects, as it remains even when there is a GEN (Tamalakaw) or OBL (Nanwang) NP actor or possessor.
3. In Nanwang these forms occur with only a very small number of kin terms, and only in the singular. It is possible that Tamalakaw non-singulars are simply copied from NOM (3SG from NEUT) and did not occur in Proto-Puyuma.
4. In Tamalakaw the forms in the PSR2 set are attached to a case-marker and precede the possessum noun. In Nanwang, they either (i) are proclitic to the possessum, forming a NOM NP, or (ii) they are enclitic to the case-marker which precedes the

possessum. I infer that in Proto-Puyuma these forms were proclitic to the possessum, but optionally preceded by a case-marker.

Appendix II. Combined forms in Atayal and Seediq (based on Huang 1993, 1995, 2006; Holmer 1996)

	Wulai Atayal	Mayrinax Atayal	Plngawan Atayal	Paran Seediq
1SG.GEN+2SG.NOM	*misu'*	*misu'*	*misu'*	*misu'*
1SG.GEN+2PL.NOM	–	–	–	*maku*
2SG.GEN+1SG.NOM	–	–	*saku'*	*saku'*
3SG.GEN+1SG.NOM	–	–	*cini'*	–
3SG.GEN+2SG.NOM	–	–	*sini'*	–
3SG.GEN+1PL.NOM	–	–	*mini'*	–
3SG.GEN+2PL.NOM	–	–	*mani'*	–

References

De Vogelaer, Gunther and Evie Coussé. 2011. The functional nature of pronominal change: Innovative plural pronouns in English and Dutch. *Neophilologus* 95.1:1–26.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press. 2nd edition.

Holmer, Arthur J. 1996. *A parametric grammar of Seediq*. Travaux de l'institut de linguistique de Lund. Travaux de l'Institut de linguistique de Lund 30. Lund, Sweden: Lund University Press.

Huang, Lillian M. 1993. *A study of Atayal syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 1995. *A study of Mayrinax syntax*. Taipei: Crane Publishing Co., Ltd.

Huang, Lillian M. 2006. Manifestations of participants in Atayal: A cross-dialectal study. Paper presented at the Tenth International Conference on Austronesian Linguistics. 17-20 January 2006. Puerto Princesa City, Palawan, Philippines. http://www.sil.org/asia/philippines/ical/papers.html

Huang, Lillian M., Elizabeth Zeitoun, Marie M. Yeh, Joy J. Wu and Anna H. Chang. 1999. A typological overview of pronouns in the Formosan languages. In *Selected papers from the Fifth International Conference on Chinese Linguistics*, ed. by H. Samuel Wang, Feng-fu Tsao and Chin-fa Lien, 165–198. Taipei: Crane Publishing Co., Ltd.

Li, Paul Jen-kuei. 1991. *taiwan nandaoyuyan de yuyin fuhao xitong* [*Orthographic systems for Formosan languages*]. Taipei: Ministry of Education. [In Chinese]

Ogawa, Naoyoshi, and Erin Asai. 1935. *Gengo ni yoru Taiwan takasago-zoku densetsu-shu* [*The myths and traditions of the Formosan native tribes (texts and notes*)]. Taihoku [Taipei]: Taihoku Imperial University. [In Japanese]

Quack, Anton. 1981. *Das Wort der Alten: Erzählungen zur Geschichte der Pujuma von Katipol (Taiwan)*. Collectanea Instituti Anthropos Vol 12. St. Augustin, Germany: Haus Völker und Kulturen.

Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In *Streams converging into an ocean: festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, ed. by Henry Y. Chang, Lillian M. Huang and Dah-an Ho, 521–563. Language and Linguistics Monograph Series W-5. Taipei: Institute of Linguistics, Academia Sinica.

Ross, Malcolm. 2008. Negative verbal clause constructions in Puyuma: Exploring constructional disharmony. In *Constructions and language change*, ed. by Alexander Bergs and Gabriele Diewald, 171-193. Trends in Linguistics: Studies and Monographs 194. Berlin and New York: Mouton de Gruyter.

Starosta, Stanley, Andrew Pawley and Lawrence A. Reid. 1981. The evolution of focus in Austronesian. Honolulu. Department of Linguistics, University of Hawai'i at Mānoa. Ms. Reprinted in *Formosan linguistics: Stanley Starosta's constributions*,

ed. by Elizabeth Zeitoun, 329–481. Language and Linguistics Monograph Series C6-2. Taipei: Academia Sinica, 2009.

Starosta, Stanley, Andrew Pawley and Lawrence A. Reid. 1982, The evolution of focus in Austronesian. *Papers from the third International Conference on Austronesian Linguistics, Vol. 2: Tracking the travellers*, ed. by Amran Halim, Lois Carrington and S. A. Wurm, 145–170. Pacific Linguistics C-75. Canberra: The Australian National University. Reprinted in *Formosan Linguistics: Stanley Starosta's contributions*, ed. by Elizabeth Zeitoun, 297–328. Language and Linguistics Monograph Series C6-2. Taipei: Academia Sinica, 2009.

Teng, Stacy Fang-ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics 595. Canberra: The Australian National University.

Teng, Stacy Fang-ching. 2009. Case syncretism in Puyuma. *Language and Linguistics* 10.4:819–844.

Teng, Stacy Fang-ching. 2011a. Noun phrase conjunction in three Puyuma dialects. *Language and Linguistics* 12.1:171–204.

Teng, Stacy Fang-ching. 2011b. Some peculiar features of personal pronouns in Tamalakaw Puyuma. Unpublished manuscript.

Teng, Stacy Fang-ching. 2014. Grammaticalization of predicative possession in Nanwang Puyuma and as a basis for reconstruction in PAn. *Oceanic Linguistics* 53.1:136–154.

Teng, Stacy F. and Malcolm Ross. 2010. Is Puyuma a primary branch of Austronesian. A reply to Sagart. *Oceanic Linguistics* 49.2:543–558.

Ting, Pang-hsin. 1978. gu beinanyu de nice [Reconstruction of proto-Puyuma phonology]. *Bulletin of the Institute of History and Philology, Academia Sinica* 49.3:321–392. [In Chinese]

Tsuchida, Shigeru. 1980. Puyuma (Tamalakaw dialect) vocabulary, with grammatical notes and texts. *Kuroshio Bunka no kai* (*Black Current Cultures Committee*), ed. by Kuroshio no minzoku, bunka, gengo (Ethnology, Cultures and Languages Along the Black Current), 183–307. Tokyo: Kadokawa Shoten. [in Japanese]

Tsuchida, Shigeru. 1995. Alienable and inalienable distinction in Puyuma? *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Dah-an Ho, Ying-kuei Huang and Chiu-yu Tseng, 793–804. Symposium series of the Institute of History and Philology, Academia Sinica No. 3. Taipei: Academia Sinica.

18 *Analogy and grammatical change: A case study of the verb of 'saying' in Mantauran Rukai*

ELIZABETH ZEITOUN

1 Introduction

Among the fifteen Austronesian languages still spoken in Taiwan,[1] Rukai, which stretches across the south of Taiwan (Map 1), consists of six dialects (Tanan, Labuan, Budai, Maga, Tona, and Mantauran). The whole Rukai population (excluding here the repartition by village) is estimated at 12,861 as of December 2014 according to the census of the Council of Indigenous Peoples, Executive Yuan (http://www.apec.gov.tw).

The internal relationships of Rukai have, to date, not been completely clarified. Based on Li's (1977) reconstruction of Proto Rukai (PR),[2] it is generally acknowledged that there are two distinct subgroups: Tanan, Labuan, and Budai on the one hand, and Maga and Tona on the other.[3] Three hypotheses have been advanced regarding the position of Mantauran: (i) It subgroups immediately with Maga and Tona (Li 1977); (ii) it is the first offshoot of the Rukai family (Tu & Cheng 1991; Tu 1994; Starosta 1994, 1995; Li 1995, 1996); and (iii) it forms a subgroup with the Tanan-Labuan-Budai cluster (Zeitoun 2003). It is this last hypothesis that I consider valid here. The position of Mantauran within the Rukai language family is difficult to assess because it has undergone many phonological and morpho-syntactic changes.

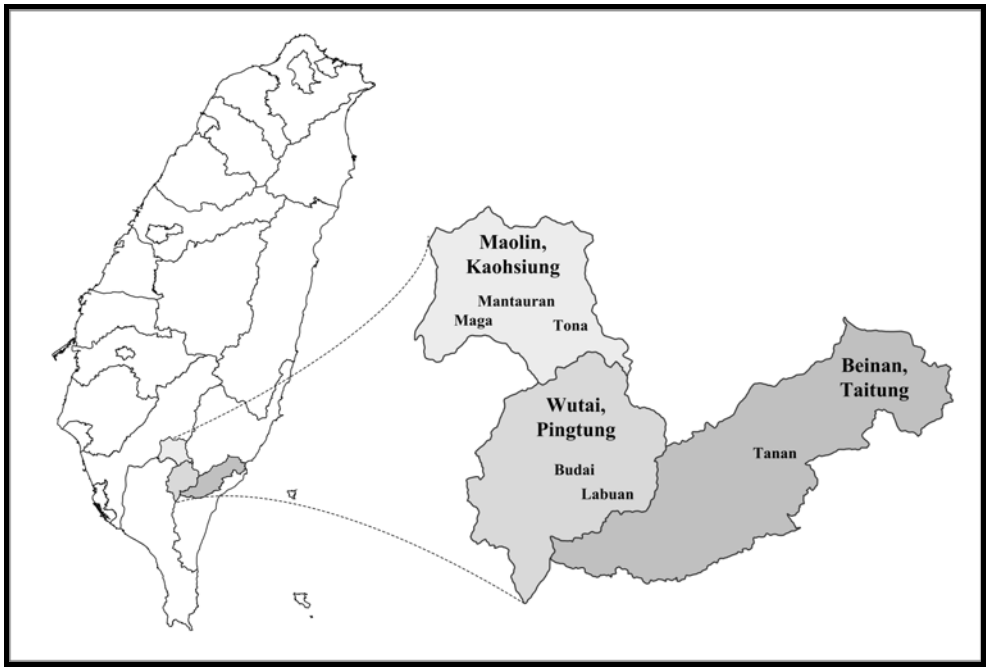
---

1 I refer here to a linguistic (rather than governmental) classification, which recognises Amis, Atayal, Saisiyat, Bunun, Kanakanavu, Kavalan, Kaxabu, Rukai, Paiwan, Puyuma, Saaroa, Seediq, Thao, Tsou and Yami as distinct languages. Among these, it is unclear whether Kaxabu, a plain-tribe language, is still actively spoken. Yami is a Philippine language, belonging to the Batanic group, and thus differs drastically from the other Formosan languages.

2 Abbreviations for languages include: Mt, Mantauran; PR, Proto Rukai; To, Tona. Abbreviations used in the examples follow those of the Leipzig Glossing Rules. Additional abbreviations include the following: ACT, active; CNTRFT, counterfactual; DYN, dynamic; FILL, filler; FIN, finite; NFIN, non-finite; NP, Noun phrase; REAL, realis; RED, reduplication; S, sentence; STAT: stative, SUBJ, subjunctive; VP, Verb Phrase.

3 Because of population movement, the present geographical distribution of the Rukai dialects does not reflect their subgrouping.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 431–450
Asia-Pacific Linguistics, 2015



Map 1: Geographical distribution of the Rukai dialects

The goals of the present paper are threefold. It will provide a discussion regarding the use of the conjunction *la* 'then' in the formation of the quotative verb 'say' in Tona (section 3) in order to highlight a number of morpho-syntactic changes that have occurred in Mantauran (section 4). This synchronic study will further serve as a basis for the reconstruction of coordinative and conjunctive markers as well as the quotative verb at the PR level (section 5). Before delving into these complexities, I first provide an overview of the sound changes undergone by Mantauran, in order to let the reader reach a better understanding of the comparative data proposed below.

2 Major sound changes in Mantauran

Mantauran has undergone a number of sound changes, and I deal in this section only with those relevant to the present study. The reader is referred to Li (1977), Zeitoun (2007), and Wolff (2010) for more details. Throughout the paper, I take Tona[4] as a representative dialect of the Rukai language family and compare it to Mantauran as shown in (1)–(5).

All the PR voiced stops have been spirantised: PR *b > Mt *v*, PR *d/*ɖ > Mt *dh*, PR *g > Mt *h*.[5]

[4] The choice of this dialect as a representative is based on two factors: (i) It is not as close to Mantauran as Labuan, Tanan and Budai but (ii) it exhibits the same characteristics as these dialects. Unless mentioned otherwise, all the examples given in Tona are extracted from the Formosan Language Archive. Data on Mantauran is drawn mostly from Zeitoun & Lin (2003), Zeitoun (2007), and unpublished folktales.

[5] For the modern languages, I use the now more conventional Romanised orthography, whereby *th* stands for *θ*, *dh* for *ð*, *lr* for *ɭ*, *ng* for *ŋ*, *'* for *ʔ*, and *e* for *ə*.

(1)		Proto Rukai	Tona	Mantauran	Gloss
	*b-	*bələbələ	*belebele*	*velevele*	‘banana’
	*-b-	*balәbalə	*balebale*	*valevale*	‘bamboo’
	*d-	*davacə	*davace*	*dhaace*	‘leave’
	*-d-	*idai	*iday*	*idhai*	‘hundred’
	*ɖ-	*ɖakəralə	*dhakeale*	*dhakerale*	‘river’
	*-ɖ-	*taloɖo	*talodro*	*talodho*	‘bridge’
	*g-	*gəməgəmə	*gemegeme*	*hemeheme*	‘hold in hand’
	*-g-	*ɭəgəɭəgə	*egeege*	*lrehelrehe*	‘mountain’

Among the PR fricatives, PR *v has become Mt *Ø*, PR *θ is reflected as Mt *s* and PR *s as Mt *’* (further triggering the merger of PR *s and PR *ʔ as Mt *’*).

(2)		Proto Rukai	Tona	Mantauran	Gloss
	*v-	*valo	*valo*	*alo*	‘bee’
	*-v-	*abo	*abo*	*avo*	‘ashes’
	*θ-	*θiabə	*thiabe*	*siave*	‘chop’
	*-θ-	*θuθu	*thotho*	*soso*	‘breast’
	*s-	*sipi	*sipi*	*’ipi*	‘dream’
	*-s-	*mabosoko	*mabosoko*	*mavo’oko*	‘drunk’

The sound change PR *k > Mt *’* has only taken place in few affixes—compare (3a) and (4)—and loan words (3b):

(3)	a.		Proto Rukai	Tona	Mantauran	Gloss
		*k-	*ki-	*ki*	*’i-*	i. ‘gather’ ii. ‘passive’
			*ka-	*ka*	*’a*	‘topic marker’
		but	*ka	*ka*	*ka*	‘ligature’
	b.		—	Tona	Mantauran	Gloss
				kaomo	*’aomo*	‘Japanese’

(4)		Proto Rukai	Tona	Mantauran	Gloss
	*k-	*kaɖoa	*kadroa*	*kaodho*	‘not exist’
	*-k-	*taka	*taka*	*taka*	‘elder sibling’

The PR vocalic sequence *oa has merged as Mt *o*:

(5)	Proto Rukai	Tona	Mantauran	Gloss
*oa-	*oakanə	*wakane*	*okane*	‘eat (FIN)’
*-oa-	*koanə	*kwane*	*kone*	‘eat (SUBJ)’

3 The use of the conjunction *la* ‘then’ in the formation of the quotative verb ‘say’ in Tona Rukai

In all the Rukai dialects except Mantauran, the conjunction *la* ‘then’ is used in the formation of the quotative construction, *la ya*, which literally translates as ‘then so’. Before dealing with the morpho-syntatic characterictics of phrasal and sentential coordination (section 3.2) and showing how the morpheme *la* ‘then’ is involved in the

formation of the above-mentioned quotative construction (section 3.3), it is important to start with a brief overview of the verbal morphology of Rukai in Tona (section 3.1).

3.1 A bird's eye view of the verbal morphology of Tona Rukai

Tona exhibits an active/passive voice system. In the active voice (realis), dynamic verbs are usually marked by *o-* (6a);[6] stative verbs are marked by *ma-*. Dynamic and stative transitive verbs can undergo agentive passivisation through the prefixation of *ki-* (6b). Non-agentive passivisation is marked by *'i-* (6c). Note that the vowels of these affixes undergo gliding in co-occurrence with the morpheme *a-* marking realis thus resulting in the bi-morphemic sequences *w-a-*, *ky-a-*, and *'y-a-*.

(6) Tona Rukai

a.	*a-nakay*	*soa'a*	*ka*	*w-a-ka'ace*	*na*	*atho.*
	TOP-this	snake	TOP	ACT-REAL-bite	OBL	dog
	'As for this snake, it bit a dog.'					
b.	*a-nakay*	*atho*	*ka*	*ky-a-ka'ace*	*na*	*soa'a.*
	TOP-this	dog	TOP	PASS-REAL-bite	OBL	snake
	'As for this dog, it was bitten by a snake.'					

c.	*'y-a-ka'ace-nga*	*nakay*	*atho.*	
	PASS-REAL-bite-already	this	dog	
	'The dog was bitten.'			
c'. *	*'y-a-ka'ace-nga*	*na*	*soa'a*	*nakay*	*atho.*
	PASS-REAL-bite-already	OBL	snake	this	dog

Tense is not grammaticalised in Tona, which exhibits, rather, a basic mood distinction between realis and irrealis. Mood closely interacts with voice and aspect. Realis is usually encoded through the prefixation of *a-*. There are different classes of verbs, so realis marking is not always overtly marked. The temporal frame (past or present) of a clause depends on the occurrence of additional aspectual morphemes, e.g. *maka-…(-nga)* 'had…(already)' (< *maka-* 'finish') as shown in (7a), *-nga* 'already', *-la* 'still' or temporal adjuncts. Irrealis, on the other hand, is overtly marked by *no-…mi-* 'IRR', which marks a projection in the future as in (7b) and *ni-* 'CNTRFCT', which encodes counterfactuality as in (7c).

(7) Tona Rukai

a.	*a-ko*	*maka-kan-ili-nga*	*na*	*do'o*	*ka*
	TOP-when.REAL	SUBJ:finish-eat-1SG.GEN-already	OBL	rice	TOP
	'abee	*kake.*			
	leave	1SG.NOM			
	'When I had finished eating, I left.'				

[6] Dynamic verbs can be classified into different verb classes. Most verbs are marked by *o-*. Some verbs are marked *mu-* as in *mw-a-baithi* (~ *baithi*) 'exchange' or *m* as in *matiavii* (~ *patiavii*) 'make a request before leaving'. Others are completely unmarked, e.g., *'aokay* 'come'. Verbs starting with a vowel are marked by *am-* as in *am-wa* (~ *wa*) 'go'.

b. *a-no* *maka-kan-ili-nga* *na* *do'o* *ka*
TOP-when.IRR SUBJ:finish-eat-1SG.GEN-already OBL rice TOP
no-'abee *mi-kake.*
IRR-leave IRR-1SG.NOM
'When I had finished eating, I left.'

c. *ni-doo* *koso* *'angi-da~do'o.*
CNTRFCT-can 2SG.NOM REFL-RED~cook
'You should be able to cook by yourself.'

Progressive, iterative, and habitual aspects are encoded through reduplication. This is illustrated in (8).

(8) Tona Rukai
w-a-do'o~do'o *'ikay* *balate.*
ACT-REAL-RED~cook exist outside
'(S/he) is cooking outside.'

Based on the following alternations, *waka'ace* 'bite/bit', (*i*)*ka'ace* '(does/did not) bit', and *koa'ace* 'bite', I assume that there is a morphological distinction between verbs marked as finite, non-finite, and subjunctive. I consider that a verb marked as realis occurs in its finite form, i.e., it heads an independent clause, which can stand on its own as a complete sentence. Finite verbs occur in sentence-initial position as in (8) though they may be preceded by a topic NP as in (6a–b) or a topic clause as in (7b); they do not take any prefixes, e.g. *no-* 'IRR', *ni-* 'CNTRFCT', *i-* 'NEG', and they do not follow any conjunction, for instance *la* 'then'. I use the term 'non-finite' to designate the verb root/stem.[7] Dynamic verbs marked by *w-a-* are unmarked in their non-finite form, cf. *w-a-ka'ace ~ ka'ace* 'bite'. Most stative verbs are marked by *ka-* in their non-finite form, e.g. *ma-dalame ~ ka-dalame* 'like, love'. A verb occurs in its non-finite form when it is preceded by a verbal prefix, e.g. *no-* 'IRR' (7b), *ni-* 'CNTRFCT' (7c) even if it occurs in sentence-initial position or if it follows a conjunction such as *la* 'then'. In a non-finite form, such verbs can be followed by pronouns marked as nominative and head independent clauses, e.g. (7b). Active (dynamic) *w-a-*marked verbs in which the first vowel of the base is *i* as in *w-a-sititi* 'beat', *o* as in *w-a-do'o* 'cook', and *e* as in *w-a-thenay* 'sing', occur as unmarked in their subjunctive form, e.g., *sititi* 'beat', *do'o* 'cook', and *thenay* 'sing'. Verbs in which the first vowel of the base is *a*, e.g. *w-a-kane* 'eat', are infixed with *-o-* (gliding to *w* in front of *a*), e.g. *k-w-ane* 'eat (SUBJ)'. Stative verbs are marked by *ma-* in their subjunctive form. A verb occurs in its subjunctive form when it appears in embedded position as V_2, as in *m-wa k-w-ane*! 'Go eating!', imperative sentences, e.g. *k-w-an-a!* 'Eat!' or in temporal/conditional/hypothetical clauses introduced by *ko* 'when.Real' or *no* 'when.Irr', as in (7a–b).

A summary of the verb classes and the verb alternations of Tona Rukai in Table 1.

[7] In the case of dynamic verbs, the stem is identical to the root, cf. *ka'ace* 'bite'. The root of stative verbs, e.g., |*dalame*| 'like, love' is bound, thus when affixed by *ka-*, what is referred to is the verb stem.

Table 1: Finite, non-finite and subjunctive verb forms in Tona Rukai

Verb class	Verb type	Finite	Non-finite	Subjunctive
	1a. *w-a-*CVCV	*w-a-*	Ø-	Ø
Dynamic	1b. *w-a-*C*a*CV	*w-a-*	Ø-	Ø-C*wa*CV
	2. *am-*VCV	*am-*	Ø-	*m-*
	3. *mw-a-*	*mw-a-*	Ø	*mu-*
	4. *m*	*m*	*p*	*m*
	5. Ø-	Ø-	Ø-	Ø-
Stative	*ma-*	*ma-*	*ka-*	*ma-*

3.2 A brief survey of coordination in Tona Rukai

In all the Rukai dialects, except Mantauran, both NP and VP coordination are marked by *si* 'and', schematised as follows: NP *si* NP and VP *si* VP and further illustrated in (9)–(10). Note that in VP coordination the second verb needs to occur in its non-finite form, as shown by the (un)grammaticality of (10a–b).

(9) Tona
takanaw ***si*** *legeai*
Takanaw and Legeay
'Takanaw and Legeay'

(10) Tona
a. *w-a-kane* ***si*** *'ongolo*
ACT-REAL-eat and NFIN.drink
'eat and drink'
b. * *w-a-kane* ***si*** *w-a-'ongolo*
ACT-REAL-eat and ACT-REAL-drink

Clauses are coordinated by (*si*) *la* '(and) then' as in (11). The verb that occurs after *la* 'then' needs also to occur in its non-finite form. The morpheme *si* is the coordinator introduced above. I treat *la* 'then' as a conjunction.

(11) Tona
a. ***si*** ***la*** *davace* *si* *la* *wa-nga* *'angi-talodro.*
and then NFIN.leave and then NFIN.go-already cross-bridge
'(They) left and crossed the bridge.'
b. * ***si*** ***la*** *w-a-davace* ***si*** ***la*** *am-wa-nga* *'angi-talodro.*
and then ACT-REAL-leave and then ACT-go-already cross-bridge

3.3 The verb *ya* '(be) so, (be) like' and the quotative construction *la ya* (lit.) 'then so' in Tona Rukai

In all the Rukai dialects, the verb *ya* '(be) so' functions mainly as a quotative verb,[8] though, as will be shown below, it also bears another meaning, '(be) like'. Like any other full-fledged verb, it exhibits the three morphological alternations mentioned above. It can occur in its finite form as *am-ya* (12a), in its subjunctive form as *m-ya* (12b), and its non-finite as *ya* (12c). The verb *ya* 'so' occurs in its non-finite form mostly when it is preceded by the conjunction *la* 'then'.

(12) Tona

a. *la* *'ikay* *aogane,* *i-ota'alebe* "*i-kibaka*
then NFIN.exist on.roof NEG-NFIN.go:below NEG-NFIN.agree
mota'alebe" *am-ya* *si....*
SUBJ:go:below ACT-so and
Lit.: 'He stayed on the roof and (he) said "(I) refuse to go down."'
(Formosan Language Archive, Text 3)

b. *i-okono* *ebel-ini* *si* *la* *'aokay* *la* *'abee.*
NEG-NFIN.find jacket-3SG.GEN and then NFIN.come then NFIN.return
"*ko-'abee* *kicibane*" *m-ya* *si* *la*
1SG.NOM-NFIN.return forget it SUBJ-so and then
'aokay *'abee.*
NFIN.come SUBJ.return
'(As) he could not find his jacket, (he) said: "Forget it, I am going back home" and he went back home.' (Formosan Language Archive, Text 2)

c. "*si* *adroo,* *sy-a-ebele* *koso*" *la* *ya*
and that wear-REAL-jacket 2SG.NOM then NFIN.so
niane *si…*
3SG.OBL and
'(But) you have your jacket on" (they) said to him…'
(Formosan Language Archive, Text 2)

The sequence *la ya* usually follows the direct quote complement (13a), though it may occur in sentence-initial position (13b).

(13) Tona

a. "'*aokay-a* *wamece* *na* *beke'e* *nosi'a* *la* *paowa*
SUBJ.come-IMP SUBJ.bring OBL pig tomorrow then NFIN.put
po-'adringi, *ni-doo* *ki* *pangetedre* *'aboale*"
put-inside CNTRFCT-NFIN.can Nom Pangetedre SUBJ.come.out
la *ya*.
then NFIN.so
'"Bring me a pig tomorrow" (s/he said) and "put it inside. Then Pangetedre will come out."' (Formosan Language Archive, Text 1)

---

[8] Note that in narrated texts, it is sometimes difficult to determine whether *ya* (and its morphological variants) refers to a direct quotation or to a thought, all the more if it refers to a first person participant.

b. *la ya nakay kodray baobao m-ya*
then NFIN.so this that young.woman SUBJ-so
niane "ko-'angi-ka~kamani 'angi-dese~desele adri
3SG.OBL 1SG.NOM-REFL-RED~STAT.be REFL-NFIN.RED~wipe NEG.IMP
pa-ti-koa~koa" la ya kodray baobao.
CAUS-make-RED~thing then NFIN.so that young.woman
'Then the young woman said: "No need (to help). I will rub myself alone. Don't do anything!"' (Formosan Language Archive, Text 7)

Note also that *la ya* 'then so' occurring at the end of a sentence can be followed immediately by *la* 'then' introducing the next sentence, as in (14a–b).

(14) Tona

a. *"'ii, am-ya nia-ni nakay ki gili,*
EXCL FIN-so Q-3SG.GEN this younger.sibling
pa-ti-adraw ami ko 'angimalelean-ini" la ya.
CAUS-make-big like.that too.much-3SG.GEN then NFIN.so
'"Oh, little sister, here you are, this is so serious" he said to her.' (Formosan Language Archive, Text 7)

b. *la wa "ko-(w)a cekele" m-ya*
then NFIN.go 1SG.NOM-go SUBJ.get.up SUBJ-so
driane baobao.
3SG.OBL young.woman
'Then he said to the young woman "I am going to (help) you get up."' (Formosan Language Archive, Text 7)

The sequence *la ya* (lit.) 'then so' as a quotative verb is particularly unusual in the Formosan languages (Hsieh 2011, 2012). Cross-linguistically, however, *like* is found as a quotative verb, sometimes as a new innovated form, as in English *like*, cf. Kohn & Franz (2009); Buchstaller & Van Alphen (2012); Vandelanotte (2012), among others.

What is interesting to note here is that in all the Rukai dialects, *ya* and its morphological variant forms do not only mean '(be) so' but also '(be) like'. Illustrative examples with *am-ya* '(be) like (FIN)', *m-ya* '(be) like (SUBJ)', and *ya* '(be) like (NFIN)' are given in (15a–c).

(15) Tona

a. *a-nakay maolay ka am-ya nakay.*
TOP-this story TOP ACT-like this
'As for this story, it goes on like this.' (Formosan Language Archive, Text 7)

b. *i-igo'o kake mani ka kaga'aoco-ini*
NEG-NFIN.know 1SG.NOM why LIG STAT:NFIN.angry-3SG.GEN
gil-ili m-ya.
younger.sibling-1SG.GEN SUBJ-like
'I do not know why my younger brother/sister is angry.'

c. "*n*(*o*)-*ya*-*mi*-*(i)ta*-*nga* *pangetedr*-*ane?*
IRR-NFIN.like-IRR.Q-1PL.INCL.GEN-already Pangetedre-OBL
i-*doo*-*nga* *'aboale!*" *la ya.*
NEG-NFIN.can-already SUBJ.come.out then NFIN.so
'"What shall we do about Pangetedre? He can't get out!" they said.'
(Formosan Language Archive, Text 1)

4 The use of the conjunction *mani* 'then' in the formation of the quotative verb 'say' in Mantauran Rukai

In this section, I discuss the use of the conjunction *mani* 'then' in the formation of the quotative verb 'say' in Mantauran by comparing this dialect with Tona and highlighting a number of morphosyntactic discrepancies. For comparison's sake, I follow the same presentation and provide, whenever possible, the same sets of examples.

4.1 Some basic comparative data

The Mantauran active/passive voice system, the interaction between voice, mood, and aspect, as well as the verbal alternations finite/non-finite/subjunctive are very similar to those of Tona Rukai, so that basic information will not be repeated in this section . What needs to be highlighted here is that Mantauran has undergone a number of phonological and morpho-syntactic changes: The active voice *o*- has fused with the realis marker *a*- yielding *o*-,[9] which marks most dynamic verbs (16a).[10] As pointed out by Zeitoun (2007:132), "the prefix *o*- is [thus] multifunctional in that it includes a combination of different verbal categories: Verb classification, verbal conjugation, mood and voice, i.e., it indicates that a verb is dynamic, finite, in the realis mood and in the active voice." The (agentive) passive form is *'i*- (16b). The non-agentive passive has not been reported in Mantauran (Zeitoun 2007). Different verb classes are found. Most dynamic verbs are marked by *o*-, but other verbs are marked by *om*- (rather than *am*- in the other Rukai dialects) as in *om*-*oa* (~ *oa*) 'go', *m*- as in *m*-*olekate* (~ *olekate*) 'be enough', *m* as in *maineke* (~ *paineke*) 'breathe' or are unmarked, e.g., *irovo* 'pile up'. Dynamic *o*-marked verbs in which the first vowel of the base is *i*, *o*, or *e* occur as unmarked in their subjunctive form, cf. *vilivili* 'pull', *coroko* 'jump', and *ke'ete* 'cut'. Most of the *o*-marked verbs in which the first vowel is *a* exhibit a vocalic *a~o* alternation:

---

[9] Note that there is no glide in Mantauran.

[10] One of the reviewers suggests that "Mantauran *o*- and Maga *u*- do not reflect Proto Rukai *u-a- but simply Proto Rukai *u-, [i.e.,] that Proto Rukai probably had an alternation between *u- 'actor voice' and *u-a- 'actor voice + realis', and Mantauran and Maga have selected the former as the basic finite actor voice, and the other dialects the latter. This is then consistent with the fact that the Mantauran form before a vowel is *om- (< *um-), Maga is either *m*- or *a*-*m*-, the other dialects *a*-*m*-." He seems to have reached a better solution, but having not yet finished my comparative study of the Rukai dialects (Zeitoun in preparation), I prefer to keep it as a possibility for further research.

(16) Mantauran

a. *'olra'a* *'a* *o-ka'ace* *ta'olro.*
snake TOP ACT.REAL-bite dog
'As for the snake, (it) bit the dog.'

b. *ta'olro* *'a* *'i-ka'ace* *'olra'a.*
dog TOP PASS.REAL-bite snake
'As for the dog, (it) was bitten by the snake.'

Table 2: Finite, non-finite and subjunctive verb forms in Mantauran Rukai (Zeitoun 2007:140)

Verb class	Verb type	Finite	Non-finite	Subjunctive
	1a. *o*-C*V*CV	*o-*	Ø-	Ø-
Dynamic	1b. *o*-C*a*CV	*o-*	Ø-	Ø-C*o*CV
	2. *om-*	*om-*	Ø-	*m-*
	3a. *m-*	*m-*	*k-*	*m-*
	3b. *m-*	*m-*	*p-*	*m-*
	3c. *m-*	*m-*	Ø-	*m-*
	4. Ø-	Ø-	Ø-	Ø-
Stative	*ma-*	*ma-*	*ka-*	*ma-*

4.2 A brief survey on coordination in Mantauran Rukai

Examples (17)–(19) represent the Mantauran counterparts of the Tona examples given in (9)–(11). Mantauran has innovated in at least two respects:

(i) NP and VP coordination is encoded through *la*, which means 'and' (and not 'then'), rather than by *si* as in the other Rukai dialects, thus yielding NP *la* NP and VP *la* VP.

(ii) It has innovated a new grammatical morpheme for 'then', viz. *mani* 'then', which has replaced the sequence *si la* 'and then'. As shown in Zeitoun (2007:430), *mani* 'then' is used to express consecutive temporal relations, while *la* 'and' coordinates verbs that represent simultaneous actions. It is easy to witness, based on these examples, that the same (non-finite) constraint on verbs preceded by *la* 'and' and *mani* 'then'. Compare the grammaticality of (18a–b) and (19a–b).

(17) Mantauran
takanaw *la* *leheai*
Takanaw and Leheai
'Takanaw and Leheai'

(18) Mantauran

a. *o-kane* *la* ***'ongolo***
ACT.REAL-eat and NFIN.drink
'eat and drink'

b. * *o-kane* *la* ***o-'ongolo***
ACT. REAL-eat and ACT.REAL-drink

(19) Mantauran

a. *mani* ***dhaace*** *mani* ***oa-nga*** *'ini-talodho.*
then NFIN.leave then NFIN.go-already cross-bridge
'(They) left and crossed the bridge.'

b. * *mani* ***o-dhaace*** *mani* ***mo-oa-nga*** *'ini-talodho.*
then ACT.REAL-leave then ACT.REAL-go-already cross-bridge

One may question the viability of my hypothesis regarding the replacement of the conjunction *la* 'then' by *mani* 'then' in Mantauran.[11] Such a claim is actually supported by the fact that *la* and *mani* are both found in the six texts (about a hundred sentences) recorded by Ogawa & Asai (1935) on Mantauran. In very few sentences (only five were spotted), *la* occurs alone (20a); in nine sentences, *la* occurs with *mani* (20b). In all the other sentences (twenty three sentences in all), it is *mani* that occurs alone (20c). In these three examples, the first line capture Ogawa & Asai's (1935) transcriptions; the following three lines (transcription, morpho-phonemic glosses and translation) are mine.

(20) Mantauran

a. *la* *kaðalamu* *savaɭü.*
la *kadhalame* *savare.*
then NFIN:like/love young man
'Then the young man loved (her).' (Based on Ogawa & Asai 1935:391, Text 5, Sentence 4)

b. *la* *mani* *atsakülaʔ-ü.*
la *mani* *'acakelae.*
then then NFIN.marry
'Then they got married.' (Based on Ogawa & Asai 1935:391, Text 3, Sentence 14)

c. *mani* *ðaʔtʂü* *anoðaðoʔa...*
mani *dha'ace* *anodhadho'a...*
then NFIN.leave two
'Then the two of them left.'
(Based on Ogawa & Asai 1935:391, Text 2, Sentence 4)

It seems sensible to assume that at one point in the history of Mantauran (perhaps over a hundred years ago), *mani* was introduced in this language (its origin is unknown at this point) and started to replace *la* in its sentential conjunction function.

[11] Stacy F. Teng (pers. comm., March 2014) raises a different possibility that cannot be disregarded lightly: It may have been the case that both *la* and *mani* co-existed and that the sequence *la mani* meant 'then so' at the PR level. While Mantauran kept *mani*, the other dialects retained *la*. However, while we have written documents from the Japanese era that show that there was a period where *la* and *mani* occurred together in Mantauran, we have no trace of such co-occurrences in the other dialects.

4.3 The quotative construction *mani ia* in Mantauran

As a consequence of the replacement of *la* by *mani* in Mantauran, it is not surprising to find the quotative construction in this dialect made up of *mani* 'then' and *ia* '(be) so' (thus *mani ia* 'then so') rather than *la ya*, found in the other Rukai dialect. Note that in Ogawa & Asai's (1935:390–393) texts, *mani* is not only found in its function as a conjunction (19c) but also as part of the sequence *mani ia* (21). The verb *om-ia* 'so' (22) was also recorded at that time. On the other hand, *la ia* is not found.

(21) Mantauran

maːtülüküŋaða	*mane*	*maütüŋaɭao...*
"*ma-ateleke-nga-dha*"	*mani ia*	"ma-'ete-nga-lrao"...
STAT.FIN-cold-already-3SG.GEN	then	NFIN.so STAT.FIN-die-already-1SG.NOM

'"…If it is cold" he said, "(then it means that) I am already dead."'
(Based on Ogawa & Asai 1935:393, Text 6, Sentence 6)

(22) Mantauran

naula	*itovoro*	*umiðijaʔü*.
"*naola*	*'itovoro*"	*om-i-dh-ia'e*.
try	ask	ACT-so-3SG.GEN-1SG.OBL

'He said to me: "I will try to ask (you)."'
(Based on Ogawa & Asai 1935:393, Text 3, Sentence 6)

The following two examples provide an illustration of *mani ia* in texts recorded in the 1990s.

(23) Mantauran

a. *mani* "*alopo-nai"* *mani ia* *dhona'i*
then NFIN.hunt-1PL.EXCL.NOM then NFIN.so that
lalake-dha.
child-3SG.GEN
'Then their children said: "We are going to hunt."' (Recorded August 1992)

b. *mani ia* *'ina'i:* "*'aliki-mita* *'i*
then NFIN.so this from-1PL.INCL.NOM
ta-teme *opalrae*" *mani i-iae.*
SUBJ.NMLZ-obstruct grindstone then NFIN.so-1SG.OBL
'(He) said this to me: "We come from a hole obstructed by a grindstone."'
(Recorded November 1998)

I assume that the morpheme *mani* is the sentential conjunction introduced above. The verb *ia* '(be) so' is the non-finite form of *om-ia* '(be) so (FIN)' (24a). As in the other Rukai dialects, the verb *ia* '(be) so (FIN)' can occur without the conjunction *mani* 'then', as shown in (24a). It can also convey one's/someone's impression or (after-)thought. In (24b), *m-ia*, for instance, can be interpreted as 'he said to himself'.

(24) Mantauran

a. *mani* "*la* *ka-ia-'-iae* *ana.*"
then and in.fact-NFIN.so-2SG.GEN-1SG.OBL this

"*ia'e,* *om-ia-lrao* *ana.*"
yes ACT-so-1SG.NOM that
'"In fact you said that to me." "Yes, I did."'

b. *mani* *oa* *m-alra* "*nao-amece*
then NFIN.go SUBJ-take 1SG.NOM-NFIN.bring
lalake-li" *m-ia.*
child-1SG.GEN SUBJ-so
'He went to take [the flower] thinking: "I am to bring (it back) to my children."'

Despite the similarities outlined above, the verb (*om-*)*ia* '(be) so', Mantauran exhibits at least two discrepancies.

