

# Journal of Universal Language

Volume 13 Number 1

March 2012

---

- Nasser Al-Horais** On the Universality of Auxiliary Verbs
- Mohammad Dabir-Moghaddam** Linguistic Typology: An Iranian Perspective
- Hemanga Dutta** Spirantization and the Notion of Phonological Strength Relations in Assamese: An Optimality Theoretic Account
- Eun-Joo Kwak** Typological Accounts for Nominal Forms
- Alan Reed Libert** The Representation of Korean and Other Altaic Languages in Artificial International Auxiliary Languages

Language Research Institute  
**Sejong University**

# Journal of Universal Language

Publisher

**Woo-hee Park**

*President of Sejong University*

Editor-in-Chief

**Hye-kyeong Jeong**, *Department of Japanese Language and Literature, Sejong University*

Managing Editor

**Silo Chin**, *Department of English Language and Literature, Sejong University*

Editorial Board

**Kee-sun Choi**, *KIST (Korean Language Engineering)*

**Bernard Comrie**, *Max Planck Institute (Syntax)*

**Matthew Crocker**, *Saarland University (Computational Linguistics)*

**Stuart Davis**, *Indiana University (Phonology)*

**Laurent Dekydtspotter**, *Indiana University (Syntax)*

**Jared Diamond**, *University of California at Los Angeles (Physiology and Biology)*

**Daniel Dinnsen**, *Indiana University (Phonology)*

**Micheal Gasser**, *Indiana University (Computational Linguistics)*

**Il-dong Han**, *Yongin University (Translation)*

**Pauline Jacobson**, *Brown University (Semantics)*

**Aravind Joshi**, *University of Pennsylvania (Computational Linguistics)*

**Andrew Large**, *McGill University (Soviet and East European Studies)*

**Se-soon Lee**, *Chungang University (Translation)*

**Alan Libert**, *University of Newcastle (Morphosyntax)*

**Christo Moskovsky**, *University of Newcastle (Syntax and Second Language Acquisition)*

**David Odden**, *Ohio State University (Phonology)*

**Steve Pulman**, *Oxford University (Philosophy and Phonetics)*

**Hyun-ju Ryu**, *Pusan University of Foreign Studies (Translation)*

**Ji-sun Shin**, *Ewha Womans University (Translation)*

**Jyunichi Tsujii**, *University of Tokyo (Computational Linguistics)*

## Submission of Manuscripts

The journal is published biannually (in spring and fall) by the Language Research Institute, Sejong University, Seoul, Korea. Submission of a paper to the journal will be taken to imply that it represents original work not previously published, that it is not being considered elsewhere for publication, and that if accepted for publication it will not be published anywhere without the consent of the Editor. Upon acceptance of an article by the journal, authors will be asked to transfer copyright of the article to the publisher.

Manuscripts should be written in English. The title page should contain the title, the name, the full postal address, and the telephone number and email of each author, and indicate which author is responsible for correspondence and proofs. Figures and tables should have descriptive titles. The manuscripts should be accompanied by an abstract of not more than 200 words, and the preferred maximum length of manuscripts is 40 typed pages. Manuscripts should be sent to the Managing Editor, Language Research Institute, Sejong University, Seoul, Korea, or to the following email address: unish@sejong.ac.kr.

## Subscriptions

Subscription prices are US \$30 per year for individuals and US \$50 per year for institutions. All issues are normally sent by surface. An additional US \$15 per year is required for air delivery. Checks and money orders are payable to Sejong University.

## Language Research Institute

Sejong University

Gunja-dong, Gwangjin-gu, Seoul, Korea 143-747, Tel: +82-2-3408-3335 Fax: +82-2-3408-3566  
<http://www.unish.org>

# **Journal of Universal Language**



# Journal of Universal Language

Volume 13 Number 1

March 2012

---

- Nasser Al-Horaïs** On the Universality of Auxiliary Verbs
- Mohammad Dabir-Moghaddam** Linguistic Typology: An Iranian Perspective
- Hemanga Dutta** Spirantization and the Notion of Phonological Strength Relations in Assamese: An Optimality Theoretic Account
- Eun-Joo Kwak** Typological Accounts for Nominal Forms
- Alan Reed Libert** The Representation of Korean and Other Altaic Languages in Artificial International Auxiliary Languages

Language Research Institute  
**Sejong University**

***Journal of Universal Language***

Volume 13 Number 1 March 31, 2012

All rights reserved. No reproduction, copy, or transmission of  
this publication may be made without written permission.

Sejong University

Gunja-dong, Gwangjin-gu, Seoul, Korea 143-747

Tel: +82-2-3408-3335

Fax: +82-2-3408-3566

(Web site) <http://www.sejong.ac.kr>

<http://www.unish.org>

ISSN 1598-6381

Journal of Universal Language 13-1  
March 2012, 7-30

## On the Universality of Auxiliary Verbs\*

**Nasser Al-Horais**  
*Qassim University, KSA*

### Abstract

Cross-linguistically, it has often been observed that there is no any specific language-independent formal definition that can be used to determine the characterization of any given element as an auxiliary verb. This paper, though it agrees to some extent with this observation, argues that there is still room to find some universal properties that help us end up with the conclusion that auxiliaries and lexical verbs are two distinct types of syntactic entities. To this end, this paper describes the characteristics

---

\* I would like to express my deepest gratitude to Qassim University for supporting this work. Here I would like to extend my sincerest thanks and appreciation to the University's Rector Prof. Khalid Al-Humodi, and his Vice-Rectorate for Graduate Studies & Research: Prof. Abdulrahman Al-Wasel for having created the "right environment" for scientific research in Qassim University. Also, I extend my thanks to Anders Holmberg, Abdelkader Fassi Fehri, Jeffrey Pool, and Maggie Tallerman for insightful comments, suggestions, and discussions on an early version of this paper. I am grateful to Mohammad Al-Moaily for his invaluable help during all phases of this work. Special thanks also go to the anonymous reviewers and editors for clarifying my ideas and for significant editorial improvements which I tried to put to use here. I am solely responsible for any mistakes or inadequacy.

Nasser Al-Horais  
Arabic Language Department, Qassim University  
P.O.Box 6611, Buraidah 51425, Saudi Arabia  
Phone: 00966556063595; Email: nasser.alhorais@gamil.com

Received September 6, 2011; Revised November 1, 2011; Accepted November 29, 2011.

## 8 On the Universality of Auxiliary Verbs

necessary for what is to count as an auxiliary verb. Having done that, the paper turns to illustrate the most common properties that can be found among languages in order to find a clear universally agreed definition of an auxiliary verb.

Keywords: auxiliary verb, lexical verb, morphological information, theta role, negation, structural property, NICE property

### 1. Introduction

Delimiting properties of auxiliary verbs vis à vis lexical verbs has been the topic of continuous debate in generative grammar. It has, as stated by Heine (1993: 26), “provided one of the most popular battlegrounds for disputes of linguistic theory.” Although it is quite widely accepted that an auxiliary verb can be defined as “an element that in combination with a lexical verb forms a mono-clausal verb phrase with some degree of (lexical) semantic bleaching that performs some more or less definable grammatical function” (Anderson 2006: 5, cf. Heine 1993: 70<sup>1</sup>), there is, in fact, no agreed formal definition that can be adopted to determine the characterization of any given element as an auxiliary verb. In this regard, Anderson (To Appear, p. 1, n. 3) considers any definition of auxiliary verb as admittedly somewhat nebulous because he strongly believes that:

There is no, and probably cannot be, any specific, language-independent criteria that can be used to determine the characterization of any given element as a

---

<sup>1</sup> Heine (1993: 70) defines auxiliary verb as “an item on the lexical verb to functional affix/particle continuum, which tends to be at least somewhat semantically bleached, and grammaticalized to express one or more of a range of salient verbal categories, most typically aspectual and modal categories, but also not infrequently temporal, negative, or voice categories.”

lexical verb (including in serialized functions) or an auxiliary verb. As in all scalar, gradual, or gradient phenomena, clines of grammaticalization and semantic bleaching will have ‘gray areas,’ where the element in question has accrued certain features generally associated with end-points or focal points on the continuum, but perhaps not others. It seems likely that the degree of grammaticalization and semantic bleaching deemed sufficient to stop calling some particular verbal element X<sub>v</sub> usages of lexical verb X<sub>lv</sub> and start calling it auxiliary verb X<sub>av</sub> will vary from researcher to researcher, even when working on the same language.