(i) First, the verbal prefix that attaches to *ia* 'so' is *om-* rather than *am-* as in the other Rukai dialects. Three hypotheses can be advanced to account for this discrepancy, though none seems really satisfactory. The first consists in supposing that the prefix *o-* 'ACT.DYN.FIN' blended with the prefix *am-*. This seems, however, somehow improbable, as the prefix *o-* represents in itself the monophthongisation of the PR bimorphemic prefix *o-a- 'ACT.DYN.FIN-REAL' and PR *am- could eventually be decomposed as *a-* 'REAL' and *m-* 'ACT.DYN'. It is difficult to imagine why these prefixes would occur together (as in **o-a-am) since they basically carry the same function. The second hypothesis, which is not strongly supported either, is to assume that the occurrence of *om-* rather than *am-* is due to language contact with Saaroa. Among all the Rukai dialects, Mantauran is known to have been in close contact with Saaroa (Ferrell 1969). In Saaroa, verbs starting with a vowel generally take the av marker *um-*, e.g. *um-a-aala* 'take (AV)', *um-arace* 'bite (AV)', *um-usalhe* 'rain (AV)', *um-u* 'eat (AV)' (Pan 2012). In Mantauran, only five verbs have been collected that start with a vowel. These five verbs take the prefix *om-* 'DYN.FIN', cf. *om-alra* 'take (DYN.FIN)', *om-oa* 'go (DYN.FIN)', *om-iki* 'exist, be at (DYN.FIN)', *om-ia* '(be) so (DYN.FIN)', *om-i'a* '(be) like (DYN.FIN)'. The verbs, with the exception of *om-alra*, are not identical to those in Saaroa and it seems somehow improbable that only the prefix *om-* was borrowed. The third hypothesis was suggested by one of the reviewers: It may happen that *om-* was replaced in all the Rukai dialects but Mantauran by *am-*, but the reason for such a change seems as improbable as the first two hypotheses. As pointed out above, *am-*, occurring only with vowel-initial verb stems, is analysed as a bimorphemic affix consisting of *a-* 'real' and *m-* 'DYN.FIN', the same way we have *o-a-* [*w-a-*] with most dynamic verbs. It may have been used as such in order to avoid homonymy with the stative prefix *ma-*.

(ii) Second, while it was shown above that in the Rukai dialects (*am-*)*ya* means either '(be) so' or '(be) like', in Mantauran, these two meanings are manifested by two distinct verbs, *om-ia* '(be) so' as opposed to *om-i'a* '(be) like'.[12] Compare (25a–b):

[12] I am not aware of the existence of a verb corresponding to *om-i'a* in other Rukai dialects. Such a discrepancy raises yet another question regarding whether Mantauran has preserved or innovated what was originally found in PR. Were there originally two distinct verbs, as in Mantauran, at the PR level and due to the fact that most dialects (except Tona) do not have a glottal stop, the two forms merged? Or is it Mantauran which has innovated a new form, distinct from the one that was earlier found in PR?

(25) Mantauran

a. *kasa-ta* "*'oponoho* *vaha-ta*"
only-1PL.INCL.GEN Mantauran language-1PL.INCL.GEN
om-ia-nai.
ACT-so-1PL.EXCL.NOM
'We just knew (lit. said) that our language was Mantauran.'

b. *om-i'a-ka-i-imite* *vaha-ni*
ACT-like-NEG-3SG.GEN-1PL.INCL.OBL language-3SG.GEN
tali-lrao~lraodho.
belong.to-RED~down(wards)
'The language spoken by the people from Pingtung county (i.e., Budai dialect) is not like ours.'

The verb *om-i'a* exhibits the same verbal alternations as *om-ia*, cf. *m-i'a* '(be) like (SUBJ)' (26a), and *i'a* '(be) like (NFIN)' (26b).

(26) Mantauran

a. *lo* *m-ore-lehe'e-lidha* *taka-palra* *m-i'a-nga* *dhona*
if SUBJ-hold-ritual-3PL.GEN with-companion SUBJ-like-already that
aitina, *ta-ka-oca-*(*a*)*e-dha.*
PL:middle-aged.woman LOCNMLZ-STAT.NFIN-person-LOCNMLZ-3SG.GEN
'If (we) recited the right prayers, then (s/he) could become a real person like her/his parent.'

b. *mani* *lriho'o-nai* *'ina* *'oponoho*
then NFIN.know-1PL.EXCL.NOM this Mantauran
"*ka-i'a-ta* *aanai*" (***o*)*m-ia****-nai.*
in fact-NFIN.like-1PL.INCL.GEN that ACT-so-1PL.EXCL.NOM
'That's how we, Mantauran, learnt that we actually are Rukai.'
(Lit. 'Then we knew that we, Mantauran, "we are actually like that."')

Besides, (*om-*)*ia* 'so' and *(om-)i'a* '(be) like' cannot be replaced one by the other. Consider the (un)grammaticality of (26b) and (27).

(27) Mantauran

* *mani* *lriho'o-nai* *'ina* *'oponoho*
then NFIN.know-1PL.EXCL.NOM this Mantauran
"*ka-ia-ta* *aanai*" (*o*)*m-i'a-nai.*
in fact-NFIN.so-1PL.INCL.GEN that ACT-like-1PL.EXCL.NOM

4.4 The ongoing replacement of *mani* 'then' and *mani ia* 'then so' by *omi* 'like that' in Mantauran

Mantauran is undergoing another major syntactic change, i.e., the rise of *omi*, tentatively glossed as 'like that', which is replacing *mani* 'then', as illustrated in (28). Note that *omi* is not found in Ogawa & Asai's (1935:390–393) texts, nor was it found in the texts collected with my late (and major) informant, Lü Yu-zhi, who passed away in

2000 at the age of seventy-six. Example (28) is drawn from a text collected in 2004 from Ching-mei Tsai, born in 1935. Because of the resemblance established between *om-ia* and *omi*, it can be hypothesised that *omi* is derived from *om-ia* 'so' by analogy.

(28) Mantauran

a. *mani* *o-valrio* *omii* *m-oa* *lamengae-dha* *omi.*
then NFIN.go-village FILL SUBJ-go spouse-3SG.GEN like.that
'Then he went back home to his wife.' (Recorded in 2004)

b. *mani* *o-valrio* *omii* *m-oa* *lamengae-dha* *mani.*
then NFIN.go-village FILL SUBJ-go spouse-3SG.GEN like.that
'Then he went back home to his wife.'

The morpheme *omi* 'like that' also tends to replace *mani ia* 'then so', as in (29). In its first occurrence, *omi* 'like that' replaces *mani ia* 'then so'. At the end of the sentence, *omi* 'like that' replaces *mani* 'then'.

(29) Mantauran

"avase-ni *o-'ange'ang-iae* *vahange-li"* *'a*
a.pity-3SG.GEN ACT-hurt-1SG.OBL belly-1SG.GEN TOP
'ini-'ange'ang-ae *omi-idhe* *omi...*
pretend-NFIN.hurt-pretend like.that-3SG.OBL like.that
'"What a pity (that) my stomach is hurting" he said to him/her pretending to suffer.'

Note that *omi* cannot precede, as *mani* does, the verb *ia*. Compare the ungrammaticality of (30a) and (30b).

(30) Mantauran

a. * "*ia'e,* [*amo-*]*lrange'e-mita*" *omi* *ia* *mani*…
yes [IRR-]NFIN.cultivate-1PL.INCL.NOM like.that NFIN.so then

b. "*ia'e,* [*amo-*]*lrange'e-mita*" Ø *ia* *omi* *mani*…
yes [IRR-]NFIN.cultivate-1PL.INCL.NOM Ø NFIN.so like.that then
'"All right, (let's start) to cultivate" (she) said.'

It is difficult to exactly pin down the meaning and function of *omi*. Unlike *mani* 'then', *omi* does not have any lexical content. It is tentatively glossed as 'like that' for want of a better translation. While *mani* is a conjunction, *omi* can, a priori, be analysed as a (quotative) verb because (i) it can be followed by a pronoun (31a), usually in the oblique case, and *mani* cannot (31b) and (ii) it shares, at least superficially, the same marking as other verbs, cf. *mi* as shown in (32a) and *i* as in (32b). Note that in both examples, it is rather difficult to determine *mi* and *i* represent the alternant form of *om-ia* '(be) so (FIN)' or whether it should be treated as deriving from *omi* 'like that'.

(31) Mantauran

a. *mani* *ia* *ana* *omi-idhe* “*ina-li*” *m-ia*
then so that like.that-3SG.OBL mother-1SG.GEN SUBJ-so
omi-idhe.
like.that-3SG.OBL
‘Then he said: “my mother”.’

b. * *mani* *ia* *ana* *mani-idhe* “*ina-li*” *m-ia*
then so that then-3SG.OBL mother-1SG.GEN SUBJ-so
mani-idhe.
then-3SG.OBL

(32) Mantauran

a. “*kani* *kapa-oa-’-iname* *’akoela?*”
why NFIN.continuously-go-2SG.GEN-1PL.EXCL.OBL SUBJ.laugh.at
mi-idhe.
NFIN.so?/like.that?-3SG.OBL
‘They asked (again): “Why did you laugh at us?”’

b. “*m-o-a* *dhoace!*” *Ø* *i-idhe* *mani*....
SUBJ-go-IMP SUBJ:leave NFIN.so?/like.that?-3SG.OBL then
‘“Go and leave!” (they) said to her and then….’

The morpheme *omi* differs from other verbs and behaves like *mani* in that:

(1) It cannot be followed by any aspectual marker, cf. **omi-nga* intended for ‘like that-already’ and **mani-nga* intended for ‘then-already’. Note that (*mani*) *ia*, on the other hand, can co-occur with *-nga* ‘already’, cf. *ia-nga omi* ‘already said then…’ vs *mani ia-nga* ‘already said’.

(2) It must be followed by a non-finite verb. This is shown by the alternation between *kama-* ‘continuously (FIN)’, which occurs in sentence-initial position as the main verb and *kapa-* ‘continuously (NFIN)’ in (33a), which follows *omi* ‘like that’. Note that *omi* in that position can be replaced by *mani* (33a’). If *omi* were a full-fledged verb, it would be followed by another verb marked as subjunctive, as shown in (33b–b’).

(33) Mantauran

a. *ka**m**a-kilakili* *omi-ilidhe*
continuously-dance.the.valiant.soldiers’.dance like.that-3PL.OBL
omi. *ka**p**a-oa* *polidhao*
like.that NFIN.continuously-go SUBJ.watch
omi-idhe *dhona’i* *titina* *omi.* [...]
like.that-3SG.OBL that middle-aged.woman like.that
‘They [=Monkeys] were dancing the dance of the valiant soldiers. The woman was watching (them). […]’

a’. *ka**m**a-kilakili* *omi-ilidhe*
continuously-dance.the.valiant.soldiers’.dance like.that-3PL.OBL
mani. *ka**p**a-oa* *polidhao…*
like.that NFIN.continuously-go SUBJ.watch
‘They [=Monkeys] were dancing the dance of the valiant soldiers. (The woman) was watching (them). […]’

b. *moa* *k**o**ne!*
SUBJ:go:IMP SUBJ:eat
'Go eat!'

b'.* *moa* *k**a**ne!*
SUBJ:go:IMP NFIN.eat
'Go eat!'

Based on the above data and analysis, it seems that in the speech of younger Mantauran speakers, (i) *mani* 'then' has been somehow re-installed in its conjunctive function though it is being replaced by *omi*, (ii) *mani ia* is being replaced by *omi*, and (iii) the two forms *mia* and *ia* (all the more when the latter is followed by a third person oblique pronoun) are analytically ambiguous: They may either be analysed as the verb (*om*)*ia* '(be) so' or (*om*)*i* 'like that'.

5 Reconstruction at the Proto Rukai (PR) level

The above discussion allows the reconstruction of a number of morphemes at the Proto Rukai level. The first morpheme which is reconstructible is the (phrasal) coordinator PR *si. I assume that it has been replaced by *la* in Mantauran. Though the morpheme *la* is also found in Tsou, cf. *ho* 'and', and Saaroa, cf. *lha* 'and' and *ɬa is reconstructed as a Proto Rukai-Tsouic morpheme by Tsuchida (1976:11), it seems more plausible to assume that (i) it was downgraded from sentential conjunction to a NP/VP coordinator or that (ii) it was borrowed from Saaroa. I believe that either hypotheses is reasonable because PR *s is reflected as *'* in Mantauran and in grammatical words, PR *k is also reflected as *'* in Mantauran (cf. the passive form *'i-*, reflex of PR *ki-). If *si 'and' had been reflected as **'i in Mantauran, it would have clashed with the *'i* which comes from the (nominative) case marker *ki*. The adoption of *la* 'and' as a means of homonymy avoidance (Hopper & Traugott 2003:102–103) after Mantauran underwent these two phonological changes seems thus plausible.

The second morpheme that can be proposed for reconstruction is the conjunction PR *la. For some reasons I cannot yet identify—after all, *la* could have been used both as a phrasal and sentential conjunction in earlier times—the conjunction *la 'then' was replaced by *mani* 'then' in Mantauran. It was shown above that both *la* and *mani* were recorded in the early 1930s, when Ogawa and Asai undertook their fieldwork on Mantauran.

I schematise these reconstructions as follows:

(34) Reconstruction of the phrasal coordinator *si* and the sentential conjunction *la* in PR

	Proto Rukai	Tona	Mantauran
a. NP/VP	*si	*si*	*la*
b. S	*la	(*si*) *la*	*mani*

The third bimorphemic sequence that can be reconstructed is *la ya 'then so' along with the verbal forms *ya, *m-ya and *am-ya. As demonstrated in section 4, we believe that Mantauran, after having undergone the replacement of *la* by *mani*, has also replaced the PR sequence *la ya 'then so' by *mani ia*. Besides, while it has preserved the verbal

form *ia, the meaning of this verb may have been narrowed down while a new verb *om-i'a* '(be) like' came in existence.

(35)	Gloss	PR	Tona	Mantauran
	a. 'then so'	*la ya	*la ya*	*mani ia*
	b. 'so, (be) like'	*am-ya ~ *m-ya ~ *ya	*am-ya ~ m-ya ~ ya*	*om-ia ~ m-ia ~ ia* 'so' *om-i'a~ m-i'a ~ i'a* '(be) like'

The morphosyntactic changes that have taken place in Mantauran can be summarised in Figure 1.

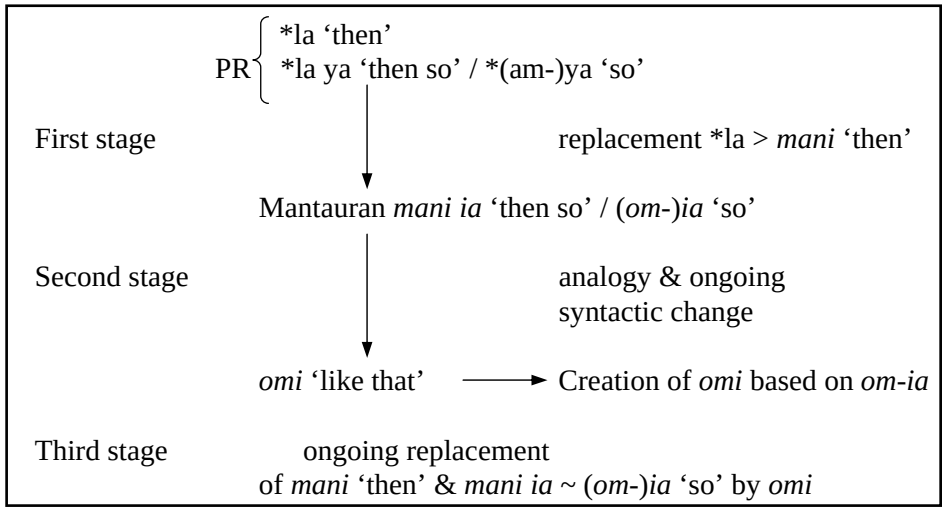


Figure 1: Evolution of the verb 'say' in Mantauran Rukai

Unlike other Rukai dialects, Mantauran has replaced the conjunction PR *la 'then' by *mani* 'then'. It has had for consequence the replacement of PR *la in the quotative construction, cf. PR *la ya > Mantauran *mani ia*. Mantauran has also innovated in the use of the prefix preceding the verb *ia* 'so', cf. *om-* instead of *am-. Following this first series of changes, Mantauran has further undergone the following innovations: (i) the creation of *omi* 'like that' based on *om-ia* 'so' (perhaps through analogy) and (ii) the ongoing replacement of *mani* 'then' by *omi* 'like that' and *mani ia* 'then so' ~ *om-ia* 'so' (verb of saying) by *omi*.

6 Conclusion

In this paper, I have tried to show the use of the conjunction in the formation of the verb of saying in Rukai. I have also highlighted a number of morpho-syntactic changes that have occurred in Mantauran Rukai and shown, through a reconstruction at the PR level, the chronology of these changes.

References

Buchstaller, Isabelle and Ingrid Van Alphen. 2012. Preface: Introductory remarks on new and old quotatives. In *Quotatives: Cross-linguistic and cross-disciplinary perspectives*, ed. by Isabelle Buchstaller and Ingrid Van Alphen, xi-xxx. Amsterdam and Philadelphia: John Benjamins.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*, Institute of Ethnology, Academia Sinica Monograph No.17. Taipei: Academia Sinica.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge Textbooks in Linguistics . Cambridge: Cambridge University Press. Second edition.

Hsieh, Fuhui. 2011. On the development of SAY verbs as stance markers in some Formosan languages. Paper read at the workshop on Stance Marking Across Languages: Typological, Diachronic and Discourse Perspectives. 18–20 July. Hong Kong: The Hong Kong Polytechnic University.

Hsieh, Fuhui. 2012. On the grammaticalization of the Kavalan SAY verb *zin*. *Oceanic Linguistics* 51.2:464–489.

Kohn, Mary Elizabeth, and Hannah Askin Franz. 2009. Localized patterns for global variants: The case of quotative systems of African American and Latino speakers. *American Speech* 84.3: 259–297.

Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* 48.1:1–92.

Li, Paul Jen-kuei. 1995. Distribution of the Formosan languages and migration of the tribes. In *Proceedings of the First International Symposium on Languages in Taiwan*, ed. by Feng-fu Tsao and Mei-hui Tsai, 1–16. Taipei: Crane Publishing Co., Ltd.

Li, Paul Jen-kuei. 1996. The pronominal systems in Rukai. In *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer, 209–230. Hamburg: Abera Verlag.

Ogawa, Naoyoshi and Erin Asai. 1935. *Gengo ni yoru Taiwan takasago-zoku densetsu-shu* [*The myths and traditions of the Formosan native tribes (texts and notes*)]. Taihoku [Taipei]: Taihoku Imperial University. [In Japanese]

Pan, Chia-jung. 2012. A grammar of Lha'aula, an Austronesian language of Taiwan. PhD dissertation. Cairns: James Cook University.

Starosta, Stanley. 1994. Proto Rukai-Tsouic: Subgroup or treetop? Paper presented at the Seventh International Conference on Austronesian Linguistics. 22–27 August. Leiden, the Netherlands.

Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa

Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng, 683–726. Symposium series of the Institute of History and Philology, Academia Sinica 3. Taipei: Academia Sinica. Reprinted in *Formosan linguistics: Stanley Starosta's contributions*, ed. by Elizabeth Zeitoun, 737–769. Language and Linguistics Monograph Series C6-2. Taipei: Academia Sinica, 2009.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No. 5. Tokyo: University of Tokyo.

Tu, Wen-Chiu. 1994. A synchronic classification of Rukai dialects in Taiwan: A quantitative study of mutual intelligibility. PhD dissertation. Urbana: University of Illinois.

Tu, Wen-chiu and Chin-chuan Cheng. 1991. A linguistic classification of Rukai Formosan. Paper presented at the Sixth International Conference on Austronesian Linguistics. 20–24 May. Honolulu: University of Hawai'i at Mānoa.

Vandelanotte, Lieven. 2012. Quotative *go* and *be like*: Grammar and grammaticalization. In *Quotatives: Cross-linguistic and cross-disciplinary perspectives*, ed. by Isabelle Buchstaller and Ingrid Van Alphen, 173–202. Converging Evidence in Language and Communication Research 15. Amsterdam and Philadelphia: John Benjamins.

Wolff, John U. 2010. *Proto-Austronesian phonology with glossary, Volume 1*. Ithaca, NY: Cornell University, Southeast Asia Program Publications.

Zeitoun, Elizabeth. 1995. Issues on Formosan linguistics. PhD dissertation. Paris, France: Université Denis Diderot. [English version]

Zeitoun, Elizabeth. 2000. Dynamic vs. stative verbs in Mantauran (Rukai). *Oceanic Linguistics* 39.2:415–427.

Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto Rukai morpho-syntax. Keynote address at the Tenth Annual Meeting of the Austronesian Formal Linguistics Association. 28–30 March. Honolulu: University of Hawai'i at Mānoa.

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran* (*Rukai*). Language and Linguistics Monograph Series, No. A4-2. Taipei: Academia Sinica.

Zeitoun, Elizabeth. in preparation. A comparative study of the Rukai dialects. Ms.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in the Formosan languages. *Oceanic Linguistics* 39.2:391–414.

Zeitoun, Elizabeth and Hui-chuan Lin. 2003. *We should not forget the stories of the Mantauran: Memories of our past.* Language and Linguistics Monograph Series A4. Taipei: Academia Sinica.

19 *Body part nomenclature and categorisation in Seediq**

AMY PEI-JUNG LEE

1 Introduction

1.1 Issues on body part terms

Body part terminology is an ideal domain for exploring lexical universals. It provides valuable linguistic resources for our understanding towards how a language differs from another in terms of representing the human body, which is both the source domain and the target domain of our perception and emotion.

Given this similar body structure, it is assumed that, in terms of nomenclature, every language has distinctive lexicon for labelling the salient parts of the body, such as HEAD, HAND, ARM, LEG, or FOOT. For example, Brown (1976) proposes twelve principles regarding the implicational universals on nomenclature of body part terms from a meronymic perspective, stating that body part meronymies rarely exceed five levels and never exceed six, and it is the third level which is generic.

Similarly, Andersen (1978) proposes lexical universals in relation to nomenclature of body part terms as the following: (1) All languages label BODY, HEAD, ARM, EYES, NOSE, and MOUTH; (2) the terms for HEAD, TRUNK, ARM and LEG will always be 'possessed by' BODY; (3) all languages will have labels for FINGER, TOE, and NAIL; (4) if a language has a distinct term for LEG and FOOT, it will have a distinct term for ARM and HAND; and (5) a language with distinct terms for individual toes implies having distinct terms for

* Funded by the National Science Council, Taiwan (NSC 101-2628-H-259-003-MY2), this paper could not have been able to be accomplished without the generous help from my language consultants. Many thanks go to Aki Pitay (aka Aki Lasi) and Tapas Tanah for Truku; Sukit Hawit and Takun Lihang for Eastern Toda; Cugi Basaw, Kadu Lulu, Kumu Tapas, and Watan Teymu for Central Toda; and Dakis Pawan, Kumu Pawan, and Tahuc Tado for Tgdaya. I dedicate this paper to Prof. Lillian M. Huang, whose guidance and encouragement have accompanied me through the years of my academic development. I am also grateful for the constructive comments from Paul Jen-kuei Li, Hsiu-chuan Liao, Stacy F. Teng, Joy Wu, and Elizabeth Zeitoun, as well as several personal communications with Apay Ai-yu Tang. The suggestions and feedback comments provided by the two anonymous reviewers are also deeply appreciated.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 451–483
Asia-Pacific Linguistics, 2015

individual fingers. With FINGER and TOE three types of labelling are identified: (1) no individual digits are labelled, (2) some individual fingers but no individual toes are labelled, and (3) a subset of fingers and toes are labelled.

Contrary to the above claims, Majid (2010) points out that not all the lexical universals as listed by Brown (1976) and Andersen (1978) hold true in languages. The following are some counterexamples. First, it is not universal that all languages have a lexical item denoting BODY. Tidore, a Papuan language spoken on the island of Tidore in the North Moluccas, has no indigenous word for BODY. The native term *mansia* is used to refer to person or human being (van Staden 2006), and the language borrows a loanword *badan* 'body' from Indonesian. Second, there is no term for HEAD in Jahai, an Aslian language of Malaysia (Burenhult 2006). The term *kuy* (and its cognates in the languages) is the closest equivalent term to be used as HEAD, yet it is considered to have a narrower sense, which should be 'top of the head'. For example, *tæc koy* 'cut head' actually means to cut one's hair, not to behead someone. Third, Majid (2010:64) also mentions that Lavukaleve (a Papuan language of the Solomon Islands) only has one general term for ARM and LEG (Terrill 2006), which contradicts the claims by Andersen (1978) and Brown (1976) that ARM and LEG are always given distinct terms.

As a reaction to the Special Issue of *Language Sciences* (2006, vol. 28, no. 2–3) "Parts of the body: Cross-linguistic categorisation", Wierzbicka (2007) puts forward seven general principles of conceptualisation of BODY and argues for a semantic methodology based on the use of the natural semantic metalanguage (NSM) for semantic typology. However, this approach is questioned by Majid (2010:63), stating that 'meaning as reduction to simpler components is not widely accepted in the cognitive and linguistic sciences today, rather many take an "embodiment" or "simulation" viewpoint instead.'

How body parts are named also leads to how the body part lexicon in a language is categorised and conceptualised. The human body is also one of the main sources for figurative language. Semantic extensions of body part terms are closely linked to the speakers' categorisation and conceptualisation of the world around them. Despite the human body being a universal model, categorisation and conceptualisation of body part terms vary among languages.

Based on a universal perspective, Brown & Witkowski (1981) present a study on figurative use of body part terms in 118 languages, covering major phyla in the world. The study shows that certain body part terms are often used figuratively in a similar way in the languages surveyed. Naming regularities draw their attention to the equation of 'pupil of the eye' as 'human (mother, child)', of 'testicle' as 'egg', of 'digit/finger' as 'kinsman', and of 'muscle' as 'mouse/lizard'. These universal tendencies are due to 'underlying marking principles, intra-language lexical constraints, physical world/perceptual givens, and language context and use considerations'. Their findings can serve as a basis for a typological comparison with the examples in the Seediq language discussed in this paper.

In the realm of Austronesian languages, Blust (2009:311[2013:322]) summarises the body part terms which are commonly extended as metaphors. These metaphors show that external body parts such as 'head', 'eye', or 'ear' are often used for designating the world of physical sense impressions, while those based on internal organs such as 'liver' or 'gall' refer to qualities of temperament or character.

Similarly, Li (2007) states that figurative use of body part terminology is channelled by three basic principles: (1) from up to down, (2) from front to back, and (3) from part to whole. These principles are reminiscent of the embodied concepts such as UP-DOWN, FRONT-BACK, PART FOR WHOLE as proposed by Lakoff & Johnson (1980).

The data from some Formosan languages, including Kavalan, Saaroa, and Mayrinax Atayal, provided by Li (2007) are especially interesting. For example, in Saaroa the seat of emotion is the belly. Sadness is expressed by 'having a spicy belly'. There have been some studies on emotion in relation to body part terms in Formosan languages, such as Huang (2002) on Tsou, Yeh (2002) on Atayal, Hsieh (2007) on Kavalan and Saisiyat, and Hsu (2008b) on Truku Seediq. However, so far there has not been a paper which studies the body part terms per se as a lexical system in a Formosan language.

Therefore, by choosing one Formosan (Austronesian) language, Seediq, as a subject for exemplification, the aim of this paper is two-fold: (1) to systematically investigate nomenclature and categorisation of Seediq body part terminology, and (2) to connect the observations in Seediq body part terms to cognitive and typological significance.

1.2 The Seediq language and its speakers

Seediq is an Austronesian language of the Atayalic subgroup (Li 1981). It is spoken in central and eastern parts of Taiwan, and comprises of three main dialects: Tgdaya, Truku, and Toda. Tgdaya is mainly spoken in Nantou County, situated in the central part of Taiwan, whereas Truku and Toda are spoken not only in Nantou but also in Hualien County, located in the eastern part of Taiwan. The language as a whole is facing endangerment to some extent, with fluent speakers usually aged in their 40s and older.[1]

The total population is estimated at around 57,500. This figure is based on the census conducted in March 2014, by the Department of Household Registration, Ministry of the Interior, Taiwan. The population of Truku in Hualien County (excluding Tgdaya and Toda) is registered as 44,734, and that of Seediq (including Tgdaya, Toda, and Truku in Nantou County) is numbered 12,788.

In the main texts and the tables of this paper the exemplified data are based on the Truku dialect spoken in Hualien County, which has the highest population of the Seediq people. However, dialectal variations are also noted wherever necessary, with Tgdaya abbreviated as (Tg), Truku as (Tk), and Toda as (Td). Strictly speaking, Toda also distinguishes Central and Eastern sub-dialects, which are abbreviated accordingly as (CTd) and (ETd). Eastern Toda is considered to have been influenced by the Truku in Hualien (Lee 2012b).[2]

---

[1] Throughout the paper only the dialect names such as Tgdaya, Truku, and Toda, are used, which are equivalent of Tgdaya Seediq, Truku Seediq, and Toda Seediq, respectively.

[2] The following orthographic symbols are adopted for transcribing the data in the three dialects: *p, b, t, d, k, g, q, ', s, x, h, c, j, m, n, ng, r, l, w, y* for the consonants, and *i, e, a, u* for the vowels. Most of the symbols are the same as their corresponding symbols in International Phonetic Alphabet, except for the following. The symbol *g* stands for the velar fricative [ɣ], *h* the pharyngeal fricative [ħ], and the apostrophe' the glottal stop [ʔ]. The symbol *c*, however, represents different phonetic values in the dialects. It represents an allophone of the phoneme /ts/ in Tgdaya and Central Toda, for example, *camac* 'wild animals or games' (Tg) or *ricah* 'plum' (Tg, CTd). In Truku and Eastern Toda [ʦ] as an allophone of the phoneme /t/ occurs at

To facilitate the presentation of the data in this paper, the following phonological differences among the dialects are mentioned in a brief passing. In terms of consonants, the prevocalic alveolar affricate *ts* (written as *c*) in Tgdaya and Central Toda corresponds to *s* in Truku and Eastern Toda, such as *qcurux* and *qsurux* 'fish'. The /w/ sound in Toda often corresponds to /g/ in the other dialects. Some examples are *bawa~baga* 'hand' or *gaga~wawa* 'there'. As for the vowels, word-final *aw and *ay have undergone monothongisation to *o* and *e* in Tgdaya (Li 1981), thus *hidaw* 'the sun' and *balay* 'very, indeed' in Truku and Toda is written as *hido* and *bale* in Tgdaya.[3]

1.3 Organisation of this paper

Nomenclature and categorisation are the two aspects focused on in this paper. To elaborate the two aspects, this paper aims to systematically explore the following contents: (1) Structural properties, including the inventory of external and internal parts, which is classified into two types of forms, simplex and complex terms, following Enfield et al. (2006), who adopt these two broad formal categories. Simplex terms are monomorphemic terms such as HEAD, ARM, and KNEE. Complex terms are 'polymorphemic and are usually semantically-motivated expressions such as KNEECAP, FINGERNAIL, INDEX FINGER' (Enfield 2006:150). Given that complex body part terms in Seediq are mostly formed by compounding, the structural and semantic compositions of the compounds are also discussed. (2) Semantic extensions, including spatial representation, metaphors, and metonymies. A study on the semantic extensions of body part terms in this language may enhance our understanding towards the correlation between embodiment and conceptualisation.

This paper is thus structured as follows. Section 2 presents the inventory of body part terms in Seediq, which is divided into eight sub-sections, including BODY, HEAD, FACE, EXTREMITIES and TORSO, INTERNAL PARTS, EFFLUVIA, ANIMAL ANATOMY, and SKIN. Note that this division of 'eight' sections does not intend to serve as an arbitrary organisation of how the Seediq people conceptualise BODY, but rather function to present the linguistic manifestation of the Seediq body part terms. The inventory of body parts is summarised in a table at the end of each subsection for a quick reference. Section 3 discusses the composition and types of complex terms as compounds. This is followed by section 4 which studies the semantic extensions of Seediq body part terms, including spatial orientation and figurative language, such as metaphors and metonymies which utilise body part terms as the source domain.

---

word-final position, such as *buut* 'bone', while the phoneme /t/ is palatalised as [tɕ] before the high vowel [i] or the palatal glide [j], as in *cyaqung* 'crow' (Tk, ETd). The symbol *j*, standing for the palatal stop [ɟ], is used in Truku and Eastern Toda, as an allophone of the phoneme /d/. As for the sonorants, the symbols *ng*, *r*, and *l* stand for the nasal velar [ŋ], the flap [ɾ], and the voiced lateral fricative [ɮ], respectively. The vowel symbol *e* stands for the mid front vowel [e] in Tgdaya but the schwa [ə] in Truku and Toda. For the sake of clarity the symbol *e* is written if occurring at penultimate position in this paper.

[3] For more details in Seediq phonology, the readers are referred to Yang (1976) and Li (1991) on Tgdaya, Tsukida (2005) and Lee (2009, 2010a) on Truku, and Lee (2012a, 2012b) on Toda.

2 The inventory of body part terms

The inventory of body part terms in Seediq is divided into the following sections: (i) BODY, (ii) HEAD, (iii) FACE, (iv) EXTREMITIES and TORSO, (v) INTERNAL PARTS, (vi) EFFLUVIA, (vii) ANIMAL ANATOMY, and (viii) SKIN. This organisation reflects the physical structure of body parts, which is culturally-independent. However, how these body parts are named and categorised linguistically vary among languages.

2.1 Body

In Seediq the concept of BODY (THE WHOLE) is encoded by the lexical term *hiyi*, which is approximately equivalent of the English word 'flesh' rather than 'body'. This is evidenced by its polysemous usage in animals and plants.

First, the term *hiyi* refers to human living body as a whole.[4] On the contrary, when *hiyi* is followed by an animal name to form a noun phrase, it mostly refers to the edible flesh (i.e., meat) of the animal rather than its living body.

(1) *iya=saku* *rebux-i* *hiyi=mu.*[5]
NEG.IMP=1SG.NOM+2SG.GEN[6] touch-UV.IMP flesh=1SG.GEN
'Don't touch me (my flesh)!'

(2) *malu/em-biyax* *bi* *ka* *hiyi=na.*
good/AV-be.energetic very NOM flesh=3SG.GEN
'His body is good (He has good health).'

(3) *m-talux* *ka* *hiyi* *rowdux.*
AV-hot NOM flesh chicken
'The chicken (meat) is very hot (in temperature).'

It is conceptually redundant to mention an animal's physical body (either living or dead), which is subsumed under its name. Therefore, while in sentence (4a) the term *hiyi* is optional, in sentence (4b) the addition of *hiyi* is not allowed. A comparison of (3) and (4b) reveals that when followed by an animal name, the term *hiyi* is only allowed in the context where it denotes the animal's edible flesh rather than its physical body.

---

[4] A human's figure or appearance is designated by the word *henegigan* 'figure, build', which is derived from *hiyi* /higig/ 'to have flesh, lean meat' > *henegigan* /h‹n›igig-an/ (flesh<PFV>-NMLZ).

[5] For interlinear morpheme-by-morpheme glosses this paper follows the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php). The other abbreviations used in this paper not included in the Leipzig Glossing Rules are: AV, actor voice; CONJ, conjunction; COS, change of state; IDEO, ideophone; LIG, ligature; LV, locative voice; PART, particle; STAT, stative; UV, undergoer voice; VBLZ, verbaliser.

[6] The pronoun *saku* is one of the portmanteau pronouns in Seediq, along with *misu* and *maku*. Their components are shown as follows (see also Holmer 1996:32).

Form	Internal components	Gloss
misu	1SG.GEN + 2SG.NOM	'as for me to you'
saku	2SG.GEN + 1SG.NOM	'as for you (SG.) to me'
maku	1SG.GEN + 2PL.NOM	'as for me to you (PL.)'

(4) a. *q‹m›ita=ku* (*hiyi*) *seejiq* (*ga*) *m-huqil* *hiya.*
<AV>see=1SG.NOM (flesh) person (PROG) AV-dead there
'I saw a person (lying) dead there.'

b. *q‹m›ita=ku* **hiyi* *kacing* (*ga*) *m-huqil* *hiya.*
<AV>see=1SG.NOM flesh ox (PROG) AV-dead there
'I saw an ox (lying) dead there.'

The term *hiyi* also contrasts with another term *siyang* in such a context; the former refers to lean meat, whereas the latter fatty meat.[7] Consider the following sentences.

(5) *se-hiyi* *balay* *ka* *babuy* *niyi.*
VBLZ-flesh very NOM pig this
'This (killed) pig has a lot of (edible) meat.'

(6) *ini* *k-siyang,* *hiyi* *bi* *kana* *ka* *babuy* *gaga.*
NEG STAT-be.fatty flesh very all NOM pig that
'That pig is not fatty. (It's got) all lean meat.'

(7) *lala* *bi* *siyang* *ka* *babuy=na;* *bilaq* *bi* *hiyi=na.*
many very fat NOM pig=3SG.GEN; little very flesh=3SG.GEN
'His pork has a lot of fat and a little lean meat.'

Moreover, *hiyi* forms a compound with *qhuni* as *hiyi qhuni* to denote FRUIT. The compound *hiyi qhuni* is literally a tree's flesh, i.e., the fleshy structure of a plant which contains seeds, and can be either edible or inedible in the raw state.

(8) *q‹em›pi=ku* *begu* *hiyi* *qhnui.*
<AV>squeeze=1SG.NOM soup flesh tree
'I am squeezing juice.'

(9) *m-n-hiyi* *risah=su* *da,* *Ubus*?
AV-PFV-flesh plum=2SG.GEN COS, Ubus
'Ubus, has your plum (tree) grown some fruits yet?'

Secondly, *hiyi* can be used for counting the number of persons. The example below suggests that *hiyi* subsumes the meaning of 'person' in such a context.

(10) *piya* *hiyi* *ka* *mcisa* *na*?
how.many flesh LIG teacher PART
'How many teachers are there?'

As the closest equivalent term for body, *hiyi* is used to refer to the whole body, person, animal flesh, as well as (edible or inedible) fruits (i.e., a tree's flesh). This case echoes Evans & Wilkins (2001, cited in Enfield et al. 2006:143) that 'terms denoting the body are diachronically unstable, and that these terms are often polysemous, being used to refer to skin, trunk, and person, as well as to body.' Therefore, Seediq contradicts

[7] The term *siyang* can also be used for plants, especially trees which are highly combustible, as shown in the following sentence:

m-siyang *bi* *ka* *qhuni* *niyi.*
AV-be.fatty very NOM tree this
'This tree burns easily.'

Brown's (1976) lexical universals that body is always labelled in languages, as in Seediq the term *hiyi* is not lexically specific to mean 'body (the whole)' only. Instead, the distribution of *hiyi* suggests that it is a case of polysemy, with the meaning of 'flesh' as basic and is metonymically extended to refer to human living body and (counting of) person.

Table 1: The body and its primary parts

Seediq term	Translation	Additional information
SIMPLEX		
baga	'hand/arm'	see section 2.4.1
dqeras	'face'	see section 2.3
hiyi	'body as a whole'	also 'flesh', '(animal's) lean meat'
papak/qaqay	'foot/leg'	see section 2.4.1
qsahur	'viscera'	see section 2.5
tunux	'head'	see section 2.2
COMPLEX		
seka hiyi	'torso'	lit. middle_flesh; see section 2.4.2

2.2 Head

The semantic universals predict that every language has a term for HEAD. The closest equivalent of HEAD in Seediq is *tunux*. When someone is beheaded, this is the term which refers to the lost part. Due to the practice of headhunting in the past, mentioning someone's head out of context is frowned upon. Therefore, a single noun phrase such as *tunux seejiq* 'a person's head' is reminiscent of the hearer that the head is no longer attached to the person. See the following sentence which is interpreted literally.

(11) *ungat ka tunux ubus. wada ngal-un ka tunux=na.*
NEG NOM head Ubus be.gone take-UV NOM head=3SG.GEN
'Ubus has no head. His head has been taken.'

In Seediq HEAD can be a meronym of FACE. In this way, *tunux* also refers to top (of the head). This is demonstrated by the following sentences, in which 'to cut head' means 'to cut one's hair' (cf. Majid 2010 on the discussion of Jahai). In such contexts, the meaning of *tunux* subsumes that of 'head hair', though *tunux* in these sentences can also be replaced by *sneunux* 'head hair' or *ubal tunux* '(hair_head) head hair'.

(12) *meha=ku p-q‹m›arik tunux=mu.*
be.going=1SG.NOM CAUS-<AV>cut head=1SG.GEN
'I am going to have my hair cut.'

(13) *aji=mu qrib-un tunux=mu.*
NEG=1SG.GEN cut-UV head=1SG.GEN
'I don't want to have my hair cut.'

(14) *iya lul-i ka tunux.*
NEG.IMP keep.hair-UV.IMP NOM head
'Don't keep (your) hair long.'

Table 2: The head and its parts

Seediq term	Translation	Additional information
SIMPLEX		
qbubul	'frontal'	also *qbubul tunux*
qpagu *qrpau* (CTd)	'hair whirl'	
qilug[8]	'hindbrain'	
quji (qudi) *qudas* (CTd)	'grey hair'	
sneunux	'head hair'	
tunux	'head'	
COMPLEX		
ayug qilug	'trough below the hindbrain'	lit. valley_hindbrain
ayuw waru (CTd)		lit. valley_neck
ubal tunux	'head hair'	lit. hair_head. cf. *ubal* 'hair of animal; feather'

2.3 Face

In Seediq FACE is labelled as *dqeras*, which includes the body parts *pngelux* or *kduruk*[9] 'forehead', *dowriq* 'eye', *muhing* 'nose', *birat* 'ear', *quwaq* 'mouth', and *bkeluy* 'jaw'. These terms constitute the meronymic level under the term *dqeras* 'face', which also encompasses the notion of CHEEK.