The problems surrounding the definition of auxiliaries, resulting in not finding a clear universally agreed definition, relate to various reasons. According to Kuteva (2001: 5), one of these reasons is that:

There exists no clear demarcation line between periphrastic verb expressions and bound morphemes—all the more so since one and the same form may often have two different synchronic statuses, as a separate word and as a bound morpheme. Even if the form is readily recognized as a separate word, it will very often be considered by some linguists as a lexical verb and by others as an auxiliary—something which is referred to as the ‘amphibian’ nature of auxiliaries.

As pointed out above by Anderson, another reason could be derived from the fact that properties of auxiliaries differ from one language to another and hence it is hard to construct a consistent set of criteria to define (and thereby identify) auxiliaries.

This paper mainly aims to address the question of what is to count as an auxiliary verb. More specifically, the current discusses

the kind of properties which have been taken into account in characterizing the auxiliary class in syntactic theory in order to find crosslinguistic criteria which set auxiliaries verbs apart from lexical verbs.

The current paper is organized as follows. Section 2 considers the grammatical properties distinguishing auxiliaries from lexical verbs by language-independent characteristics. Section 3, by contrast, shows that defining auxiliaries is further complicated by the fact that there are some properties of auxiliaries that are specific to some languages, and do not extend to other languages auxiliaries. This can be seen in English, some Uralic languages (e.g., Finnish), and Tamil. Section 4 concludes the paper, and shows that although it is extremely difficult to decide what a good auxiliary should look like, there is still room to find some universal properties that help us argue that auxiliaries and lexical verbs are two distinct types of syntactic entities.

## 2. Some Properties of Auxiliaries

Cross-linguistically, the formal characteristics which delimit auxiliary from lexical verbs can be divided into two types: *structural* and *thematic*. Both of these types are different from one language to another, even between languages of the same family. Let us first begin by the structural properties.

### 2.1. Structural Properties

One of the most distinguishing structural properties of verbs which can function as auxiliary verbs is that auxiliaries are strongly associated with verbal rather than with nominal elements of a sentence. “They ‘fill out’ add to or further specify the predicate rather than its arguments” (Berman 1980: 17). Accordingly, it has been argued that auxiliaries must be in combination with a

lexical verb, forming a mono-clausal verb phrase (Anderson 2006: 5). This can be noticeably seen in a number of languages such as Korean (1), Tamil (2), and Khalkha Mongolian<sup>2</sup> (3) where, in these languages, no element can intervene between the auxiliary and the lexical verb.

- (1) a. John-un chayk-ul ppali ilk-e peli-ess-ta.  
John-Top book-Acc quickly read throw away-Past-Decl  
“John quickly completed reading a book.”
- b. \*John-un chayk-ul ilk-e ppali peli-ess-ta.  
John-Top book-Acc read quickly throw away-Past-Decl  
(Suh 2000: 107)
- (2) a. avan̄ porudkal-ai vāŋki mudi-t-t-ān̄.  
he-Nom groceries-Acc bought-Inf has-Perf-3ms  
“He has bought groceries.”
- b. \*avan̄ vāŋki porudkal-ai mudi-t-t-ān̄.  
he-Nom bough-Inf groceries-Acc has-Perf-3ms  
[Thanks to Mohana Ramasamy for the Tamil data.]
- (3) a. Egč nom uns-an̄.  
elder:sister book read-NPast  
“The elder sister reads a book.”

---

<sup>2</sup> In the interests of consistency, some of the examples are reproduced in this paper in a form slightly different from that in which they appear in the sources cited. Abbreviations used in this paper are as follows:

1, 2, 3—first, second, third person; Acc—accusative; Aux—auxiliary; Conn—connective ending; Decl—Declarative; f—feminine; Fut—future; Gen—genitive; Indic—Indicative; Inf—infinitive; m—Masculine; Neg—Negative; Nom—nominative; NP—Noun Phrase; p—Plural; PastPart—Past Participle; Part—Particle; Perf—perfective; Pot—Potential; Pres—Present; s—singular; tA—Infinitive with—tA(*non-finite verb morphology in Finnish*); Top—topic; V—verb

12 On the Universality of Auxiliary Verbs

- b. Egč nom unš-ij bai-na.  
elder:sister book read-Conn be-NPast  
“The elder sister is reading a book.”
- c. Ehč nom uns-aad bai-na.  
elder:sister book read-Conn be-NPast  
“The elder sister keeps reading a book.”
- d. Egč nom unš-ij gar-na.  
elder:sister book read-Conn get:out-NPast  
“The elder sister begins to read a book.” (Kwon 1998: 2)

The puzzling fact, however, is that this conclusion cannot be generalized to other languages. In Arabic and Hebrew; for example, the subject occurring between the auxiliary and the lexical verb is marked but still grammatical. Hence, the two verbs don't form a compound structure as shown in (4) for Arabic and (5) for Hebrew.

- (4) a. ?Sbaha-t hind-un ta-takalam-u l-engeliziat-a.  
become.3fs Hind-Nom 3fs.speak-indic the-English-Gen  
(Intended) “Hind has started to speak English.”
- b. hind-un ?aSbaha-t ta-takalam-u l-engeliziat-a.  
Hind-Nom become.3fs 3fs.speak-indic the-English-Gen  
(Al-Horais 2009: 93)
- (5) Be yaldut-o haya Eli poter  
in childhood-his would.Past.3ms Eli solve.Part.3ms  
targil-ey matemática be kalut.  
exercises mathematics in ease

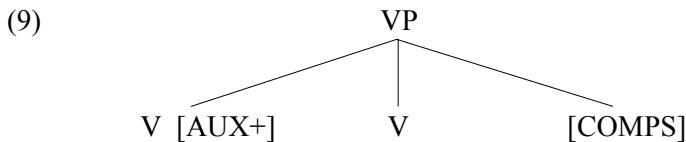
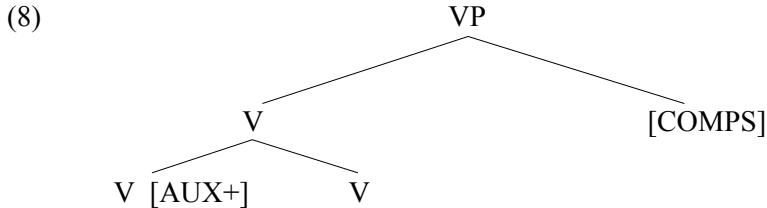
"In his childhood, Eli would solve math exercises easily."  
 (Falk 2004: 242)

More complicatedly, the two different pictures can be found in a single language family such as Romance languages. In Romanian, for instance, the auxiliary and the lexical verb have an internal hierarchical structure and thus they must be in monoclausal relation. According to Monachesi (2005), Romanian does not present any construction in which the subject or the verbal complements can intervene between the auxiliary and the lexical verb as illustrated in (6) and (7), taken from Monachesi (2005: 137).

- (6) a. Mama a făcut o prăjitură.  
 mum has made a cake  
 "Mum has made a cake."
  - b. \*A mama făcut o prăjitură.  
 has mum made a cake  
 (Intended) "Mum has made a cake."
- (7) \*Am filme bune văzut.  
 have films good seen  
 "I have often seen good films."

In each of these examples, the auxiliary and the lexical verb in Romanian form a unit which cannot be separated by other elements. Hence, it seems that the two verbs are in the compound structure (8), not flat structure (9), and hence the former is the most appropriate to capture auxiliary verb constructions in Romanian (see Monachesi 2005: 102).

14 On the Universality of Auxiliary Verbs



The situation is different in Italian where subject can occur between the auxiliary and the lexical verb.<sup>3</sup> This can be shown by the following example in (10):

- (10) Avendo Martina deciso di partire, possiamo usare  
 having Martina decided to leave, can use  
 la sua stanza.  
 the her room  
 “Martina having decided to leave, we can use her room.”  
 (Monachesi 2005: 137)

The same behaviour is observed with French auxiliaries in which subject clitics can occur between the auxiliary and the lexical verb, as shown by the subject clitic *tu* in (11):

---

<sup>3</sup> One of the anonymous reviewers pointed out that in Italian, it is not only the subjects that can occur between the auxiliary and the lexical verb, but also other elements like adverbs also can do so. S/he cited the following example:

Io ho semplicemente deciso così.  
 I have simply decided so.

- (11) L'as-tu           mange?  
 it have-you eaten  
 “Have you eaten it?” (Monachesi 2005: 137)

Interestingly enough, in some languages, like Korean, auxiliary verb constructions can be formed with a lexical verb followed by more than one auxiliary verb. Consider the following examples, taken from Yoo (2003: 414):

- (12) a. Nay-ka sakwa-lul       mek-nun-ta.  
          I-Nom apple-Acc       eat-Pres-Decl  
          “I eat an apple.”
- b. Nay-ka sakwa-lul       mek-e   po-ass-ta.  
          I-Nom apple-Acc       eat       do.as.a.try-Past-Decl  
          “I tried to eat an apple.”
- c. Nay-ka sakwa-lul       mek-e   po-ci           anh-key  
          I-Nom apple-Acc       eat       do.as.a.try   not  
          toy-ess-ta.  
          come.to-Past-Decl  
          (Lit.) “I came to not try to eat an apple.”

Another debatable property of the auxiliary verb comes from the observation that auxiliary verbs can occur as main verbs (Conrad 1988: 92). For example, the typical two English auxiliary verb *have* and *be* can function as a lexical verb<sup>4</sup> as in (13b) and (14b). The same behavior can be observed in the Arabic auxiliary verb *kaana* as in (15b), but not in the other Arabic auxiliaries as

---

<sup>4</sup> By contrast, ‘get’ does not have the properties that English auxiliaries are supposed to have. It doesn’t have the NICE properties (see section 4), and hence it is labelled a ‘lexical verb.’ For more discussion about ‘get,’ see Haegeman (1985), Fleisher (2006).

exemplified in (16) below.

- (13) a. I don't have enough money. (AUXILIARY *have*)  
 b. I have not enough money. (LEXICAL *have*, see Collins 2009: 12)
  
- (14) a. She is waiting. (AUXILIARY *be*)  
 b. She is a doctor. (COPULAR *be*, a lexical verb)  
 (Payne 2011: 262)
  
- (15) a. kaana ahmad-u taajir-an.  
 was.3ms Ahmad-Nom tradesman-Acc  
 “Ahmad was a tradesman.”
  
- b. kaana taajir-un.  
 existed.3ms tradesman-un  
 “There existed a tradesman.” (El-Rakhawi 1982: 260)
  
- (16) a. kaada-t hind-un ta-njah-u fii  
 [almost]<sup>5</sup>-3f Hind-Nom 3fs-pass-Indic in  
 l-ixtibar-i.  
 the-exam-Gen  
 (Intended) “Hind almost pass the exam.”
  
- b. Saara aqalj-u ya-taSaGTu bi-kaqafat-in.  
 started.3ms the-snow-Nom 3m-falling heavily-Gen  
 “The snow started falling heavily.”
  
- c. baat-uu ya-tafahamuuna l-muŠkilat-a.  
 become.3mp 3m-understand.3mp the-problem-Acc  
 (Intended) “They have become aware of the problem.”  
 (Al-Horais 2009: 96)

---

<sup>5</sup> I tried to find an equivalent verb in English to the verb *kaada*, but I couldn't. Therefore, I used the adverb ‘almost’ to convey the meaning of *kaada*.

Another significant structural property of auxiliaries is that of carrying all morphological information. In languages with rich inflectional morphology, auxiliaries carry all morphological information relating to a lexical verb such as person/number agreement, tense, aspect, or mood affixes, leaving a lexical verb in a non-finite participial form. This is best reflected in Finnish as shown in (17), taken from Koskinen (1998: 70).

- (17) a. Hely-n täyty-y lähte-ää.  
Hely-Gen must-3s leave-tA  
“Hely must (=has to) leave.”
- b. Hely-n täyty-i lähte-ää.  
Hely-Gen must-Past.3s leave-tA  
“Hely had to leave.”
- c. Hely-n täyty-ne-e lähte-ää.  
Hely-Gen must-Pot-3s leave-tA  
“Hely probably has to leave.”

In certain languages, like Arabic; a language with a heavily inflectional morphology, both the lexical verb and auxiliary verb carry the same morphological information, in particular, phi-features as shown in (18). A somewhat similar patterning is found in Hebrew, Hindi-Urdu, Manam,<sup>6</sup> and Persian, as respectively illustrated in (19)-(22), below:

- (18) a. kaan-a ya-ktub-u r-risaalat-a.  
was-3ms 3ms-write-indic the-letter-Acc  
“He was writing the letter.”

---

<sup>6</sup> Manam is Austronesian language spoken by about 6,000 people on Manam Island (see Turner 1986: 10).

- b. hum        kaan-uu        ya-ktub-uu-na        r-risaalat-a.  
     they-Nom were-3mp 3-write-3mp-indic the-letter-Acc  
     “They were writing the letter.” (Al-Horais 2009: 97)
- (19) Ruti        hayta        toferet        smalot.  
     Ruti        be-Past-3fs        sew-fs        dresses  
     “Ruti was sewing/used to sew dresses.”  
     (Shlonsky 1997: 61)
- (20) Rahul        kitaab        parh-taa        thaa.  
     Rahul.m-Nom book.f read-Hab.ms be.Past.ms  
     “Rahul used to read a/the book.” (Bhatt 2005: 759)
- (21) raʔa'na        ?u-em=emaʔ-i'-be        ?u-so'a?i  
     what        2-redpl=do-3obj-and 2-aux  
     “What are you doing?”  
     (Lichtenberk 1983: 198, as cited in Anderson 2006: 222)
- (22) a. Dâšt-am        qazâro        *mi*-xord-am.  
       had-1s.        food        asp-eat-Past-1s  
       “I was eating the food.”
- b. dâr-am        qazâro        *mi*-xor-am.  
       have-1s        food        asp-eat-Pres-1s  
       “I am eating the food.” (Taleghani 2006: 27-28)

## 2.2. Thematic Properties

One important and perhaps agreed criterion which singles out the auxiliary verb from the lexical verb has to do with argument structure. While the former loses its argument structure, which makes it unable to thematically associate with any of the NPs in the sentence, the latter preserves argument structure, assigning

thematic theta roles<sup>7</sup> to the NPs in subject and object position. That is, auxiliary verbs are completely devoid of any argument structure and hence their role in the sentence is only functional, supplying or modifying information about tense, aspect, modality, and polarity (Payne 1985, Heine 1993), and this, as stated by Cowper (1990: 85), is “at least part of what defines as an auxiliary verb.”<sup>8</sup> This, however, can be precisely seen in all the examples cited above, where the auxiliary verbs there never assign any theta roles to the subjects but contribute only functional information (cf. Pollock 1997: 239, Lasnik 2000).

Being unable to assign a theta role also distinguishes auxiliary verbs constructions from other compound constructions like V1-V2 compound formation<sup>9</sup> (23-24), light verb constructions<sup>10</sup> (25), and serial verb constructions<sup>11</sup> (26). In these constructions, the two verbs share the same role in determining the assignment of theta-roles to arguments,<sup>12</sup> while in auxiliary verbs constructions,

<sup>7</sup> In generative grammar, a *theta role* or  $\theta$ -*role* is the formal device for representing syntactic argument structure (the number and type of noun phrases) required syntactically by a particular verb (for more information, see Carine 2006).