The parts related to EYE, such as EYELID, SCLERA, EYEBALL, EYEBROW, and EYELASH, are complex terms. Note that PUPIL in the Seediq dialects has different versions. A simpler version is to use the colour term *qalux* 'black' to form a compound: *qalux dowriq*, as opposed to *bhegay dowriq* '(white_eye) sclera'.[10] Another complex term *buwax dowriq* is also found, which is commonly used in daily conversations than its synonymous simplex term, suggesting that PUPIL is likened as a grain of uncooked rice ('*buwax*') in the eye.

In Truku a simplex term for PUPIL is *qusaw*, which also means 'the thin layer on top of porridge'. In Tgdaya the word *utux* 'spirit' also refers to pupil of the eye. This is reminiscent of Brown & Witkowski's (1981:599) survey that in Fijian (Austronesian) 'pupil of the eye' is literally 'spirit master or spirit chief', and in Nukuoro (Austronesian) it is 'infant ghost'. The naming of 'pupil of the eye' is figuratively associated with the spiritual world. According to my Tgdaya consultant, 'pupil of the eye' is *utux* because it catches the image of a person through reflection. Similarly, Tgdaya speakers also

[8] In Truku the word-final *g* is pronounced as [w] if following [u] or as [j] if following [i]. In Tgdaya the word is written as *qilu*.

[9] The body part term *kduruk* 'forehead' is only found in Tgdaya and Toda.

[10] Nowadays it requires a native speaker with higher language proficiency to be able to know the simplex term for 'pupil of the eye', thus *qalux dowriq* is a simpler version.

consider photographs as *utux*, thus 'to take a photo' is expressed by *m-angal utux* (AV-take_spirit).

As for the parts related to NOSE and EAR such as ALAR and EARLOBE, in Truku they are conceptualised as the soft regions of NOSE and EAR, henceforth *rbenaw muhing* '(soft_nose) alar' and *rbenaw birat* '(soft_ear) earlobe'. In Tgdaya EARLOBE is the sagging part of the ear, hence *tuduy birac*.

However, there are asymmetrical lexical distinctions regarding the mouth and its parts. In Truku, only 'upper lip' is lexically distinct as *pdahung*, whereas its counterpart is a compound *lubuy quwaq* '(bag_mouth) lower lip'. In Toda, 'lips' are formed by the compound *pdahung quwaq*, and no lexical distinction is made for upper and lower lip. The compound *pdahung quwaq* in Toda suggests a connection between MOUTH and LIPS. Given this dialectal variation, it is not clear whether *pdahung* is conceptually equivalent to the notion of LIPS in Truku.

Table 3: The face and its parts

Seediq term	Translation	Additional information
SIMPLEX		
birat	'ear'	
bkeluy	'jaw'	
dowriq	'eye'	
dqeras	'face'	including 'cheek'
gupun	'teeth'	
rupun (Tg, Td)	'teeth'	
hema	'tongue'	
kadang	'molar tooth'	
muhing	'nose'	
ngudus	'moustache, beard'	also means '(animals') whiskers'
pdahung	'upper lip'	also means 'jaw' in Central Toda
pngelux	'forehead'	
kduruk (Tg, Td)	'forehead'	
qusaw	'pupil'	also means 'a thin layer on top of porridge'
utux (Tg)	'pupil'	also means 'spirits'
quwaq	'mouth'	
waqit[11]	'canine tooth'	

(continued)

[11] My consultants in Central Toda consider that only animals but not humans have *waqit*. Human teeth are divided into two kinds: those in the front are called *rupun* and those from the canine teeth to the back *kadang*.

Table 3: *(continued)*

Seediq term	Translation	Additional information
COMPLEX		
klabuy dowriq	'eyelid'	lit. cover_eye
klabi doriq (Tg, Td)	'eyelid'	lit. cover_eye
qalux dowriq	'pupil'	lit. black_eye
buwax dowriq	'pupil'	lit. uncooked rice_eye
bhegay dowriq	'sclera'	lit. white_eye
quru dowriq	'eyeball'	lit. round_eye
bubali doriq (Tg)	'eyeball'	lit. bullet_eye
ubal baraw dowriq	'eyebrow'	lit. hair_above_eye
ubal dowriq	'eyelash'	lit. hair_eye
ubal muhing	'nasal hair'	lit. hair_nose
ubal quwaq (CTd)	'moustache'	lit. hair_mouth
rbenaw muhing	'nose wing; alar'	lit. soft_nose
tudu muhing	'nasal bridge'	lit. ridge_nose[12]
duwiaq muhing (Td)	'nasal bridge'	lit. mountain_nose
lubuy dowriq	'under-eye bags'	lit. bag_eye
lubuy quwaq	'lower lip'	lit. bag_mouth; a mouth's bag[13]
rbenaw birat	'earlobe'	lit. soft_ear
tuduy birac (Tg)	'earlobe'	lit. sag_ear
rulun gupun	'front tooth'	lit. middle_tooth

2.4 Extremities and torso

This section is organised into two parts: extremities, including the upper and the lower limbs, and torso.

2.4.1 The upper and the lower limbs

Structurally speaking, upper extremities include ARM and HAND, and lower ones LEG and FOOT. In Seediq the term *baga* mostly denotes HAND, but its extension also includes the parts of ARM, as shown in the following compounds, such as *utux baga* '(spirit_hand) pulse', *pungu baga* '(joint, knot_hand) elbow, knuckle', or *rumu baga* '(soft _hand) triceps, upper arm'.

Similarly, the notion of LEG and that of FOOT in Seediq are also overlapping. Two terms, *qaqay* and *papak*, are used for FOOT/LEG in Seediq, with dialectal variations. In Tgdaya only *papak* is used. In Central Toda, *qaqay* tends to be used on human, while *papak* is used on animals, as shown in (15a–c). However, a human vs nonhuman distinction for the two terms does not really hold true in all circumstances. For example, in Central Toda *qaqay* is used to mean 'table leg' (15d). On the other hand, for some of

[12] A similar expression *bgebug muhing* is found.

[13] The reverse compound *quwaq lubuy* means 'the mouth of bag'.

my consultants, such a distinction is blurred, thus in most contexts *papak* and *qaqay* are interchangeable.[14]

(15) a. *qaqay* *ima*?
leg who
'Whose leg?'

b. *papak* *manu*?
leg what
'What (animal's) leg?'

c. * *papak* *ima*?
leg who
'What animal's leg?'

d. *qaqay* *handay*
leg table
'table leg'

The lexical term *kulaw* (*kulo* in Tgdaya) forms a compound with *baga* and *papak* (or *qaqay*) to denote the notions of ARM and LEG, respectively. It appears that there is an overlap in the concept of *baga* and *kulaw baga* as parallels to *papak* and *kulaw papak*. The lexical term *kulaw* is polysemous. It refers to the part of a long object between two nodes, so that the part of a limb between two joints is a *kulaw* of a *baga* or of a *papak*. The term *kulaw* is also extended to refer to a section of bamboo between two nodes, which is similar to the shape of the refereed body parts.

Most consultants consider that *kulaw baga* covers a bigger region of the body than *baga*, in the sense that the former denotes the whole upper limb including the hand, while the latter the hand only. However, *kulaw baga* as a linguistic form is part of *baga*, as the term *kulaw* cannot stand alone in such a context. This is demonstrated by the following sentences.

(16) a. *kingal* *kenlegan* *baga* *ka* *kulaw* *baga.*
one part hand/arm NOM section hand/arm
'The arm is part of the hand.'

b. * *kingal* *kenlegan* *kulaw* *ka* *kulaw* *baga.*
one part section NOM section hand/arm
'The arm is part of the section .'

(17) a. *niq-an* *deha* *kulaw* *baga* *ni* *papak* *ka* *seejiq.*
exist-LV two section hand/arm and leg NOM person
'A person has two arms and legs.'

[14] A conceptual variation exists in this context. My Toda consultant in Hualien considers that *papak* should be used for referring to 'table leg'. My Truku consultant in Hualien also considers *papak handay* as the correct expression. However, for my Truku consultant, the distinction between human and nonhuman exists in *papak* and *qaqay* when referring to 'footprint', hence *papak pada* 'a pygmy deer's footprint' and *qaqay seejiq* 'a human's footprint'.

b. * *niq-an* *sepat* *kulaw* *ka* *seejiq.*
exist-LV four section NOM person
'A person has four limbs.'

Therefore, there is no primary lexeme or simplex term for ARM or LEG in Seediq. In terms of nomenclature, a limb is labelled only for the upper part (above elbow of the arm or knee of the leg), not for the lower part (between the elbow and the wrist, or the knee and the ankle). The body term *hiqur* is the name for 'upper arm', and *bteriq* for 'thigh, or upper leg'.[15] There is no equivalent term in Seediq for the English 'forearm' or 'lower leg'.

In Truku another term related to *baga* and *papak/qaqay* is *tduling* 'digit', hence *tduling baga* means 'finger' and *tduling papak* 'toe'.[16] This is contradictory to the ninth principle proposed by Brown (1976:408) stating that FINGER and TOE are always labelled. In Seediq, *tduling* and *baga* or *papak* forms a meronymic relationship. My consultants also agree that 'digits of hand or foot' is part of the upper limb/hand or the lower limb/foot.

(18) *kingal* *kenlegan* *baga* *ka* *tduling.*
one part hand/arm NOM digit
'The digits are part of the hand/the foot.'

(19) *kingal* *kenlegan* *papak* *ka* *tduling.*
one part foot/leg NOM digit
'The digits are part of the foot.'

(20) *kingal* *kenlegan* *kulaw* *baga* *ka* *tduling.*
one part section hand/arm NOM digit
'The digits are part of the arm.'

In Tgdaya there is no simplex term for DIGIT. The compound *wawa baga* is ambiguous depending on the perspectives. In terms of a hand, it refers to all the fingers, whereas in terms of a finger, it refers to little finger. The same is applied to *wawa papak* 'toe/little toe'. One meaning of the word *wawa* in Seediq is 'little animal', as in *wawa rowdux* '(small_chicken) chick', thus FINGER or TOE in Tgdaya are conceptualised as the 'small one' of the hand or foot.

The naming of THE THUMB and of BIG TOE in Seediq is associated with kinsmen (cf. Brown & Witkowski 1981). In Truku and Toda both are related to *tama* 'father' as in *tama baga* 'the thumb' and *tama qaqay* 'big toe'. Yet in Tgdaya they are *baki baga* and *baki papak*, literally 'grandfather of hand' and 'grandfather of leg'. It is interesting to find that different dialects in a language adopt different male figures (i.e., father or grandfather) to represent the first digit of the hand/foot.

As for NAIL, the notion is overlapping with that of CLAW and HOOF and is expressed by the term *kukuh*, which refers to both human nails and animal claws and hooves, namely,

---

[15] In the Seediq dialects spoken in the Nantou area, 'elbow' is labelled as *hiqur baga* (Tk, Tg) or *hiqur bawa* (Td). In Truku one act related to the upper arm *hiqur* is when someone wants to tell off another one for being 'good for nothing', he or she uses the right hand to tap this part and repeats '*hiqur, hiqur*' at the same time.

[16] This term is *tluling* in Toda.

the hard parts located at the extrusive point of a limb. Since there are no primary lexemes for FINGER and TOE in Seediq, these body part terms are formed through compounding with the primary lexemes *baga* and *papak*, so that *kukuh baga* means 'fingernail' and *kukuh papak* 'toenail', thus *kukuh* forms a meronymic relationship with *baga* and *papak*/*qaqay* rather than with *tduling* 'digit'.

(21) *kingal* *kenlegan* (*kulaw*) *baga* *ka* *kukuh.*
one part section hand/arm NOM nail
'Nails are part of the hand (arm).'

(22) *t-kekuh-an* *ka* *tduling.*
VBLZ-nail-LV NOM digit
'Digits (fingers and toes) are where nails grow.'

The naming of lower extremities is not entirely parallel to that of the upper ones, showing that what is considered as salient parts of the upper extremities is different from those of the lower ones. For example, 'upper leg' or 'thigh' is labelled as *bteriq*, and 'knee' as *bqelit* instead of **pungu qaqay* (cf. *pungu baga* 'elbow').[17] The part behind the knee is also labelled with a simplex term *heduq*. Similar to the naming of upper extremities, 'lower leg' is not labelled by a simplex term.

In Seediq 'muscle' is named through compounding in a figurative way. The body part 'biceps' is associated with the animal name 'mouse', hence *qowlit baga* (lit. mouse_ hand/arm, Tgdaya). Brown & Witkowski (1981:602) have found that 'mice and rats predominate among animals entering into this metaphor'. Seediq is another language with this figurative naming, along with other genetically unrelated languages such as Mixe (Mesoamerican Zoquean), Colorado (Macro-Chibchan Barbacoan), and Chraw (Mon-Khmer).

In Truku, the other parts of 'muscle' is conceptualised as something soft. The word *rumu* 'soft' is imported to form a compound with a body part term to denote MUSCLE of the body part, hence *rumu baga* '(soft_hand/arm) triceps', *rumu bteriq* '(soft_thigh) thigh muscle', and *rumu papak* '(soft_foot/leg) calf of the leg'. It is also used in the context related to rice cooking. For example, *rumu idaw* means '(soft_cooked rice) porridge'.

[17] The compound *pungu qaqay* or *pungu papak* is considered as 'ankle' rather than 'knee' by my Truku consultant. However, in other dialects 'knee' can also be expressed by *pungu papak*, *pungu qaqay*, or *pungu bqelic*.

Table 4: The extremities and the related parts

Seediq term	Translation	Additional information
SIMPLEX		
baga	'hand/arm'	
bqelit	'knee'	cf. 'elbow'
bteriq	'thigh'	
dutuy	'tiptoe'	
heduq	'knee pit'	also *heduq papak*
hiqur	'upper arm'	from elbow above
kunug	'heel'	
papak	'foot/leg'	
qaqay	'foot/leg'	
qapal	'sole'	
tduling	'digit'	
tluling (Td)	'digit'	
COMPLEX		
dgiyaq papak	'shinbone'	lit. mountain_foot/leg
klekah-an papak	'sole'	lit. step-NMLZ_foot/leg
kukuh baga	'fingernail'	lit. nail_hand/arm
kukuh papak	'toenail'	lit. nail_foot/leg
kulaw baga	'arm; upper limb'	lit. section _hand/arm
kulaw qaqay	'leg; lower limb'	lit. section _foot/leg
pungu baga	'elbow'	lit. joint_hand/arm
hiqur bawa (CTd)	'elbow'	lit. upper arm_hand/arm
pungu papak (Tg)	'knee'	lit. joint_foot/leg
pungu bqelic (CTd)	'knee'	lit. joint_knee
qowlit baga (Tg)	'biceps'	lit. mouse_hand/arm
rumu baga	'triceps'	lit. soft_hand/arm
rumu bteriq	'thigh muscle'	lit. soft_thigh
rumu qaqay/papak	'calf of the leg'	lit. soft_foot/leg
hiyi qaqay (CTd)		lit. flesh_foot/leg
seka tduling baga	'middle finger'	lit. middle_digit_hand/arm
seka tduling papak	'middle toe'	lit. middle_digit_foot/leg
s-m-ciyu tduling baga	'index finger'	lit.VBLZ-AV-point_digit_hand/arm[18]
s-m-tama tduling papak	'big toe'	lit. VBLZ-AV-father_digit_foot/leg
tama baga	'the thumb'	lit. father_hand/arm
tduling baga	'finger'	lit. digit_hand/arm
wawa baga (Tg)	'finger/little finger'	lit. small_hand/arm
laqi bawa (CTd)	'finger'	lit. child_hand/arm
tduling qaqay	'toe'	lit. digit_foot/leg
wawa papak (Tg)	'toe/little toe'	lit. small_foot/leg
truma qduan	'armpit'	lit. under_armpit; cf. *heduq* 'knee pit'
wawa baga	'little finger'	lit. small_hand/arm
wawa tduling papak	'little toe'	lit. small_digit_foot/leg

[18] The prefix *s-* combined with the active voice *m-* derives a verb from a noun, sometimes with a slight change of meaning (Tsukida 2009:261). The examples *sn-ciyu* and *sn-tama* are thus derived from /s-m-tiyu/ 'VBLZ-AV-point' and /s-m-tama/ 'VBLZ-AV-father', respectively. Due to syncope and nasal homorganic rule occurred in the prefixes, the nasal /m/ is changed to [n] and is reflected in the written form (see also Lee 2010a).

2.4.2 Torso

In Seediq the meaning of TORSO is not lexically distinct. It is expressed by the compound *seka hiyi* '(middle_body) torso' in Truku, suggesting that the body part is considered as the central or middle area of the whole.

The body part terms which refer to the area between the head and the rest of the body include *waru* (Tk, Td)/ *nduyung* (Tg) 'neck', *hudung* 'back neck' (Td), as well as *hirang* (Tk, Td)/ *ahing* (Tg) 'shoulder'.

The body part terms *berah* 'chest' and *bukuy* 'back' are related to spatial representations, which have evolved into spatial grams FRONT-REGION and BACK-REGION, following Svorou's (1994) proposal. See also section 4.1.

(23) *t-em-bqelit* *berah* *bubu=na* *ka* *laqi.*
VBLZ-AV-knee chest mother=3SG.GEN NOM child
'The child knelt down in front of his/her mother.'

(24) *wada* *t-berah* *ka* *laqi=mu.*
be.gone VBLZ-chest NOM child=1SG.GEN
'My child has gone forward.'

(25) *gaga* *t-k-bukuy* *ka* *laqi=mu.*
PROG VBLZ-STAT-back NOM child=1SG.GEN
'My child is (lagging) behind.'

The labelled simplex body parts below CHEST excluding lower extremities are *embuyas* 'abdomen',[19] *pega* 'navel', *hwinuk* 'waist', *bunuh* 'lower belly', *rayuh* 'groin', *psaniq* 'genitals', *puyaq* 'anus', and *sulay* 'hip'.

As in other languages, simplex body part terms related to 'genitals' tend to be taboo words, which usually have equivalent euphemisms. In Truku, *psaniq* is a generic term for the taboo words including *pipi* 'vulva', *pulut* 'clitoris',[20] *utas* 'penis', and *balung* 'testicle', which also means 'egg'. The euphemisms for *pipi* and *utas* are *psaniq kuyuh* '(genitals_woman)' / *qnqaya kuyuh* '(thing_woman)' and *psaniq senaw* '(genitals_man)' / *qnqaya senaw* '(thing_man)'. They can also be alternatively called as *n-namu* 'yours', as shown in sentence (26).

(26) *niq-an* *bteriq,* *papak,* *n-namu* *ka* *tuma* *embuyas.*
exist-LV thigh leg POSS-2PL NOM under belly
'Below the belly there are the thighs, the legs, and the private parts.'

A comparison of Tgdaya and Truku reveals semantic changes in the body part terms, as shown in the Table 5. In Tgdaya the term *pusaniq* means 'taboos' in general, including anything prohibited to act or to mention in the Seediq society and culture.[21] Its meaning is shifted to refer to genitals in Truku, which is conceptually connected, as genitals are usually prohibited to be mentioned. In Tgdaya there are euphemisms for mentioning

---

[19] This body part term is often omitted to be *buyas*, though in other dialects *nbuyas* is also heard.

[20] In Toda, *pulut* is an orthophemism which has a euphemism *riwa* and a dysphemism *peyluh*.

[21] A synonym in Tgdaya for the word *pusaniq* is *s-en-gaya*, where *gaya* refers to social constraints, rituals, beliefs, or the ways people live by which are considered appropriate in the culture. *Gaya* is a cultural term which requires elaborative anthropological definitions.

women's genitals, but none for men's. Another euphemism for 'a woman's private parts' is *henegigan*, derived from *hei* 'body', which also means 'a woman's figure'.

Table 5: Semantic differences of private body part terms in Seediq dialects

body part terms	Truku	Toda	Tgdaya
balung	egg, testicle	egg, testicle	egg
bunuh	lower belly	hat	A euphemism especially used by men to refer to women's private parts.
psaniq	genitals	taboo	taboo
puyaq	anus	anus	hip; buttocks
sulay/*sule* (Tg)	hip area	hip area	A taboo word meaning women's private parts. This word can only be used by women to mention their own body parts.

On the other hand, in Truku and Toda 'testicle' is lexically the same as 'egg' probably due to their similarities in shape. Brown & Witkowski (1981:603) have found that twenty-one languages use egg for testicle, and suggest that it is because 'since eggs, like testicles, are associated with biological fertility'. Other languages in case include Taiwan Mandarin, Spanish, Ainu, or Tuvalu (Austronesian).[22] See the following example.

(27) *wada=mu* *tena* *gm-beleng-i* *ka* *babuy* *gaga.*[23]
be.gone=1SG.GEN already VBLZ-egg- UV.IMP NOM pig that
'I have castrated the (male) pig.'

Table 6: The torso and its parts

Seediq term	Translation	Additional information
SIMPLEX		
balung	'testicle'	also 'egg'
begax (Tg)	'testicle'	
berah	'chest'	also 'in front of'; 'front part of the body' (Tg)'
teerang (Tg)	'chest'	'rib' (Td)
bukuy	'back'	also 'behind'
bunuh	'lower belly'	'hat' (CTd)
embuyas	'abdomen; belly'	
hudung	'back neck'	
hirang	'shoulder'	
ahing (Tg)	'shoulder'	
hwinuk	'waist'	
psaniq	'genitals'	'taboo' (CTd)
pipi	'vulva'	taboo word
pega	'navel'	

[22] See the website: http://language.psy.auckland.ac.nz/austronesian/word.php?v=98.

[23] The form *gm-beleng-i* is derived from *balung*, in which the two vowels are reduced due to suffixation.

pulut	'clitoris'	taboo word
Seediq term	Translation	Additional information
puyaq	'anus'; 'hip (Tg)'	
rayuh	'groin'	
ruku (Tg)	'glans'	also *tunux eyu*
sulay	'pelvis; hip area'; 'vagina (Tg)'	taboo word (Tg)
tpaqan	'side hip region'	also *tpaqan puniq* 'butt of gunstock'
unuh	'breast'	
nunuh (Tg, Td)	'breast'	
utas	'penis'	taboo word
eyu (Tg), *masa* (CTd)	'penis'	taboo word
waru	'neck'	
unduyung (Tg)	'neck'	also *nuduyung*
COMPLEX		
puqu waru	'Adam's apple'	lit. round_neck
pungu waru (CTd)	'Adam's apple'	lit. joint_neck
sari sulay	'hip muscle'	lit. taro_hip
seka hiyi	'trunk, torso'	lit. middle_body
tunux unuh/nunuh	'nipple'	lit. head_breast
tunux utas	'glans'	lit. head_penis
tunux eyu (Tg)	'glans'	lit. head_penis
ubal psaniq	'pubic hair'	lit. hair_genitals

2.5 Internal parts

Internal body parts include internal organs, bone structure, and its related parts. Unlike external body parts, internal body parts in normal circumstances cannot be directly perceived. Perception of their types and placements inside the abdomen usually comes from dissection of captured animals. Therefore, the internal organ terms listed in Table 7 are used for both humans and animals.

Table 7: Internal parts

Seediq term	Translation	Additional information
SIMPLEX		
baraq	'lung'	cf. 'heart'
buut	'bone'	also 'seed'
gelu	'throat'	
iraq	'intestine'	
jiyay	'vocal cord'	cf. *k-jiyay* 'cicada'
luqi	'marrow'	
pahung	'gall'	
psaqux	'larynx'	
qsahur	'viscera'	a generic term
quru	'kidney'	
rmune (Tg)	'kidney'	

(continued)

Table 7: *(continued)*

Seediq term	Translation	Additional information
rketu	'stomach'	
lhebun (Tg, CTd)	'stomach'	
rkebul (ETd)	'stomach'	also 'animal stomach' (Tk)
rkebun (Tg)	'animal stomach'	
rumul	'liver'	
tkerang	'rib'	
teerang (Td)	'rib'	'chest' (Tg)
hwiras (Tg)	'rib'	'love handles' (ETd)
tlanga	'pancreas'	
idas (Tg)	'pancreas'	also 'moon'
towcin	'cheek bone'	also 'hammer'
tua'an	'womb'	
tudu	'spine'	also 'ridge'
urat	'vein; blood vessel; nerve; artery'	
COMPLEX		
buut tunux	'skull'	lit. bone_head
luqi tunux	'brain'	lit. marrow_head
luqi buut	'bone marrow'	lit. marrow_bone
p-henang-an	'vocal cord'	lit. CAUS-sound-LOC
pusu baraq	'trachea'	lit. origin_lung
qthur iraq	'large intestine'	lit. fat_intestine
iraq paru (Tg)	'large intestine'	lit. intestine_big
qling iraq	'small intestine'	lit. thin_intestine
iraq biciq (Tg)	'small intestine'	lit. intestine_small
rebu-an shemu	'bladder'	lit. wrap-LOC_urine[24]
cubu rebu (Tg)	'bladder'	lit. balloon_urine[25]
yup-un (Td)	'bladder'	lit. blow-UV
tama baraq	'heart'	lit. father_lung; cf. 'lung'
toma bahak (Tg)	'heart'	lit. under_lung

Like SKIN (see section 2.8) which covers all over the body, marrow also runs all over the body. In Seediq, BRAIN is conceptualised as a kind of marrow which fills inside the head. Therefore, Seediq distinguishes two kinds of marrow, *luqi tunux* '(marrow_head) brain' and *luqi buut* '(marrow_bone) bone marrow'.

Among the internal organs, STOMACH is named differently in the three dialects. It is also the only internal organ which makes a lexical distinction of human and animal. In Truku 'stomach' is *rketu*, while in Eastern Toda it is *rkebul*. However, in Central Toda

[24] The form *rebu-an* is derived from /rabu-an/, where the antepenultimate vowel is reduced due to suffixation.

[25] There was no rubber balloon in ancient Seediq societies. Balloons or balls were made from pigs' bladders.

the term *rketu* refers to animal stomach, whereas 'human stomach' is *lhebun*. In Tgdaya, *rkebun* (*rkebul* in Eastern Toda) is 'animal stomach' and *lhebun* 'human stomach'. This is summarised in Table 8.

Table 8: 'Stomach' in different Seediq dialects

meaning / *dialect*	human stomach	animal stomach
Truku	*rketu*	*rkebul*
Eastern Toda	*rkebul*	
Central Toda	*lhebun*	*rketu*
Tgdaya	*lhebun*	*rkebun*

The meaning of the term *lhebun* has become rather abstract and figurative, which also refers to mind or conscience. See the following examples.

(28) *ungat* *lhebun*!
NEG stomach
'No conscience.'

(29) *ini* *huya.* *iya* *s-lhebun*!
NEG why NEG VBLZ-stomach
'It's nothing. Don't worry!'

The lexical term *buut* is generic, which not only refers to an animal's bone, but also a fruit's core or seed, thus it is defined as 'the hard part of an organic element'. For example, *buut tunux* is 'skull', but *buut apu* is 'seed of persimmon'. Compare the following sentences.

(30) *lala* *bi* *buut=na* *ka* *sagas.*
many very seed=3SG.GEN NOM watermelon
'A watermelon has many seeds.'

(31) *lala* *bi* *buut=na* *ka* *seejiq.*
many very bone=3SG.GEN NOM person
'A person has many bones.'

2.6 Bodily fluid and effluvia

The lexical items related to bodily fluid and effluvia are considered as a category given that they are all produced by the body. Bodily fluid includes SALIVA, TEARS, BLOOD, SWEAT, and bodily effluvia for the rest.

In Seediq there are two lexical items for TEARS, *biyuq* and *rusuq*, thus *biyuq dowriq* is the same as *rusuq dowriq*. Both of the two terms also refer to sap, hence *biyuq qhuni* and *rusuq qhuni*. However, only the former has the sense of 'milk', as in *biyuq unuh* 'breast milk (milk_breast)', but not **rusuq unuh*.

(32) *m-tucing* *biyuq*/*rusuq* *dowriq=na* *da.*
AV-fall tear eye=3SG.GEN COS
'His/her tears dropped.'

The domain of bodily effluvia is linked with taboo words and euphemisms. Seediq is no exception. For example, the word *qasay* 'menstruation blood' is a taboo word, which is usually replaced by the euphemism *kensapil* 'be padded', derived from *sapil* 'to pad, to insert a mat', as in *spil-an* 'chopping board, floor mat'.

(33) *gisu=ku* *k-m-sapil.*[26]
PROG=1SG.NOM STAT-AV-pad
'I have my period.'

Table 9: Bodily fluid and effluvia

Seediq term	Translation	Additional information
SIMPLEX		
biyuq	'tears'	also 'sap, milk'
dara	'blood'	
harus	'saliva'	
hurulas (Tg)	'saliva'	
mering	'sweat; perspiration'	
ngusul	'nasal discharge'	
pusiq	'eye mucus'	
qasay	'menstruation blood'	taboo word
qpuji	'earwax'	
quci	'faeces'	
rusuq	'tears'	also 'sap, wax candle'
shemu	'urine'	
rebu (Tg)	'urine'	
utaq	'semen; vagina discharge'	
COMPLEX		
pucaq doriq (CTd)	'eye mucus'	lit. mucus_eye
ruciq muhing	'booger'	lit. dirty_nose
pukah muhing (CTd)	'booger'	lit. hard_nose

2.7 Animal anatomy

Animal body part terms share similar nomenclatural patterns with human ones, including *tunux* 'head', *dqeras* 'face' and its related parts, *baga* 'hand/arm', *papak* 'foot/leg', and *kukuh* 'nail/claw/hoof'. These shared body part terms have an inherent concept, so that the conceptual image can be projected to both humans and animals. For example, *kukuh* when used on animals means 'claw/hoof', while on humans means 'fingernail/toenail', thus essentially it is defined as 'the hardest part at the front end of a mammal's limb or of a human's digit'. The same body part for birds is lexically specified as *texil*.

The body part FORELEG of an animal is referred to as *berah papak* or *bteriq baga*. The former is semantically transparent, whereas the latter is rather idiomatic, which is paired

[26] Due to nasal homorganic rule the active voice *m-* is pronounced as [n] when followed by alveolar consonants (Lee 2010a), hence the written form as *knsapil*.

with *bteriq papak* ‘hind leg’. The body part term *bteriq* ‘thigh’, compounded with *baga*, is used to refer to the upper and wider part of a quadruped’s limb, suggesting that a front leg is conceptually parallel to the position of a human arm. This linguistic representation reveals that conceptualisation of animal anatomy in Seediq follows an anthropological rather than a zoomorphic model (cf. Svorou 1994).

(34) *niq-an* *deha* *bteriq* *ka* *kacing.*
exist-LV two thigh NOM ox
bteriq *kacing* *’u* *m-niq* *bukuy.*
thigh ox TOP AV-exist back
‘An ox has two thighs. The thighs of an ox are at the back.’

Table 10: Animal body part terms

Seediq term	Translation	Additional information
SIMPLEX		
kukuh	‘animal claw, hoof’	also ‘nail’
ngudus	‘whisker’	
ngungu	‘tail’	cf. *m-ngungu* ‘be timid’
pahing	‘wing’	
palic (Tg, Td)	‘wing’	
qerung	‘cockscomb’	also ‘ear fungus’
sneli	‘bird crop (esp. chicken)’	
texil	‘bird claw’	
ubal	‘feather’	also ‘hair’
urung	‘horn’	
tnguraw (ETd)	‘antlers’	also ‘male deer, buck’
waqit	‘chela (of crab, lobster)’	also ‘canine tooth’
COMPLEX		
berah papak	‘foreleg’	lit. chest_foot/leg
bawa qaqay (CTd)	‘foreleg’	lit. hand_foot/leg
bteriq baga	‘foreleg’	lit. thigh_hand/arm
bteriq papak	‘hind leg’	lit. thigh_foot/leg
bukuy papak	‘hind leg’	lit. back_foot/leg

2.8 Skin

How can SKIN as a body part term be categorised? As the material which covers all over the surface of a body, SKIN is conceptualised as an integral part by itself, regardless of which parts it forms the outer layer. Given this nature, it cannot be categorised as belonging to a certain part of a body. For example, when SKIN is mentioned, it is the body part rather than the skin which is encoded.

(35) *m-rbenaw* *bi* *dqeras=su.*
AV-soft very face=2SG.GEN
‘(The skin of) your face is very soft.’

(36) ?? *malu* *bi* *qraqil=su.*
good very skin=2SG.GEN
‘Your skin is nice.’

It is thus understandable that in Seediq SKIN is labelled in terms of different 'kinds' rather than different 'parts', and there is no equivalent term of SKIN in Seediq. The language distinguishes 'animal (including human) skin' from 'discarded or unwanted skin'. The former, labelled as *qraqil*, is human skin or peeled animal skin used for clothing, and the latter, *rehiq* (*rehaq* in Tgdaya), is generated by snakes, trees, corns, or the foreskin of penis. It also refers to egg shell, i.e., outer layer of an egg. Therefore, *qraqil* and *rehiq* are hyponyms of SKIN, which is not lexically encoded in Seediq.

3 Compounds and idiomatic expressions

Most complex body part terms in Seediq share the structure of being a compound. Compounding is the most prevalent morphological process used to coin new expressions in Seediq.

In terms of semantic classification, body part complex terms in Seediq exhibit two types of compounds: endocentric and exocentric. Endocentric compounds are headed. In Seediq body part terms both right-hand headed (RHH) and left-hand headed (LHH) compounds can be found. The following two tables list the examples of the two types of endocentric compounds, showing that an endocentric compound is a kind of its head.

Table 11: Complex body part terms as endocentric LHH compounds

Seediq term	Gloss	Semantic component
kukuh *baga*	fingernail	nail_hand/arm
kukuh *papak*	toenail	nail_foot/leg
luqi *tunux*	brain	marrow_head
luqi *buut*	marrow	marrow_bone
tduling *baga*	finger	digit_hand/arm
tduling *papak*	toe	digit_foot/leg
ubal *tunux*	head hair	hair_head
ubal *dowriq*	eyebrow	hair_eye

Table 12: Complex body part terms as endocentric RHH compounds

Seediq term	Gloss	Semantic component
berah ***papak***	foreleg	chest_foot/leg
bteriq ***baga***	foreleg	thigh_hand/arm
bteriq ***papak***	hind leg	thigh_foot/leg
bukuy ***papak***	hind leg	back_foot/leg
iril ***baga***	left hand	left_hand/arm
nanat ***baga***	right hand	right_hand/arm
qling ***iraq***	small intestine	thin_intestine
qthur ***iraq***	large intestine	fat_intestine
seka ***tduling baga***	index finger	middle_(digit_hand/arm)
seka ***tduling papak***	middle toe	middle_(digit_foot/leg)
s-m-tama ***tduling papak***	big toe	VBLZ-AV-father_(digit_ hand/arm)
wawa ***tduling papak***	little toe	small_(digit_foot/leg)

There is another kind of endocentric compound in Seediq body part complex terms which exhibits a meronymic relationship with its modifier. This type of compound is also termed as a 'meronymic compound', which in Seediq is mostly left-hand headed. Therefore, as an entire N1-N2 compound represents the part of N2, N1 is modified by N2, which represents the whole.

Table13: Complex body part terms as LHH meronymic compounds

Seediq term	Gloss	Semantic component
bhegay *dowriq*	sclera	white_eye
buut *tunux*	skull	bone_head
klabuy *dowriq*	eyelid	cover_eye
kukuh *baga*	fingernail	nail_hand/arm
kukuh *papak*	toenail	nail_foot/leg
quru *dowriq*	eyeball	round_eye
rbenaw *muhing*	alar; nose wing	soft_nose
rbenaw *birat*	earlobe	soft_ear
tudu *muhing*	nasal bridge	spine_nose
tduling *baga*	finger	digit_hand/arm
tduling *papak*	toe	digit_foot/leg
tunux *unuh*	nipple	head_breast
tunux *utas*	glans	head_penis

The data in Table 13 which overlap with those in Table 11 are *tduling baga* 'finger', *tduling papak* 'toe', *kukuh baga* 'fingernail', and *kukuh papak* 'toenail'. In these compounds the head is in taxonomic relation with the whole compound, and in meronymic relation with its modifier. For example, *tduling baga* 'finger' is a kind of *tduling* 'digit', as opposed to *tduling papak* 'toe'. On the other hand, *tduling baga* is also a part of *baga* 'hand/arm'.

There are two examples which are considered as RHH meronymic compounds: *seka hiyi* 'trunk; torso' and *sari sulay* 'buttocks'. As discussed in section 2.4.2, *seka hiyi* is headed by *hiyi*, and the whole compound represents part of the head *hiyi*. Similarly, the word *sari* 'taro' in *sari sulay* is a modifier of the head *sulay*, based on the similarity in shape.

Table14: Complex body part terms as RHH meronymic compounds

Seediq term	Gloss	Semantic component
seka ***hiyi***	trunk; torso	middle_body
sari ***sulay***	buttocks	taro_hip

Contrary to an endocentric compound, an exocentric one is headless. An exocentric compound refers to a compound of which the meaning is not directly obtained from any of its constituents. The constituents in an exocentric compound do not have a head-modifier relationship. Instead, the semantic head of an exocentric compound is unspecified. Therefore, the meaning of an exocentric compound tends to be opaque, which must be listed in the lexicon. For example, *tama baraq* 'heart' is composed of the

constituent *tama* ‘father’ and *baraq* ‘lung’. None of these constituents is related to the concept of the organ HEART.

The semantic opacity is strengthened when the constituents of a compound are polysemous. Although one of the constituents signals the region of the body part, the meaning of the whole compound is not transparent. For example, given that upper arm and upper leg are named as *rumu baga* and *bteriq*, respectively, *kulaw baga* and *kulaw papak* are usually referred to lower arm and lower leg.

This distribution shows that body part nomenclature in Seediq is not parallel but based on special features of body parts. The word *rumu* means ‘soft’, as in *rumu idaw* ‘porridge; (lit. soft_rice)’, thus *rumu baga* refers to the soft part of an arm, which is the triceps, while *rumu papak* the soft part of a leg, i.e., calf of the leg. There is also *rumu bteriq* which refers to thigh muscle.

As for the rest of the examples, *dgiyaq papak* refers to shinbone, which is located at the front part of a leg. None of the two constituents is connected to the concept of BONE. The compounds *tama baga* and *wawa baga* are fingers, i.e., ‘the thumb’ and ‘little finger’, respectively. The word *tduling* ‘digit’ is omitted.

Table 15: Complex body part terms as exocentric compounds

Seediq term	Gloss	Semantic component
dgiyaq papak	shinbone	mountain_foot/leg
kulaw baga	arm; lower arm	section_hand/arm
kulaw papak	leg; lower leg	section_foot/leg
pusu baraq	trachea	origin_lung
rumu baga	triceps	soft_hand/arm
rumu btriq	thigh muscle	soft_thigh
rumu papak	calf of the leg	soft_foot/leg
tama baga	the thumb	father_hand/arm
tama baraq	heart	father_lung
wawa baga	little finger	small_hand/arm

The verbal phrase *psiyus qmuci* ‘(to make) flatulence’ is considered as an idiomatic expression. Sentence (38) shows that an enclitic pronoun is attached to the first constituent, suggesting that the idiomatic expression is a phrasal structure rather than a compound.

(37) *ima* *namu* *ka* *psiyus* *q‹m›uci*?
who 2PL.NOM NOM IDEO <AV>faeces
‘Who among you has broken wind?’

(38) *tayal* *kenux* *ka* *psiyus=su* *q‹m›uci.*
much odour NOM IDEO=2SG.NOM <AV>faeces
‘The wind you broke was smelly.’

The term *ayug qilug*, however, is both a compound and an idiom. As a compound, it refers to the body part ‘the trough below the hindbrain’. As an idiom, it means ‘be stingy’, which is a synonym of *bserat* ‘be stingy’.

(39) *m-narux* *ka* *ayug* *qilug=mu.*
AV-ill NOM valley hindbrain=1SG.GEN
‘My back neck hurts.’

(40) *ayug=su* *qilug* *wah.*
valley=2SG.NOM hindbrain IDEO
'You are so stingy!'

4 Semantic extensions

This section discusses two aspects of semantic extensions of Seediq body part terms, including spatial representation and metaphorical expressions. The aim of this section is to connect the observations to cognitive and typological significance.

4.1 Spatial representation

Based on Svorou's framework (1994), two spatial grams in Seediq, FRONT-REGION and BACK-REGION, are derived from body part terms. The FRONT-REGION gram evolves from *berah* 'chest', and the BACK-REGION gram from *bukuy* 'back' and *sulay* 'hip, buttocks'. See the following examples.

(41) *t-em-bqelit* *berah* *bubu=na* *ka* *laqi.*
VBLZ-AV-knee chest mother=3SG.GEN NOM child
'The child knelt down in front of his/her mother.'

(42) *gaga* *bukuy* *ka* *laqi=mu.*
PROG back NOM child=1SG.GEN
'My child is behind (us).'

(43) *iya* *p-s-sulay* *t‹m›abuy* *m-kesa*!
NEG CAUS-VBLZ-hip <AV>go.downhill AV-walk
'Don't walk backwards while going downhill.'