<sup>8</sup> The same conclusion has been reached by Kenesei (2001: 79). After considering criteria for auxiliaries in Hungarian, she states that “an auxiliary is an independent word that has a complement structure in terms of categories, but has no argument structure, i.e., a capacity to assign thematic roles.”

<sup>9</sup> Both V1 and V2 are lexical verbs, and hence they assign a theta role to their shared object. For more discussion about this phenomenon, see Fukushima (2008).

<sup>10</sup> Light verb constructions are constructions in which an active/patient-denoting verbal noun or adjective combines with a member of a very restricted class of verbs to form a compound verb (Han & Rambow 2000).

<sup>11</sup> As defined by Johnson (2006: 40), “Serial Verb Constructions consist of two verbs (or verb phrases) that occur in sequence without an intervening conjunction (subordinating or coordinating) between the verbs.” Baker (1989) provides an analysis of serial verbs in which is often criticized for only accounting for serial verbs which share their objects. Shared objecthood, however, is a defining feature for serial verbs, and hence they must determine theta-role assignment.

<sup>12</sup> As for light verbs, there is disagreement among syntacticians regarding whether

the auxiliaries play no part in determining the assignment of thematic roles to arguments.

- (23) a. Ken-ga musuko-o nadē-sobire-ta.  
 Ken-Nom son-Acc stroke-fail-Past  
 "Ken failed to stroke his son."
  - b. Ken-ga musuko-o nadē-mawasi-ta.  
 Ken-Nom son-Acc stroke-fondle-Past  
 "Ken caressed his son."  
 (Japanese, Hashimoto & Bond 2005: 146)
- (24) a. Taro-nun yek-e kele-kass-ta.  
 Taro-Top station-Goal walk-go-Past  
 "Taro walked to the station."
  - b. Taro-nun yek-e tallye-kass-ta.  
 Taro-Top station-Goal run-go-Past  
 "Taro ran to the station." (Korean, Tanaka 2002: 422)
- (25) a. mwulka-ka halak-i TOY-ess-ta.<sup>13</sup>  
 price-Nom drop-Nom BECOME-Past  
 "Prices went down."
  - b. sintaylywuk-i Columbus-ey-uyhay palkyen-i  
 new continent-Nom Columbus-by discovery-Nom  
 TOY-ess-ta.  
 BECOME-Past

---

light verbs determine theta-role assignment or not. For a discussion about this issue, see Han & Rambow (2000), Bowern (2004), Seiss (2009).

<sup>13</sup> As illustrated by Sato (1993), the Korean light verb *toytā* 'BECOME' can only occur in light verb constructions with theme subjects. For more discussion about the role of light verbs in determining theta-role assignment, see Sells (1989), Seiss (2009).

“America was discovered by Columbus./America became discovered by Columbus.” (Korean, Sato 1993)

- (26) a. i bai klos gi im pikin.  
he buy clothes give his child  
“He bought some clothes which he gave to his child.”
  - b. a tek n<sup>g</sup>f kut di bred.  
I take knife cut the bread  
“I cut the bread with a knife.”
- (Krio,<sup>14</sup> Finney 2004, cited in Johnson 2006: 40)

### 3. Language Specific Properties

As shown above, recognizing a class of auxiliaries distinct from lexical verbs can be distinguished by language-independent characteristics that look very different in different languages. In this section, I illustrate that this recognition can be distinguished also by language specific properties.

In Tamil and some Uralic languages like Finnish, for example, negation is uttered by a negative marker functioning as an auxiliary verb, while in English the division between auxiliaries and lexical verbs seems relatively clear-cut by using morphosyntactic “tests” for distinguishing both classes from each other (i.e., auxiliaries have the NICE properties, while lexical verbs do not). These tests involve four syntactic properties deemed the NICE properties<sup>15</sup> by Huddleston (1980: 333, see also Huddleston & Pullum 2002: 92-112). In the following, I explain these specific properties.

---

<sup>14</sup> Krio is the most widely spoken language in Sierra Leone (Simpson 2008).

<sup>15</sup> NICE is an acronym for Negation, Inversion, Code, and Emphasis.

### 3.1. Auxiliation and Negation

Cross-linguistically, it has been observed that some certain languages negate a clause via a negative marker which has the properties of a finite auxiliary followed by a lexical verb in a non-finite participial form. This is best reflected in Finnish and Tamil, in which a negative marker is an auxiliary verb.

#### 3.1.1. Finnish

In Finnish, negative sentences verbal categories are expressed on a negative auxiliary while the lexical verb is expressed in some non-finite form (Sulkala & Karjalainen 1992, Koskinen 1998). The most common negative marker in Finnish is the negation word **e-**, which functions as an auxiliary verb, carrying the finite person/number agreement marking which in affirmative sentences is attached to the main verb as previously exemplified in (17). More interestingly, the negative never bears voice or finite tense or mood morphology, all of which always appear on the main verb. The negation can combine with all tenses and moods (except the past tense form, using the perfect participle instead) (Sulkala & Karjalainen 1992, Koskinen 1998). The following data, in which **e-** in bold, demonstrate this interesting strategy of expressing sentential negation in Finnish:

- (27) **e-n**        osta        talo-a.  
             neg-1s     buy        house-Part  
             “I do not buy a/the house.” (Miestamo 2000: 66)

- (28) Minä        **e-n**        lue.  
             I.Nom     neg-1s    read  
             “I’m not reading./I don’t read.” (Koskinen 1998: 69)

- (29) Kerron ett-e-t puhu suomea.  
say-1s that-neg-2s speak Finnish  
“I say that you don’t speak Finnish.” (Mitchell 2006: 231)
- (30) Hän e-i tul-lut.  
s/he neg-3s come-Past-Part  
“S/he didn’t come.” (ibid.)

### 3.1.2. Tamil

As explained in Steever (2000: 107-108), negation in Tamil<sup>16</sup> is expressed in two ways. The first one is morphological negation in which the negative marker *ā-*, exemplified in (31), is a part of the derivational morphology of the verb. The second one, which is relevant to our discussion here, is syntactic negation expressed by two negative auxiliary verbs: *illai* and *mātt*. The former is only compatible with present and past tense interpretation as shown in (32), whereas the latter is only compatible with the future tense interpretation as in (33). All the examples below are taken from Steever (2000: 108).

- (31) katitam inke var-ā-tu.  
letter-Nom here come-neg-es  
“The letter will not come here.”
- (32) avan vara.v illai.  
he-Nom come-Inf neg  
“He did/does not come.”

---

<sup>16</sup> Tamil is a southern Dravidian language. It is spoken mainly in South India. It is the official language in the Indian state of Tamil Nadu. Tamil is also spoken in north-western Sri Lanka, but the kind of Tamil spoken there is slightly different from the kind used in South India (Steever 2000: 31).

- (33) *avan vara.v mattan.*  
 he-Nom come-Inf Fut-neg-3s  
 “He will not come.”

Interestingly enough, the gloss in (31) and (32) draws a significant surprising distinction between the two negative auxiliary verbs *illai* and *mātt*. While *illai* does not inflect for the agreement feature, *mātt* does. It is not clear from Steever (2000) why *illai*, unlike *mātt*, does not carry any inflectional element to show at least the agreement feature since the negative is considered as an auxiliary verb produced by a language with quite rich morphological inflections. Such a fact may add more complexity to the question of what a good auxiliary should look like.

### 3.2. The NICE Properties of Auxiliaries

In English, auxiliaries in general are characterized by the NICE properties, by which auxiliary verbs in English are distinguished from lexical verbs. These properties refer to English auxiliaries occurring with negation, inversion, code, and emphatic affirmation. In the following subsections, I explain these properties, relying on Hauge (2003) (for further discussion about the NICE properties, see Huddleston 1980, Haegeman 1985, Downing 1996, Huddleston & Pullum 2002, Payne 2011).

#### 3.2.1. Negation

As a property unique to English auxiliaries and drawing a dividing line between auxiliaries and main verbs, English auxiliaries have distinct negative forms with the negation (not) undergoing cliticization to form negative auxiliaries such as *isn’t*, *haven’t*, and *shouldn’t*. Consider the following example:

- (34) a. I shouldn't go to London.  
       b. \*I gon't to London.