The term *berah* in *berah baga* 'palm' also functions as a FRONT-REGION gram.[27] Unlike English, however, 'back of the hand' in Seediq is not **bukuy baga* but *pakux baga*, where *pakux* means 'the opposite side; turn over'.

In Eastern Toda *berah* 'chest' rather than *baraq* 'lung' is used to refer to lung disease in the following sentence. It is interesting that the noun *baraq* 'lung' is not accepted in the sentence.

(44) a. *m-narux* *berah=nia.* (ETd)
AV-ill chest=3SG.GEN
'He has lung disease.'

b. * *m-narux* *baraq=nia.* (ETd)
AV-ill lung=3SG.GEN
'He has lung disease.'

[27] In Tgdaya, however, 'palm' is *qapan baga*, literally 'sole of the hand'.

The body part term *berah* as a FRONT-REGION gram can also be used metaphorically as a temporal noun meaning 'before'. However, such semantic extension is not found in the BACK-REGION gram *bukuy*.

(45) *m-huqil* *berah na* *siida* *ka* *huling* *'u* *m-t-bowyak* *balay.*
AV-die chest GEN when NOM dog TOP AV-VBLZ-boar very
'All dogs before (its) time of death are struggling (like a boar) very much.'

(46) *nasi=ku* *saw* *m~mowsa* *berah* *na,*
if=1SG.NOM be.like FUT~go chest PART
m-hedu=ku *mita* *patas* *ga* *da.*
AV-complete=1SG.NOM see book that COS
'Before I was leaving, I would have already finished reading the book.'

The body part terms *berah* and *bukuy* are also used for directions. This spatial representation is based on the speaker's point of view, and reflects the person's location and his/her surrounding environment. 'The South' is *berah dgiyaq* 'front_mountain', whereas 'the North' is *bukuy dgiyaq* 'back_mountain'. 'The East' is also the right *narat*, and 'the West' the left *iril*.[28]

4.2 Metaphors and metonymies

Body part lexicon is a rich repertoire for figurative languages. In Seediq many metaphors and metonymies come from body part terms, such as *tunux* 'head', *dowriq* 'eye', or *baga* 'hand'. Semantic extensions of body part terms sometimes involve affixation in order to place a body part term at predicate position.[29]

In Seediq one's intelligence depends on whether his/her head is 'good' or 'bad', thus someone who is intelligent has 'a good head', whereas someone who is not has 'a bad head'. This metaphorical expression is similar to that in Mandarin Chinese, where 'the brain is good' is used to describe someone who is smart.

(47) *naqih* *tunux=na.*
bad head=3SG.GEN
'He is stupid.'

(48) *malu* *bi* *tunux=na.*
good very head=3SG.GEN
'He is smart.'

---

[28] The spatial term is also used to mean 'be different, not compatible', as shown below.
me-iril *lenglung-an=na.*
AV-left think-LOC=3SG.GEN
'He is different. He does not fit in.'

[29] An exhaustive illustration of Seediq body part terms as metaphorical and metonymic sources is beyond the scope of this paper. This section thus provides a glimpse at some interesting examples of figurative uses of body part terms, based on my first-hand data. A similar topic is studied by Hsu (2008b), who presents how Truku body part terms are metaphorically used to express the concepts related to EMOTION.

Moreover, one's head can be personified to 'know' or 'not know' anything, indicating metaphorically one's ability about doing something. If one's head 'knows a lot', that means someone is clever.

(49) *m-kela* *bi* *tunux=na*.
AV-know very head=3SG.GEN
'He/She is clever.'

(50) *ini* *kela* *ana* *manu* *tunux=na*.
NEG know INDF what head=3SG.GEN
'He/She is not clever; he/she knows nothing.'

The body part *tunux* 'head' is semantically extended to mean 'the protruding point of an object', which is applied to the other body parts such as *tunux unuh* '(head_breast) nipple' and *tunux utas* '(head_penis) glans'.

In the Seediq dialects spoken in Nantou, the exocentric compound *tanah tunux* 'red_head' refers to the Japanese, in which *tunux* 'head' is metonymically extended to mean 'person'. The colour red does not signify the colour of the soldiers' hair but that of the military hats they wore when entering the villages in the early 1900s.

Similarly, the body part *muhing* also metaphorically refers to corner and peak, as in *muhing sapah* 'ceiling corner', *muhing dxegal* 'floor corner', *muhing cukuey* 'desk corner', or *muhing dgiyaq* 'mountain peak'.

The term *baga* 'hand/arm' is metaphorically associated with stealing and pickpocketing. Therefore, the metaphor HAVING HANDS IS STEALING is reminiscent of the phrase *san-chi-sou* 'three hands' in Mandarin Chinese, which means 'thief, pickpocket', as *sou* 'hand' in this case is a metonym.

(51) *em-baga* *bi* *ka* *baga=na*.
AV-hand/arm very NOM hand/arm=3SG.GEN
'He likes stealing; he often steals.'

(52) *ini=ku* *em-baga* *ka* *yaku*.
NEG=1SG.NOM AV-hand/arm NOM 1SG.NOM
'I am not a thief.'

The internal organ, *baraq* 'lung', is prefixed to mean 'yell, shout loudly'. This semantic extension is probably based on the association of the organ with vital capacity, which is necessary for producing loud voice.

(53) *t-em-baraq* *bi* *r‹m›engaw* *ka* *senaw* *gaga.*
VBLZ-AV-lung very <AV>speak NOM man that
'That man speaks very loudly.'

(54) *t-baraq* *bi* *m-lawa*!
VBLZ-lung very AV-call
'Call loudly!'

(55) *iya* *t-baraq* *r‹m›engaw*.
NEG VBLZ-lung <AV>speak
'Don't speak so loudly.'

Some body part terms are also metaphorical sources for expressing emotion. In Seediq, the body part term *pahung* ‘gall’ is associated with braveness and fierceness, which is also linked with ANGER. See the following examples.

(56) *em-pitu* *pahung=na* *ka* *hiya.*
AV-seven gall=3SG.GEN NOM 3SG.NOM
‘He/She is very brave.’ (Lit. ‘His gall is seven.’)

(57) *em-pahung* *bi* *ka* *laqi* *niyi.*
AV-gall very NOM child this
‘This child is bad-tempered; the child often gets angry.’

(58) *em-pahung* *bi* *kari=na.*
AV-gall very word=3SG.GEN
‘He often yells (angrily); he often tells (people) off.’

(59) *em-pahung* *lenglung-an=na.*
AV-gall think-LOC=3SG.GEN
‘He is sulking.’ (Lit. ‘His mind, place of thinking, is in anger.’)

Similar to Mandarin Chinese, the metonymies A BIG GALL STANDS FOR BRAVENESS and A SMALL GALL STANDS FOR COWARDICE are also found in Seediq, as shown below.

(60) *paru* *pahung=mu* *ka* *yaku.*
big gall=1SG.GEN NOM 1SG.NOM
‘I am very brave.’

(61) *bilaq* *bi* *pahung=na.*
small very gall=3SG.GEN
‘He/she is a coward.’

Another body part term from animals, *ngungu* ‘tail’, is associated with the emotion FEAR. This linguistic connection probably comes from the image of canine behaviour that the tail is put between the back legs when being intimidated or in fear.

(62) *m-ngungu* *balay* *m-usa* *bqerus* *ka* *lituk.*
AV-tail very AV-go cemetery NOM Lituk
‘Lituk is afraid of going to cemeteries.’

(63) *ini* *k-ngungu* *quyu* *ka* *biyang.*
NEG STAT-tail snake NOM Biyang
‘Biyang is not afraid of snakes.’

The term *dowriq* ‘eye’ constitutes the metonym LAUGHING EYES STANDS FOR HAPPINESS and the metaphor HAVING TOO MANY EYES IS CHOOSY, as shown in (64) and (65), respectively. It also serves as a metonym in the compound *qusaw dowriq* ‘coloured_eye; foreigner’ or *dome doriq* ‘(blue_eye) foreigner, (especially of) European (ancestry) (Tg)’, where *dowriq* stands for a person with coloured or blue eyes, hence a foreigner.

(64) *m-hulis* *dowriq=na* *ka* *uwa* *niyi.*
AV-laugh eye=3SG.GEN NOM young.woman this
'The young woman is very happy.'

(65) *lala* *bi* *dowriq=na* *ka* *risaw* *gaga,*
many very eye=3SG.GEN NOM young.man that
kiya ni *ini* *angal* *kuyuh* *bitaq sayang*
therefore NEG take woman until now
'That young man is very choosy; that's why he is still not married now.'

Our ears are the main receptor of ambient acoustic stimuli, which can also lead to annoyance and headache. In Seediq the prefix *s-* is attached to the root *birat* 'ear', *quwaq* 'mouth', and *tunux* 'head' to denote the meaning of 'to be noisy, to annoy' and 'be annoyed (e.g., by surrounding noise)'. A similar use of the prefix is found with the body part term *embuyas*, and the prefixed form means 'to have diarrhea', as shown in (71). It appears that the prefix in this context is related to uneasiness or discomfort associated with the body parts.[30]

(66) *s-em-b~birat* *bi* *ka* *seejiq* *gaga.*
VBLZ-AV-RED~ear very NOM person that
'That person is making a lot of noise.'

(67) *s-brat-an=misu* *da*?
VBLZ-ear-LV=1SG.GEN+2SG.NOM COS
'Have I disturbed you?'

(68) *s-qwaq-an* *ka* *qbheni* *'u* *qduriq.*
VBLZ-mouth-LV NOM bird TOP escape
'Birds will fly away once disturbed.'

(69) *s-em-tunux* *balay huling* *ga* *ni,* *ini* *p-taqi.*
VBLZ-AV-head very dog that CONJ NEG CAUS-sleep
'That dog makes so much noise that (people) are annoyed and cannot sleep.'

(70) *iya* *s-tunux* *hini,* *m-narux* *baki=mu.*
NEG VBLZ-head here AV-ill grandfather=1SG.GEN
'Do not make noise here; my grandpa is ill.'

(71) *gisu=ku* *s-em-buyas.*
PROG=1SG.NOM VBLZ-AV-belly
'I am having diarrhea.'

---

[30] It is apparent that the meaning of this prefix is different from that of the prefix *s-* as discussed in Hsu (2008a). One prefix *s-* attaches to a noun to mean 'to have big X', such as *s-buyas* 'to have a big belly'. Another also attaches to a noun plus optional mono-syllabic reduplication to mean 'to smell of X', such as *s-b~begus* 'to smell of urine' (see also Lee 2010b). The other prefix *s-* attaches to a noun to mean 'to produce, to grow, to have', as in *s-dara* 'to bleed' < *dara* 'blood'; *s-ngudus* 'to grow moustache' < *ngudus* 'moustache, beard'; or *s-huda* 'to snow' < *huda* 'snow'.

Another body part term associated with 'ear' is *qpuji* 'earwax', which is used metaphorically to mean 'be disobedient', thus constitutes the metaphor TO HAVE EARWAX IS DISOBEDIENCE, as shown below.

(72) *maa=su* *m-qpuji* *bi* *rengag-an*?
why=2SG.NOM AV-earwax very speak-NMLZ
'Why don't you listen to what was said?'

Body part terms are often used figuratively to denote physiological and cognitive activities such as intelligence, discomfort, and emotion. For example, being clever or stupid is determined by the condition, good or bad, of *tunux* 'head'. Mental or bodily distress is associated with *tunux* 'head', *birat* 'ear', *quwaq* 'mouth', and *embuyas* 'belly'. The following table summarises the body part terms associated with mental abilities and activities found in Seediq.

Table 16: Figurative uses of Seediq body part terms

body part	gloss	bodily connection	figurative language
tunux	'head'	intelligence	A GOOD HEAD IS INTELLIGENCE
			A BAD HEAD IS STUPIDITY
		annoyance	AN AFFECTED HEAD IS ANNOYANCE
birat	'ear'	annoyance	AN AFFECTED EAR IS ANNOYANCE
qpuji	'earwax'	disobedience	TO HAVE EARWAX IS DISOBEDIENCE
dowriq	'eye'	happiness	LAUGHING EYES STANDS FOR HAPPINESS
		choosiness	TO HAVE MANY EYES IS CHOOSY
quwaq	'mouth'	annoyance	TO HAVE (BIG) MOUTH IS TALKATIVE/NOISY
baga	'hand'	stealing	TO HAVE HANDS IS STEALING
pahung	'gall'	anger	TO HAVE (SEVEN) GALL IS ANGER
		braveness	A BIG GALL STANDS FOR BRAVENESS
		cowardice	A SMALL GALL STANDS FOR COWARDICE
embuyas	'belly'	diarrhea	AN AFFECTED BELLY IS DIARRHEA
ngungu	'tail'	fear	AN AFFECTED TAIL IS FEAR

5 Concluding remarks

This paper presents a study on body part terminology in Seediq, with a focus on two aspects—nomenclature and categorisation, which are captured through an exploration of their structural properties and semantic extensions. With structural properties the inventory of body part terms is divided into eight categories. In each category the lexical items are grouped into simplex and complex terms. This is followed by a discussion on the morpho-semantic types of compounds exhibited by the complex body part terms. As for semantic properties, the discussion is focused on spatial grams evolved from body part terms and figurative usage of body part terms as metaphors and metonyms.

Nomenclature of Seediq body part terms echoes some of the lexical universals proposed by Andersen (1978) except that all languages will have labels for FINGER, TOE, and NAIL. It is found that many body part terms in Seediq are polysemous. The lexical

item *hiyi* (or written as *hei* in Tgdaya), as an equivalent term of 'human body', also refers to person, animal's meat, and pulp or flesh of fruits.[31] The parts ARM and HAND share the same body part term *baga*, so as LEG and FOOT *papak*. FINGER and TOE are not lexically labelled, and are conceptually realised as 'small one' or 'digit' of *baga* 'hand/arm' and *qaqay* or *papak* 'foot/leg'. The body part *kukuh* 'nail' in Seediq also lexically overlaps with 'claw or hoof' of mammals.

Although some body part terms are shared both by humans and animals, such as *waqit* (or *waqic*) ('canine tooth' and 'chela') or *kukuh* ('nail', 'claw', 'hoof'), certain body parts have distinctive terms used for humans and nonhumans, including FOOT/LEG and STOMACH. Some speakers of Toda and Truku make a distinction between human *qaqay* and nonhuman *papak*. The lexical items related to STOMACH such as *rketu*, *rkebul* (*rkebun*), or *lhebun* in different dialects as shown in Table 8 also reflect the semantic change of these terms.

Nearly two hundred Seediq body part terms have been collected so far. In total about 180 terms are listed, including around 110 simplex terms (59%) and 70 complex terms (41%). Almost all the complex terms are formed by compounding. These complex terms are categorised into three types: (i) left-hand headed endocentric compound, (ii) right-hand headed endocentric compounds, and (iii) exocentric compounds. The endocentric compounds also contain several meronymic compounds.

A moderate degree of inter-dialectal variation is also demonstrated in this study, as shown in the tables. In comparison, for certain body parts Tgdaya exhibits more lexical items which are different from those in Truku and Toda. Generally, Truku and Toda are more similar to each other in body part nomenclature than each of them is to Tgdaya.

Three body part terms in Seediq have evolved into FRONT-REGION and BACK-REGION spatial grams, following Svorou's (1994) framework. The FRONT-REGION spatial gram is represented by *berah* 'chest', and the BACK-REGION one by both *bukuy* 'back' and *sulay* 'hip'.

Semantic extensions of body part terms in Seediq contradict partly from Blust's (2009:311[2013:322]) observations that '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 16 shows that HEAD is figuratively associated with both physical senses (e.g., ANNOYANCE, DISTURBANCE) and qualities of temperament (INTELLIGENCE/STUPIDITY). Similarly, EYES can also be connected with emotion (HAPPINESS) and personal character (CHOOSINESS). The internal organ GALL is associated with both BRAVENESS/COWARDICE and the emotion ANGER. Finally, the association of TAIL with FEAR probably stems from the image of a dog's behaviour that its tail is put between the back legs when being intimidated, reflecting a close relationship between people and dogs in the Seediq culture.

[31] In Tgdaya and Toda the lexical item *wawa* is used to mean 'meat of games or wild beasts'.

References

Andersen, Elaine S. 1978. Lexical universals of body-part terminology. *Universals of human language, Volume 3: Word structure*, ed. by Joseph H. Greenberg, 335–368. Stanford: Stanford University Press.

Berlin, Brent and Paul Kay. 1969. *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Blust, Robert. 2009[2013]. *The Austronesian languages*. Pacific Linguistics 602. Canberra: The Australian National University. [2013 revised edition online at: http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf]

Brown, Cecil H. 1976. General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist* 3.3:400–424.

Brown, Cecil H. and Stanley R. Witkowski. 1981. Figurative language in a universalist perspective. *American Ethnologist* 8.3:596–615.

Burenhult, Niclas. 2006. Body part terms in Jahai. *Language Sciences* 28.2–3:162–180.

Enfield, Nick J., 2006. Elicitation guide on parts of the body. *Language Sciences* 28:148–157.

Enfield, N. J., Asifa Majid, and Miriam van Staden. 2006. Cross-linguistic categorisation of the body: Introduction. *Language Sciences* 28.2–3:137–147.

Evans, Nicholas, and David Wilkins. 2001. The complete person: Networking the physical and the social. In *Forty years on*: *Ken Hale and Australian languages*, ed. by Jane Simpson, David Nash, Mary Laughren, Peter Austin, & Barry Alpher, 493–521. Pacific Linguistics 512. Canberra: The National Australian University.

Holmer, J. Arthur. 1996. *A parametric grammar of Seediq*. Travaux de l'institut de linguistique de Lund 30. Lund, Sweden: Lund University Press.

Hsieh, Fuhui. 2007. Language of emotion and thinking in Kavalan and Saisiyat. PhD dissertation. Taipei: National Taiwan University.

Hsu, Lowking Wei-Cheng. 2008a. tailugeyu goucifa yanjiu [Word formation in Truku]. MA thesis. Hsinchu: National Hsinchu University of Education. [In Chinese]

Hsu, Lowking Wei-Cheng. 2008b. tailugeyu qingxu gainian yongfa yanjiu [The study of emotion concepts in Truku]. Paper read at the Habul Workshop on Formosan Languages, 6 December. Hsinchu: National Hsinchu University of Education. [In Chinese]

Huang, Shuanfan. 2002. Tsou is different: A cognitive perspective on language, emotion, and body. *Cognitive Linguistics* 13.2:167–186.

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by.* Chicago: University of Chicago Press.

Lee, Amy Pei-jung. 2009. Dorsal consonant harmony in Truku Seediq. *Language and Linguistics* 10.3:569–591.

Lee, Amy Pei-jung. 2010a. Phonology in Truku Seediq. *Taiwan Journal of Indigenous Studies* 3.3:123–168.

Lee, Amy Pei-jung. 2010b. Reduplication and odor in four Formosan languages. *Language and Linguistics* 11.1:99–126.

Lee, Amy Pei-jung. 2012a. The segment /w/ and contrastive hierarchy in Paiwan and Seediq. *Concentric: Studies in Linguistics* 38.1:1–37.

Lee, Amy Pei-jung. 2012b. Contact-induced sub-dialects in Toda Seediq. Paper presented at the 12th International Conference on Austronesian Linguistics (12-ICAL), 2–6 July. Bali, Indonesia: Udayana University.

Li, Paul Jen-kuei. 1981. Reconstruction of proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52.2:235–301.

Li, Paul Jen-kuei. 1991. Vowel deletion and vowel assimilation in Sediq. In *Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, ed. by Robert Blust, 163–169. Pacific Linguistics C-117. Canberra: The Australian National University.

Li, Paul Jen-kuei. 2007. renti ge buwei mingcheng zai yuyan shang de yingyong. [Usage of body part terminology] *Language and Linguistics* 8.3:711–722. [In Chinese]

Majid, Asifa. 2010. Words for parts of the body. In *Words and the mind: How words capture human experience*, ed. by Barbara C. Malt and Phillip Wolff, 58–73. Oxford: Oxford University Press.

Svorou, Soteria. 1994. *The grammar of space*. Typological Studies in Language 25. Amsterdam: John Benjamins.

Terrill, Angela. 2006. Body part terms in Lavukaleve, a Papuan language of the Solomon Islands. *Language Sciences* 28.2-3:304–322.

Tsukida, Naomi. 2005. Seediq. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 291–325. Routledge Language Family Series. London and New York: Routledge.

Tsukida, Naomi. 2009. Sedekku-go-no bunpōu [A grammar of Seediq]. PhD dissertation. Tokyo: The University of Tokyo [in Japanese]

van Staden, Miriam. 2006. The body and its parts in Tidore, a Papuan language of Eastern Indonesia. *Language Sciences* 28.2-3:323–343.

Yang, Hsiu-fang. 1976. saideyu wushe fangyan de yinyun jiegou [Phonological structure of the Paran: A dialect of Sediq] *The Bulletin of the Institute of History and Philology, Academia Sinica* 47.4:611–706. [In Chinese]

Yeh, Maya Yuting. 2002. Emotion concepts in Squliq Atayal. MA thesis. Taipei: National Taiwan University.

Wierzbicka, Anna. 2007. Bodies and their parts: An NSM approach to semantic typology. *Language Sciences* 29.1:14–65.

20 *Metaphorical expressions of anger and happiness in Truku Seediq and English**

APAY AI-YU TANG

1 Introduction

Does emotion language consist only of literal emotion words like *fear*, *anxiety*, *anger*, *love*, *happiness*, and so forth? Can these abstract emotion concepts be expressed figuratively? Is the way in which abstract emotional concepts are expressed different cross-culturally? Figurative (metaphoric and metonymic) terms were once considered merely decorative and ornamental in nature. However, a more recent approach views figurative speech as pervasive and significant in language and thought (e.g., Kövecses 1990; Lakoff 1987, 1993; Lakoff & Johnson 1980). These studies claim that although human emotions are abstract in nature, they are extensively expressed in figurative terms. Therefore, a major challenge in emotion language research is understanding how our abstract emotional reality is constituted and how emotions are expressed cross-culturally and cross-linguistically.

Cognitive linguists claim that metaphors have mappings that are based on widely shared experiences and perception even in unrelated languages like English (Lakoff & Johnson 1980), Hungarian (Bokor 1997), and Chinese (Yu 1995). Metaphors can be motivated by universal correlations in our bodily experience; therefore, when we are joyful, we tend to be up, move around, be active, jump up and down, rather than being static or inactive (Kövecses 2005).

This paper primarily attempts to compare the metaphorical expressions of ANGER and HAPPINESS in English and Truku Seediq. The Seediq language is an Austronesian (AN) language spoken by one of the indigenous groups who live in the northeastern part of

* This paper is dedicated to Prof. Lillian Huang, who has taken a leading role in the study of Formosan languages. I am grateful to two anonymous reviewers for their instructive comments, and the feedback from the audience at the Mental Architecture for Processing and Learning of Language Conference, 3–4 August 2013, Kwansei Gakuin University, Osaka, Japan. I would like to thank Elizabeth Zeitoun, Stacy F. Teng, Joy J. Wu, Hsiu-chuan Liao, Amy P. Lee, Rik De Busser, and Maya Yuting Yeh for their comments and suggestions. I am solely responsible for the remaining errors.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 485–512
Asia-Pacific Linguistics, 2015

Taiwan. There are three major dialects including Teuda, Tkdaya, and Truku. The current study focuses on Truku dialect, which contains the largest number of speakers among these dialects. The Truku population is around 29,410. Nevertheless, Truku currently shows strong signs of cross-generational decline and further erosion in the youth.[1]

Truku speakers inhabit mountainous areas, and hunting is traditionally the men's main activity and responsibility in Truku society. Like most AN-speaking people, Truku are agriculturalists, and the traditional economy was based on cultigens such as yam, several types of taro, sugarcane, and banana. The most widespread domesticated animals are dogs, pigs, and fowls. Correlated with the agricultural cycle, headhunting expeditions were traditionally important (Blust 2009).

For linguistic data resources, this paper uses English example sentences from studies by Lakoff & Johnson (1980), Yu (1995), and Kövecses (2000, 2005).[2] Truku Seediq data come from narratives in the work of Pisaw et al. (2005) and Nowmaw & Pisaw (2007), and from my conversations mainly with Yudaw Pisaw in Qowgan village.[3]

In terms of the paper's organization, I begin with a brief explanation of the motivation of the study in section 2. In section 3, types of emotion language and the relationship of metaphor and emotional concepts are discussed. The metaphorical conceptualization of ANGER in both languages is described in section 4. I then present the metaphorical conceptualization of HAPPINESS in section 5. Section 6 provides a brief general discussion, and section 7 concludes the paper.

2 Motivation

The motivation of this study is threefold. First, little scholarly attention has been paid to emotional figurative expressions among Formosan languages or their cross-cultural comparison with those in other language families. Previous studies of the Truku dialect mainly deal with descriptions of morphosyntax (Tsukida 2005), phonology (Lee 2009, 2010) and phonetics (Chiang & Chiang 2005). In addition, Lee (2011) provides a thorough description of metaphorical euphemisms underlying the categories of RELATIONSHIP and DEATH in Kavalan, Paiwan, and Truku Seediq. Regarding emotional language research, Yeh (2002) gives an account of emotion expressions in Squliq Atayal. Lin (2006) explores the language of emotion concepts in Kavalan, and Hsieh (2007, 2011)

---

1 The evidence of such linguistic decline was demonstrated in Tang (2011:42–146) http://www.ling.hawaii.edu/graduate/Dissertations/ApayTangFinal.pdf.

2 However, based on the limited scope of the examples given in this paper, I assume that other expressions exist in English.

3 The elder Yudaw Pisaw (eighty-nine years old at the time of this study) is a widely acknowledged competent Truku speaker who has served as one of the members of the Truku Dictionary Committee from 2005 until the present. From 1953 to 1963, he translated the New Testament of the Bible with Dr. Ralph R. Covell, a fluent Truku speaker and an American missionary in the Seediq speech community, and Teylung Liduk, a senior pastor in Qowgan village, Hualien County, Taiwan. The data were additionally confirmed by senior speakers including Reykong Huwat, Yuki Kumus, and Bowtung Yudaw (sixty-seven, sixty-nine, and fifty-five years old, respectively, at the time of the study), all living in Qowgan village, Hualien, eastern Taiwan.

investigates how language conceptualises emotions in Kavalan, Saisiyat, and Paiwan. Earlier studies have not systematically discussed the topic of emotional language in Truku Seediq.

Second, relatively little attention has been paid to the role of figurative language in comparing the conceptualization of emotion cross-culturally, although several studies have been conducted to investigate the conceptualization of emotion in non-Indo-European languages. For instance, Munro (1991) studies anger in Wolof, Yu (1995) compares metaphorical expressions of anger and happiness in English and Chinese, Matsuki (1995) studies metaphors of anger in Japanese, Bokor (1997) analyzes the use of anger in Hungarian concepts, and Taylor & Mbense (1998) investigate the concept of anger in Zulu. Therefore, by focusing on one of the Austronesian languages, this paper intends to provide further insight into the nature of the relationship between language and emotion cross-linguistically.

Third, an integrated account of emotional phenomena in human language as a whole is still lacking. LeDoux (1996), a neurobiologist, claims that emotion language consists only in literal emotion words, such as *fear*, *happiness*, and so forth. However, Kövecses (2000) provides a new synthesis in the study of emotion; he attempts to integrate language, culture, and body in such a way as to achieve a more complete and incorporated account of emotional phenomena in human beings.[4] Following his approach to the study of emotion, this paper intends to serve as a preliminary investigation of language and emotion concepts in Truku Seediq not only from a cognitive linguistic viewpoint, but also from social and cultural perspectives.

The paper focuses on anger and happiness because they are two of the emotion concepts that have received attention from a variety of scholars.[5] Kövecses (2000:20) claims that they are prototypical and occur on most lists of "basic emotions" (e.g., anger, fear, happiness, sadness). He further indicates that research by scholars using a cognitive linguistic framework, such as Lakoff & Johnson (1980), Lakoff (1987), and others, have taken figurative language most seriously in the study of human conceptual systems.

3 Language and emotion concepts

3.1 Types of emotion language

How do we think about and convey emotions? Kövecses (2000:20) suggests that emotion-related language can be expressed in three ways, through expressive, descriptive, and figurative terms, as depicted in Figure 1. First, expressive emotion words can convey emotions such as enthusiasm with *wow!* or disgust with *yuk!*

---

[4] In his attempt to gain a coherent and holistic view of emotion language, Kövecses (2000:xii) brings together three threads, including "the research done on how the human body behaves in an emotional state, the research on how cultural and social factors influence and shape emotional experience, and the research on emotional language from a cognitive linguistic perspective."

[5] Basic emotions such as anger, fear, sadness, disgust, happiness, contempt, and shame represent universal categories of subjective emotional experience and universal categories used in the recognition of emotional signals (Izard 1977).

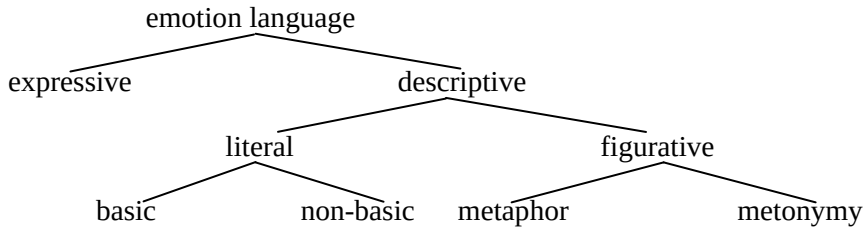


Figure 1: Summary of types of emotion language

Second, descriptive terms like *I love you* or *I am angry* describe particular emotions. They signify and are seen as "more or less basic." For instance, speakers feel *anger* to be more basic than *annoyance*. *Anger*, the "basic level" emotion category, is in between the superior superordinate-level category *emotion* and the subordinate-level category of *annoyance*, as seen in Figure 2.

superordinate-level:	*emotion*
middle (basic) level:	*anger*
subordinate-level:	*annoyance*

Figure 2: Levels of emotion terms in a vertical hierarchy

Third, figurative terms denote different aspects of emotional concepts, such as intensity, cause, and so forth. Although figurative expressions may be the largest of these categories, they have received the least attention in the study of emotion language. As noted earlier, some scholars believe that emotion language consists only of literal emotion words. Nevertheless, an increasing number of scholars do not accept this oversimplified view of how emotion and language interact; instead, they believe that abstract concepts like emotion can generally be expressed only through the mediation of metaphor (Gibbs 1994; Holland & Kipnis 1995; Yu 1995, among others). That is to say, abstract concepts are not literally conceptualised but metaphorically constituted. In the following subsections, conceptual metaphor and metonymy, as well as meaning theories and emotions, are briefly discussed.

3.2 Conceptual metaphor

Lakoff (1993) argues that conceptual metaphor is pervasively employed when we talk about abstract notions. Lakoff & Johnson (1980:5) argue that "the essence of metaphor is understanding and experiencing one kind of thing in terms of another." For instance, "time" can be spent, possessed, invested, or budgeted just as if it were a physical object. Hence, the abstract notion of "time" is expressed metaphorically in terms of properties of a physical object. Lakoff (1987) further argues not only that abstract notions can be described in terms of concrete concepts, but that abstract source domains can be conceived of or mapped onto in terms of concrete target domains. For example, the abstract concept of ANGER can be conceived of and expressed in terms of the concrete source domain HOT FLUID IN A CONTAINER (e.g., "boil with anger," "simmer down," or

"seethe with anger"). The conceptual mapping of this correspondence can be characterised as in Figure 3 (Kövecses 2006:123).

Source: hot fluid in a container		Target: anger
the physical container	→	the angry person's body
the hot fluid inside the container	→	the anger
the degree of fluid heat	→	the intensity
the cause of increase in fluid heat	→	the cause of anger

Figure 3: Conceptual mapping of the correspondence

As can be seen here, metaphorical linguistic expressions make the conceptual metaphor (i.e., the metaphor in the mind) manifest, and the components of conceptual metaphors include: (i) source domain, (ii) target domain, (iii) basis of metaphor, (iv) neural structures that correspond to i and ii in the brain, (v) relationships between the source and the target, (vi) metaphorical linguistic expressions, (vii) mapping, (viii) entailment, (ix) aspects of source and target, (x) blends, (xi) nonlinguistic realizations, and (xii) cultural models (Kövecses 2006:116).[6] Kövecses suggests that "metaphor is a linguistic, conceptual, social-cultural, neural, and bodily phenomenon and that it exists on all of these different levels at the same time" (ibid.:126).

3.3 Metaphor and metonymy

Although both metaphor and metonymy are closely linked to how our emotional reality is constituted, they differ from each other (Kövecses 2000:4, 2006:126–127). First, while a metaphor provides understanding for a more abstract concept (the target domain) through a more concrete one (the source domain), metonymy is a process that provides mental access to an entity through another entity. Second, metaphor involves two discrete domains, whereas metonymy involves a single domain. Third, metaphor deals with a relationship between entities that are far from each other in conceptual space (e.g., "life" and "journey"), although they may "come close together" as a result of correlation in experience. However, the elements in a metonymic relationship are conceptually close to each other (e.g., the "place" and the "institution").

Metaphor and metonymy may express any aspect of emotion concepts (Kövecses et al. 2003). Some metaphors reveal universal image schemata (e.g., the idea that anger is conceptualised as pressure in a container), and metonymies may convey universal aspects of emotions as well (e.g., the idea that anger is physiologically felt as redness, a rise in body temperature, and loss of rationality). Other metaphors and metonymies can be unique to a specific culture. For example, Zulus become wet in anger, but Americans do not (Taylor & Mbense 1998), perhaps because sweating is considered more disgusting and embarrassing by Americans, or because it would signify that a person loses control. Hence, Kövecses et al. (2003) suggest that emotion concepts are motivated neither by the human body nor the social structure, but by both human physiology and a specific social

[6] Kövecses (2005:23–24) states that "conceptual metaphors are ensembles of neurons in different parts of the brain connected by neural circuitry. The ensembles of neurons located in different parts of the brain are the source and target domains, and the physical neural circuitry that connects them are the mappings."

and cultural environment. They further conclude, "[d]ifferent cultural and social environments bring different values to bear on the formation of emotion concepts" (153). The following sections discuss mainly these two distinct views—*prototype* and *social constructionist*—of meaning theories and emotion.

3.4 Meaning theories and emotions

3.4.1 The "prototype" view

Kövecses et al. (2003:139) describe the "prototype" view of emotional meaning as based on the following characteristics. First, it is related to the best examples of basic-level categories mentioned in section 2.1 above. In English, for instance, the best examples of the category include anger, fear, sadness, happiness, and love. As we try to specify the structure and content of the basic-level categories, we are working within this view of emotional meaning.

Next, the structure of emotion concepts is viewed as a scenario, script, or model by many scholars. For example, according to Lakoff & Kövecses (1987), anger is conceptualised as a five-stage scenario: (i) cause of anger, (ii) anger exists, (iii) attempt at controlling anger, (iv) loss of control over anger, and (v) retribution. In other words, the series of events makes up the structure of the prototypical concept of any given emotion. Third, Kövecses (1990) suggests that emotions such as love, fear, and happiness have multiple prototypes, not just one; several members can be the best example within an emotion category, because one can be typical, another can be salient, still another can be ideal, and so forth.

3.4.2 The "social constructionist" view

Several scholars view emotion concepts as social constructions. For instance, Lutz (1988) offers a model to account for the emotion *song* (roughly corresponding to anger) in Ifaluk, part of Yap State in the Caroline Islands of Micronesia. She claims that the structure of an emotion concept depends upon particular aspects of a specific society and culture and gives the following account of anger in Ifaluk: (i) There is a rule or value violation, (ii) it is pointed out by someone, (iii) this person simultaneously condemns the act, (iv) the perpetrator reacts in fear to that anger, and (v) the perpetrator amends his or her ways.

Kövecses et al. (2003) conclude that it would be oversimplified to view the concept of anger as only motivated either by human physiology or by social construction. Rather, they suggest it is motivated by both spheres. In other words, social structure must have some basis in universal bodily experiences, and bodily motivations must acquire specific social-cultural content and interpretation. Based on this blending view, the current paper compares and explores two sets of prototypical basic emotion language—for anger and happiness—in English and Truku Seediq.

4 Figurative expressions of anger in English and Truku Seediq

In English, both emotion words and descriptive emotion terms are available; some emotion words can express an emotion without naming it. Examples are *shit* or *hey* (with emphasis), said when angry. Descriptive emotion terms like *I am angry with him!* can express particular emotion, in which the word *angry* is used to describe and express the emotion of anger.[7]

Similarly, both words and descriptive terms of emotion are also found in Truku Seediq. Emotion words like *cik* 'sound that one produces when angry' convey the emotion of anger. Four specific descriptive emotion terms of anger are *saang* 'to be literally or generally angry' as in (1a); *arih* 'to be angry but refrain from saying it' as in (1b); *teuqu* 'to be upset because of being mistreated by others' as in (1c); and *beteraw* 'to be so angry that she leaves from the exact spot, specifically referring to a daughter-in-law's being angry and leaving home due to her mother-in-law' as in (1d).[8]

(1) a. *rinah* *me-**saang*** *ka* *hiya*[9]
even more AV-angry NOM 3SG[10]
'He is even angrier.'

b. *cse-**arih*** *kari=na* *ka* *hiya.*
like-angry word=3SG.GEN NOM 3SG
'His words seem angry.'

c. ***teuqu*** *ka* *Tumun* *da,* *yasa* *biq-u*[11] *bilaq* *bi* *iyang.*
resent NOM Tumun already because give-GV1 little very meat
'Tumun resents, because she was given very little meat.'

d. ***beteraw*** *me-saang* *ka* *Tumun.*
leave AV-angry NOM Tumun
'Tumun leaves because of anger.'

---

[7] Compared to English, Truku Seediq seems to have more expressions that contain the form *saang* 'angry' or *qaras* 'happy' from the examples given in this paper. For example, descriptive emotion words of anger can be seen in (1a)–(1d). In terms of happiness, emotion words like *wah* 'sound that one produces when happy' expresses the emotion of happiness. Descriptive emotion terms of happiness can be conveyed by *qaras* 'to be literally or generally happy' as in the sentence *m-**qaras** ku balay* 'I am very happy' and the examples in (46)–(53), except for (49).

[8] The generic word for emotion is *lenglung-an* 'think-NMLZ', where the suffix *-an* deverbalises the verb 'think' and makes it an abstract noun; this is consistent with other Formosan languages like Squliq Atayal (Yeh 2002:14).

[9] To distinguish the position of schwa /ə/ written as the letter *e* in a word, the agent voice (AV) is written as *me-* as in (1a) or *em-* as in (4c).

[10] This paper follows the Leipzig Glossing Rules in glossing the example sentences with the following additions: AV, agent voice; EXC, exclamation; GV, goal voice; PCL, particle. In addition, the verbal prefixes *t-*, *p-*, *k-*, and *s-*, explained more fully in footnote 16, are glossed as T-, P-, K-, and S-.

[11] I agree with Tsukida (2005:299), who treats GV as a cover term for *-un* and *-an*; *-un* is called GV1 and *-an* is called GV2. There is an overlapping distribution in their use, and I suggest elsewhere that the difference between them is in the degree of transitivity (Tang 2010:14).

As can be seen from these examples, expressive and descriptive emotion terms to express anger are adequately available in both languages. In her study of emotion language in Kavalan, Saisiyat, and Paiwan, Hsieh (2011:65) shows that languages having emotion nouns have comparatively rich metaphorical expressions for emotion, whereas those languages where linguistic mechanisms to encode abstract nouns are limited dis-preferred metaphorical strategies. Hence, it is postulated that both English and Truku Seediq, having a wide selection of emotion nouns, may provide language users with more choices of metaphorical strategies in conceptualizing emotion events.[12] In the following sections, figurative expressions including metaphorical and metonymic expressions of anger are discussed for both languages.

4.1 Metaphorical expressions of anger

In general, Kövecses (2006:172) claims that the metaphorical system that each culture produces will in general be coherent with embodiment, culturally defined experience, and certain cognitive preferences and styles. The following section discusses two generic metaphorical categories of anger expressions—ANGER IS HEAT and ANGER IS DANGER—in English and Truku Seediq.

4.1.1 ANGER IS HEAT

Conceptual metaphors of anger have been studied first in English, then Chinese, Japanese, Hungarian, Wolof, Zulu, and Polish. Kövecses (2005:39) concludes that in all these languages, the conceptual metaphor THE ANGRY PERSON IS A PRESSURISED CONTAINER is found, either with or without heat. There are systematically corresponding mappings between source and target domains. He further argues that these mappings produce a scene or situation for anger in which there is a force inside a person and the force causes the person to act in certain ways that should be suppressed.

In English, Lakoff & Kövecses (1987:197) suggest that the basis of the most general metaphor for anger is ANGER IS HEAT, in which a heated container maps to an angry individual, heat maps to anger, smoke/steam maps to signs of anger, explosion maps to extreme, and uncontrolled, anger (see Table 1). This central metaphor has two subcategories. On one hand, when heat is applied to solids, the metaphor is ANGER IS FIRE as in (2a)–(2d).

(2) ANGER IS FIRE (English)
a. These are *inflammatory* remarks.
b. He's doing a slow *burn*.
c. His anger is *smoldering*.
d. *Smoke* was pouring out of his ears.

---

[12] The specific emotion nouns in Truku Seediq are *saang* 'anger', *kliyus* 'angry with clenched teeth', *pahung* 'bad temper; meanness', *arih* 'upset', *teuqu* 'resentment', *beteraw* 'anger with leaving', *isug* 'fear', *ngungu* 'fear', *siqa* 'shyness', *kmsteutux* 'awe', *kisi* 'disgust', *riqu* 'dislike', *qaras* 'joy or happiness', *udaw* 'pride', *smpraw* 'pride', *qnnaqih* 'guilt', *punu* 'craziness', *bsrat* or *bsu* 'stinginess', and *gealu* 'sympathy' (data taken from Nowmaw & Pisaw 2007, and Truku e-dictionary 2013).

On the other hand, when heat is applied to fluids, the metaphor is ANGER IS A HOT FLUID IN A CONTAINER. As illustrated in Figure 3 above, the physical container is the source and the angry person's body is the target domain. The hot fluid inside the container maps to anger, the degree of fluid heat maps to intensity, and the cause of increase in fluid heat maps to the cause of anger, as in (3a–3d).

(3) ANGER IS A HOT FLUID IN A CONTAINER
a. She is *boiling* with anger.
b. His pent-up anger *welled up* inside him.
c. You make my blood boil.
d. James is just *blowing off* steam.

The ANGER IS HEAT METAPHOR can be found not only in English but also in Truku Seediq. However, rather than being divided into ANGER IS FIRE and ANGER IS A HOT FLUID IN A CONTAINER as in English, the metaphor has three subcategories in Truku Seediq: (1) ANGER IS FIRE, (2) ANGER IS A BODILY FLUID IN A CONTAINER, and (3) ANGER IS GAS IN A CONTAINER. Three differences can be found between the two languages.

First, although heat can be applied to solids with ANGER IS FIRE in both languages, the conceptualization of anger is less diverse in English than it is in Truku Seediq. In English, the heat-related words for anger are from burning, smoldering, and smoking as in (2a–2d) above. In Truku Seediq, in addition to the heat-related words *sqama* 'burn' as in (4a) and *sranaq* 'flame' as in (4b), the word *bngux* 'ash' is also used to show the intensity of anger, as in (4c).[13] Moreover, the sound of burning fire *em-bengbing* can be employed to indicate the immediateness and intensity of anger, as in (4d). In other words, the entire process of burning, including smoldering, burning, the sound of blazing, and the resulting ash, can be employed and activated in a Truku speaker's mind through metaphor. This prevalent metaphor for anger might be influenced by their social-cultural experience, since the mountain-inhabiting Truku depend greatly on fire in their everyday life.

(4) ANGER IS FIRE

a.	*ma*	*su*	*s-qama*	*saang*	*hug*?
	how	2SG	S-burn	angry	Q
	'How come you are so angry?' (Lit. 'How come you have burning anger?')				
b.	*ma*	*su*	*s-ranaq*	*saang*	*hug*?
	how	2SG	S-flame	angry	Q
	'How come you are so angry?' (Lit. 'How come you have flame-like anger?')				
c.	*ma*	*su*	*em-bngux*	*me-saang*	*hug*?
	how	2SG	AV-ash	AV-angry	Q
	'How come you are so angry?' (Lit. 'How come you have ash-like anger?')				
d.	*ma*	*su*	*em-bengbing*	*me-saang*	*hug*?
	how	2SG	AV-sound of burning fire	AV-angry	Q
	'How come you are so angry?'				
	(Lit. 'How come you are angry so that your sound is like burning fire?')				

[13] The ash-like anger described here is similar to smoldering anger in English (Yudaw Pisaw, pers. comm., January 2013).

Next, in contrast to the second English subcategory, in which ANGER IS HEAT applies to fluids, the Truku Seediq version applies to bodily fluids, and is without heat. Hence the metaphorical concept derived is ANGER IS A BODILY FLUID IN A CONTAINER. On one hand, speakers of English greatly rely on the physical image of hot fluid, as in *boiling with anger*, *well up*, and so forth, to show the intensity of anger as in (3a–3d) above. Anger refers to a hot substance, and an angry person is a container filled with a hot substance.

On the other hand, although THE ANGRY PERSON IS A PRESSURISED CONTAINER metaphor is applicable to Truku Seediq, it is different from English in three aspects. First, the substance in the container is not hot but is body temperature as in (5a–5c). Next, unlike speakers of English who activate the physical image of hot fluid in an *external* container such as a kettle or pot, Truku speakers tend to rely on the image of bodily fluid pressurised in an *inner* container, which is the human body itself. The image schema emerging from this bodily experience manifests as less intense than it is in English. Third, the forms or states of the bodily fluids, such as *halus* 'saliva' in (5a), *buwa* 'bubble of saliva' in (5b), and *blus* 'babbling' in (5c), are the natural results of anger, and their different expressive uses seem to suggest a gradual intensification of the notion of anger. In addition, the Truku expressions do not imply that the person is shouting angrily, although they refer to the mouth and saliva; they are used when no shouting co-occurs with anger.

(5) ANGER IS A BODILY FLUID IN A CONTAINER

a. *ma* *su* *me-s-halus* *me-saang* *hug*?
why 2SG AV-S-saliva AV-angry Q
'How come you are so angry?' (Lit. 'How come you are so angry that your saliva appears around your mouth?')

b. *ma* *su* *em-buwa* *me-saang* *hug*?
how 2SG AV-bubble AV-angry Q
'How come you are so angry?' (Lit. 'How come you are so angry that you are frothing at the mouth?')

c. *ma* *su* *t-blus* *me-saang* *hug*?
how 2SG T-babbler AV-angry Q
'How come you are so angry?' (Lit. 'How come you are so angry that you are babbling and fluid comes out?')

The third difference between these two languages is that Truku Seediq may use *gas* in its conventionalised expression of anger. As in (6), the word *t-peruq* 'break from within' shows that the emotion of anger is mapped onto *gas* and the force of the inner pressure brings the ultimate consequence of explosion.

(6) ANGER IS GAS IN A CONTAINER

ma	*su*	*t-peruq*	*saang*	*hug*?[14]
how	2SG	T-break	angry	Q

'How come you are so angry?'
(Lit. 'How come you are so angry as if your inner gas breaks out?')

4.1.2 ANGER IS DANGER

In addition to the central metaphor ANGER IS HEAT in English, Kövecses (2000:21) provides the list in (7)–(13) of metaphorical expressions and linguistic examples illustrating the mapping between source and target domains. These source domains address many aspects of the concept of anger. For instance, FUNCTIONING MACHINE places the stress on the angry person, PHYSICAL ANNOYANCE on the cause of anger, AGGRESSIVE ANIMAL BEHAVIOR on the angry behavior, and so forth.

English

(7) ANGER IS A CAPTIVE ANIMAL: He *unleashed* his anger.
(8) ANGER IS AGGRESSIVE ANIMAL BEHAVIOR: Don't *snarl* at me!
(9) THE CAUSE OF ANGER IS PHYSICAL ANNOYANCE: He's *a pain in the neck*.
(10) ANGER IS A NATURAL FORCE: It was a *stormy* meeting.
(11) THE ANGRY PERSON IS A FUNCTIONING MACHINE: That really *got* him *going*.
(12) ANGER IS A BURDEN: He *carries* his anger *around* with him.
(13) ANGER IS A SOCIAL SUPERIOR: His actions were completely *governed* by anger.

With English, Truku Seediq shares two similar metaphorical expressions other than those of ANGER IS HEAT. Nevertheless, they are culture-specific. First, both languages share the generic-level structure ANGER IS DANGER. However, in contrast to the English version in which ANGER IS DANGER is expressed in metaphors of a captive animal, as in (7), and aggressive animal behavior as in (8), the Truku Seediq alternative metaphor uses a specific non-human fierce animal, *huling* 'dog' or *kumay* 'bear', as in (14).

(14) ANGER IS AGGRESSIVE ANIMAL BEHAVIOR

masu	*saw*	*huling/kumay*	*me-saang*	*kn-an*?
how	2SG	like dog/bear	AV-angry	1SG-GV2

'How come you are very angry with me?'
(Lit. 'How come your anger is dog/bear-like with me?')

Another similar but culturally specific metaphorical expression in both languages is related to the ANGER IS A NATURAL FORCE metaphor. While the emotion of anger is mapped onto *storm* as in (10) in English, it can be applied to both *bruwa* 'thunder' as in (15a) and *dowras* 'cliff' as in (15b) in Truku Seediq. The metaphorical use of the word

[14] The prefix *t-* implies that the direction of the force is from inside to out, as in the word *t-peruq* 'break out from within', whereas the prefix *p-* is a causative that indicates the direction of the force is from the outside, as in the word *p-samu* 'CAUS-nail/prick', according to an informant, Yudaw Pisaw (pers. comm., November 2012). In addition, the prefix *k-* denotes increasing or emphasizing something, as in the word *t-k-kran* 'shake much' in (24a), and the prefix *s-* before a verb signifies the act of augmenting, as in the word *me-s-bitaw* 'hop extendedly' in (24b) or *me-s-pakux* 'turn down greatly' in (26d).

cliff might be strongly influenced by Truku speakers' living environment because they live and work in the mountains.

(15) ANGER IS A NATURAL FORCE

a. *ma* *su* *bruwa* *me-saang* *hug*?
how 2SG thunder AV-angry Q
'How come you are so angry?'
(Lit. 'How come your anger is like the thunder?')

b. *ma* *su* *dowras* *me-saang* *hug*?
how 2SG cliff AV-angry Q
'How come you are so angry?'
(Lit. 'How come your anger is like the cliff?')

4.1.3 INTENSITY OF ANGER IS QUANTITY OF SUBSTANCE IN A CONTAINER VS QUALITY OF A CHARACTER

One striking difference between the two languages in the metaphorical expression of anger is the use of the metaphorical domain of INTENSITY OF ANGER IS QUANTITY (OF SUBSTANCE IN A CONTAINER) in English (e.g., *He is full of anger*), in contrast to INTENSITY OF ANGER IS QUALITY OF A CHARACTER in Truku. Compared to English speakers who do not seem to utilise either mother or father to specify the intensity of an emotion, Truku speakers tend to use the word *bubu* 'mother' to indicate stronger intensity of an emotional concept, and the noun can function as an intensifier modifying the adjective *saang* 'angry' as in (16).

According to Yudaw Pisaw (pers. comm., 2012), the Truku word *bubu* 'mother' is often associated with the strong intensity of a certain emotion such as *bubu msaang* 'very angry', *bubu mhulis* 'laugh loudly', or *bubu mingis* 'cry greatly'.[15]

(16) *bubu* *saang* *ka* *seediq* *ga*.
mother angry NOM person that
'That person is very angry.' (Lit. 'That person's anger is mother-like.')

In contrast, rather than being associated with any type of emotional concept, the Truku word *tama* 'father' can only be attached to the physical body part *baraq* 'lung', in the Truku term for heart, *tama baraq*. Truku is a patriarchal society, and a baby's caretaker is mostly his or her mother during the process of socialization. As Crawford et al. (1992:35) claim, "women's position within the social structures may lead to their constructions of emotions and moral evaluations, which differ from those of men." I suggest that the role of a Truku mother is so intimate, significant, and salient that speakers cannot help but associate emotional concepts with a mother's image. In addition, a mother's emotion in a Truku home can be shown in a passionate or extreme way, so it is conceptually more accessible than the father's image for a Truku speaker's mind to

[15] Two types of expressions are employed when Truku speakers express the intensity of anger. First, they commonly use a general intensifier such as *bi* or *balay* 'very' after the word *saang* 'anger'. Second, the older people tend to use *bubu* 'mother' before the word 'anger' to show their intensity of anger.

associate the intensity of any emotional concept with this specific image (Yudaw Pisaw, pers. comm., November 2012).

4.2 Metonymic expressions of anger

In addition to the metaphorical concepts discussed above, both languages make use of the general metonymic principle that the physiological effects of an emotion stand for the emotion. In fact, many scholars, especially Kövecses and his colleagues, show that emotions are largely understood metaphorically and that the physiological effects of emotion are metonymically related to those emotions.

The relationship between a given emotion and a particular physiological reaction is taken to be causal: Emotion causes a physiological reaction (Radden 2003:430). The mapping between these can be illustrated by the conceptual correspondences in the ANGER IS HEAT metaphor and in the metonymy that links emotions to physiological effects as in Table 1.[16] For example, a heated container maps to an angry individual, heat maps to anger, and so forth.

Table 1: Conceptual correspondences in the ANGER IS HEAT metaphor and in the metonymy linking emotions to physiological effects

SOURCE	TARGET	
"physical events"	"emotions"	"physiology"
container	person	person
heat	anger	body heat
steam	sign of anger	perspiration, redness
explode	show extreme anger	acute shaking, loss of physiological control
boiling point	highest degree of emotion	

The similarities and differences of metonymic expressions of anger in both languages can be discussed in terms of two major categories: (i) increase in body temperature and pressure, and (ii) body parts, agitation, and perception, to which we now turn.

4.2.1 Increase in body temperature and pressure

As many studies of the physiology of anger show, increases in skin temperature and blood pressure are universal physiological correlates of anger **(**Kövecses 2006:171). These effects of anger appear in English metaphors, as in (17)–(19).

English

(17) BODY HEAT STANDS FOR ANGER

a. Don't get hot under the collar.

b. They were having a heated argument.

[16] This table is provided by Turner & Fauconnier (2003:473).

(18) REDNESS IN FACE AND NECK AREA STANDS FOR ANGER
   a. She was scarlet with rage.
   b. He was flushed with anger.

(19) INTERNAL PRESSURE STANDS FOR ANGER
   a. He almost had a hemorrhage.
   b. When I found out, I almost burst a blood vessel.

Although the selection of linguistic expressions for a specific physical experience is different, Truku Seediq apparently observes the same metonymic principle, and these metonymic expressions of anger are found very commonly in everyday use, as in (20) to (22). The differences between the languages can be illustrated by the different phrases in (19)–(22), which both express the consequences of anger caused by inner pressure. The English expression selects *hemorrhage* but the Truku Seediq one chooses *hair smokes*. Nevertheless, the speakers of both languages refer to the same physiological effects of anger.

(20) BODY HEAT STANDS FOR ANGER

m-t-talux	*saang=na*	*ka*	*tama.*
AV-T-hot	angry=3SG.GEN	NOM	father

'(Our) father is very angry.' (Lit. 'Father's (heart) is hot and he feels angry.')

(21) REDNESS IN FACE STANDS FOR ANGER

a.

tayal	*saang*	*ka*	*Ukah*	*wah,*	*banah*	*kana*	*ka*
very	angry	NOM	Ukah	EXC	red	all	NOM

deqeras=na	*da.*
face=3SG.GEN	PCL

'Ukah is so angry that his face is red.'

b.

tayal	*banah*	*birat=su*	*wah*!
very	red	ear=2S.GEN	EXC

'You look very angry.' (Lit. 'Your ears are so red.')

(22) HAIR SMOKING STANDS FOR ANGER

t-qerengul	*kana*	*ka*	*seneunux=su*	*da.*
T-smoke	all	NOM	hair=2SG.GEN	PCL

'You are so angry.' (Lit. 'Your hair smokes.')

4.2.2 Agitation, perception, and body parts

Radden (2003:430) also notices that the relationship between a given emotion and a specific physiological reaction is based on one's observation and other people's behaviors. Therefore, when we see a person shaking or jumping, we may conclude that s/he is angry.

Although expressions in this metonymic domain are based on the same physiological effects of anger, four linguistic expressions for a particular physical experience show differences between the languages. First, as manifested in (23) and (24), Truku Seediq is more inclined to specify body parts in its expressions of AGITATION STANDS FOR ANGER than English is. In the English example in (23), no specific body part is mentioned. No

body part is positively specified in (23a) or (23b). Truku Seediq, however, specifies specific body parts including hair as in (24c) and hand as in (24d).

English

(23) AGITATION STANDS FOR ANGER
a. She was shaking with anger.
b. I was hopping mad.

Truku Seediq

(24) AGITATION STANDS FOR ANGER
a. *tayal* *saang* *ka* *Ukah* *wah,* *t-k-kran* *kana* *ka*
very angry NOM Ukah EXC T-K-shake all NOM
hiyi=na.
body=3SG.GEN
'Ukah is so angry that his body shakes.'

b. *ma* *su* *saw* *me-s-bitaw* *me-saang.*
how 2SG like AV-S-hop AV-angry
'You are hopping mad.' (Lit. 'You are so angry that you hop.')

c. *me-hiyug* *kana* *ka* *seneunux=su* *da.*
AV-stand all NOM hair=2SG.GEN PCL
'You are so angry.' (Lit. 'Your hair is all standing straight up.')

d. *saw* *karit* *me-saang* *ka* *seediq* *ga.*
like tight fist AV-angry NOM person that
'You are so angry that your fist is tight.'

Second, the most remarkable difference between these two languages is that, while English speakers do not seem to have an expression that uses an organ to stand for anger, Truku speakers do, using *pahung* 'gallbladder' to stand for anger as in (25).[17]

(25) A WHOLE BODY PART STANDS FOR ANGER
tayal *pahung=su.*
very gallbladder/bad-tempered=2SG.GEN
'You are so angry.' (Lit. 'Very your gallbladder.')

Third, in contrast to English, in which DISTORTED FACE STANDS FOR ANGER seems to be infrequent, Truku Seediq makes a great use of distorted facial expressions to represent anger including metaphors with *face* as in (26a–c), *mouth* as in (26d), and *whites of eyes* as in (26e).

(26) DISTORTED FACE STANDS FOR ANGER
a. *masu* *panguy/suling* *ta-an* *hug*?
how 2SG distorted see-GV2 Q
'Why are you so angry?' (Lit. 'Why does your face look distorted?')

[17] It is common for Truku speakers to use a body part to refer to an abstract notion. For instance, the word *rumul* 'liver' in the sentence *tayal rumul=su* 'You are so brave.' (Lit. 'You are liver.') stands for bravery.

b. *ma* *su* *saw* *ngasul* *me-saang* *hug*?
how 2SG like distorted-face AV-angry Q
'Why are you so angry?' (Lit. 'Why does your face look distorted?')

c. *saw* *kalih* *me-saang* *ka* *seediq* *ga.*
like fierce AV-angry NOM person that
'That person is fiercely angry.'
(Lit. 'That person looks fierce and feels very angry.')

d. *ma* *me-s-pakux* *quwaq=su*?
how AV-S-turn down mouth=2SG.GEN
'Why are you angry?' (Lit. 'Why is your mouth turned down?')

e. *ma* *su* *me-riqaw* *me-saang*?
how 2SG AV-white eyes AV-angry
'Why are you angry?'
(Lit. 'Why are you so angry that you are showing the whites of your eyes?')

Another notable descriptive difference is using the behavior POURING WATER ONTO FIRE STANDS FOR ANGER to express a person's rage in Truku Seediq. Unlike in English-speaking cultures, in which fire is not commonly used in everyday life, Truku people particularly rely on fire in their mountainous life. Hence, it is very likely that Truku speakers base this linguistic expression on their particular cultural experience as in (27).

(27) POURING WATER ONTO FIRE STANDS FOR ANGER
tblus *saang* *ka* *tama.*
pouring water angry NOM father
'(Our) father is very angry.'
(Lit. 'Our father is very angry as if someone is pouring water onto the fire.')

The last difference I wish to discuss is about employing INSECT'S BITING STANDS FOR ANGER to refer to one's intense anger in Truku Seediq as in (28a) and (28b). Unlike English, in which a bug's or caterpillar's biting does not represent the notion of anger, Truku, whose speakers are agriculturalists who live in mountainous areas and see bugs or caterpillars frequently, tends to use this linguistic expression, which is based on the speakers' specific life experience.

(28) INSECT'S BITING STANDS FOR ANGER

a. *ma* *su* *saw* *kliyus* *me-saang*?
how 2SG like bug AV-angry
'Why are you so angry?' (Lit. 'Why is your anger like bug's (biting)?')

b. *ma* *su* *saw* *pngasuy* *me-saang*?
how 2SG like caterpillar AV-angry
'Why are you so angry?' (Lit. 'Why is your anger like caterpillar's (biting)?')

Despite the differences described so far, both languages observe a similar metonymic principle that describes the notion of anger by referring to its related physiological effects,

like INTERFERENCE WITH ACCURATE PERCEPTION STANDS FOR ANGER as in (29) and (30).

English

(29) INTERFERENCE WITH ACCURATE PERCEPTION STANDS FOR ANGER
a. I was beginning to see red.
b. I was so mad I couldn't see straight.

Truku Seediq

(30) *tayal saang ka tama wah, ana elug ini na*
very angry NOM father EXC even road NEG 3SG.GEN
kela-i m-ita.
know-INF AV-see
'(Our) father is so mad that he is unable to see the road.'

In sum, compared to English, four other selections of linguistic expressions for a particular physical experience are unique in Truku Seediq, including A WHOLE BODY PART STANDS FOR ANGER, DISTORTED FACE STANDS FOR ANGER, POURING WATER ONTO FIRE STANDS FOR ANGER, and INSECT'S BITING STANDS FOR ANGER. Most examples are based on embodiment and specific cultural experience in Truku society, in which hunting is a main activity for men, fire is greatly relied on, and a mountain-dwelling lifestyle is fundamental. This is consistent with Cauquelin's observation (2008:36), where she found that metonymies are used for hunting activities and for ritual objects, and the real meaning of certain concepts is dissimulated by metaphors.

5 Figurative expressions of *happiness* in English and Truku Seediq

Truku Seediq shares several conceptual metaphors for happiness with English, of which five stand out in importance: HAPPY IS LIGHT, HAPPY IS UP, HAPPINESS IS BEING OFF THE GROUND, HAPPINESS IS A FLUID IN A CONTAINER, and HAPPINESS IS VITALITY, to which we now turn.

5.1 Metaphorical expressions of happiness

5.1.1 HAPPY IS LIGHT

On one hand, both languages share the same metaphorical concept of happiness, which is characterised by several source domains including LIGHT as in (31) and (32). Compared to the Truku word *redax* 'bright' in (32a), *durah* in (32b) connotes 'bright and shining', suggesting the difference between these two sentences is one of degree; namely, the intensity of happiness is stronger in (32b) than it is in (32a). In addition, as might be predicted from the existence of the HAPPY IS LIGHT metaphors, both languages conceptualise unhappiness in terms of DARK.[18]

[18] For UNHAPPINESS IS DARK, examples in English and Truku Seediq are as follows.

(31) English
a. She radiates joy.
b. She brightened up at the news.
c. He was gleaming.
d. You are the sunshine of my life.

(32) Truku Seediq
a. *r‹em›edax* *bi* *ta-an* *ka* *deqeras=na.*
bright-AV very see-GV2 NOM face=3SG.GEN
'He is happy.' (Lit. 'His face looks shining.')

b. *durah* *bi* *ta-an* *ka* *deqeras=na.*
radiant very see-GV_2 NOM face=3SG.GEN
'He is happy.' (Lit. 'His face looks radiant.')

5.1.2 HAPPY IS UP

Second, the two languages share the same metaphorical concept of HAPPY IS UP. Nevertheless, whereas the speakers of English have a number of conventionalised lexical expressions that express happiness as going up, especially with the abstract word 'spirit' as in (33b) to (33d), Truku speakers directly use the emotion word *qaras* 'happy' with this metaphorical concept, as in (34).

(33) English
a. We had to cheer him up.
b. My spirits rose.
c. That boosts my spirits.
d. They were in high spirits.

(34) Truku Seediq
me-karaw *bi* *ka* *qaras=na.*
AV-climb very NOM happy=3SG.GEN
'He felt high/happy.' (Lit. 'His happiness climbed up.')

5.1.3 HAPPINESS IS BEING OFF THE GROUND

Next, being off the ground to show happiness is considered desirable and it is positively mapped onto happiness for both languages, unlike in Chinese, whose speakers consider being off the ground to manifest happiness in a way that is negatively out of self-control or lacking steadiness (Yu 1995:74).

(35) English
a. I am six feet off the ground.

(i) a. He feels gloomy.
b. She is in a dark mood.
(ii) *me-kuung* *bi* *ta-an* *ka* *deqeras=na.*
AV-dark very see-GV2 NOM face=3SG.GEN
'He is unhappy.' (Lit. 'His face looks dark.')

b. I was so happy my feet barely touched the ground.

(36) Truku Seediq

me-s-bowlung	*me-qaras*	*ka*	*laqi.*
AV-s-jump	AV-happy	NOM	child

'The child is jumping with happiness.' (Lit. 'The child jumps to show his or her happiness.')

5.1.4 HAPPINESS IS A FLUID IN A CONTAINER

Another major metaphor conceptualizing happiness is HAPPINESS IS A FLUID IN A CONTAINER.[19] Within the limited scope of the examples given here, the emotion word *joy* or *happiness* appears with this metaphor in both languages, but one distinction can be found. That is, although it is apparent that happy feeling can be mapped onto a fluid inside a container in both English and Truku speakers' minds, the fluid can flow from the inside out as in (37a) or from the outside in as in (37b) in English, while it tends to flow out in Truku expressions, as in (38). Moreover, an indication of the degree of intensity can be detected here; compared to the expression with the noun *lingis* 'tear' in (38a), the expression with *dengu* 'dry' in (38b) implies that one is even happier, so happy that the bodily fluids, including tears, are used up.

(37) English

a. He was overflowing with joy.
b. His heart is filled with happiness.

(38) Truku Seediq

a.

l‹em›ingis	*me-qaras*	*ka*	*seediq*	*ga.*
<AV>cry	AV-happy	NOM	person	that

'That person is so happy.'
(Lit. 'That person is crying with joy/happiness.')

b.

me-dengu	*me-qaras*	*ka*	*seediq*	*ga.*
AV-dry	AV-happy	NOM	person	that

'That person is so happy.'
(Lit. 'That person is so happy that his or her body fluid is dry.')

5.1.5 HAPPINESS IS VITALITY

The concept of happiness in both languages is also characterised by some limited source domains like vitality. While in English the emotion word *joy* is directly expressed as in (39), in Truku Seediq, a common behavioral reaction to emotion, a feeling of something beating, is observed in the description of happiness in (40).

(39) English

He was alive with joy.

---

[19] Although emotions like anger and happiness seem to be FLUID IN A CONTAINER in Truku Seediq, this metaphor does not seem to be applied to other basic emotions such as fear or sadness (Yudaw Pisaw, pers. comm., January 2013).

(40) Truku Seediq
s-bowlung bi ka lenglung-an=mu.[20]
S-jump very NOM think-GV2=1SG.GEN
'I am very happy.' (Lit. 'My emotion is beating hard.')

5.1.6 INTENSITY OF HAPPINESS IS QUALITY OF A CHARACTER

As noted earlier, Truku speakers tend to use the word *bubu* 'mother', but not *tama* 'father', to indicate stronger intensity of any type of emotional concept, including happiness, and the noun 'mother' functions as an intensifier modifying *qaras* 'happy' as in (41).

(41) Truku Seediq
bubu me-qaras ka seediq ga.
mother AV-happy NOM person that
'That person is very happy.' (Lit. 'That person's happiness is like the happiness of a mother.')

5.2 Metonymic expressions of happiness

Kövecses (1991) observes that some behavioral responses are associated with the emotion of happiness and they are metonymic. Although the same metonymic principles can be observed in both languages examined here, there are two descriptive differences between them. One has to do with behavioral reactions, and the other with changes of body status, posture, or shape, as discussed in the following subsections.

5.2.1 Behavioral reactions (mild happiness)

In general, metaphors with behavioral reactions such as jumping, dancing, smiling, and bright eyes that are associated with the emotion of happiness can be observed both in English, as in (42)–(45), and Truku Seediq as in (46)–(50). However, unlike English, in which bright eyes are regarded as expressing a notion of happiness as in (45), Truku Seediq tends to consider bright eyes as an indicator of beautiful or healthy appearance as in (49).

English

(42) JUMPING STANDS FOR HAPPINESS
a. He jumped for joy.
b. He was leaping with joy.

(43) DANCING STANDS FOR HAPPINESS
a. We were dancing with joy.
b. They kicked up their heels.

(44) SMILING STANDS FOR HAPPINESS
a. She was smiling with happiness.

[20] This Truku sentence is from Lowking Nowbung's field notes (27 June 2007).

b. He was all teeth.

(45) REACTIONS IN EYES
a. His eyes were shining.
b. His eyes were sparkling like diamonds.

Truku Seediq

(46) JUMPING STANDS FOR HAPPINESS
saw sbowlung me-qaras ka laqi.
like jump AV-happy NOM child
'The child is jumping with happiness.'
(Lit. 'The child is so happy that she jumps.')

(47) DANCING STANDS FOR HAPPINESS
tayal qaras ka laqi ga, pesa-un=na baraw ka qaqay= na.
very happy NOM child that put-GV1=3SG up NOM foot=3SG.GEN
'The child is so happy.'
(Lit. 'The child put his foot up to show his or her happiness.')

(48) SMILING STANDS FOR HAPPINESS
hulis me-qaras ka laqi.
smile/laugh AV-happy NOM child
'The child smiles with happiness.' (Lit. 'The child is so happy that s/he smiles.')

(49) REACTIONS IN EYES
r‹em›edax bi ta-an dowriq=su da.
<AV>bright very see-GV2 eye=2SG PCL
'You are better now than you were before.' = speaker's feeling of happiness after seeing someone's bright eyes (Lit. 'Your eyes look very bright.')

(50) CHEERING STANDS FOR HAPPINESS
me-jiras me-qaras ka seediq niyi.
AV-cheer AV-happy NOM person this
'This person is so happy.' (Lit. 'This person is so happy that s/he cheers.')

5.2.2 Change of body state, posture, or shape (intense happiness)

In contrast to English, Truku Seediq tends to utilise changes of physical state, posture, or shape to express the intensity of happiness. If one is extremely happy, one's breathing can be stopped as in (51), one's body posture can change as in (52), and one's body shape can be transformed as in (53). These expressions of happiness are apparently rooted in common bodily experience and reinforce the observation, made in the discussion of metonymic expressions of anger, that Truku Seediq tends to use more visible facial or bodily experience in describing notions of emotion.

(51) STOPPING BREATHING STANDS FOR HAPPINESS
stuq qaras ka seediq ni.
stop happy NOM person this
'This person is so happy.'
(Lit. 'This person is so happy that s/he almost stops breathing.')

(52) THE BACK OF ONE'S HEAD TOUCHING THE GROUND STANDS FOR HAPPINESS

t-qilu	*me-qaras*	*ka*	*seediq*	*ni.*
T-the back of the head	AV-happy	NOM	person	this

'This person is so happy.'
(Lit. 'This person is so happy that the back of his/her head touches the ground.')[21]

(53) FLATTENED BODY SHAPE STANDS FOR HAPPINESS

tg-tmaq	*me-qaras*	*ka*	*seediq*	*ni.*
ART-press flat	AV-happy	NOM	person	this

'This person is so happy.'
(Lit. 'The person is so happy that his or her body is almost pressed flat.')[22]

6 Discussion

In sum, a descriptive difference between English and Truku Seediq that is apparent is that Truku tends to utilise more body parts than English does in depicting anger and happiness. The figurative expressions of anger are shown in Table 2. Overall, while both languages are similar in that they use a number of metaphors for emotional concepts, Truku Seediq tends to have more metonymic expressions than English.

Table 2: Figurative expressions of *anger* in English and Truku Seediq

Metaphor	English	Truku Seediq
Anger is heat	(1) Anger is fire	(1) Anger is fire
	(2) Anger is a hot fluid in a container	(2) Anger is a bodily fluid in a container
Anger is danger		(3) Anger is gas in a container
	(1) Anger is aggressive animal behavior	(1) Anger is aggressive animal behavior
Intensity of anger	(2) Anger is a natural force	(2) Anger is a natural force
	(3) Anger is a captive animal	
	Intensity of anger is quantity (of substance in a container)	Intensity of anger is quality of a character

[21] According to the informant Yudaw Pisaw, if a person laughs very hard because of happiness, his or her whole body will lie flat on the ground, so that the lower back of the head touches the ground.

[22] Similar to THE BACK OF ONE'S HEAD TOUCHING THE GROUND STANDS FOR HAPPINESS; if a person laughs very hard because of happiness, he or she cannot stand up but lies flat on the ground, which consumes or exhausts his or her energy, so that the body is almost pressed flat.

Metaphor	English	Truku Seediq
Metonymy	(1) Body heat stands for anger (2) Redness in face and neck area stands for anger (3) Internal pressure stands for anger (4) Agitation stands for anger (5) Interference with accurate perception stands for anger	(1) Body heat stands for anger (2) Redness in face stands for anger (3) Hair smoking stands for anger (4) Agitation stands for anger (5) Interference with accurate perception stands for anger (6) A whole body part stands for anger (7) Distorted face stands for anger (8) Pouring water onto fire stands for anger (9) Insect's biting stands for anger

Similar to their expressions for anger, both languages have almost equivalent amounts of metaphorical expressions for happiness; nevertheless, Truku Seediq has more metonymic expressions than English does, as shown in Table 3.

Table 3: Figurative expressions of *happiness* in English and Truku Seediq

Metaphor	English	Truku Seediq
	(1) Happy is light (2) Happy is up (3) Happiness is being off the ground (4) Happiness is a fluid in a container (5) Happiness is vitality (6) Jumping stands for happiness	(1) Happy is light (2) Happy is up (3) Happiness is being off the ground (4) Happiness is a fluid in a container (5) Happiness is vitality (6) Jumping stands for happiness (7) Intensity of happiness is quality of a character
Metonymy	(1) Dancing stands for happiness (2) Smiling stands for happiness (3) Reactions in eyes	(1) Dancing stands for happiness (2) Smiling stands for happiness (3) Reactions in eyes (4) Cheering stands for happiness (5) Stopping breathing stands for happiness (6) The back of one's head touching the ground stands for happiness (7) Flattened body shape stands for happiness

Five main points can be supported by the data presented and discussed in this paper. These points are based on Kövecses' views on universality and variation in metaphor (2005:292–294). First, data from both languages confirm that the variation of metaphors is based on bodily experience, social and cultural experience, and cognitive preferences and styles. Take the ANGER IS A HOT FLUID IN A CONTAINER metaphor in both languages as an example. Metaphors derived from both languages are coherent with their bodily experience of anger and social-cultural experience. This coherence derives from the fact

that Truku Seediq utilises more heat-related words than English because Truku speakers depend greatly on fire in daily life; the detail and variety of fire-related metaphorical expressions suggest that these are culturally constrained. Their metaphorical use is coherent within a particular cognitive system, one that has a preferential experiential basis for understanding the components of both heat and pressure. And although English prefers ANGER IS A *HOT* FLUID IN A CONTAINER and Truku Seediq prefers ANGER IS A *BODILY* FLUID IN A CONTAINER, both of these metaphorical domains belong to conceptual systems in which heat far outweighs pressure in the conceptualization of anger.

Second, the central metaphors like ANGER IS HEAT and ANGER IS DANGER are all at a general level in both languages, but the variation observed between the languages is culture-specific. For example, unlike English in the use of the ANGER IS HEAT metaphor, Truku Seediq uses a number of different fire-related words to express the intensity of anger. The data support that a generic-level conceptual metaphor is instantiated in culture-specific ways at a specific level.

Third, embodiment itself can be of two kinds. First, metaphorical conceptualization can be on the basis of simple physical reactions or effects, such as increased pressure in blood vessels or the head, as found in both languages' expressions. The second type is purely based on experience, and that is just as much cultural as it is physical, such as in the metaphors where INSECT'S BITING or POURING WATER ONTO FIRE STANDS FOR ANGER.

Next, the data also suggest that the human mind is equally the product of embodiment and culture. In other words, the study of human cognition should be based on observations of both embodiment and culture. Because of their specific lifestyles and living experience, Truku speakers make a greater use than English speakers of expressions that are associated with fire, facial expressions, mountain-dwelling, and bodily fluids in their figurative expressions. Further, using *bubu* 'mother' to indicate the stronger intensity of any type of emotion makes sense in a way that is specific to Truku society, based on the roles and behaviors expected of mothers in this society. In addition, speakers select the words that are most accessible to them because of their specific life experiences, so, for example, Truku's hunting society evokes dogs and bears to express anger.

Last, Truku Seediq tends to utilise more of both internal and external bodily experiences than English in conceptualizing anger and happiness. Such metaphors in both languages are coherent with certain physical aspects of the human body (Kövecses 2005:285). Nevertheless, Truku and English speakers utilise bodily experiences in distinct ways; Truku may be more attuned to different aspects of their bodily functioning in relation to a target domain. For example, FLATTENED BODY SHAPE or STOPPING BREATHING metaphors are selected for conceptualizing happiness, perhaps because Truku can be so expressive that they utilise the whole body to experience the notion of happiness.

7 Conclusion

The main purpose of this paper is to compare metaphorical expressions of anger and happiness in Truku Seediq and English. It shows that Truku Seediq like English has

many emotion nouns, and supports the argument that languages with this variety of choice may also provide language users more metaphorical strategies for conceptualizing emotion events.

The data suggest that three main components—bodily experience, social-cultural experience, and cognitive preferences and styles—play crucial roles in accounting for both the universality and the variation of metaphors. It is too simplistic to suggest that the expression of emotion concepts is entirely based on either human physiology or social construction. Rather, both spheres motivate emotion expressions. Nevertheless, this paper can only be a very modest beginning in dealing with the issue of variation in metaphorical use between English and Truku Seediq. The universality and variation in figurative expressions in the Formosan languages and cultures and within individual languages and cultures are significant, and their in-depth investigation and comparison are worthwhile topics of future research.

References

Blust, Robert. 2009. *The Austronesian languages*. Pacific Linguistics 602. Canberra: The Australian National University. [2013 revised edition online at: http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf]

Bokor, Zsuzsanna. 1997. *Body-based constructionism in the conceptualization of anger*. C.L.E.A.R. Series 17. Budapest: Department of English, Hamburg University and the Department of American Studies, ELTE.

Cauquelin, Josiane. 2008. *Ritual texts of the last traditional practitioners of Nanwang Puyuma*. Language and Linguistics Monograph Series A-23. Taipei: Academia Sinica.

Chiang, Wen-yu and Fang-mei Chiang. 2005. Vowel dispersion in Truku. *Concentric: Studies in Linguistics* 31.1:21–45.

Crawford, June, Susan Kippax, Jenny Onyx, Una Gault and Pam Benton. 1992. *Emotion and gender: Constructing meaning from memory*. Newbury Park, CA: SAGE Publications.

Gibbs, Raymond W. Jr. 1994. *The poetics of mind: Figurative thought, language, and understanding*. New York: Cambridge University Press.

Holland, Dorothy and Andrew Kipnis. 1995. American cultural models of embarrassment: The not-so-egocentric self laid bare. In *Everyday conceptions of emotion: An introduction to the psychology, anthropology, and linguistics of emotion*, ed. by José-Miguel Fernández-Dols, Antony S.R. Manstead and J.C. Wellenkamp, 181–202. NATO ASI series, Series D. Behavorial and social sciences No. 81. Dordrecht: Kluwer Academic Publishers.

Hsieh, Fuhui. 2007. Language of emotion and thinking in Kavalan and Saisiyat. PhD dissertation. Taipei: National University of Taiwan.

Hsieh, Fuhui. 2011. The conceptualization of emotion events in three Formosan languages: Kavalan, Paiwan, and Saisiyat. *Oceanic Linguistics* 50.1:65–92.

Izard, Carroll E. 1977. *Human emotions*. New York: Plenum Press.

Kövecses, Zoltán. 1990. *Emotion concepts*. New York: Springer-Verlag.

Kövecses, Zoltán. 1991. Happiness: A definitional effort. *Metaphor and Symbolic Activity* 6.1:29–46.

Kövecses, Zoltán. 2000. *Metaphor and emotion: Language, culture, and body in human feeling*. Studies in emotion and social interaction. Cambridge and New York: Cambridge University Press; Paris: Editions de la Maison des sciences de l'homme.

Kövecses, Zoltán. 2005. *Metaphor in culture: Universality and variation*. Cambridge: Cambridge University Press.

Kövecses, Zoltán. 2006. *Language, mind, and culture: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Zoltán, Gary B. Palmer and René Dirven. 2003. Language and emotion: the interplay of conceptualisation with physiology and culture. In *Metaphor and metonymy in comparison and contrast*, ed. by René Dirven and Ralf Pörings, 133–159. Berlin and New York: Mouton de Gruyter. 2nd edition.

Lakoff, George. 1987. *Women, fire, and dangerous things*. Chicago and London: University of Chicago Press.