### 3.2.2. Inversion

The second of the NICE properties is inversion of the auxiliary and the subject in certain types of constructions. The most common type of construction where this occurs is in interrogative sentences, such as (35a). Subject-auxiliary inversion is also found in certain types of conditional adverbial clauses, such as (35b), and with some adverbs (expressing negative meaning) that are in initial position, as in (35c):

- (35) a. Is the clown coming?  
       b. Had I known about the appetite of seals, I'd never  
           have shared my lunch.  
       c. Seldom had they seen such an animal.

By contrast, lexical verbs do not exhibit this property (Payne 2011: 263-264):

- (36) a. \*Eats she kimchi?  
       b. \*Broke the workers the vase?  
       c. \*What broke the workers? (trying to mean “What did  
           the workers break?”)  
       d. \*Where live we?

### 3.2.3. Code

The third NICE characteristic of an auxiliary is code, referring to sentences in which a lexical verb is subsequently picked up by an auxiliary, in a similar manner to a noun being picked up by a pronoun. This is often the case in constructions containing *and so*. Examples of this are found in (37)-(39), taken from Payne

(2011: 264). The ungrammatical examples illustrate the fact that lexical verbs do not have this property:

- (37) a. I should see the doctor, and so should she.  
b. \*I saw the doctor, and so saw she.
- (38) a. Who should eat kimchi? She should.  
b. Who ate kimchi? \*She ate.
- (39) a. We were eating kimchi, and so was she.  
b. \*We eat kimchi and so eats she.

### **3.2.4. Emphatic Affirmation**

The last NICE property is emphatic affirmation, where stress is on the auxiliary. In English there can be stress on any verbal form for focus purposes. What is particular about stress on the auxiliary is that it is used for emphatic affirmation of a doubtful statement or denial of a negative statement, as can be seen in (40) below.

- (40) a. I **can** come. (You are wrong to think that I cannot come!)  
b. We **did** see them. (You thought we did not see them!)  
(Hauge 2003: 59)

## **4. Conclusions**

A concrete fact can be drawn from the preceding discussion is that variation in the properties of auxiliaries among natural languages is precisely observed and therefore having a universal standard definition of the auxiliary verb seems elusive. Nevertheless, there is still room for arguing that at least two universal properties must co-occur in order to distinguish the

auxiliary verb from other syntactic categories. (i) Auxiliation should be understood as the development of constructions into markers of tense, agreement, modality, and perhaps aspect. (ii) Auxiliary verbs do not enter into a thematic relation with the arguments in the sentence, leaving this job to the lexical verbs that auxiliaries tend to occur separately from. This may constitute the standard syntactic argument that auxiliaries and lexical verbs are two distinct types of syntactic entities.

## References

- Al-Horais, N. 2009. The Syntax of Arabic Negation Marker Laysa. Ph.D Dissertation. Newcastle University.
- Anderson, G. 2006. *Auxiliary Verb Constructions*. Oxford: Oxford University Press.
- \_\_\_\_\_. To Appear. Auxiliary Verb Constructions in Old Turkic and Altai-Sayan Turkic. In M. Erdal (ed.), *Studies in Old Turkic Linguistics*. Wiesbaden: Harrassowitz.
- Baker, M. 1989. Object Sharing and Projection in Serial Verb Constructions. *Linguistic Inquiry* 20, 513-553.
- Berman, R. 1980. On the Category of Auxiliary in Modern Hebrew. *Hebrew Annual Review* 4, 15-37.
- Bhatt, R. 2005. Long Distance Agreement in Hindi-Urdu. *Natural Language & Linguistic Theory* 23, 757-807.
- Bowern, C. 2004. Bardi Verbal Morphology in Historical Perspective. Ph.D Dissertation. Harvard University.
- Carnie, A. 2006. *Syntax: A Generative Introduction*. Malden, MA: Wiley-Blackwell.
- Collins, P. 2009. *Modals and Quasi-Modals in English*. Amsterdam: Rodopi.
- Conrad, R. (ed.) 1988. *Lexikon sprachwissenschaftlicher Termini*. Leipzig: VEB Bibliographisches Institut.

- Cowper, E. 1990. Thematic Underspecification: The Case of ‘Have.’ Ms., University of Toronto.
- Downing, A. 1996. The Semantics of *Get*-Passives. In R. Hasan et al. (eds.), *Functional Descriptions* 179-205. Amsterdam: John Benjamins.
- El-Rakhawi, T. 1982. Aspects of Sentential Negation in Arabic: A Contribution to the Typology of Negation. Ph.D Dissertation. University of Birmingham.
- Falk, Y. 2004. The Hebrew Present-Tense Copula as a Mixed Category. In M. Butt & T. King (eds.), *Proceedings of the LFG04 Conference* 226-246. Stanford, CA: CSLI Publications.
- Fleisher, N. 2006. The Origin of Passive ‘Get.’ *English Language and Linguistics* 10.2, 225-252.
- Fukushima, K. 2008. On the Type-Wise Productivity of Lexical V-V Compounds in Japanese: A Thematic Proto-Role Approach. *Gengo Kenkyu* 134, 119-140.
- Haegeman, L. 1985. The *Get*-Passive and Bruzio’s Generalization. *Lingua* 66, 53-77.
- Han, C. & O. Rambow. 2000. The Sino-Korean Light Verb Construction and Lexical Argument Structure. *Proceedings of the 5<sup>th</sup> International Workshop on Tree Adjoining Grammars and Related Formalisms* 221-226. Universite Paris 7.
- Hashimoto, C. & F. Bond. 2005. A Computational Treatment of V-V Compounds in Japanese. In S. Muller (ed.), *Proceedings of the 12<sup>th</sup> International Conference on Head-Driven Phrase Structure Grammar* 143-156. University of Lisbon.
- Hauge, H. 2003. Towards a Unified Representation of English and Norwegian Auxiliaries. *Nordic Journal of English Studies* 2.1, 53-74.
- Heine, B. 1993. *Auxiliaries, Cognitive Forces, and Grammaticalization*. New York: Oxford University Press.
- Huddleston, R. 1980. On Palmer’s Defense of the Distinction between Auxiliaries and Main Verbs. *Lingua* 50, 101-115.
- Huddleston, R. & G. Pullum. 2002. *The Cambridge Grammar of*

- the English Language*. Cambridge: Cambridge University Press.
- Johnson, S. 2006. Revisiting the Structure of Serial Verb Constructions. *LSO Working Papers in Linguistics* 6, 39-48.
- Kenesei, I. 2001. Criteria for Auxiliaries in Hungarian. In I. Kenesei (ed.), *Argument Structure in Hungarian* 73-106. Budapest: Akadémiai Kiadó.
- Koskinen, P. 1998. Features and Categories: Non-Finite Constructions in Finnish. Ph.D Dissertation. University of Toronto.
- Kuteva, T. 2001. *Auxiliation: An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University Press.
- Kwon, J. 1998. Auxiliary Verb Constructions in Korean and Khalkha Mongolian. Paper Presented at the 3<sup>rd</sup> International Symposium on Mongolology 17-24. Inner Mongolia University.
- Lasnik, H. 2000. *Syntactic Structures Revisited: Contemporary Lectures on Classic Transformational Theory*. Cambridge, MA: MIT Press.
- Miestamo, M. 2000. Towards a Typology of Standard Negation. *Nordic Journal of Linguistics* 23.1, 65-88.
- Mitchell, E. 2006. The Morpho-Syntax of Negation and the Positions of NegP in the Finno-Ugric Languages. *Lingua* 116.3, 228-244.
- Monachesi, P. 2005. *The Verbal Complex in Romance: A Case Study in Grammatical Interfaces*. Oxford: Oxford University Press.
- Payne, J. 1985. Negation. In T. Shopen (ed.), *Language Typology and Syntactic Description: Volume I, Clause Structure* 197-242. Cambridge: Cambridge University Press.
- Payne, T. 2011. *Understanding English Grammar: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Pollock, J-Y. 1997. Notes on Clause Structure. In L. Haegeman (ed.), *Elements of Grammar: A Handbook of Generative Syntax* 237-279. Dordrecht: Kluwer.
- Sato, Y. 1993. Complex Predicate Formation with Verbal Nouns in Japanese and Korean: Argument Transfer at LF. Ph.D

- Dissertation. University of Hawaii.
- Seiss, M. 2009. On the Difference between Auxiliaries, Serial Verbs, and Light Verbs. In M. Butt & T. King (eds.), *Proceedings of the LFG09 Conference* 501-519. Stanford, CA: CSLI Publications.
- Sells, P. 1989. More on Light Verbs and Theta-Marking. Ms., Stanford University.
- Shlonsky, U. 1997. *Clause Structure and Word Order in Hebrew and Arabic: An Essay in Comparative Semitic Syntax*. Oxford & New York: Oxford University Press.
- Simpson, A. 2008. *Language and National Identity in Africa*. Oxford: Oxford University Press.
- Steever, S. 2000. *The Tamil Auxiliary Verb System*. London & New York: Routledge.
- Suh, Y. 2000. A Study of the Auxiliary Verb Construction and Verb Serialization in Korean. Ph.D Dissertation. University of Washington.
- Sulkala, H. & M. Karjalainen. 1992. *Finnish*. London & New York: Routledge.
- Taleghani, A. 2006. The Interaction of Modality, Aspect, and Negation in Persian. Ph.D Dissertation. Arizona University.
- Tanaka, E. 2002. A Japanese Compound Verb V-te-iku and Event Structure. *Proceedings of Pacific Asia Conference on Language and Computation* 16, 421-432.
- Turner, B. 1986. *A Teaching Grammar of the Manam Language*. Ukarumpa: Summer Institute of Linguistics.
- Yoo, E-J. 2003. Case Marking in Korean Auxiliary Verb Constructions. *Proceedings of the 9<sup>th</sup> International Conference on HPSG* 413-438. Stanford University.

Journal of Universal Language 13-1  
March 2012, 31-70

## Linguistic Typology: An Iranian Perspective\*

**Mohammad Dabir-Moghaddam**

*Allameh Tabataba'i University*

### Abstract

The Iranian languages spoken in Iran reveal a very intriguing typological peculiarity. They all strongly benefit from agreement as a typological parameter. In this paper, I will begin with Comrie (1978) in which he has proposed the five possible language types based on case-marking and verb-agreement and will address the

---

\* The data from the varieties of Kurdish spoken in Iran as well as Talyshi and Davani are taken from an extensive corpus which I have collected through a questionnaire that I have designed for my research on Modern Iranian languages and dialects. The audio-recorded materials collected this way, supplemented with the frequent follow up interviews and consultations with the same informants, have been used in my analysis of the data. I am grateful to my informants for their valuable information they shared with me and made this research possible. The data from other sources are directly quoted without any change in the transcriptions used by the authors.

The author also wishes to express his sincere thanks to the anonymous reviewers of the JUL for their valuable comments and recommendations. However, he should be held responsible for any remaining shortcomings.

Mohammad Dabir-Moghaddam  
Department of Linguistics, Allameh Tabataba'i University  
Chamran Highway, Pol-e Modiriyat, Allameh Tabataba'i Str., Tehran, IRAN  
Phone: 982188038013; Email: [dabirmoghaddam@yahoo.com](mailto:dabirmoghaddam@yahoo.com)

Received October 18, 2011; Revised November 23, 2011; Accepted December 12, 2011.

status of several varieties of Kurdish, as well as a dialect of Talyshi and Davani with respect to agreement. I will show that type (d), i.e., the Tripartite system in Comrie's terminology where S, A, and P each has a distinct marking, and type (e), for which he has not proposed any name but is a type in which A and P are identically marked, are highly productive and stable systems in the mentioned Iranian languages. The observations reported have implications for the notion of type, language change, and linguistic variation.

Keywords: agreement system, Kurdish, Talyshi, Davani, clitic

## 1. Introduction

The notion of ‘type’ has always been a basic preoccupation of the typologists. The typologists of the eighteenth and nineteenth centuries practiced ‘the holistic typology,’ equally called ‘the *individualizing* approach’ to typological classification by Greenberg (1974), and typologized languages on the basis of individual morphological parameters. The typologists of the twentieth century and the present time have practiced ‘the partial typology,’ equally called by Greenberg (*ibid.*) ‘the *generalizing* approach’ to typological classification, as they aim at typologizing languages based on specific structural strategies. In this paper, I intend to shed more light on the notion of type by discussing the status of the Iranian languages with respect to agreement as a structural strategy. This discussion has implications for language change and variations as well as linguistic theory.

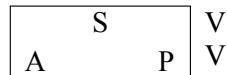
## 2. Agreement

Bernard Comrie in his seminal article entitled “Ergativity”

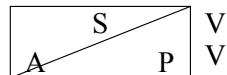
(1978) provides Figure 1 below to illustrate the five possible language types based on case-marking systems (p. 332):

Figure 1. Case-Marking Systems for S, A, and P<sup>1</sup>

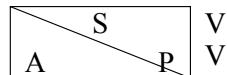
(a) Neutral



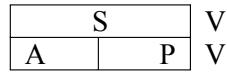
(b) Nominative-Accusative



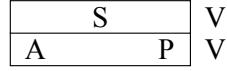
(c) Ergative-Absolutive



(d) Tripartite



(e)



Comrie provides examples for types (a-c). As for type (d), he mentions that “Type (d), with three different morphological markers, is relatively *rare* across the languages of the world. Some

<sup>1</sup> The symbols S, A, and P (also called O) stand for the subject of the intransitive verbs, the Subject/Agent of the transitive verbs, and the Patient/Object, respectively. Furthermore, there are a number of other abbreviations used in this paper which are listed below:

CLC	clitic
p/pl./PL	plural
PST	past
s/sg.	singular
SAP	Speech Act Participant

languages have this system of case-marking for a *limited* number of noun phrases: in the Australian language Dyirbal, for instance, the interrogative pronoun ‘who?’ has distinct forms for S (*wan<sup>y</sup>a*), A (*wan<sup>y</sup>d<sup>y</sup>u*), and P (*wan<sup>y</sup>una*) (Dixon 1972: 53), but this is not true of noun phrases in general . . . Motu, an Austronesian language of New Guinea, might appear to be a type (d) language from the data cited by Capell (1969: 36, 43, 54), with the postposition *na* for S, *ese* for A, and *no overt* marker for P . . . Motu, however, is *not a pure* type (d) language, since the A postposition *ese* is in fact optional, and the conditions on the occurrence of the S postposition *na* are more complex (and vary from dialect to dialect), so that some instances of S also lack any overt marker.” (p. 333-334). Comrie, then, refers to type (e) as a “logical possibility” and claims that “Type (e), with the same marker for both A and P and a different marker for S, seems *not to occur* as an *attested* case-marking system . . .” (p. 334). It is worth mentioning that Comrie has not given any name to type (e).

In another section of his article, Comrie adds a new dimension to the notion of ‘type’ which is my concern in this paper. He announces that “Just as case-marking can operate in accordance with the five logically possible systems of Figure 1 (though, as already noted, (e) seems not to occur in practice), so too can verb-agreement.” (p. 338). Comrie seems to believe that all the five logically possible types are actually attested. Most interesting to us is what he says about type (e). In his words

“Even type (3) [actually (e)] is occasionally found with verb-agreement, as in the Iranian dialect Dānesfāni (Yar-Shater 1969: 204), where the past participle agrees with S, but with no constituent of a transitive sentence:

- (1) Hasan                    buma.  
Hasan-(Masc.)        came-Masc.  
'Hasan came.'

- (2) Zeynaba                bumia.  
     Zeynaba-(Fem.)    came-Fem.  
     ‘Zeynaba came.’
- (3) Hasan                /Zeynaba                šet-  
     Hasan-(Masc.)    /Zeynaba-(Fem.)    milk-(Masc.)  
     eš                   uxa.  
     -Aux.-3sg.A        drank  
     ‘Hasan/Zeynaba drank the milk.’
- (4) Hasan                /Zeynaba                āwa-  
     Hasan-(Masc.)    /Zeynaba-(Fem.)    water-(Fem.)  
     š                   uxa.  
     Aux.-3sg.A        drank  
     ‘Hasan/Zeynaba drank the water.’