Lakoff, George. 1993. The contemporary theory of metaphor. In *Metaphor and thought*, ed. by Andrew Ortony, 202–251. Cambridge: Cambridge University Press. 2nd edition

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by: What categories reveal about the mind*. Chicago and London: University of Chicago Press.

Lakoff, George and Zoltán Kövecses. 1987. The cognitive model of anger inherent in American English. In *Cultural models in language and thought*, ed. by Dorothy Holland and Naomi Quinn, 195–221. New York: Cambridge University Press.

LeDoux, Joseph. 1996. *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.

Lee, Amy Pei-jung. 2009. Dorsal consonant harmony in Truku Seediq. *Language and Linguistic*s 10.3:569–591.

Lee, Amy Pei-jung. 2010. Phonology in Truku Seediq. *Taiwan Journal of Indigenous Studies* 3.3:123–168.

Lee, Amy Pei-jung. 2011. Metaphorical euphemisms of RELATIONSHIP and DEATH in Kavalan, Paiwan, and Seediq. *Oceanic Linguistics* 50.2:351–379.

Lin, Dong-yi. 2006. The language of emotion in Kavalan. MA thesis. Taipei: National Taiwan University.

Lutz, Catherine A. 1988. *Unnatural emotions: everyday sentiments on a Micronesian atoll and their challenge to Western theory*. Chicago: University of Chicago Press.

Matsuki, Keiko. 1995. Metaphors of anger in Japanese. In *Language and the cognitive construal of the world*, ed. by John R. Tayor and Robert E. MacLaury, 137–151. Trends in Linguistics: Studies and Monographs 82. Berlin and New York: Mouton de Gruyter.

Munro, Pamela. 1991. ANGER is HEAT: Some data for a cross-linguistic survey. Manuscript, Department of Linguistics, UCLA.

Nowmaw, Pusi, and Yudaw Pisaw. 2007. *Kari Truku* [*Truku language*]. Hualien: Council of Indigenous People. [In Truku and Chinese].

Pisaw, Yudaw, Iyang Taying and Iyuq Ciyang. 2005. *Patas Suyang Kari Truku* [*Truku Bible*]. Taipei: The Bible Society in Taiwan.

Radden, Günter. 2003. How metonymic are metaphors? In *Metaphor and metonymy in comparison and contrast*, ed. by René Dirven and Ralf Pörings, 407–434. Berlin and New York: Mouton de Gruyter. 2nd edition.

Tang, Apay Ai-yu. 2010. The interpretation and transitivity of *-un*/*-an* in Truku. *University of Hawai'i at Mānoa Working Papers in Linguistics* 41.2:1–15.

Tang, Apay Ai-yu. 2011. From diagnosis to remedial plan: A psycholinguistic assessment of language shift, L1 proficiency, and language planning in Truku Seediq. PhD dissertation. Honolulu: University of Hawai‘i at Mānoa.

Taylor, John R. and Thandi G. Mbense.1998. Red dogs and rotten mealies: How Zulus talk about anger. In *Speaking of emotions: Conceptualisation and expression*, ed. by Angeliki Athanasiadou and Elżbieta Tabakowska, 191–226. Cognitive Linguistics Research, No 10. Berlin and New York: Mouton de Gruyter.

Truku e-dictionary. 2013. Hualien: Council of Indigenous People. Retrieved from http://e-dictionary.apc.gov.tw/trv/Search.htm

Tsukida, Naomi. 2005. Seediq. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar, and Nikolaus P. Himmelmann, 291–325. Routledge Language Family Series. London and New York: Routledge.

Turner, Mark, and Gilles Fauconnier. 2003. Metaphor, metonymy, and binding. In *Metaphor and metonymy in comparison and contrast*, ed. by René Dirven, and Ralf Pörings, 469–487. Berlin and New York: Mouton de Gruyter. Second edition.

Yeh, Maya Yuting. 2002. Emotion concepts in Squliq Atayal. MA thesis. Taipei: National Taiwan University.

Yu, Ning. 1995. Metaphorical expressions of anger and happiness in English and Chinese. *Metaphor and Symbolic Activity* 10.2:59–92.

21 *The lexical system of Yilan Creole**

YUEHCHEN CHIEN

1 Introduction

1.1 Yilan Creole

As a result of Japan's pre–World War II colonial expansion, language contact between the Japanese language and local languages in Taiwan occurred over a fifty-year period (1895–1945), and consequently is responsible for various types of language change. For example, contact between Japanese and two Austronesian languages, Atayal and Seediq, produced a little-known creole.

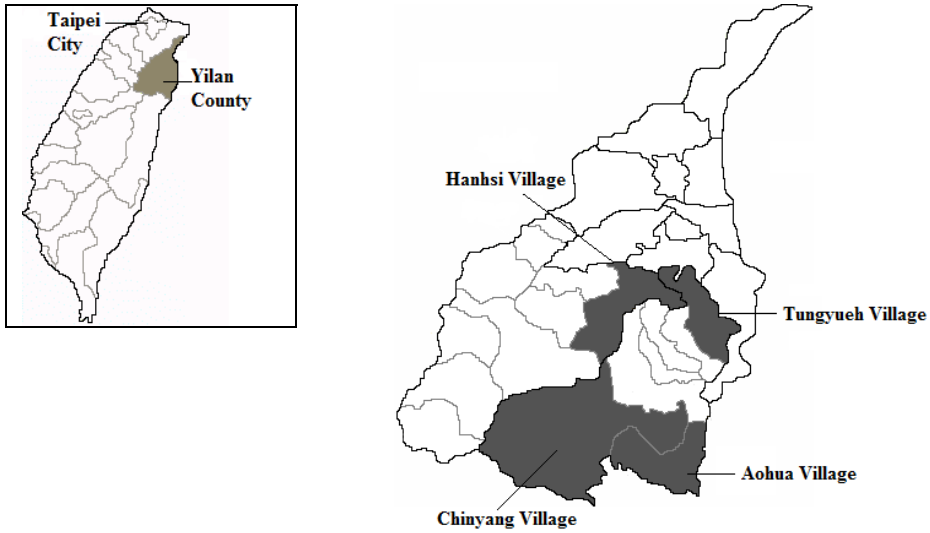
This creole is spoken by Atayal and Seediq people living in four villages of Yilan County, Taiwan.[1] They are Tungyueh Village, Chinyang Village, Aohua Village, and Hanhsi Village, as illustrated in Map 1.

Chien & Sanada (2010a, 2010b) named the creole "Yilan Creole" for the place where it is spoken. It is supposed that Yilan Creole has been used as a first language for the Atayal and Seediq people since the 1930s. Nowadays, in Tungyueh Village, most of those born from the 1980s onward can't speak Yilan Creole, while in Aohua Village we found young fluent speakers born in the 1990s. Although there is variation in the level of endangerment between the four villages, Yilan Creole is threatened by the dominant language, Mandarin.

* My thanks to L.M. Huang for her research on the creole used in Hanhsi Village. I am grateful for the generosity and hospitality of the consultants from Tungyueh Village for the data of this paper. The constructive suggestions from the three anonymous reviewers are deeply appreciated. This paper also benefited from helpful comments from the editors, Elizabeth Zeitoun, Stacy F. Teng, and Joy Wu. This paper was partly funded by the Ministry of Science and Technology, Taiwan.

[1] Atayalic is subdivided into Atayal and Seediq. Two Atayal dialects (Squliq Atayal and C'uli' Atayal) and one Seediq dialect (Tosa Seediq) were spoken in the four villages.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, *513–532*
Asia-Pacific Linguistics, 2015



Map 1: Distribution of Yilan Creole in Yilan County, Taiwan

The population of the four villages is about 3,000, and the number of speakers of Yilan Creole is unknown. Except for the older people, who cannot speak Mandarin fluently, most speakers of Yilan Creole are bilingual in Yilan Creole and Mandarin.[2]

Yilan Creole has nineteen consonants /p, t, d, k, ʔ, ts, ß, s, z, x, ɣ, h, m, n, ŋ, l, r, w, j/ and five vowels, /i, e, a, o, u/. The sound system retains Atayal features. However, the consonant /d/ is due to the influence of Japanese. Both Japanese and Atayal have the same five vowels, but the /u/ is an unrounded [ɯ] in standard Japanese, while it is rounded in Atayal. Yilan Creole's /u/ is rounded like that of Atayal.

Yilan Creole has SOV as its basic word order, like the superstrate language, Japanese. I have also observed some SVO word order sentences, especially in younger generations' speech, which can be considered a result of the influence of the adstrate language, Mandarin, or a universal tendency of creoles.

1.2 Research on Yilan Creole

Yilan Creole has been used for more than a century, but almost no investigation of it has been carried out. The existence of this creole is not mentioned even in the article Ehrhart & Mühlhäusler (2007), which focuses on pidgins and creoles in the Pacific.

Sanada & Chien (2007) first outlined the sociolinguistic environment and described the syntactic structure of this creole. Since then, Sanada & Chien (2008a, 2008b) have dealt with various aspects of Yilan Creole's origin, use, and linguistic properties. Further works by Chien & Sanada (2010a, 2010b) discuss the sociohistorical background and the structural characteristics to establish that it is indeed a creole.

Yilan Creole has caught the attention of other linguists in recent years. Abe et al. (2008) discussed words of Japanese origin in the variety used in Hanhsi Village based on an analysis of a textbook. Otani & Huang's (2009) study concerned the Japanese impact on phonology in the Hanhsi Village variety. Chien & Sanada (2011) examined the

[2] Some Yilan Creole speakers can speak Southern Min, too.

negation system of Yilan Creole in Tungyueh Village. These studies contribute to our knowledge of Yilan Creole. However, the study of Yilan Creole has only just begun; its structure has not yet been fully investigated.

As Roberts (1999:317) noted:

> Creoles are particularly interesting because they represent an extreme of language change, but it is the mechanisms of language change, which are ubiquitous in the history of every language and every language family, that have made creoles what they are.

Many theories of language change based on the study of creoles have been proposed, for example, the language bioprogram hypothesis (Bickerton 1984), the relexification hypothesis (Lefebvre 1986), and the founder principle (Mufwene 1996), among others. However, the main problem with these theories is that the creoles they consider all descend from Indo-European lexifier languages and emerged under similar circumstances. In order to elucidate the mechanisms of language change precisely, dealing with the non-Indo-European creoles is absolutely necessary.

Yilan Creole is the only known Japanese-lexicon creole in the world. For this reason alone, it is interesting and important to clarify the similarities and differences between Yilan Creole and the other creoles in the world.

This paper aims to describe the lexical system of Yilan Creole. It begins by giving a brief account of the data collection (section 2). Section 3, on basic vocabulary, gives a comprehensive account of the sources of lexical items in Yilan Creole. In section 4, on word formation processes, derivation as well as compounding and clipping are discussed. Semantic change is the subject of section 5, which looks at how the three major sources of the Yilan Creole lexicon—retentions from the substrate language, input from the superstrate language, and internal innovations—have interacted.

2 Data

The data come from fieldwork conducted by a team from Dong Hwa University, Taiwan, consisting of myself, two of my students, and Professor Shinji Sanada of Nara University, Japan. The student team members are Habaw Watan, who is a native speaker of Yilan Creole, and Ipay Yuraw, who is an Atayal native speaker. Data for this paper were collected during fieldwork that took place on several short visits between 2008 and 2010.

We have found clear differences between the four villages' varieties of Yilan Creole. This paper will focus on one village, and a cross-dialectal comparison will be made in our further work.

The data presented below come from consultations with residents in Tungyueh Village. The consultants for this paper are five native speakers of Yilan Creole born between 1947 and 1974.[3] All were born and brought up in Tungyueh Village, and they are bilingual speakers of Yilan Creole and Mandarin. None of them were educated in Japanese.

---

[3] The five consultants include a 1947-born female, a 1951-born male, a 1964-born male, a 1966-born female, and a 1974-born female.

Our elicitation list is based on the Swadesh list of basic vocabulary,[4] with the addition of other related words. We interviewed the consultants in Chinese and asked them to answer in Yilan Creole.

The Atayal language has two major subgroups: Squliq and C'uli'. We have been investigating Yilan Cyoli Atayal, a C'uli' dialect spoken in Yilan County, for our baseline data. We chose a nearby village, Chinyueh Village, to collect data on Yilan Cyoli Atayal.

Atayal and Yilan Creole examples are presented in the Atayal writing system promulgated by the Council of Indigenous Peoples, Taiwan since 15 December 2005. Glottal stop is marked by an apostrophe. For instance, /laʔi/ is written *la'i*.

3 Basic vocabulary of Yilan Creole

This section looks at the composition of the lexicon.

3.1 The sources of lexical items

Table 1 presents a summary of the sources of the Yilan Creole words for 219 items in the Swadesh list of basic vocabulary. In Tables 1 and 2, columns headed by more than one language name indicate that more than one Yilan Creole vocabulary item exists for the Swadesh list items in that category. For example, the third column is "Atayal & Japanese." In this category are seventy-four items to which Yilan Creole can refer by using either of two terms, one from Atayal and one from Japanese.

As can be seen in Table 1, the proportion of Atayal-derived[5] words in the Yilan Creole basic vocabulary is 18.3%, and that of Japanese-derived words is 35.6%.[6] Both Atayal-derived words and Japanese-derived words can be used for 33.8% of the items.[7] There are a few Mandarin-derived and Southern Min–derived words as well. Japanese is the basis for the lexicon and can be called the "lexifier language" for Yilan Creole. Atayal is the "substrate language," and Mandarin and Southern Min are "adstrate languages."

---

4 The Swadesh list we used is from Kamei et al. (1996). For the semantic fields, we follow Tsuchida (2000).

5 In creole linguistics (creolistics), the term "derived" is used with a different meaning than in other branches of linguistics (e.g., historical linguistics). "Derived" in creole studies describes the source language for the word in question. For instance, "French-derived vocabulary" in a creole language comes from French.

6 Atayal-derived words are, for example, *muyox* 'blood', *puga* 'navel', and *helu* 'smoke'. Japanese-derived words are, for example, *isi* 'stone', *hana* 'flower', and *omoy* 'heavy'.

7 An example of this category (in which Atayal-derived and Japanese-derived words are used as variant ways to express the same meaning), is 'rain', which can be expressed by Atayal-derived *walax* as well as Japanese-derived *ame* in Yilan Creole. Some variants are intra-individual and others inter-individual.

Table 1: Sources of Yilan Creole basic vocabulary

Origin	Atayal	Japanese	Atayal & Japanese	Atayal & Mandarin	Atayal & Japanese & Mandarin	Japanese & Mandarin	Japanese & Southern Min	Japanese & Southern Min & Mandarin	Total
Yilan Creole	40	78	74	6	8	11	1	1	219
%	18.3%	35.6%	33.8%	2.7%	3.6%	5.0%	0.5%	0.5%	100.0%

3.2 Distribution by semantic field

As noted above, Yilan Creole has its lexical base mostly in Japanese. But the substrate language, Atayal, contributes to the lexicon as well. This subsection looks at the distribution of word source and semantic field.

As Table 2 shows, both Atayal-derived words and Japanese-derived words are used in almost all semantic fields, except for in the fields of temporal words, pronouns, and others (mostly function words), which mainly use Japanese-derived words.

It is noteworthy that Atayal contributed most of the words for body parts, which has been said to be an area of the lexicon that shows both particular loss and particular resistance to loss in distinct situations of language death (Craig 1997:262).

The other semantic fields that retain many terms from the substrate language are nature, plants, animals, verbs, and adjectives. In the field of nature, the Atayal-derived words *puni* 'fire', *helu* 'smoke', *bulit* 'ash', *urow* 'earth', and *timu* 'salt' are used. Fire plays a very important role in Atayal and Seediq traditional life, and Atayal words related to fire survive in Yilan Creole. In the field of plants, the Atayal-derived words *koni* 'tree' and *yula* 'bark' have survived. For verbs, the Atayal-derived words *luyak* 'hunt' and *laka* 'fly' are retained. These words describe concepts closely associated with Atayal and Seediq traditional life. To sum up, Atayal-derived words for culturally relevant things often survived in the creole.

Also interesting is that all words that describe parts of animals' bodies come from Atayal, such as *ngungu* 'tail', *paleh* 'wing', and *kumus* 'fur, feather'. In contrast, animal names either have both Atayal- and Japanese-derived variants or only a Japanese-derived term. For example, creole speakers express 'dog' with Atayal-derived *hoying* as well as Japanese-derived *inu*, and 'snake' with Japanese-derived *hebi*.

Table 2: Semantic fields and contributing languages

Origin / Semantic field	Atayal	Japanese	Atayal & Japanese	Atayal & Mandarin	Atayal & Japanese & Mandarin	Japanese & Mandarin	Japanese & Southern Min	Japanese & Southern Min & Mandarin	Total
Body parts	16	4	8						28
Culture words	1	1	1						3
Persons			6		2				8
Nature	5	10	6		3	3			27
Plants	2	2	2	1	1	1		1	10
Animals	3	3	5	2		1			14
Verbs	7	25	17		2	2			53
Adjectives	6	9	22	3		2			42
Space			2			2			4
Temporal		3							3
Numerals			5						5
Pronouns		14					1		15
Others		7							7
Total	40	78	74	6	8	11	1	1	219

It is noteworthy that, as Table 2 indicates, there are a large number of words (seventy-four) for which both Japanese and Atayal variants co-exist. For instance, the Atayal-derived *cyoli* and the Japanese-derived *ninggen* are both used to mean 'person', and the Atayal-derived *ya'eh* and the Japanese-derived *waruy* are both used to mean 'bad'. Some of the variants appear in free variation, and others have a complementary distribution (see section 5.2.3).

Comparing two older speakers' and a younger speaker's[8] word choices in this category, I found that they all use both Atayal- and Japanese-derived words for twenty-four of the seventy-four items. For twenty-four more items, the older speakers use both Atayal-derived and Japanese-derived words, while the younger speaker uses only Atayal-derived words. And there are twenty items for which the older speakers use both Atayal- and Japanese-derived words, while the younger speaker uses only Japanese-derived words. Although I was unable to find any regular pattern based on semantic field in their usage, it is clear that the older speakers use more Atayal and Japanese variants than the younger speaker. This generational difference seems to indicate that available variants decreased and the system of Yilan Creole became more stable between the times that the older speakers and the younger speaker acquired the language.

It is also notable that, to a lesser extent, some Mandarin-derived words are employed in all fields except for body parts, culture words, temporal words, numerals, and others. When a Mandarin variant exists, it tends to be the younger speakers who use it. For example, older speakers use Atayal-derived *puying* 'root', while younger speakers use Mandarin-derived *ken* 'root'. And older speakers use both Atayal-derived *abow* and Japanese-derived *hapa* to mean 'leaf', while younger speakers use Mandarin-derived *yehtzu*. This indicates that the influence of Mandarin is increasing. Note that Mandarin-derived words become non-tonal in Yilan Creole.

3.3 Survival of dialectal usages

Yilan Creole also preserves some words of Atayal and Japanese dialectal origins.

First, Yilan Creole's substrate is considered to be the Yilan Cyoli dialect. But it also contains forms drawn from the Squliq dialect, although the number is small. For example, 'house' in Yilan Creole is *ngasang*, derived from the Squliq dialect, instead of *sali* from the Cyoli dialect. According to Utsushikawa et al. (1935), in 1931, the population of the Atayal Nan-ao group[9] was 1,692. Squliq people made up 44.4% of the population, while Cyoli people made up 41.3%, and Tosa people, 14.3%.[10] The different groups had settled in Nan-ao as early as the 18th, and the social setting and circumstances led to dialect contact as well as language contact.

Second, western Japanese dialectal forms have also been preserved in Yilan Creole. Some examples are *taru* 'enough', *oru* 'exist', *nasubi* 'eggplant', *yasay taku* 'cook

---

[8] The older speakers are a 1947-born female and a 1951-born male. The younger speaker is a 1974-born female.

[9] Nan-ao is in the southern part of Yilan County. Several Atayal groups that live in the Nan-ao area, including Squliq, Cyoli, and Tosa people, are together called the Nan-ao group.

[10] Squliq and Cyoli are Atayal dialects, and Tosa is a Seediq dialect.

vegetables', *cigo* (negative form), and *-ng* (negative affix).[11] These words can be attributed to the migration of many settlers, including policemen, teachers, government employees, and peasants, from western Japan during 1895–1945.

During Japan's occupation of Taiwan, immigrants made up 5% of Taiwan's population; 70% of the immigrants came from western regions in Japan (Chien 2006). Standard Japanese was widely spoken, but residents from the western regions in Japan used regional components unconsciously when they spoke standard Japanese.[12] As Nishioka (1936) and Saito (1943) note, this speech style had an influence in Taiwan.

As Chien (2006) points out, in the early 2000s, most members of the older generations—people over age seventy—in Taiwan had been educated in Japanese using this speech style. Yilan Creole has thus inherited western Japanese regional features.

In at least one case, a Japanese dialectal variant in contrast with a standard Japanese variant has come to express a grammatical distinction in Yilan Creole. Chien and Sanada (2011) explain that in Yilan Creole, the Japanese-derived negatives *-nay* (standard Japanese form) and *-ng* (western Japanese dialectal form) express realis and irrealis. This aspect of the negation system of Yilan Creole demonstrates a restructuring of Japanese forms to fit Atayal grammatical categories.

4 The formation processes of lexical items of Yilan Creole

The main processes of word formation in Yilan Creole include derivation, compounding, and clipping.

Reduplication is said to be a productive mechanism for word formation in many creoles (Hancock 1980; Holm 2000). However, it is not observed in Yilan Creole.[13]

4.1 Derivation

Yilan Creole has two derivational morphemes that form new words. The following discussion deals with the two affixes *-suru*, a verbal suffix (originally meaning 'to do' in Japanese), and *-rasyeru*, indicating causativity.

4.1.1 The Verbal Suffix *-suru*

In Yilan Creole, the suffix *-suru* is productively used to create verbs. This process is considered to be drawn from the superstrate language.

---

[11] The standard Japanese forms are *tariru* 'enough', *iru* 'exist', *nasu* 'eggplant', *yasai o niru* 'cook vegetables', *janai* (negative form), and *-nai* (negative affix).

[12] Even nowadays, many people in western Japanese regions use regional components unconsciously when they speak standard Japanese. It seems that some dialectal forms are difficult for these speakers to perceive.

[13] Meanings expressed by reduplicated forms in Atayal are rendered in Yilan Creole by suffixes or analytic expressions. For example, 'children' in Atayal (Yilan Cyoli dialect) is *l-la'i*, but in Yilan Creole is *la'i-taci*. And 'smaller' in Atayal is *t-tipik*, but is *moto cipik* ('more' + 'small') in Yilan Creole.

In Japanese, *suru* is a verb that has its own independent meaning, 'do'. *Suru* also can attach to nouns to form compound verbs, as in the examples in (1).

(1) Japanese

undō	'exercise' (N)	*undōsuru*	'exercise' (V)
pasu	'pass' (N)	*pasusuru*	'pass' (V)

Yilan Creole also utilizes *-suru* to form verbs, as shown in (2).

(2) Yilan Creole

Lexical items		**Source language**	**Verbs**	
kusi	'comb' (N)	Japanese	*kusisuru*	'comb' (V)
pakyaw	'gambling' (N)	Southern Min	*pakyawsuru*	'gamble' (V)
baka	'fool' (Adj)	Japanese	*bakasuru*	'look down on' (V)
cayox	'ashamed' (Adj)	Atayal	*cayoxsuru*	'feel ashamed' (V)
acuy	'hot' (Adj)	Japanese	*acuysuru*	'warm up' (V)
kilox	'hot' (Adj)	Atayal	*kiloxysuru*	'warm up' (V)
horu	'dig' (V)	Japanese	*horusuru*	'dig' (V)
yasumu	'take a rest' (V)	Japanese	*yasumusuru*	'take a rest' (V)
naga	'wait' (V)	Atayal	*nagasuru*	'wait' (V)
okosu	'waken' (V)	Japanese	*okosuru*	'waken' (V)

From (2), we can see, for example, that *-suru* is attached to a noun *kusi* 'comb' to form a related verb, *kusisuru* 'to comb'. It appears to behave like *-suru* in Japanese, but there is an important difference. As mentioned above, in Japanese *suru* is an independent verb 'do' and also combines with nouns to form compound verbs. But in Yilan Creole, *-suru* cannot be used as a verb meaning 'do'; it is a bound morpheme without content meaning. In other words, *-suru* in Yilan Creole acts only as an affix and not as a verb. We can say that in Yilan Creole, lexical items like *kusisuru* are derivative words rather than compounds.

Moreover, of particular interest is that *-suru* can attach to superstrate-derived forms as well as substrate-derived forms and adstrate-derived forms, and *-suru* can attach not only to nouns but also to adjectives and verbs, e.g., *kilox-suru* (Adj + *suru*), *horu-suru* (V + *suru*). In Yilan Creole, *-suru* has broadened its functional range.

Forms like *horu-suru* (V + *suru*) and *yasumu-suru* (V + *suru*) are used widely among the younger generations. But the older generations criticize these forms as mistakes. This shows that such V + *suru* forms can be considered innovations. Moreover, verbs ending with *-su* are shifting toward ending with *-suru*, e.g., *okosu* 'waken' is shifting to *okosuru*. This suggests again that *suru*, reanalyzed as a suffix, has come to function as a verbal marker in Yilan Creole.

4.1.2 The causative marker *-rasyeru*

Another affix, *-rasyeru*, is one of the most characteristic features of Yilan Creole. Its source is likely the Japanese causative suffix *-(s)aseru*. (3) gives some examples.

(3)	**(Consonant-ending Verb)**[14]		**Gloss**	**Japanese**	**Yilan Creole**
	kaeru	'go home'	'let/make…go home'	*kaeraseru*	*kaerasyeru*
	kaku	'write'	'let/make…write'	*kakaseru*	*kakarasyeru*
	(Vowel-ending Verb)				
	taberu	'eat'	'let/make…eat'	*tabesaseru*	*taberasyeru*
	miru	'see'	'let/make…see'	*misaseru*	*mirasyeru*

As shown in (3), in Japanese, causative verbs are made by attaching suffixes to roots. Verbs are made into the causative form differently depending on the verb type. The causative suffix *-aseru* is attached to verb stems ending in consonants, e.g., *kaer-u* 'go home' becomes *kaer-aseru* 'make/let…go home', and *kak-u* 'write' becomes *kak-aseru* 'make/let…write'. For verb stems ending in vowels, *-saseru* is attached, e.g., *tabe-ru* 'eat' becomes *tabe-saseru* 'make/let…eat', and *mi-ru* becomes *mi-saseru* 'let/make…see'.

In Yilan Creole, the causative suffix *-rasyeru* is attached to verbs that end in both consonants and vowels. I assume that the Yilan Creole form *-rasyeru* comes from the causative form of verbs that end in /r/, which are the majority of Japanese verbs ending in consonants. An example is *kaer-aseru* 'make/let…go home'. Taking Japanese verbs that end in /r/ as the model, Yilan Creole adopted the causative suffix, reanalyzing the verb-stem-final /r/ as suffix initial in accord with Yilan Creole's CV syllable structure, and also palatalizing the /s/, perhaps due to the influence of western Japanese dialect. Thus *-aseru* becomes *-rasyeru.*

4.2 Compounding

Compound words are composed of two or more elements. Some examples of compound words used in Yilan Creole are given in (4).

(4) Yilan Creole

Form	**Source language**	**Literal meaning**	**Gloss**
hopa-la'i	Atayal-Atayal	'big_child'	'first-born child'
hopa-tenki	Atayal-Japanese	'big_weather'	'fine weather'
kako-balay	Japanese-Atayal	'figure_very'	'act big'
naka-lukus	Japanese-Atayal	'inside_clothes'	'underwear'
naka-pangcyu	Japanese-Japanese	'inside_pants'	'panties, shorts'
unme-zyoto	Japanese-Japanese	'destiny_fine'	'good luck'

These compound words can be classified into four types according to their sources:

Type 1: Atayal-derived word + Atayal-derived word (e.g., *hopa-la'i*)
Type 2: Atayal-derived word + Japanese-derived word (e.g., *hopa-tenki*)
Type 3: Japanese-derived word + Atayal-derived word (e.g., *naka-lukus, kako-balay*)
Type 4: Japanese-derived word + Japanese-derived word (e.g., *naka-pangcyu*, *unme-zyoto*)

Compounds of type 1 (Atayal-derived word + Atayal-derived word) can occur in Atayal. However, compounds of type 4 (Japanese-derived word + Japanese-derived word)

[14] There are two main verb types in Japanese, those whose stems end with consonants and those whose stems end with vowels. For example, the stem of the verb *tomaru* is *tomar*, and that of *taberu* is *tabe*. *-u* and *-ru* are inflectional suffixes for consonant-ending verbs and vowel-ending verbs, respectively.

cannot occur in Japanese. The compounds of types 2, 3, and 4 are considered innovations in Yilan Creole.

Compounding is a productive mechanism for word formation in many creoles (Mühlhäusler 1985; Holm 2000). Compounding does exist in Yilan Creole, but it seems not to be as productive as in other creoles.

4.3 Clipping

Clipping is "a type of word formation in which a new word is derived by shortening another word. Examples include *exam* from *examination* and *ad* from *advertisement*" (Crystal 1994:67). (5) shows Yilan Creole words derived through clipping.

(5) Yilan Creole

Form	**Japanese as source language**	**Gloss**
baku	*tabako*	'tobacco'
cumuli/muli	*katatsumuri*	'snail'
kosi	*sukoshi*	'a little'
osi	*oishī*	'delicious'

As shown in (5), for example, *baku* 'tobacco' is a clipped form of *tabako.* In all of these words, it is the first syllable that has been cut, and, interestingly, they tend to become two moras.

Moreover, clipping occurs in plural pronouns as well. As shown in Table 5 of section 5.3, the first plural pronouns *watasitaci* and *wasitaci* can be produced as *wataci*, and the third plural pronoun *aretaci* can be shortened to *ataci*. These clipped pronouns are innovative forms very different from those in their source language.

5 Development of the Lexical System of Yilan Creole

Many of both superstrate-derived and substrate-derived lexical items have their phonological shapes and meanings fairly well preserved in Yilan Creole. However, many are also modified in various ways. This section will examine the interaction between the substrate influences, superstrate influences, and internal developments.

5.1 Relexification

This section discusses Yilan Creole lexical items whose forms are derived from Japanese but whose semantic properties are derived from Atayal. Some examples are given in (6).

The Yilan Creole verbs in (6) all have two or more meanings. For example, the verb *haku* means 'put on' and can be applied to putting things on the head, the trunk, the lower half of the body, the arm, or the finger, as shown in (7). Its form is derived from that of the corresponding Japanese lexical item *haku*. However, *haku* in the Japanese language, while it also means 'put on', can only be used to express the action of putting things on the lower half of the body. The corresponding Atayal lexical entry, however, has the same meanings as the Yilan Creole one.

(6)

Yilan Creole	**Japanese**	**Atayal**
haku	*haku*	*posa*
‘put on’ (for the action of putting things on the head, trunk, lower half of the body, hand, or arm)	‘put on’ (for the action of putting things on the lower half of the body)	‘put on’ (for the action of putting things on the head, trunk, lower half of the body, hand, or arm)
neru	*neru*	*abi*
‘sleep’	‘sleep’	‘sleep’
‘stay’		‘stay’
toru	*toru*	*agal*
‘take’	‘take’	‘take’
‘meet’		‘meet’

(7) Yilan Creole

Form	**Gloss**	**Part of the body**
bosi haku	‘put on a hat’	head
lukus haku	‘put on clothes’	trunk
pancyu haku	‘put on trousers’	lower half of the body
kucyusita haku	‘put on socks’	lower half of the body
toke haku	‘put on a wristwatch’	arm
king haku	‘put on a ring’	finger

In Japanese, putting things on the lower half of the body is expressed by *haku* ‘put’, and putting things on other parts of the body besides the lower half is expressed by other verbs (*kaburu* ‘put’, *kiru* ‘put’, *tsukeru* ‘put’, *hameru* ‘put’) as (8) shows.

(8) Japanese

Form	**Gloss**	**Part of the body**
pantsu o haku	‘put on trousers’	lower half of the body
kutsushita o haku	‘put on socks’	lower half of the body
bōshi o kaburu	‘put on a hat’	head
fuku o kiru	‘put on clothes’	trunk
tokei o tsukeru	‘put on a wristwatch’	arm
yubiwa o hameru	‘put on a ring’	finger

As the examples in (8) demonstrate, in Japanese, there is a restricted selection between the verbs and the parts of the body.

How do we explain the usage in Yilan Creole? Let us take a look at the Atayal forms in (9).

(9) Atayal

Form	**Gloss**	**Part of the body**
posa bosi	‘put on a hat’	head
posa lukus	‘put on clothes’	trunk
posa pancyu	‘put on trousers’	lower half of the body
posa kucyusita	‘put on socks’	lower half of the body
posa tokey	‘put on a wristwatch’	arm
posa king	‘put on a ring’	finger

As (9) makes clear, *posa* 'put on' is used for all parts of the body. This shows that the Atayal lexical item is the source of the semantic properties of the Yilan Creole lexical item. In other words, the lexical entry has a phonological representation derived from the superstrate language but a semantic property derived from the substrate language. This can be accounted for by the relexification hypothesis (Lefebvre 1986; Lumsden 1999).

Lumsden (1999:225) defines relexification as "a mental process that allows a language learner to create a new vocabulary of lexical categories (i.e., nouns, verbs, adjectives, prepositions and adverbs) by linking new phonological forms with syntactic and semantic information that is already established in the lexicon of his native language."

5.2 Semantic changes

As seen in the preceding section, substrate influences can account for some semantic properties of Yilan Creole forms. But they cannot explain all semantic change types observed in Yilan Creole. This section is concerned with the development of semantic changes.

5.2.1 Substrate influence and semantic broadening

Certain lexical features in Yilan Creole seem to be a result of relexification from the substrate, but in fact they cannot be fully explained by relexification. (10) shows two examples of such forms. To take the first, in Japanese, *otoko* means 'male', but in Yilan Creole, *otoko* refers not only to a male person but also to a male animal. Furthermore, it also means 'husband'.

(10)

Yilan Creole	**Japanese**	**Atayal**
otoko	*otoko*	*likuy*
'male (human)'	'male (human)'	'male (human)'
'male (animal)'		'male (animal, especially dog)'
'husband'		'husband'
onna	*onna*	*nerin*
'female (human)'	'female (human)'	'female (human)'
'female (animal)'		'female (animal)'
'wife'		'wife'

Otoko in Yilan Creole takes its phonological shape from Japanese *otoko*, but it displays a meaning different from the corresponding Japanese lexical item *otoko*. In Yilan Creole, *otoko* has broadened its range of meaning.

It appears that the substrate language is the source of the semantic properties of Yilan Creole *otoko*. The comparison of Atayal and Yilan Creole forms in (11) and (12) provides clear evidence for this.

(11)

Atayal	**Yilan Creole**
likuy hoying	*otoko no hoying*[15]
male dog	male GEN dog
'male dog'	'male dog'

[15] *no* is a genitive case particle derived from Japanese.

(12)	**Atayal**	**Yilan Creole**
	Hiya ga likuy mu.	*Kore ga wasi no otoko.*
	he TOP husband my	he TOP my GEN husband
	'He is my husband.'	'He is my husband.'

(11) and (12) show that the meaning of *otoko* in Yilan Creole is likely influenced by Atayal. The usage can be accounted for by the relexification hypothesis, as mentioned above. However, several usages of *otoko* cannot be explained in terms of substrate influence, as (13) makes clear.

As (13) shows, in Atayal, for example, to refer to a 'male dog', *likuy* is used as the adjective modifier to make a noun phrase, *likuy hoying*. For 'male chicken; cock', *tingan* is the adjective; for 'male pig; boar', it is *tohok*. However, **tingan hoying* and **tohok hoying* are ungrammatical. Similarly, **tohok welung* is not accepted. That is, in Atayal, there are restrictions on the co-occurrence of specific adjective modifiers and head words.

(13)	**Atayal**	**Yilan Creole**
	likuy hoying 'male dog'	*otoko no hoying* 'male dog'
	tingan welung 'cock'	*otoko no welung* 'cock'
	tohok biyak 'boar'	*otoko no biyak* 'boar'

However, in Yilan Creole, the modifier *otoko* 'male' is used with *hoying* 'dog', *welung* 'chicken', and *biyak* 'pig'. The Yilan Creole word *otoko* has had its meaning extended to refer not only to dogs, but also to chickens and pigs.

In this case, Yilan Creole lexical properties have developed from an interaction between substrate influence and a strategy of semantic change, i.e., semantic broadening.

5.2.2 Semantic generalization

Semantic generalization is a process that "involves relating lexemes in terms of one or more shared specifications" (Hancock 1980:77).

The process of semantic generalization in Yilan Creole is evident in the items in (14).

(14) shows that the word *'ba* can refer to not only 'hand' but also 'finger', 'wrist', 'elbow', and 'arm' in Yilan Creole. It is an Atayal-derived form, but obviously it has a very different semantic range from the original Atayal word *'ba*. In Atayal, *'ba* means only 'hand', while 'finger' is expressed by *tloling*, 'wrist' by *ho'i*, 'elbow' by *hiku*, and 'arm' by *piying*. Yilan Creole *'ba* has a more general meaning, referring to all of these body parts.

(14)	**Meaning**	**Yilan Creole**	**Japanese**	**Atayal**
	'hand'	*'ba*	te	*'ba*
	'finger'	*'ba*	yubi	*tloling*
	'wrist'	*'ba*	tekubi	*ho'i*
	'elbow'	*'ba*	hiji	*hiku*
	'arm'	*'ba*	ude	*piying*

We also can say that Yilan Creole *'ba* has a broadened meaning. This kind of semantic change can be considered one type of semantic extension process (Hancock 1980).

Other similar examples appear in (15).

(15)

Meaning	**Yilan Creole**	**Japanese**	**Atayal**
'foot, leg'	*kakay*	ashi	*kakay*
'thigh'	*kakay*[*]	momo	*muyiy*
'mouth	*ngowa*	kuchi	*ngowa*
'lip'	*ngowa*	kuchibiru	*prahung*
'tree'	*koni*	ki	*khoni*
'branch'	*koni*	eda	*ara*

[*]Older speakers use *muyiy* instead of *kakay*.

As (15) shows, *kakay* 'foot' refer to the entire limb, including 'thigh', not just to the extremity. *Ngowa* means not just 'mouth', but also 'lip'. *Koni* means not just 'tree', but also 'branch'.

5.2.3 Semantic narrowing

In contrast with semantic broadening, the narrowing of the semantic range of some words has occurred in Yilan Creole, too. (16) provides some examples.

(16)

Yilan Creole	**Gloss**	**Source language**	**Gloss**
ici	'one' (cardinal number)	Japanese	'one'
utux	'one' (numerative)	Atayal	'one'
ni	'two' (cardinal number)	Japanese	'two'
sa'ing	'two' (numerative)	Atayal	'two'
araw	'wash (face, hands, and objects except clothes)'	Japanese	'wash (any object)'
pemasuru / memasuru	'wash (body)'	Atayal	'wash (face, body)'
mahusuru	'wash clothes'	Atayal	'wash clothes'
tobu	'jump'	Japanese	'fly, jump'
lakasuru	'fly'	Atayal	'fly'
koni	'tree'	Atayal	'tree, wood'
maki	'wood'	Japanese	'wood'

Ici and *utux* are of Japanese origin and Atayal origin respectively, and both of them mean 'one'. However, in Yilan Creole, *ici* is a cardinal number (i.e., *one*, *two*, etc.), that is, it is used for listing the numbers or answering the question "How many?" while *utux* is a numerative (e.g., *one person*, *one dog*, etc.), as in (17) and (18).

(17) *utux* *cyoli*
one person
'one person'

(18) *utux* *hoying*
one dog
'one dog'

Ici and *utux* were coexisting variants for the same item 'one' until semantic differentiation occurred, so that their meanings and functions diverged. That is, two sets of numbers now exist, and they are in complementary distribution. Their meanings have become less general than their earlier meanings. This process is also known as specialization.

Both Hancock (1980) and Holm (2000) note that restriction of a word's meaning seems to be less frequent than extension in creole languages. However, I have found several examples of semantic narrowing in Yilan Creole. This difference between Yilan Creole and other creole languages is another topic to be investigated in future research.

5.3 Superstrate input and simplification

As Winford (2003:338) notes, "internally driven innovations in creole grammar seem to arise from the restructuring of superstrate and substrate input, regulated by universal principles of acquisition (the need for economy and simplicity)." A case in point in Yilan Creole is the development of the pronominal system.

Before going on to the main subject, we will take a brief look at the pronominal systems of Atayal and Japanese. Table 3 illustrates the pronominal system of the substrate language, Atayal. As Table 3 shows, Atayal makes a distinction between bound pronouns and free pronouns. Pronouns indicate number and case, and the first person plural pronouns distinguish between inclusive and exclusive.