In all such instances known to me, however, the verb agrees with S, and shows no agreement whatsoever with either A or P; moreover, the examples with which I am familiar are all in languages where the ergative system is breaking down, being replaced by a nominative-accusative or neutral system. It seems likely that at an intermediate stage in the development from ergative-absolute to nominative-accusative, a situation can be reached where the conflict between moribund ergative-absolute morphology and nascent nominative-accusative morphology is resolved by simply omitting all morphological markers, giving rise to the type illustrated above from Dānesfāni: in the intransitive sentence there is no conflict, and the participle agrees with S; in the transitive construction, there is conflict as to whether the participle should agree with A or with P. The compromise reached is for it to agree with neither. Thus we expect type (e) to arise only as a result of conflict of this kind.” (p. 341-342).

There are a couple of points that I like to raise in regard to Comrie's examples (3) and (4). First, the mentioned examples *do show* 'agreement,' also called 'cross-referencing' and 'indexation,' with the A. The agreement/cross-referencing/indexation marker in (3) and (4) is realized as an enclitic and the direct objects 'milk' and 'water' host the enclitic. My assumption that cross-referencing/indexation enclitics are actually 'agreement' markers is compatible with the following proposal which I quote from Anderson (2005: 239-240): "I propose to regard clitic pronominals as in fact a form of agreement, differing from verbal agreement only in whether the functional content is realized as the morphology of a phrase or a word. This is not a novel proposal . . . The overt manifestation of agreement material by pronominal special clitics can appear in various places, as we have already seen. Clitics may appear with reference to the beginning of the clause—in second position . . ."

With this assumption, then I will be prepared to raise my second point. I suggest that sentences (1)-(4) are *not* examples of type (e) in Figure 1. They represent type (d), the tripartite type: S is encoded by agreement suffixes on the verb. A is encoded through pronominal enclitics which are hosted by the direct object, and the verb does not agree with the direct object. So the P is left without agreement.

### **3. Agreement in the Iranian Languages**

The Iranian languages spoken in Iran show a very intriguing peculiarity. They all contain a rich agreement system. So, I propose that agreement should be viewed as a strong typological parameter in our characterization of these languages. The manifestation of the mentioned parameter in the Iranian languages as well as the range of *variations* that these languages show in this parameter will be relied on to shed more light on the notion of

type in *general* and of type in the Iranian languages in *particular*.

### **3.1. Kurdish**

I begin this discussion with a review of this parameter in Sorani Kurdish, equally called the Central group of Kurdish, which embraces the varieties of this language spoken in the Kurdistan Province in Iran and Suleimaniyah area in Iraq. According to McCarus (2009) “Linguistically, Kurdish as a whole occupies an intermediate position between North-Western and South-Western Iranian dialects.” (p. 587).

McCarus (*ibid.*) in a chapter on Kurdish which is devoted to “Sulaimani Kurdish” (p. 589) reports an interesting observation without providing any explanation for that. In a section entitled “Verb-Only Phrases” he notes:

“There are specific rules for sequencing past agent suffixes and patient endings after simple verb forms. The basic rule requires that the agent suffixes be attached directly to the verb form, followed by the personal ending marking the patient: VB + agent suffix + patient ending . . .” (p. 616)

Following a table called “PRETERITE AGENT AND PATIENT” in which he has listed what he calls “Agent/Enclitic” and “Patient/Ending” he presents representative examples for the above mentioned quotation (p. 617). His examples are directly quoted below. Next to each example I have repeated the example with a morphemic segmentation in which clitics/CLC are separated by an equal sign.

- (5) bīnī-m-īt; [bīnī=m-īt]  
     'I saw you-2s'
- (6) bīnī-t-im; [bīnī=t-im]  
     'you-2s saw me'
- (7) bīnī-mān-in; [bīnī=mān-in]  
     'we saw you-2p'
- (8) bīnī-tān-īn; [bīnī=tān-īn]  
     'you-2p saw us'
- (9) bīnī-tān-in; [bīnī=tān-in]  
     'you-2p saw them'
- (10) bīnī-yān-in; [bini=yān-in]  
       'they saw you-2p'

There are two points about these examples and Sulaimani Kurdish in general which deserve mentioning: (i) The Agent enclitic and the Patient suffix for the first person singular are homophones (*im* ~ *m*) (see examples (5) and (6)). (ii) The second and third person plural patient suffixes are syncretized as *-in* (see examples (9) and (10)).

McCarus adds the following remark to his aforementioned rule about the order of the Agent enclitic with respect to the Patient suffixes.

"However, there are two exceptions, to this basic rule: (i) the 3s agent suffix *-ī* [my notation =*ī*] is always second; (ii) the 1s-*im* [my notations =*im* ~ =*m*], whether agent suffix or patient ending, precedes any *plural* subject or object:

3s agent second:

- (11) b̄n̄-m-̄; [b̄n̄-m=̄]  
‘he saw me’
- (12) b̄n̄-t-̄; [b̄n̄-t=̄]  
‘he saw you-2s (-̄(t))’
- (13) b̄n̄-n-̄; [b̄n̄-n=̄]  
‘he saw us/you-2p/them’  
(b̄n̄-n < b̄n̄-̄n, -in, -in)

1s before plural patient and agent:

- (14) b̄n̄-m-in; [b̄n̄=m-in]  
‘I saw you-2p’
- (15) b̄n̄-m-in; [b̄n̄=m-in]  
‘I saw them’
- (16) b̄n̄-m-tān; [b̄n̄-m=tān]  
‘you-2p saw me’
- (17) b̄n̄-m-yān; [b̄n̄-m=yān]  
‘they saw me’

The rule of 1s + plural agents may not apply to all varieties of Kurdish, as other varieties have different rules. In all other combinations, the agent comes before patient.”

It is interesting to note that McCarus (1958) has also reported the mentioned observation discussing Kurdish. He says “In the preterite, present perfect, and past perfect tenses, the subject and

object suffixes are both suffixed.” (p. 68). The examples he provides from the preterite tense are exactly the ones we witnessed in (5)-(17) above. The examples he presents for the present perfect and the past perfect tenses are directly quoted in (18)-(25) below (p. 69):

Present Perfect:

- (18) xwardúmə [xward-ú=m-ə]  
ate-participle-1sg.CLC-3sg.  
‘I have eaten [it]’
- (19) xwardúmin [xward-ú=m-in]  
ate-participle-1sg.CLC-2pl./3pl.  
‘I have eaten you (pl.)/them’
- (20) xwardútanə [xward-ú=tan-ə]  
ate-participle-2pl.CLC-3sg.  
‘you (pl.) have eaten [it]’
- (21) xwardútanim [xward-ú=tan-in]  
ate-participle-2pl.CLC-1pl.  
‘you (pl.) have eaten us’

Past Perfect:

- (22) xwárdibum [xwárd-ibu=m]  
ate-perfect-1sg.CLC  
‘I had eaten’
- (23) xwárdibumi [xwárd-ibu-m=i]  
ate-perfect-1sg.-3sg.CLC  
‘he had eaten me’

- (24) xwárdibumit [xwárd-ibu=m-it]  
 ate-perfect-1sg.CLC-2sg.  
 ‘I had eaten you (sg.)’
- (25) xwárdibutanin [xwárd-ibu=tan-in]  
 ate-perfect-2pl.CLC-1pl.  
 ‘you (pl.) had eaten us’

MacKenzie (1961) has also reported this peculiarity in Suleimaniyah Kurdish.