Table 3: The pronominal system of Atayal[16]

Case	Bound pronoun		Free pronoun	
Number/Person	Subjective	Genitive	Locative	Neutral
Singular				
1	*cu*	*mu*	*knan*	*ku'ing/kun/knan*
2	*su*	*su*	*isu*	*isu*
3	*hiya*	*nya*	*naha*	*hiya*
Plural				
1 (inclusive)	*ta*	*ta*	*ita*	*ita*
1 (exclusive)	*cyamin*	*myan*	*cyamin*	*cyamin*
2	*cimu*	*mo*	*cimu*	*cimu*
3	*naha*	*naha*	*naha*	*naha*

Next let us consider the pronominal system of the superstrate language, Japanese, which is presented in Table 4. As can be seen in Table 4, the pronouns of Japanese indicate number, gender (*kare* 'he' vs *kanozyo* 'she'), and degree of politeness (e.g., *antatachi* 'you [plural, informal]' vs *anatagata* 'you [plural, honorific]').

[16] The data are from Yawi Nomin (2009) and a 2014 interview with Yawi Nomin.

Table 4: The pronominal system of Japanese

Number/Person	Politeness			
	Plain form	Polite form		Informal form
		Honorific	Humble	
Singular				
1	*watashi*		*watakushi*	*atashi/boku/ore*
2	*anata*	*anatasama/otaku*		*anta/kimi/omae*
3	*anohito/kare/kanozyo*	*anokata*		*are/aitsu*
Plural				
1	*watashitachi*		*watakushidomo*	*atashira/bokura/orera*
2	*anatatachi*	*anatagata*		*antara/kimira/omaera*
3	*anohitotachi*	*anokatagata*		*aitsura*

The pronominal system of Yilan Creole is summarized in Table 5. Contrasting the forms in Table 5 with those in Table 3, we find that except for number, the distinctions made in Atayal are not found in the pronouns of Yilan Creole.

Table 5: The pronominal system of Yilan Creole

Number/Person	Pronoun
Singular	
1	*watasi*/*wasi*
2	*anta* (*nta*)
3	*are*
Plural	
1	*watasitaci*/*wasitaci* (*wataci*)
2	*antataci* (*antaci*)/*ntataci* (*ntaci*)
3	*aretaci* (*ataci*)

Contrasting the forms in Table 5 with those in Table 4, it is clear that most of the pronoun forms of Yilan Creole are derived from Japanese. But the distinctions of gender and politeness are lacking in Yilan Creole. Yilan Creole has a simpler system than Japanese. It seems that the simplification of the Japanese system in Yilan Creole is due to substrate influence, since Atayal does not have polite forms and gender distinctions.

It is notable that an important distinction in Atayal, the one between inclusive and exclusive, is also not made in Yilan Creole. Many Pacific creoles with an Austronesian substrate have a striking parallel with Austronesian languages in the existence of pronoun forms for several numbers (singular, dual, trial, plural) and the distinction between inclusive and exclusive (Holm 1989, 2000; Lorenzino 1993, among others). On the other hand, Malay-based creoles have lost the inclusive/exclusive distinction.[17] Further investigation should be directed toward understanding this difference. What can be pointed out here is that Yilan Creole has a simpler pronominal system than Atayal, too.

The pronoun forms of Yilan Creole come mainly from standard Japanese, for example, *watashi* 'I', *anta* 'you', *-tachi* (plural suffix). There is also a western Japanese dialect form, *washi* 'I'. In addition, there is one form that is likely influenced by the adstrate

[17] My thanks to a reviewer for providing this information.

language: It is the variant *han no* for the first person possessive, which apparently combines Southern Min *guan* 'my' and Japanese *no*.[18]

To sum up, through the restructuring of superstrate, substrate, and adstrate input, Yilan Creole has created a simplified pronominal system, which indicates number (singular and plural), but does not make distinctions of case, bound and free, inclusive and exclusive, gender, and politeness.

6 Conclusion

This paper has considered the lexical system of Yilan Creole, based on data collected through consultations with residents in Tungyueh Village.

Examining the basic vocabulary clarifies that Yilan Creole is a Japanese-lexicon creole. Atayal is its substrate, and Mandarin and Southern Min can be considered adstrates. The paper analyzes the source of lexical items by semantic field, finding that the substrate language contributed most of the lexical items for body parts and culturally relevant items.

As for word formation, the paper discusses three main processes, which are derivation, compounding, and clipping, and examines various ways in which lexical items that come from both the superstrate and the substrate language have been modified in Yilan Creole.

Several lexical items have phonological representations derived from the superstrate language but semantic properties derived from the substrate language. This is accounted for by the relexification hypothesis. But there are several factors that cannot be explained by the relexification hypothesis. The paper discusses different types of semantic change, such as semantic broadening, semantic generalization, and semantic narrowing, as well as considering simplification, to explore the processes of interaction between the three major sources of the Yilan Creole lexicon—retentions from the substrate language, input from the superstrate language, and internal innovations.

[18] The pronouns of Yilan Creole can serve as possessive adjectives by combining with the genitive case particle *no*, which has its origins in Japanese, for example, *watasi no* 'my', *anta no* 'your', and *are no* 'his'.

References

Abe, Seiya, Shigeru Tsuchida and Sumino Niida. 2008. Atayarugo no kankei hōgen ni haitta nihongo: Taiwan genzyūmin gengo nōryoku shiken mondai ni okeru [Hanxi dialect: A Japanese-based creole in the Formosan Austronesian language 'Atayal']. *Tōyōbunka Kenkyū* [*Journal of Asian Cultures*] 10:29–63. [In Japanese]

Bickerton, Derek. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7.2:173–188.

Chien, Yuehchen. 2006. Taiwan zanson nihongo ni mirareru hiteiji *-nai* to *-n*: Karenken o fīrudo ni [Negation *-nai* and *-n* in Taiwan Japanese: The case of Hualien prefecture]. *Nihongo Kagaku* [*Japanese Linguistics*] 20:5–25. [In Japanese]

Chien, Yuehchen and Shinji Sanada. 2010a. Yilan Creole in Taiwan. *Journal of Pidgin and Creole Languages* 25.2:350–357.

Chien, Yuehchen and Shinji Sanada. 2010b. dongtaiwan taiyazu de yilan keliaoer [Yilan Creole of the Atayal people in eastern Taiwan]. *taiwan yuanzhumin yanjiu jikan* [*Taiwan Journal of Indigenous Studies*] 3.3:75–89. [In Chinese]

Chien, Yuehchen and Shinji Sanada. 2011. Taiwan no giran kureōru ni okeru hiteiji: *-nay* to *-ng* no henyō o megutte [Negation in Taiwan's Yilan Creole: Focusing on *-nay* and *-ng*]. *Gengo Kenkyu* [*Journal of the Linguistic Society of Japan*] 140:73–87. [In Japanese]

Craig, Colette Grinevald. 1997. Language contact and language degeneration. In *The handbook of sociolinguistics*, ed. by Florian Coulmas, 257–270. Blackwell Handbooks in Linguistics 4. Oxford: Blackwell.

Crystal, David. 1994. *An encyclopedic dictionary of language and languages*. London: Penguin.

Ehrhart, Sabine and Peter Mühlhäusler. 2007. Pidgins and Creoles in the Pacific. In *The vanishing languages of the Pacific rim*, ed. by Osahito Miyaoka, Osamu Sakiyama and Michael E. Krauss, 118–143. Oxford: Oxford University Press.

Hancock, Ian F. 1980. Lexical expansion in creole languages. In *Theoretical orientations in Creole studies*, ed. by Albert Valdman and Arnold Highfield, 63–88. New York: Academic Press.

Holm, John. 1989. *Pidgins and creoles, Vol. 2: Reference survey*. Cambridge Language Surveys. Cambridge: Cambridge University Press.

Holm, John. 2000. *An introduction to pidgins and creole*s. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Kamei, Takashi, Rokurō Kōno and Eiichi Chino. 1996. *Gengogaku daijiten* [The Sanseido encyclopaedia of linguistics]. Tokyo: Sanseido. [In Japanese]

Lefebvre, Claire. 1986. Relexification in creole genesis revisited: The case of Haitian Creole. In *Substrata versus universals in Creole genesis*, ed. by Pieter Muysken and Norval Smith, 279–300. Amsterdam: John Benjamins.

Lorenzino, Gerardo A. 1993. African vs. Austronesian substrate influence on the Spanish-based creoles. In *Atlantic meets Pacific: A global view of pidginization and creolization*, ed. by Francis Byrne and John Holm, 399–408. Creole Language Library 1. Amsterdam: John Benjamins.

Lumsden, John S. 1999. The role of relexification in creole genesis. *Journal of Pidgin and Creole Languages* 14.2: 225–258.

Mufwene, Salikoko S. 1996. The founder principle in creole genesis. *Diachronica* 13.1:83–134.

Mühlhäusler, Peter. 1985. The lexical system of Tok Pisin. In *Handbook of Tok Pisin (New Guinea Pidgin)*, ed. by Stephen A. Wurm and Peter Mühlhäusler, 423–440. Pacific Linguistics C-70. Canberra: Australian National University.

Nishioka, Hideo. 1936. Taiwan to hōgen [Taiwan and dialects]. *Seionkyōiku* [*Phonetic Education*] 3.1:79–83. [In Japanese]

Otani, Yuki and Lillian M. Huang. 2009. Japanese impact on Han-xi Creole: A phonological perspective. Poster presented at the Tenth National Conference on Linguistics, 27–28 November. Taoyuan: Yuan Ze University.

Roberts, Ian. 1999. Verb movement and markedness. In *Language creation and language change: Creolization, diachrony, and development*, ed. by Michel DeGraff, 287–327. Cambridge, MA: MIT Press.

Saito, Gishichirō. 1943. Taiwan ni okeru kotoba o megutte [On languages in Taiwan]. *Kotoba* [*Language*] 5.1:40–54. [In Japanese]

Sanada, Shinji and Yuehchen Chien. 2007. Taiwan Atayaruzoku ni okeru nihongo kureōru [The Japanese creole of the Atayal people in Taiwan]. In *Proceedings of the 2007 Spring Conference of the Society for Japanese Linguistics*, ed. by the Society for Japanese Linguistics, 167–174. Osaka: Kansai University. [In Japanese]

Sanada, Shinji and Yuehchen Chien. 2008a. Taiwan ni okeru nihongo kureōru ni tsuite [On a Japanese creole in Taiwan]. *Nihongo no Kenkyu* [*Studies in the Japanese Language*] 4.2:69–76. [In Japanese]

Sanada, Shinji and Yuehchen Chien. 2008b. Taiwan no nihongo kureōru [A Japanese creole in Taiwan]. *Gengo* [*Language*] 37.6:94–99. [In Japanese]

Tsuchida, Shigeru. 2000. Nokotta nokotta! no gengogaku (Gengo ga kieru toki: Taiwan genzyūmin no gengo no baai) [Linguistic survival (When a language disappears: The case of Taiwan indigenous languages)]. Paper presented to the research society “Sennin no kai”, 17 Dec. Tokyo: Hosei University. [In Japanese]

Utsushikawa, Nenozō, Nobuto Miyamoto and Tōichi Mabuchi. 1935. *Taiwan takasagozoku keitō shozoku no kenkyū* [*The Formosan native tribes: A*

genealogical and classificatory study]. Institute of Ethnology, Taihoku Imperial University. Tokyo: Tōkōshoin. [In Japanese]

Winford, Donald. 2003. *An introduction to contact linguistics*. Language in Society 33. Oxford: Blackwell.

Yawi, Nomin (ed.). 2009. *Tipik khu cnyan patas ke na Cyoli Tayal bih Llyung Klesan* [*An Yilan Cyoli dialect of Atayal language dictionary with illustrations*]. Yilan: Yilan County Government. [In Atayal and Chinese]

22 *A corpus-based approach to the classification of Yami emotion**

VICTORIA RAU, YI-HSIN WU AND MENG-CHIEN YANG

1 Introduction

The point of departure for an investigation of emotion in a language is its lexicon (Saucier & Goldberg 1996). Although emotion concepts, such as happiness, anger, sadness and fear, are intuitively clear and can be found in various languages, we cannot assume all languages have the same set of emotions.[1] Church et al. (1998) found that the "hypercognised"[2] (Levy 1984) emotions in Filipino[3] are anger, anxiety/fear, happiness, contentment, sadness, and arousal, whereas the relatively "hypocognised" emotion domains include feeling tired, guilty, surprised, contemptuous, and aspiring. They also recommended that the terms in all three subcategories of the affective conditions class (i.e., pure affective, affective-behavioral, and affective-cognitive states) in Clore et al.'s (1987) taxonomy of emotion terms be viewed as referring to emotions. Although Church

* This study is partially supported by two NSC grants: "A typological study of Austronesian languages in Taiwan and their revitalisation" (NSC100-2420-H-194-011-MY3), 1 November 2011–31 October 2014, and "Yami ontology: Yami lexical semantics and sociogrammar" (NSC 100-2410-H-194-104), 1 August 2011–31 July 2012.Various parts of our research have been presented at the 2012 CLDC (Rau, Wu, Yang & Hu 2012), the 2012 International Conference on Landscape, Seascape, and the Spatial Imagination (Rau, Wu & Yang 2012), and the International Conference on Asian Language Processing (Yang, Rau & Wu 2012). We would like to thank the three anonymous reviewers for their helpful comments and the rigorous work of the editors of this volume, which has significantly improved the technical writing of this paper. All remaining errors are ours.

1 Emotion can be used as either an uncountable or countable noun, depending on whether the collective concept of emotion is intended (singular) or individual references of emotion are intended (plural).

2 Hypercognised and hypocognised emotions refer to the dichotomy between most important/maximally lexicalised and least important/minimally lexicalised emotion terms.

3 The word 'Filipino' is used by Church et al. (1998). Filipino is the official name of the national language of the Philippines, primarily based on Tagalog (http://www.ethnologue.com/country/PH/languages). Although Tagalog is more frequent in common parlance, Filipino or Pilipino underlines its role as the national language. The name Pilipino is also used in the textbook title *Pilipino through Self-Instruction* (Wolff et al. 1991) to emphasise its role as a widely used second language for inter-group communication.

Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu, eds.
New Advances in Formosan Linguistics, 533-554
Asia-Pacific Linguistics, 2015

et al. used a quantitative experimental method to validate Clore et al.'s emotion model, they relied on a translation of emotion terms from English rather than searching directly in Filipino. The translation approach is based on the researchers' assumptions of universals in emotion. Although this has its place to provide "etic" data, one cannot be certain that the equivalents can be found in translation, not to mention that there is always something lost in translation. Thus to yield an "emic"[4] perspective of emotion from a language, it is imperative to extract emotion terms directly from the language, but the problem is how to determine what constitutes an emotion term in a language.

Cognitive linguists have proposed to construe emotion based on a grammatical model for Formosan languages, such as Tsou (Huang 2002). Following Talmy's (2000) suggestion that emotion events are inherently causal, Hsieh (2011) examined emotional causality in Kavalan, Paiwan, and Saisiyat. Their theoretical approach has provided a framework for the present study to explore Yami emotion. As Yami is the only Philippine language in Taiwan, it is also important to compare the results of Yami emotions with the results of Filipino obtained by Church et al. (1998).

How is emotion defined? Wierzbicka (1992) suggests that emotion concepts can be defined in terms of universal primitives, such as 'good', 'bad', 'do', 'happen', 'know', and 'want', and prototypical scripts formulated in terms of 'thoughts', 'wants', and 'feelings'. In other words, an emotion event involves someone's cognition, affection, or feelings about something that happened to someone or the fact that someone did something. The thoughts/wants/feelings can be evaluated as positive or negative. To discover what meets the definition of emotion from an emic perspective, this study aims at a grammatical model encoded by the prefix *ika-* 'the reason/cause for a certain feeling' to conceptualise Yami emotion concepts, complemented by an ontological approach to compare Yami emotions with Filipino. As recent interdisciplinary investigations on Yami fish names (Hu & Rau 2013) and Yami fish ontology (Tai et al. 2008; Rau et al. 2009) have brought us a better understanding of how metaphors are used in describing fish names and place names, we intend to adopt a similar interdisciplinary approach to investigating the classification and ontology of Yami emotion with the goal of building a complete ontology of the targeted language and culture.

Our aim is to explore the following three questions:

1. What are the hypercognised and hypocognised emotions in Yami?
2. Are there more positive or negative emotion terms in Yami?
3. Do Yami emotions share the same distinct domains as Filipino emotions?

The organisation of this paper is as follows. After the brief introduction above, an introduction to Yami morphology, with a focus on *ka-* and *ika-*, is presented. After that, the results of classification of Yami emotion are presented, followed by a comparison with Filipino emotion based on the results of a cluster analysis of Yami emotion. This paper ends with an evaluation of the application of this approach to the study of emotion.

---

[4] The terms 'emic' and 'etic' are used in anthropology to contrast an insider's view from an outsider's view. They are derived from the linguistic distinction between 'phonemic' and 'phonetic'. Different sounds which are 'phonetically' different may be perceived by native speakers of a language as either the same or different 'phonemically' depending on the phonological system of the language.

2 Yami morphology: *ika-*

To understand why the prefix *ika-* was chosen as the point of departure for the investigation of emotion, a brief discussion of Yami morphology is in order. Yami verbs are classified either as dynamic or stative. Transitive verbs can be marked by one of the four focus markers: agent focus (AF) *m-*, patient focus (PF) *-en*, locative focus (LF) *-an*, and instrument focus[5] (IF) *i-*. Stative verbs, which are mostly intransitive (i.e., agent focus) (see Table 1), are marked with the *ma-* prefix. Note that among the three types of *ma-*, only the second type is stative.

Table 1: Yami *ma-* verbs

(1) *ma-*	agent focus	*ma-cimoy*	'rain'
		ma-ngay	'go'
(2) *ma-*	agent focus	*ma-saray*	'happy'
	stative verb	*ma-tava*	'fat'
(3) *ma-*	patient focus	*ma-cita*	'can see, visible'
	potentive verb	*ma-hap*	'can get'

Dynamic transitive verbs (see Table 2), on the other hand, are marked by the *p-* prefix. Here are some contrasting examples of verb forms with the *m-* and *p-* prefixes.

Table 2: Yami dynamic verbs

Intransitive	Transitive
mi-palit	*pi-palit-en*
(AF *mi-*the root is *palit* 'change')	(*pi-*root-PF.*en*)
'exchange'	'exchange'
*ma**n**azang*	*pa**n**azang-an*
(AF *maN-* the root is *sazang* 'buy')	(*paN-*root-LF.*an*)
'buy'	'place where one bought'
maka-vonas	*paka-vonas-en*
(AF *maka-* the root is *vonas* 'remove')	(*paka-*root-PF.*en*)
'can remove'	'must remove'
maci-vazay	*paci-vazay-an*
(AF *maci-* the root is *vazay* 'work')	(*paci-*root-LF.*an*)
'engage in work'	'engage in work with someone'

The stative *ka-* as described by Zeitoun & Huang (2000) only appears when the transitive forms are affixed with the instrument focus prefix *i-* to indicate the O argument is in IF (instrument or reason of the action), as illustrated in Table 3. Note that the Yami prefix *ka-* is polysemous.[6] It can form nouns (e.g., *ka-tangked* 'nearby', *ka-paganam* 'dance'), verbs with a separate inflectional paradigm from the regular focus system (e.g., *ka-doa* 'two in total', *ka-tangara* 'looked up just now', *ka-lavi* 'why cry (blaming)', *ka-teneng* 'then understand', exclamatory sentences (e.g., *ka-zakat* 'Go to hell!', *ka-lowlaw* 'so

[5] Instrument Focus covers instrument, beneficiary, and reference.

[6] A preliminary discussion of the functions of the *ka-* prefix can be found in Rau & Dong (2006:132–134). For a more current description of *ka-*, consult Rau & Dong (2010, forthcoming).

bloody good!'), or stative verb stems to derive verbs of instrument focus (e.g., *i-ka-rahet* 'consider something bad because …'). The stative *ka-* is invisible in intransitive verbs, e.g., *marahet* 'bad', but when a transitive verb prefixed with *i-* is formed from a stative root, the *ka-* prefix marks its stative root, e.g., *i**ka**rahet* 'consider something bad because ...'.

Table 3: Yami stative verbs

Intransitive	Transitive
mazakat 'killed, dead'	*i-ka-zakat* 'killed because…'
marahet 'bad'	*i-ka-rahet* 'consider something bad because…'
mam'ing 'smile'	*i-ka-m'ing*[7] 'amused because…'.

The emotion-related *ika-*[8] served as a promising point of departure for searching Yami feeling and emotion based on a bottom-up approach because it helped us find the majority of Yami emotion roots. The stative roots (defined as roots that can co-occur with the *ka-* prefix) identified as emotion led us to find other derived verbs. For example, *ikangsah* 'feel bored because of such and such (IF)' can lead to other derived forms, such as *mangsah* 'feel bored (AF)' or *angsahen* 'feel impatient about someone (PF)' with the same root *angsah* 'bored'. Take *ika'oya* 'feel angry because of such and such (IF)' as another example. We can find several other derived verbs with the same root *'oya* 'angry': *m'oya* 'angry with someone (AF)', *mi'oya'oya* 'very upset (AF)', *'oyan* 'reason to be angry (LF)', and *i'oya* 'get upset with someone (IF)'. As illustrated above, this bottom-up approach, rooted in the form of the language, provides a reliable basis to search for the iconic relationship of isomorphism, i.e., same form, same function.

3 Methods

This study adopted a corpus-based approach to find Yami emotion terms from the *ika-* construction and compare the classification of Yami emotion terms with that of Filipino based on a cluster analysis. The methods are described in the following four steps.

[7] As the orthography of the laryngeal features of /h/ and glottal stop in Yami remain to be worked out, currently there are various ways of spelling *i-ka-m'ing* and *mam'ing*, such as *ikamiying*, *ikamihing*, *mamiying*, and *mamihing*. For an updated version of Yami phonology and orthography, see Rau & Dong (forthcoming).

[8] The other homophonous, monomorphemic *ika-* encoding ordinal number, such as *ika-dwa* 'the second' is not related to emotion and was excluded from our discussion.

Step 1: Search for tokens with *ika-* from the Yami language documentation website

To find potential Yami emotion terms, we began our study by extracting all 1763 tokens of *ika-* from our Yami corpus, consisting of sixty-three texts and the New Testament data from the Yami-language documentation website.[9] The extraction of all the *ika-* tokens helped us identify the "construction" meaning (Goldberg 1995:4) of the grammatical model *ika-* as 'the reason/cause for a certain feeling'. Although *ika-* is composed of the instrument focus *i-* and the prefix *ka-* in stative verbs, the frequent pairing of the form *ika-* with the meaning of 'the reason/cause for a certain feeling' has created a new "construction" which cannot be predicted from the composition of *i-* and *ka-* in Yami. Under this cause frame (Dirven 1997), we noticed that *ika-* can be prefixed with a wide range of word classes, from pronouns (e.g., *ikaiya* 'he is the cause/reason') to negation markers (e.g., *ikabeken, ikabo* 'reason for being not') and stative verbs (e.g., *ikamo* 'reason to feel embarrassed'). Overall, *ika-* is most frequently prefixed to roots of stative verbs (e.g., *masaray* 'happy', *ma'oya* 'angry') and bare-root attributive modifiers (e.g., *aro* 'many', *apia* 'good') to form the most prototypical cause construction (e.g., *ikasaray* 'reason to be happy', *ika'oya* 'reason to be indignant', *ikaro* 'reason to be abundant', and *ikapia* 'reason to be good'). However, the wide range of words co-occurring with the *ika-* construction still made it difficult to classify emotion according to Clore et al.'s (1987) taxonomy. Thus we decided to set aside the *ika-* tokens temporarily and search the emotion lexicon by analyzing twenty narrative texts (Rau & Dong 2006) with clear story lines, as the narrative context made the task of identifying and coding emotion words much easier.

Step 2: Search for Chinese translations of emotion expressions

The search for emotion terms based on reading the twenty Yami texts with Chinese translations (Rau & Dong 2006) helped us identify 258 potential emotion expressions, not all of which included *ika-*. We coded the 258 terms into nine categories. The first six categories fit Clore et al.'s (1987) framework, but the other three categories include interjections, curses, and metaphors/metonyms. The coding was jointly decided by the first two authors. Clore et al. (1987) made a distinction between internal and external conditions. As we did not find any token that would fit nicely in the category of "external" conditions, defined as (1) "subjective" evaluations of character or stable characteristics (e.g., attractive, trustworthy) and (2) "objective" conditions, such as things done to a person (e.g., abandoned, insulted), "external" conditions were excluded from the study.

The remaining internal conditions were further divided between mental and non-mental states, as shown in Figure 1.

---

[9] The three Yami websites are Yami language documentation (http://yamiproject.cs.pu.edu.tw/yami), Yami e-learning (http://yamiproject.cs.pu.edu.tw/elearn), and Yami online dictionary (http://yamibow.cs.pu.edu.tw).

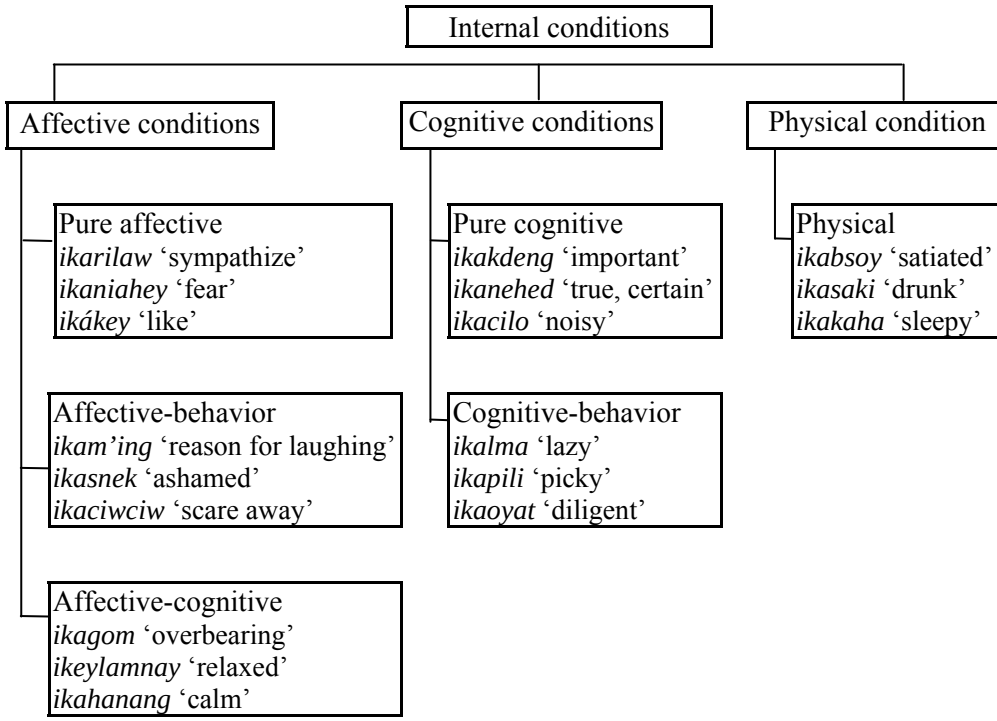


Figure 1: Classification of Yami emotion based on Church et al. (1998)

The internal non-mental states refer to physical and bodily states (e.g., sleepy, seasick). According to Clore et al. (1987), internal mental states consist of affective conditions and cognitive conditions. Under the category of affective conditions, we further separated pure affective states (e.g., afraid, angry, happy) from affective-behavioral states and affective-cognitive states, depending on whether the affective emotion is followed by an action (e.g., scare away, fight) or a cognitive consequence of the emotion (e.g., impatient, sorrowful). The category of cognitive conditions was similarly further divided into pure cognitive states and cognitive-behavioral states, with the former referring to the internal mental states in which cognition is dominant (e.g., smelly, stuffy) and the latter being followed by an action (e.g., picky, discreet). The other categories of emotion identified from the twenty texts included interjections/curses (e.g., ouch, damn) and metaphors/metonyms (e.g., the body is as healthy as light feathers or someone being as despised as goats). We can see in Table 4 that the emotion expressions from the twenty texts are divided into nine categories. Note that the first six categories contain the *ika-* prefix, while the last three do not.

Table 4: Coding categories for Yami emotion

Category	Yami	English[10]
1. Physical and bodily states	*ikakaha*	'sleepy because…'
2. Pure affective states	*ikaniahey*	'afraid because…'
3. Affective-behavioral states	*ikavozoaw*	'scare away because…'
4. Affective-cognitive states	*ikaotok*	'impatient because…'
5. Pure cognitive states	*ikangot*	'smelly because…'
6. Cognitive-behavioral states	*ikapili*	'picky because…'
7. Interjections	*ananay*	'ouch'
8. Curses	*mo kavazat*	'Damn!'
9. Metaphors/metonyms[11]	*nimananat so velek*	'terrified (in the stomach)'

Step 3: Identification of the final set of Yami emotion terms

As steps 1 and 2 led us to ascertain that the *ika-* prefix is really the key area to search for emotion terms in Yami, we began our final search to find all the Yami emotion terms in 166 texts (including the twenty texts mentioned above and 146 other narratives from the three Yami websites). The final search yielded 126[12] Yami emotion terms with the *ika-* prefix to serve as the database for categorisation and analysis in the present study. After the linguistic classification of the emotion terms was completed, a diagram was drawn using the Protégé program to represent the Yami emotion ontology.

Step 4: A cluster-analysis of Yami emotions for comparison with Filipino

To explore the possibility of comparing our results with the previous study on Filipino, a hierarchical cluster analysis was conducted to produce distinct domains of Yami emotion.

Following Church et al.'s (1998:78) procedure, a between-clusters linkage algorithm was calculated to produce comparable dendrograms for comparison with Filipino emotions. The cross-relationship between the emotion terms was calculated to build the hierarchical structure, using the knowledge extracted from our proposed ontological computation procedure. In contrast to Church et al.'s questionnaire methods for data collection, our study used a bottom-up corpus approach to create and grow the ontologies of the emotion concepts manually. These factors were used to calculate judgment values for evaluating whether an *ika-* emotion term could be put into a specific English emotion

[10] As the English translations of the *ika-* 'the reason/cause for a certain feeling' examples are all 'feel such and such because…', we will not repeat the same frame but only translate the emotion terms in the rest of the paper.

[11] Several body parts have been identified as related to emotion in metaphors. Due to the scope of this paper, we only list some examples here with the keywords bolded and will leave a systematic study on Yami "embodiment" in cognitive linguistics (Lakoff & Johnson 1999) for future investigation.

(1) *ji anisomalap o* ***pahad*** *na*. 他的魂嚇壞了. 'His **soul** has indeed flown away.'

(2) *malaw no* ***velek*** *a kalawan*. 憂心 'worry to the **stomach**, i.e., very worried'

(3) *do keyngeyngen na no* ***oo***. 感到非常頭痛 'sick to the **head**, i.e., have a headache'

[12] The 258 potential emotion expressions in step 2 contained both *ika-* words and three other categories (exclamations, curses, and metaphors). After the 132 expressions without *ika-* were excluded, this yielded the final 126 terms. As our focus is on classification of different types of emotion, it is not our concern to report the tokens of each individual type.

cluster. The clusters in each domain are represented by the key words and the frequency numbers of Yami *ika-* words with similar semantics (near synonyms) shown in parentheses after the *ika-* word. The relative distance from each *ika-* emotion term with the English translation was calculated using these factors. Finally, dendrograms were drawn to visualise the clustering results.[13]

4 Results

Based on the analysis of the 126 Yami emotion terms with the *ika-* prefix, the following section first presents the quantitative results of the most and least important emotion terms in Yami, followed by the distribution of positive and negative emotion terms, and ontological representation of semantic categories of Yami emotion. The second part of the results demonstrates the similarities and differences between Yami and Filipino.

4.1 Most important and least important emotions in Yami

What is considered the most or least important emotion in Yami is based on an interpretation of the frequency distribution of the 126 types. Overall, the pure cognitive category (e.g., good, bad, intelligent, difficult) constitutes the majority of Yami emotion terms, while affective-cognitive (e.g., lonely) and cognitive-behavioral categories (e.g., lazy) are the least frequent. The discovery of the most and the least important emotion domains in Yami generally matches Church et al.'s (1988) findings in Filipino data, except that there are no external conditions in Yami. As shown in Figure 1, Yami emotion based on the 126 terms with the *ika-* prefix constitutes three internal conditions, affective, cognitive, and physical, with the three affective conditions and the two cognitive conditions "hypercognised". This finding supports Church et al.'s recommendation that the terms in all three subcategories of the affection conditions class in Clore et al.'s (1987) taxonomy of emotion terms be viewed as referring to emotions.

As shown in Table 5 and Figure 2, almost half of the emotion terms are cognitive (44% pure cognitive, e.g., *ikacilo* 'noisy') or cognitive related (5% cognitive-behavior, e.g., *ikalma* 'lazy'), one third are affective (13% pure affective, e.g., *ikarilaw* 'sympathise'; 13% affective-behavioral, e.g., *ikami'ing* 'amused'; and 6% affective-cognitive, e.g., *ikagom* 'overbearing'), and less than one fifth are physical (19% pure physical, e.g., *ikabsoy* 'satiated'). Interestingly, if we had not investigated the *ika-* construction, we would not have discovered the saliency of the cognitive categories in Yami emotion.

[13] The procedure reduplicated the illustrative dendrograms in Church et al.'s study for comparison. The detailed steps of our ontological simulation for processing the *ika-* emotion terms supporting Church et al.'s study is described in Yang et al. (2012).

Table 5: Distribution of 126 Yami emotion terms (*ika-* verbs)

Internal conditions	Pure affective	Affective-behavioral	Affective-cognitive	Pure cognitive	Cognitive-behavioral	Pure physical
N=126	17	16	7	56	6	24
100%	13%	13%	6%	44%	5%	19%

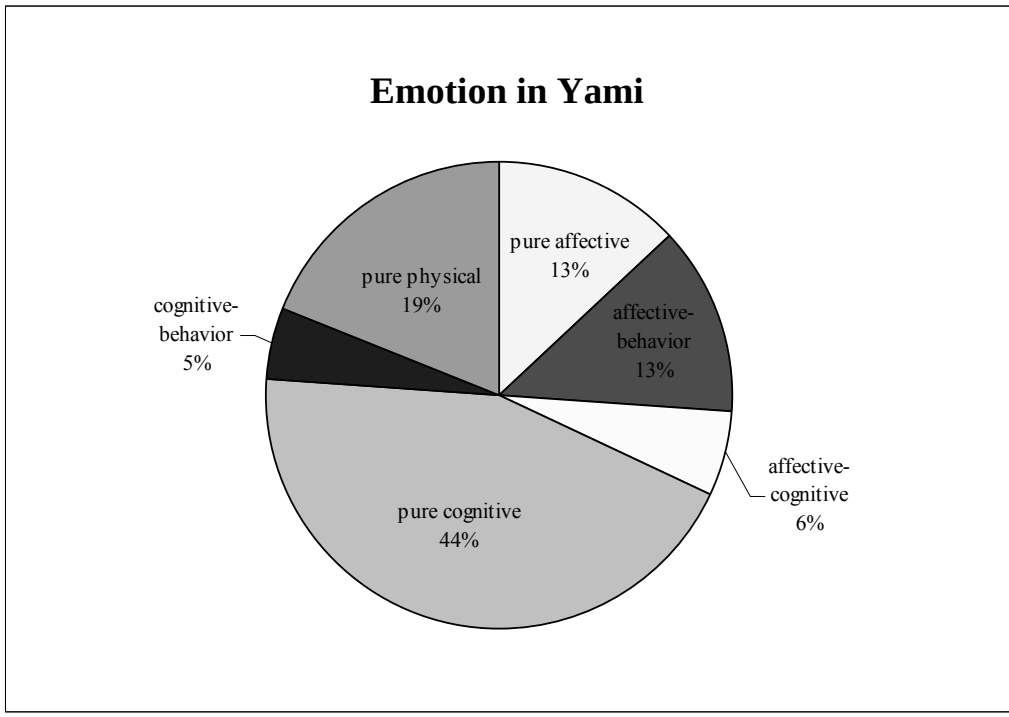


Figure 2: Classification of Yami Emotion

4.2 Positive and negative Yami emotion

As 'good' and 'bad' are part of the universal primitives in Wierzbicka's (1992) definition of emotion, we assume that emotion can be coded as a continuum of positive and negative feelings. Based on the Chinese translations of these emotion terms, we assigned all the Yami emotions to five levels: positive, relatively positive, neutral, relatively negative, and negative (see appendix). Relatively positive and relatively negative evaluations are determined in relation to prototypes of the two ends. For example, *ikasaray* 'happy' is taken as prototypically positive while *ika'oya* 'angry' prototypically negative. Relatively negative and relatively positive evaluations describe the positive and negative emotions with lesser degrees in comparison with the two extremes. For example, *ikanig* 'embarrassed, ashamed' is classified as relatively negative in relation to the negative end of 'angry'. Similarly, *ikabsoy* 'satiated' is classified as relatively positive in relation to the positive end of 'happy'.

Some emotion terms may not be "valenced" or evaluated as positive or negative. For example, 'surprise' and 'amazement' do not imply anything good or bad (Wierzbicka

1992:550); therefore, it is necessary to have a neutral category when it comes to evaluation of emotion, such as *ikakaha* 'sleepy'.

The results indicate that Yami has more negative emotions, as shown in Figure 3. The distribution of Yami emotion is negative 47% (23% negative, e.g., *ika'oya* 'angry'; 24% relatively negative, e.g., *ikanig* 'embarrassed, ashamed'), neutral 30% (e.g., *ikakaha* 'sleepy'), and positive 23% (12% positive, e.g., *ikasaray* 'happy'; 11% relatively positive, e.g., *ikabsoy* 'satiated').

This assignment of postive and negative emotion based on Chinese–English translation remains tentative, and requires further corroboration with native Yami speakers. In particular, the neutral category and the fine-grained classifications of the relatively positive and relatively negative evaluations of Yami emotions may require a future field study to elicit the 'emic' judgments from the Yami speech community to validate our classification. However, the generalisation that Yami has more prototypically negative (23%) than positive type of emotion probably still holds, given the low percentage of positive emotion (12%).

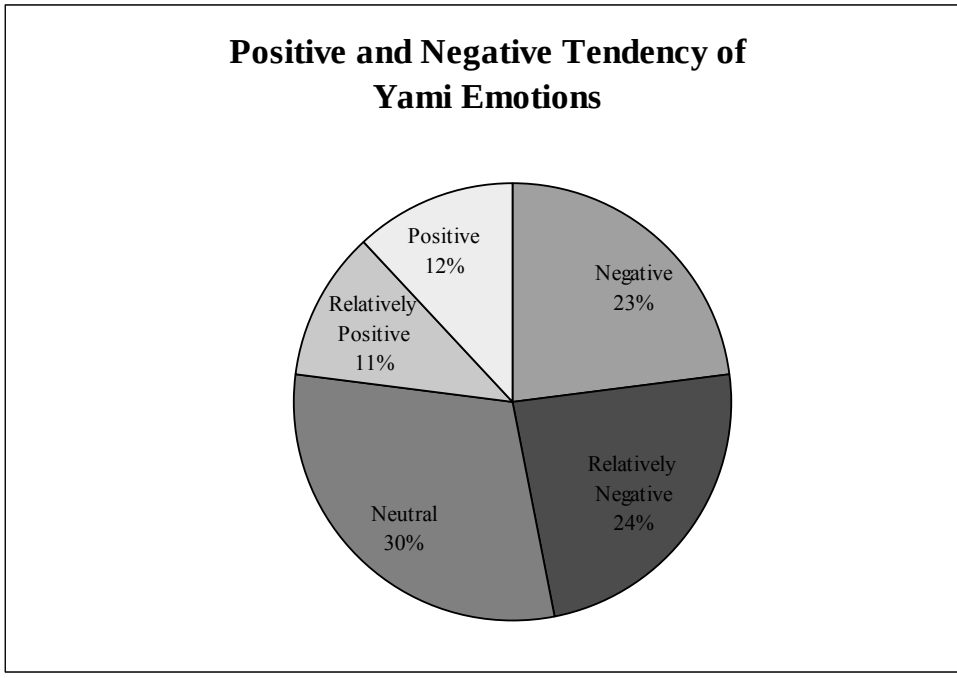


Figure 3: Distribution of positive and negative emotion terms in Yami

4.3 Semantic categories of Yami emotion

The same set of emotion data was further categorised based on near synomyms. Table 6 lists all the semantic categories of Yami emotion, divided into Clore et al.'s three categories: (i) affective conditions, (ii) cognitive conditions, and (iii) physical and bodily states.

Table 6: Semantic categories of Yami emotion

Clore et al.'s categories	Semantic categories of Yami emotion
Affective conditions	Anger, fear, happiness, longingness, love, nervousness, sorrow, sympathy, worry, evaluation of size, curse, forgetfulness, jealousy, noise/calmness, shame, stinginess, danger, overbearingness, relaxation
Cognitive conditions	Age, evaluation of size, boredness, certainty, cleverness, cold/heat, curiosity, danger/safety, darkness, difficulty, relaxation, external states of things, distance, fortune, goodness/badness, greatness, importance, independence, weight, noise/calmness, quantity, correctness, smoothness, states of human body, taboo, taste/smell, watchmacallit, diligence/laziness
Physical and bodily states	Disagreement, nitpicking, scheme

The same table can be further represented by the diagram drawn using the Protégé program to represent the Yami emotion ontology, as shown in Figure 4. The diagram also helps us visualise the overrepresentation of cognitive conditions in encoding Yami emotion. In addition, several semantic categories show cross-sectional representations, the most important of which is 'fear', occurring not only in pure affective and affective-cognitive conditions, but also in physical and bodily states.

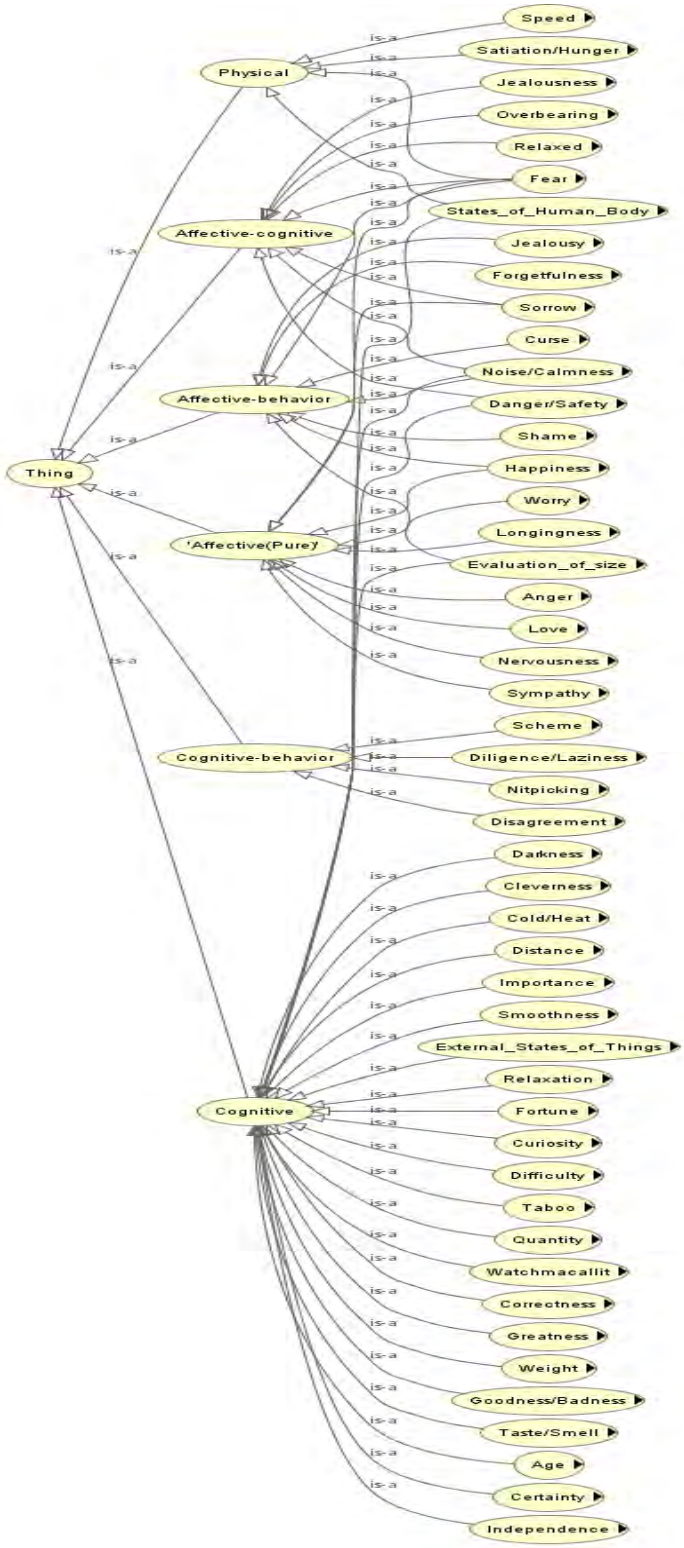


Figure 4: Ontology describing categories of near synonyms

4.4 Comparison of Yami emotions with Filipino emotions

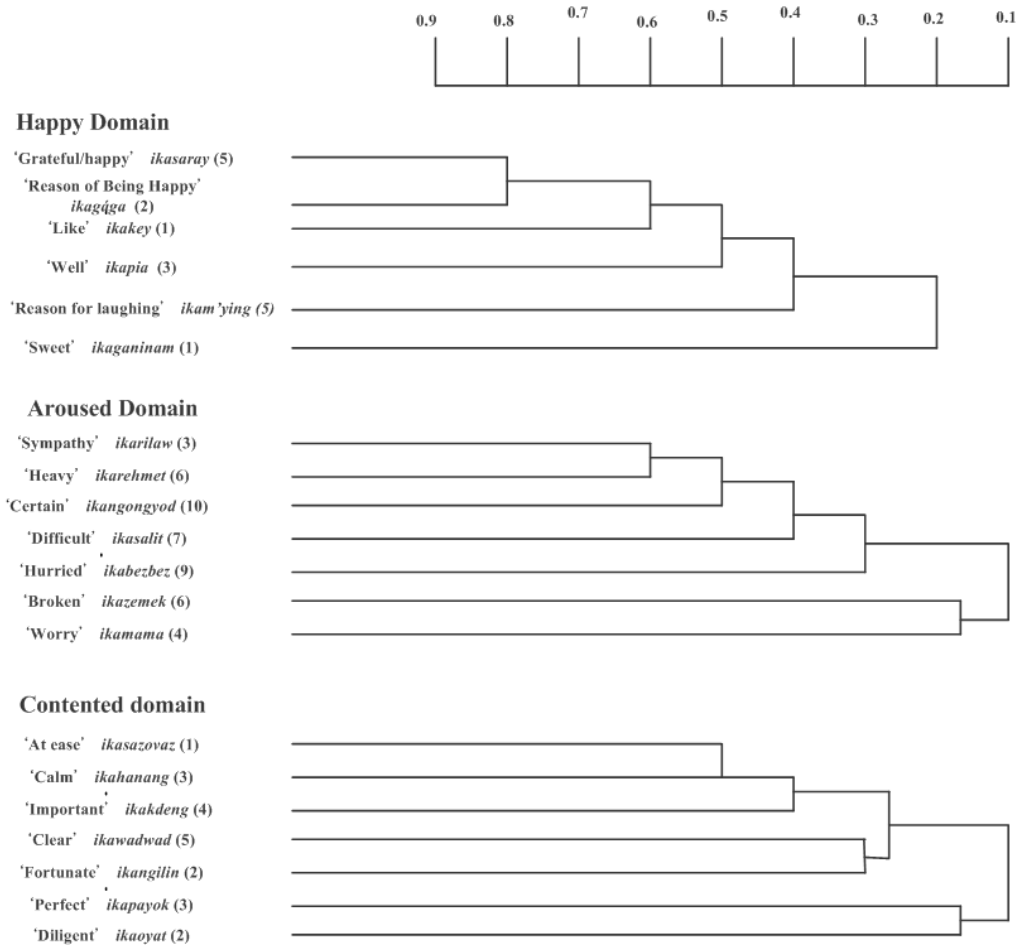
Our hierarchical cluster analysis of Yami emotion terms yielded eleven distinctive domains (see Figure 5), i.e., happy, aroused, contented, emotionless, contemptuous, angry, sad, tired, quiet/shy, anxious, and aspiring, almost identical to the Filipino emotion terms, except that the guilty domain is lacking in Yami. However, this does not mean Yami does not have the guilty domain. The Yami word for guilty *miraraten* (*mi-raraten*) is derived from the stem *raraten* 'guilt, sin' (< *rahet* 'bad'). As the *ika-* prefix would derive *ikarahet* 'consider something bad because…', instead of guilty, this explains why the guilty domain is missing from the cluster analysis.

How well these domains are clustered can be evaluated by the weight. Figure 5 depicts the Yami emotion clusters derived from the ontological calculation. The dendrograms of this figure show the selected clusters in each domain. The scale shown is a simulation calculation derived by the weight function of semantic distance between each cluster in the emotional ontology. This simulation is used to emulate the judgment process in Church et al.'s study. If the weight of the lexical word is close to 0.9, this indicates the word fits well in its semantic domain. On the other hand, if the weight is close to 0.1, the word does not fit the domain well.

The domains in Figure 5 form a hierarchy of fitness: happy (0.8), angry (0.8), anxious (0.7), aroused (0.6), sad (0.6), tired (0.6), emotionless (0.5), contented (0.5), contemptuous (0.4), quiet/shy (0.3), and aspiring (0.3). Thus, the cluster analysis provides further confirmation of our identification of the eight most important (or hypercognised) emotion categories and the three least important (or hypocognised) emotion categories in Yami, if we use 0.5 as an arbitrary threshold.

Although cluster analysis is a useful quantitative tool to yield preliminary results in our analysis, we cannot avoid cross-sectional reprentations. Some emotion terms, albeit a negligible minority, were categorised into two different domains. For example, *ikazoay* 'feel proud' was put in both "emotionless" and "aspiring" domains. This points out that assignment of lexical items to different semantic categories by the third author also awaits future validation by the speech community members.

In summary, our cluster analysis of Yami emotion identified all eleven emotion domains found by Church et al. (1998), except the guilty domain. The reason, as explained previously, is that *ika-rahet* 'consider something bad because …' only encodes the emotion of anxiety/worry, whereas guilty is encoded by *mi-raraten*. The most important emotions in Yami are anxiety/fear, arousal, contentment, anger, happiness, sadness, and "emotionless" (i.e., feeling bored and other cognitive conditions) whereas the least important emotion domains include feeling tired, quiet/shy, contemptuous, and aspiring. Our findings also confirm that the terms in all three subcategories of the affection conditions class in Clore et al.'s (1987) taxonomy of emotion terms are viewed as referring to emotions.


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Happy Domain
'Grateful/happy' ikasaray (5)
'Reason of Being Happy' ikagága (2)
'Like' ikakey (1)
'Well' ikapia (3)
'Reason for laughing' ikam'ying (5)
'Sweet' ikaganinam (1)
Aroused Domain
'Sympathy' ikarilaw (3)
'Heavy' ikarehmet (6)
'Certain' ikangongyod (10)
'Difficult' ikasalit (7)
'Hurried' ikabezbez (9)
'Broken' ikazemek (6)
'Worry' ikamama (4)
Contented domain
'At ease' ikasazovaz (1)
'Calm' ikahanang (3)
'Important' ikakdeng (4)
'Clear' ikawadwad (5)
'Fortunate' ikangilin (2)
'Perfect' ikapayok (3)
'Diligent' ikaoyat (2)

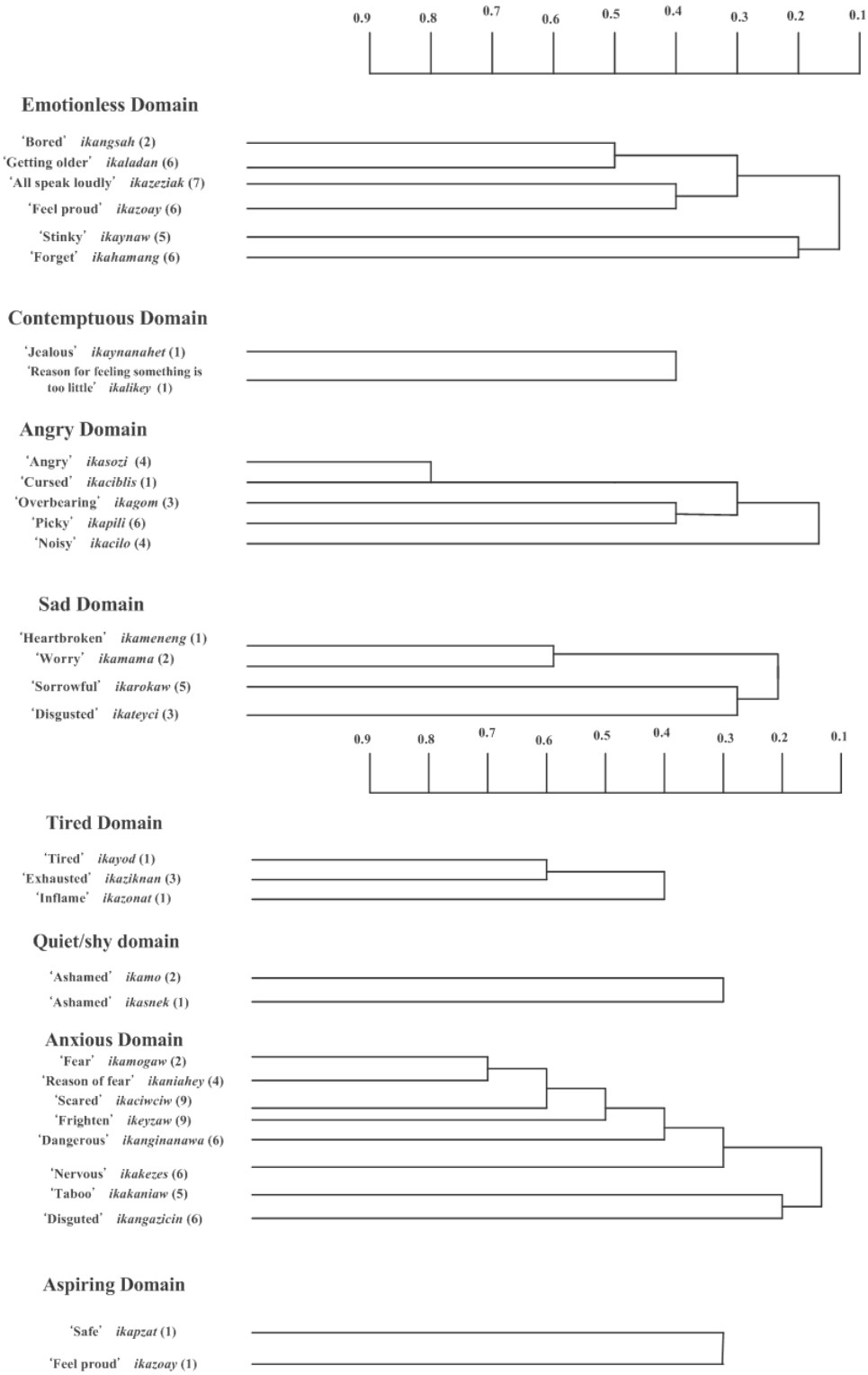


Figure 5: Hierarchical cluster analysis of Yami *ika-* emotion terms

4.5 Evaluation of the current methods

Before we conclude, an evaluation of the corpus linguistics–cum–ontology methods in the study of Yami emotion is in order. The strength of this interdisciplinary approach to investigate Yami emotion lies in the identification of a grammatical model, i.e., Yami prefix *ika-* 'the reason/cause for a certain feeling', as a keyword for systematic search in the online corpus. A detailed analysis of Yami narratives in step 2 resulted in further data on emotion that overlapped with the *ika-* words to yield the final 126 emotion terms in our analysis.

The usefulness of the computerised representations of the Yami ontology as shown in Figures 4 and 5 in computational linguistics depends solely on the accuracy of the corpus linguistic analysis of the Yami data. This prompted us to further validate our classification in future fieldwork. However, a word of caution is necessary here. Although the assignment of the emotion terms into different categories was determined by the first two authors solely on the basis of clear operational definitions, this was probably the best approach to our study given the circumstances. An attempt was made for the first author to bring the list of emotion terms (in the appendix) to Orchid Island for the community members to either confirm or reject the validity of our classification. However, this task incited a debate on what emotion is, and no consensus was reached. Nonetheless, the preliminary results we found in the study have paved the way for two immediate follow-up studies: (1) an evaluation of positive and negative Yami emotion by community members using a five-point Likert scale and (2) a classification of emotion terms into the eleven domains in the cluster analysis by community members.

5 Conclusion

This study has classified Yami emotion into six internal conditions. Like Filipino, the most important emotions in Yami are anxiety/fear, arousal, contentment, anger, happiness, sadness, and emotionless (e.g., feeling bored), whereas the least important emotion domains include feeling tired, quiet/shy, contemptuous, and aspiring. In general, Yami emotions basically share the same distinct domains as Filipino emotions and similarly the language contains more negative emotion terms than positive ones.

This study, albeit preliminary, has demonstrated how to use an interdisciplinary approach to study Yami ontology, using emotion as a semantic domain. Future investigation can apply the same methods to cover wider semantic domains in preparation for building a complete ontology of a targeted language and culture.

Appendix

Classification of the 126 Yami emotion terms

Positive		
Yami	English	Category (Church et al. 1988)
ikagága	'happy'	pure
ikáglaw	'love'	pure
ikaizay	'great'	cognitive
ikakey	'like (something)'	pure
ikákza	'like (someone, something)'	pure
ikamiying, ikami'ing	'laugh'	affective-behavior
ikapatak	'competent'	cognitive
ikapayok	'perfect'	cognitive
ikapia	'good, well'	cognitive
ikarilaw	'sympathetic'	pure
ikasaray	'thankful/happy'	affective-behavior
ikasi	'sympathetic'	pure
ikateneng	'intelligent'	cognitive
ikaveyvow	'independent'	cognitive
ikazoay	'feel proud, glorious'	affective-behavior

Relatively Positive		
Yami	English	Category (Church et al. 1988)
ikabsoy	'satiated'	physical
ikacigzang	'strong, hard'	physical
ikahanang	'calm'	affective-cognitive
ikakdeng	'important'	cognitive
ikalamnay	'relaxed'	cognitive
ikamoay	'plump'	cognitive
ikaoyat	'diligent'	cognitive-behavior
ikapzat	'safe'	cognitive
ikaraevaes	'fitting'	cognitive
ikasazovaz	'relaxing'	affective-behavior
ikasingat	'important/expensive'	cognitive
ikasonong	'smooth'	cognitive
ikawadwad	'clear, clean'	cognitive
ikeylamnay[14]	'relaxed'	affective-cognitive

[14] It may be a variant of *ikalamnay* 'relaxed'.

Neutral		
Yami	English	Category (Church et al. 1988)
ika'amang	'fainted'	physical
ikabezbez	'hurried'	physical
ikaganinam	'sweet'	cognitive
ikahamang	'forget'	affective-behavior
ikahangno	'smells fragrant'	cognitive
ikahep	'dark'	cognitive
ikahithitkahen	'love to sleep'	physical
ikakaha	'sleepy'	physical
ikakoan, ikaikikoan	'watchmacallit'	cognitive
ikakoat	'burning heat'	cognitive
ikakopad	'bitter'	cognitive
ikaladan	'older'	cognitive
ikalak	'curious'	cognitive
ikalavayo	'young'	cognitive
ikanehed	'true/certain'	cognitive
ikangilin	'lucky'	cognitive
ikangongyod	'real, certain'	cognitive
ikangot	'smelly, stinky'	cognitive
ikangsah	'bored'	cognitive
ikanoyong	'real'	cognitive
ikaotok	'bored'	cognitive
ikapait	'salty'	cognitive
ikapaw	'miss (someone, something)'	pure
ikarehmet	'heavy'	cognitive
ikarekmeh	'cold (weather)'	cognitive
ikasagpaw	'heavy'	cognitive
ikasinasina	'divergent'	cognitive-behavior
ikasngen	'too close'	cognitive
ikasoliket	'sticky'	cognitive
ikateleh	'deaf'	cognitive
ikavaheng	'black'	cognitive
ikavaw	'cool'	cognitive
ikavokay	'dry (powder)'	cognitive
ikaynaw/ ikeynaw	'stinky, fishy'	cognitive
ikazemek	'broken'	cognitive
ikazeziak	'all speak loudly'	affective-behavior
ikehma	'soft'	cognitive
ikeyngen	'muscle ache'	physical

Relatively Negative		
Yami	English	Category (Church et al. 1988)
ikaa	'gluttonous'	physical
ikaciciaw	'too talkative'	cognitive
ikacilo	'noisy'	cognitive
ikagolang	'thin'	physical
ikagom	'overbearing'	affective-cognitive
ikahango	'seasick'	physical
ikahen	'very cold'	cognitive
ikakaram	'as small as a mouse'	cognitive
ikakcin	'hungry'	physical
ikalanan	'gluttonous'	physical
ikalikey	'too small'	affective-behavior
ikalinlin	'faint, too weak to stand'	physical
ikalita	'scheme'	cognitive-behavior
ikalotoy	'bulging stomach'	physical
ikamez	'chilly'	cognitive
ikamo	'embarrassed'	affective-behavior
ikanig	'ashamed, embarrassed'	affective-behavior
ikapereh	'few'	cognitive
ikapili	'picky'	cognitive-behavior
ikaraway	'become ugly'	cognitive-behavior
ikasaki	'drunk'	physical
ikasnek	'ashamed'	affective-behavior
ikasngisngit	'sharp pain'	physical
ikaspet	'dangerous or complicated'	cognitive
ikatahaw	'weak'	physical
ikateyci	'feel disgusted'	physical
ikayod	'tired'	physical
ikazazomay	'sick of eating something'	physical
ikaziknan	'exhausted'	physical
ikazonat	'inflame'	physical

Negative		
Yami	English	Category (Church et al. 1988)
ikaciblis	'cursed'	affective-behavior
ikaciwciw	'scared away'	affective-behavior
ikakaniaw	'taboo'	cognitive
ikakezes	'nervous'	pure
ikalag	'taboo'	cognitive
ikalas	'wrong'	cognitive
ikálaw	'worried'	pure
ikalma	'lazy'	cognitive-behavior
ikaloit	'dirty'	cognitive
ikamáma	'worried'	pure

Negative		
Yami	English	Category (Church et al. 1988)
ikameneng	'heartbroken'	pure
ikamogaw	'afraid'	pure
ikananat	'scared, as if the stomach is lifted'	physical
ikangazicin	'disgusted'	physical
ikanginanawa	'dangerous'	affective-cognitive
ikaniahey	'afraid'	pure
ika'ogto	'frightened'	pure
ika'oya	'angry'	pure
ikarahet	'bad/feel upset'	cognitive
ikararaten	'stingy'	affective-behavior
ikarokaw	'lonely, sorrowful'	affective-cognitive
ikasalit	'difficult'	cognitive
ikasozi	'angry (violent)'	pure
ikatamoad	'embarrassing'	affective-behavior
ikavozoaw	'scared away'	affective-behavior
ikaynanahet	'jealous'	affective-behavior
ikaywam/ ikeywam	'afraid'	pure
ikeynanahet	'selfish/jealous'	affective-cognitive
ikeyzaw	'frightened'	affective-cognitive

References

Church, A. Timothy, Marcia S. Katigbak, Jose Alberto S. Reyes and Stacia M. Jensen. 1998. Language and organisation of Filipino emotion concepts: Comparing emotion concepts and dimensions across cultures. *Cognition and Emotion* 12.1:63–92.

Clore, Gerald L., Andrew Ortony and Mark A. Foss. 1987. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology* 53.4:751–766.

Dirven, René. 1997. Emotions as cause and the cause of emotions. In *The language of emotions: Conceptualization, expression, and theoretical foundation*, ed. by Susanne Niemeier and René Dirven, 55–83. Amsterdam and Philadelphia: John Benjamins.

Dong, Maa-Neu, Victoria D. Rau and Ann Hui-Huan Chang. (eds.) 2012. *dawu yu cidian* [*Yami* (*Tao*) *dictionary*]. Taipei: National Taiwan University Press.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Cognitive Theory of Language and Culture Series. Chicago: University of Chicago Press.

Hsieh, Fuhui. 2011. The conceptualization of emotion events in three Formosan languages: Kavalan, Paiwan, and Saisiyat. *Oceanic Linguistics* 50.1:65–92.

Hu, Jackson and Victoria D. Rau. 2013. dawu yuming de shengtai tixian han renzhi yuyi [Ecological embodiment and cognitive semantics of Tao fish names] In *Analyzing language and discourse as intercultural and intracultural mediation,* ed. by Ming-Yu Tseng, 187–227. Kaohsiung: National Sun Yat-sen University. [In Chinese]

Huang, Shuanfan. 2002. Tsou is different: A cognitive perspective on language, emotion, and body. *Cognitive Linguistics* 13.2:167–186.

Lakoff, George and Mark Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Levy, Robert I. 1984. The emotions in comparative perspective. In *Approaches to emotion*, ed. by Klaus R. Scherer & Paul Ekman, 397–412. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rau, D. Victoria and Maa-Neu Dong. 2010. Serves you right: *Ka-* as an attitudinal marker in Yami. Paper presented at the 4th Conference on Language, Discourse, and Cognition. 1-2 May 2010. Taipei: National Taiwan University.

Rau, D. Victoria and Maa-Neu Dong. Forthcoming. *dawu yu yufa cankao shu* [*Yami reference grammar*] [In Chinese]

Rau, D. Victoria and Maa-Neu Dong. 2006. *Yami texts with reference grammar and dictionary*. Language and Linguistics Monograph Series A-10. Taipei: Academia Sinica.

Rau, D. Victoria, Yi-Hsin Wu and Meng-Chien Yang. 2012. What can *ika-* tell us about Yami emotion? Paper presented at the 2012 International Conference on

Landscape, Seascape, and the Spatial Imagination. 2–3 November, Kaohsiung: National Sun Yat-sen University.

Rau, D. Victoria, Yi-Hsin Wu, Meng-Chien Yang and Jackson Hu. 2012. What can *ika-* tell us about Yami emotion? Poster presented at the *2012 CLDC*, 4–6 May. Taipei: National Taiwan University.

Rau, D. Victoria, Meng-Chien Yang, Hui-Huan Ann Chang and Maa-Neu Dong. 2009. Online dictionary and ontology building for Austronesian languages in Taiwan. *Journal of Language Documentation and Conservation* 3.2:192–212. http://nflrc.hawaii.edu/ldc/.

Saucier, Gerard and Lewis R. Goldberg. 1996. The language of personality: Lexical perspective on the five-factor model. In *The five-factor model of personality: Theoretical perspectives*, ed. by Jerry S. Wiggins, 21–50. New York: Guilford Press.

Tai, Yin-Sheng, D. Victoria Rau and Meng-Chien Yang. 2008. A semantic study on Yami ontology in traditional songs. IJCNLP 2008. In *Proceedings of the 6th Workshop on Asian Language Resources*, 81–84, 11–12 January. Hyderabad, India. http://www.aclweb.org/anthology/I/I08/I08-7011.pdf.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Yang, Meng-Chien, D. Victoria Rau and Yi-Hsin Wu. 2012. Analyzing and classifying the Yami emotion phrases using ontological structure and computation. Paper presented at the International Conference on Asian Language Processing (IALP), 13–15 November. Hanoi, Vietnam.

Wierzbicka, Anna. 1992. Defining emotion concepts. *Cognitive Science* 16.4:539–581.

Wolff, John U., with Maria Theresa C. Centeno and Der-Hwa V. Rau. 1991. *Pilipino through self-instruction*. Ithaca, NY: Cornell University Press.

Zeitoun, Elizabeth and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391–414.

Editors and Contributors

Editors

• ELIZABETH ZEITOUN is Research Fellow at the Institute of Linguistics, Academia Sinica (hsez@gate.sinica.edu.tw). Her major research interests lie in Austronesian linguistics, morphosyntax, typology, and historical linguistics. She has published a number of sketch grammars in Chinese, two books in English on Mantauran Rukai (Zeitoun & Lin 2003; Zeitoun 2007) and another on Saisiyat (Zeitoun et al 2015). She specializes in Rukai, Saisiyat, and Tsou and has worked on nearly all the Formosan languages. She was Acting Editor and later Executive Editor of *Language and Linguistics* from April 2008 through June 2013.

• STACY F. TENG is Associate Research Fellow at the Institute of Linguistics, Academia Sinica (stacyt@gate.sinica.edu.tw). Her major research interests lie in Austronesian linguistics, morphosyntax, typology, and historical linguistics. She has published her Ph.D dissertation (Teng 2008) as well as a couple of papers on the comparative study of the Puyuma dialects. She specializes in Puyuma and has started to work on the Tsouic languages (in particular Kanakanavu).

• JOY J. WU is Associate Professor at the Department of English, National Taiwan Normal University (jjwu@ntnu.edu.tw). Her major research interests lie in Austronesian linguistics, morphosyntax, typology and functional linguistic theory (Role and Reference Grammar, RRG). She specializes in Amis and was co-editor of *Concentric: Studies in Linguistics* from 2009 to 2010, and 2014.

Contributors

• HENRY Y. CHANG is an Associate Research Fellow at the Institute of Linguistics, Academia Sinica and a jointly appointed Associate Professor at the Graduate Institute of Linguistics, National Tsing Hua University (henryylc@gate.sinica.edu.tw). He specializes in Austronesian formal syntax and semantics. He has been an Associate Editor for *Language and Linguistics* since July 2013. Details of his personal information and publications are available at http://idv.sinica.edu.tw/henryylc.

• YUEHCHEN CHIEN is Associate Professor at the Department of Indigenous Languages and Communication, National Dong Hwa University (chienyc@mail.ndhu.edu.tw). Her major research interests lie in contact linguistics, sociolinguistics, and Japanese linguistics. She has published a number of articles on Taiwan Japanese and on Yilan Creole, and two books, one on the Japanese language as it is spoken outside of Japan as a result of colonialism and migration, and the other specifically on its use as a lingua franca in Taiwan.

• FUHUI HSIEH is Professor at the Department of Applied Foreign Languages, Tatung University (hsiehfh@ttu.edu.tw). Her major research interests lie in

Austronesian linguistics, syntax, semantics, pragmatics, and discourse analysis. She specializes in Kavalan, Paiwan, and Saisiyat.

• HUI-CHUAN J. HUANG is Associate Research Fellow at the Institute of Linguistics, Academia Sinica and a jointly appointed Associate Professor at the Graduate Institute of Linguistics, National Tsing Hua University (hcjhuang@gate.sinica.edu.tw). Her major research interests lie in Austronesian linguistics, phonology, phonetics, and historical linguistics. She has published a number of articles on the metrical and syllabic structures of Bunun and Atayal and on issues relating to glides.

• AMY PEI-JUNG LEE is Associate Professor at the Department of Indigenous Languages and Communication, National National Dong Hwa University (aplee@mail.ndhu.edu.tw). Her major research interests lie in phonology, morphology, typology, and cognitive linguistics of Austronesian languages. She has published papers on Seediq and Paiwan phonology, reduplication, metaphorical euphemism, and the language of olfaction in several Formosan languages.

• PAUL JEN-KUEI LI is Adjunct Research Fellow at the Institute of Linguistics and Academician at Academia Sinica (paulli@gate.sinica.edu.tw). He is a leading specialist on Formosan languages. He has made a lifelong study of all these languages, and has published 16 monographs, including dictionaries on the endangered Pazih and Kavalan languages.

• DONG-YI LIN is a postdoctoral researcher at the Department of Linguistics (English Studies), Ghent University (dongyi.lin@ugent.be). His research focuses on the morphosyntax of Austronesian languages in Taiwan, especially Amis and Kavalan. His current project investigates the implications of interrogative and adverbial verbs for the Cartographic approach to Syntax.

• GUJING LIN is Assistant Professor at Tzu Chi University (paicxeevi@mail.tcu.edu.tw). She works extensively on the Tsou language and her interests lie in argument structure, complex predicate formation, construction grammar, and morpho-phonology. She is currently engaged in analyzing stem alternation in the Tsou voice paradigm.

• HUI-SHAN LIN is Professor at the Department of English, National Taiwan Normal University (hslin@ntnu.edu.tw). Her major research interests lie in Austronesian phonology, Chinese phonology and Optimality Theory. She has worked on reduplication in Pazeh, Kavalan, Tsou, and Atayal and was co-editor of Conc*entric: Studies in Linguistics* from 2010 to 2011.

• ADLAY KUN-LONG LIU is Section Chief of Experiment & Research at the Academic Affairs Office, Taipei Municipal Shilin High School of Commerce (adlay.k.liu@gmail.com). His major research interests lie in Austronesian morphosyntax, information structure and theoretical syntax (in particular Lexical-Functional Grammar & the Minimalist Program). He specializes in Atayal.

• CHIA-JUNG PAN is Lecturer at the School of Literature, Nankai University, China (chiajung.pan216@gmail.com). His major research interests lie in morphosyntax, linguistic typology, Basic Linguistic Theory, language contact, and language obsolescence. He has published his MA thesis (Pan 2010) as well as some papers on Saaroa. He is specialized in Saaroa and Tsou, and has started to work on Kanakanavu.

• VICTORIA RAU is Professor of linguistics at the Institute of Linguistics, National Chung Cheng University (Lngrau@ccu.edu.tw). Her research interests include Austronesian linguistics, sociolinguistics, and applied linguistics. She has published Yami texts with reference grammar and vocabulary (2006), Yami dictionary (2012), and three websites on Yami language documentation, e-learning, and online dictionary (2005-2008) in collaboration with Maa-neu Dong.

• MALCOLM ROSS is an Emeritus Professor in Linguistics in the College of Asia and the Pacific at the Australian National University (Malcolm.Ross@anu.edu.au). His research area is the reconstruction of the histories of the Austronesian languages of New Guinea and the Pacific and of the Papuan languages of New Guinea. Until recently he also worked on the early history of Austronesian, focusing on the Formosan languages. He has published numerous articles and edited several books on the Austronesian language family.

• LI-MAY SUNG is Associate Professor at the Graduate Institute of Linguistics, National Taiwan University (limay@ntu.edu.tw). The focus of her research has been in Formosan syntax and its interactions with morphology and semantics. She is also a field linguist with interests in Formosan language documentation, lexicography, typology and its applications. In recent years, she has been examining the nature of exclamatives in various Formosan languages and the nature of stance that is relevant to natural language syntax. Her other interests that have led to presentations and publications include cross-linguistic studies on comparative constructions, existentials and demonstratives, saying verbs and peripheral particles, and the nature of reflexivity/reciprocity.

• APAY AI-YU TANG is Assistant Professor at the Department of Indigenous Languages and Communication, National Dong Hwa University (apay@mail.ndhu.edu.tw). She is a speaker of Truku Seediq and her major research interests lie in language documentation and conservation, as well as Austronesian linguistics.

• WEI-TIEN DYLAN TSAI received his PhD degree in linguistics from Massachusetts Institute of Technology in 1994, and is currently Professor at the Graduate Institute of Linguistics, National Tsing Hua University (wttsai@mx.nthu.edu.tw). His recent research explores issues on the syntax-semantics interface under the cartographic approach, with an emphasis on Chinese, Austronesian and Slavic languages from a comparative point of view. These studies lead to a series of publications on the inner-outer dichotomy of adverbials, applicatives and light verbs, as well as an evolutional analysis of a cluster of phenomena collectively known as "conjunctive reduction". He has been International Editor of *Gengo Kenkyu* (Journal of

the Linguistic Society of Japan), and is now Co-editor of International *Journal of Chinese Linguistics*.

• NAOMI TSUKIDA completed her PhD degree in linguistics at the University of Tokyo and is currently Professor at Aichi Prefectural University, Japan. Her primary interests are in the fields of morphosyntax and typology and she specializes in Truku Seediq.

• YI-HSIN CATHERINE WU is a graduate student at the Institute of Linguistics, National Chung Cheng University (yscatherine@gmail.com). Her major research interests are cognitive linguistics and second language acquisition. She has been a research assistant for Prof. Victoria Rau, investigating Yami ontology and cognition.

• MENG-CHIEN YANG is Associate Professor at the Department of Computer Sciences and Communication Engineering, Providence University (mcyang2@pu.edu.tw). His research interests are digital archiving, computational technologies for Austronesian languages, and game-based learning design. He collaborated with Prof. Victoria Rau in Digital Archiving Yami Language Documentation, sponsored by SOAS, University of London in 2005-2007.

• MARIE M. YEH is an Associate Professor in the Institute of Taiwan Languages and Language Teaching at National Hsinchu University of Education (mlyeh@mail.nhcue.edu.tw). Her research interest lies mainly in the area of Austronesian cognitive functional linguistics. Recently, she has been working on the grammaticalization and discourse functions of demonstratives and PCU verbs in Saisiyat.

• MAYA YUTING YEH is an ethnic Atayal and is currently a postdoctoral researcher at the Institute of Linguistics, Academia Sinica (mayasquliqatayal@gmail.com). Her research interests center on Austronesian linguistics, cognitive linguistics, semantics, morphosyntax, and discourse analysis.

The present volume is a festschrift in honour of Lillian M. Huang, who, in a very few years, became a leading figure in Formosan linguistics after she obtained her PhD degree in 1987. It covers data from nearly all the extant Formosan languages, Atayal, Amis, Bunun, Kanakanavu, Kavalan, Rukai, Paiwan, Puyuma, Saaroa, Saisiyat, Seediq, Thao, Tsou, and Yami. It includes papers on new advances on the diachronic and synchronic phonology as well as the morpho-syntax of Formosan languages. New topics are explored, including an analysis of discourse in Saisiyat, and an investigation of the lexicon, from different perspectives: semantic extension, language contact and corpus linguistics.

Editors

- ELIZABETH ZEITOUN, Research Fellow at the Institute of Linguistics, Academia Sinica (hsez@gate.sinica.edu.tw)
- STACY F. TENG, Associate Research Fellow at the Institute of Linguistics, Academia Sinica (stacyt@gate.sinica.edu.tw)
- JOY J. WU, Associate Professor at the Department of English, National Taiwan Normal University (jjwu@ntnu.edu.tw)

Contributors

- HENRY Y. CHANG, Associate Research Fellow at the Institute of Linguistics, Academia Sinica and jointly appointed Associate Professor at the Graduate Institute of Linguistics, National Tsing Hua University (henryylc@gate.sinica.edu.tw)
- YUEHCHEN CHIEN, Associate Professor at the Department of Indigenous Languages and Communication, National Dong Hwa University (chienyc@mail.ndhu.edu.tw)
- FUHUI HSIEH, Professor at the Department of Applied Foreign Languages, Tatung University (hsiehfh@ttu.edu.tw)
- HUI-CHUAN J. HUANG, Associate Research Fellow at the Institute of Linguistics, Academia Sinica and jointly appointed Associate Professor at the Graduate Institute of Linguistics, National Tsing Hua University (hcjhuang@gate.sinica.edu.tw)
- AMY PEI-JUNG LEE, Associate Professor at the Department of Indigenous Languages and Communication, National National Dong Hwa University (aplee@mail.ndhu.edu.tw)
- PAUL JEN-KUEI LI, Adjunct Research Fellow at the Institute of Linguistics and Academician at Academia Sinica (paulli@gate.sinica.edu.tw)
- DONG-YI LIN, postdoctoral researcher at the Department of Linguistics (English Studies), Ghent University (dongyi.lin@ugent.be)
- GUJING LIN, Assistant Professor at Tzu Chi University (paicxeevi@mail.tcu.edu.tw)
- HUI-SHAN LIN, Professor at the Department of English, National Taiwan Normal University (hslin@ntnu.edu.tw)
- ADLAY KUN-LONG LIU, Section Chief of Experiment & Research at the Academic Affairs Office, Taipei Municipal Shilin High School of Commerce (adlay.k.liu@gmail.com)
- CHIA-JUNG PAN, Lecturer at the School of Literature, Nankai University, China (chiajung.pan216@gmail.com)
- VICTORIA RAU, Professor of linguistics at the Institute of Linguistics, National Chung Cheng University (Lngrau@ccu.edu.tw)
- MALCOLM ROSS, Emeritus Professor in Linguistics in the College of Asia and the Pacific at the Australian National University (Malcolm.Ross@anu.edu.au)
- LI-MAY SUNG, Associate Professor at the Graduate Institute of Linguistics, National Taiwan University (limay@ntu.edu.tw)
- APAY AI-YU TANG, Assistant Professor at the Department of Indigenous Languages and Communication, National Dong Hwa University (apay@mail.ndhu.edu.tw)
- WEI-TIEN DYLAN TSAI, Professor at the Graduate Institute of Linguistics, National Tsing Hua University (wttsai@mx.nthu.edu.tw)
- NAOMI TSUKIDA, Professor at Aichi Prefectural University, Japan
- YI-HSIN CATHERINE WU, graduate student at the Institute of Linguistics, National Chung Cheng University (yscatherine@gmail.com)
- MENG-CHIEN YANG, Associate Professor at the Department of Computer Sciences and Communication Engineering, Providence University (mcyang2@pu.edu.tw)
- MARIE M. YEH, Associate Professor in the Institute of Taiwan Languages and Language Teaching at National Hsinchu University of Education (mlyeh@mail.nhcue.edu.tw)
- MAYA YUTING YEH, postdoctoral researcher at the Institute of Linguistics, Academia Sinica (mayasquliqatayal@gmail.com)