“When the verbal ending is other than 3rd sg., it normally follows the Agential suffix. A general exception to this is the 3rd sg. Agential suffix -*i* [i.e., Agent enclitics] which always follows the verbal ending. Particular exceptions are noted below.” (p. 112, section 228) [my emphasis]

MacKenzie, then, presents a table in which examples that support the mentioned quotation can be found (p. 113). The examples are from the verbs *dīn* ‘see’ and *nārdin* ‘send.’ In examples (26)-(30) below, I have quoted examples which show the past tense conjugation of the verb *dīn*. Morphemic segmentation is added by me. In examples (26)-(28) the Agent enclitics occur before the patient agreement suffixes.<sup>2</sup>

---

<sup>2</sup> However, MacKenzie’s hesitation in referring to P, or in his terminology Direct Affectee, verbal suffixes as agreement is voiced in the following statement:

“A Direct Affectee when present, whether expressed by a nominal form or not, is always manifested in a verbal ending of the appropriate tense . . . It is, however, an over-simplification to state that the verb ‘agrees’ with such a Direct Affectee, as is demonstrated by the frequent intrusion of the Agential suffix between verbal stem and personal ending . . .” (1961: 110)

- (26) dī=m-īt  
       'I saw thee'

- (27) dī=m-in  
       'I saw you/them'

- (28) dī=tān-īn  
       'you saw us'

However, in examples (29) and (30), the Agent enclitics occur *after* the patient agreement suffixes.

- (29) dī-t=ī  
       'He saw thee'

- (30) dī-n=ī  
       'He saw you/them'

Another interesting dimension is added to this discussion by MacKenzie. He noticed that "The 3 pl. Agential suffix -yān [i.e., =yān] may either precede or follow the verbal ending, more commonly the former." (1961: 114, section 5). Examples (31)-(34), which I quote from the aforementioned table (*ibid.*: 113) illustrate this alteration.

- (31) dī=yān-im ~ dī-m=yān  
       'they saw me'

- (32) dī=yān-ī(t) ~ dī-t=yān  
       'they saw thee'

- (33) dī=yān-īn ~ dī-n=yān  
       'they saw us'

- (34) dī=yān-in ~ dī-n=yān  
          ‘they saw you/them’

Haig (2008) has also discussed some of the examples which MacKenzie has brought to our attention in regard to the past transitive verbs in “Suleimani” Kurdish (p. 288). This is what he says about the function of the Agent clitics:

“. . . the A-past clitic in fact exhibits the features of an agreement marker, i.e., it *obligatorily* cross-references a different constituent, and is prosodically dependent rather than independent. As Corbett (2003) points out, the distinction between agreement markers and pronouns is often a gradual one; the A-past clitics of Suleimani are a case in point. Below I will point out further features of the A-past clitics which bring them closer to a canonical form of agreement.” (p. 288-289) [my emphasis]

In discussing the construction whose sole constituent is the past transitive verb and both the Agent clitic and the Patient/Object agreement suffix are encoded in the verb, Haig states the following:

“On the assumption that the verbal agreement suffix is a suffix, and the A-past marker is a clitic, one could reasonably expect to find that the clitic would attach outside the suffix. Unfortunately, this is not the case. It is more often the supposedly clitic—A past marker which attaches directly to the verb stem, while the O-past agreement suffix follows it . . .” (p. 291)

The example which he quotes from Fattah (1997: 220) is given in (35).

- (35) sand=im-in  
 take-PST=1s:CLC-3PL  
 ‘I took them’

At this stage Haig refers to MacKenzie (1961) and the relevant pages where I quoted examples (26)-(34) from and announces that:

“It turns out that the order of argument cross-referencing markers is not always Verb-A-O, as in (329) [i.e., example (35) above]. Instead, it interacts with the *person* of the arguments. The relevant person categories are SAP (Speech Act Participants) vs. non-SAP, whereby non-SAP is further divided into singular and plural.” (p. 292) [my emphasis]

I would consider this quotation as Haig’s explanation about the *ordering* of the argument cross-referencing on the verb in Suleimani Kurdish. Haig’s final step in this discussion is that he attempts to put the behavior of this variety of Kurdish in the perspective of the known facts in the typological literature. His remarks are cited below:

“Thus we find in Suleimani a pocket of what is, according to Nichols (1992), a cross-linguistically rare alignment type. The agreement facts of Suleimani of course reflect a well-known typological generalization, noted for example by Croft (2001: 318-319): The typologically least marked constellation of A and O is when A is a SAP and O is a non-SAP, in Croft’s notation SAP → non-SAP (SAP acts on non-SAP). When these conditions are met in Suleimani, the A-past clitic appears to act as a suffix rather than a clitic.” (p. 293)

Haig’s discussion of the “Suleimani Morphosyntax” begins with

the examples (36)-(39) below which “were supplied by Sorani native speakers<sup>3</sup>” (p. 278). The first two examples contain intransitive past verbs. As it is expected, in these examples, the S is cross-referenced with agreement suffixes.

- (36) min hāt-im                bō erā  
     1s come:PST-1s      to here  
     ‘I came here’ (Haig 2008: 279, ex. (293))

- (37) ewā hāt-in                bō erā  
     2PL come:PST-2PL      to here  
     ‘you (PL) came here’ (ibid., ex. (294))

However, in examples (38) and (39) whose verbs are transitive past, the As are cross-referenced by the Agent clitics and the direct objects are the hosts for them. The P arguments themselves are *not* cross-referenced as the verb has appeared in its past stem form.

- (38) min     ewā=m                bīnī  
     1s        2PL=1s:CLC      see:PST  
     ‘I saw you (PL)’ (ibid., ex. (295))

- (39) ewā     min=tān                bīnī  
     2PL      1s=2PL:CLC      see:PST  
     ‘you (PL) saw me’ (ibid., ex. (296))

If we take examples (36)-(39) into consideration, then undoubtedly the following position on Suleimani is quite understandable:

---

<sup>3</sup> In a footnote Haig informs the reader on his informants: “The speakers who kindly supplied the information are a married couple, in their late 30’s, who were born in Suleimanye and spent most of their lives there. They have been living in Germany for the last six years.” (p. 278, n. 1)

“Personally I prefer to analyze agreement in past tenses as tripartite: S, A, and O each determines a distinct, though partially overlapping, type of agreement.” (p. 302)

Haig’s position is reiterated in the “Summary of the Central Group” section which I quote below:

“Just how the Past Transitive construction should be classified in terms of the accepted taxonomies of alignment systems *remains an open question*. My tentative conclusion is that case marking is neutral, while agreement is tripartite.”  
 (p. 305) [my emphasis]

I believe that a detailed analysis of the examples presented in McCarus (1958) clearly reveals the actual type of Suleimaniyah Kurdish. Examples (40) and (41) contain intransitive verbs. Examples (42)-(44) have transitive verbs formed with present stems. In all of them S and A are encoded on the verb by agreement suffixes. Morphemic segmentation and glosses are provided by me.

- (40) min-iš ye-m  
 I-too come-1sg.  
 ‘I’ll come too’ (McCarus 1958: 97, part 5.231, ex. (2))

- (41) hát-im  
 Came-1sg.  
 ‘I came’ (ibid.: 59)

- (42) kürdi' ?ə-zán-i(t)  
 Kurdish Imperfect-know-2sg.  
 ‘Do you know Kurdish?’ (ibid.: 58)

- (43) ?ə-xo'-n  
 Imperfect-eat-3pl.  
 'they eat' (McCarus 1958: 58)

- (44) də́ st-it      mác      ?ə-kə-m  
 hand-your kiss      Imperfect-do-1sg.  
 'I kiss your hand' (ibid.: 97, part 5.231, ex. (3))

In examples (45)-(50), both A and P are encoded in the verb. The A appears as a verbal suffix and the P is realized as an oblique enclitic whose host is the imperfect aspect/Indicative mood prefix ?ə-. The examples are cited from McCarus (ibid.: 68). I have provided the segmentation and glosses.

- (45) ?ə=m-nás-e  
 Imperfect-1sg.CLC-know-3sg.  
 'he knows me'

- (46) ?ə=t-nás-e  
 Imperfect-2sg.CLC-know-3sg.  
 'he knows you'

- (47) ?ə=y-nás-e  
 Imperfect-3sg.CLC-know-3sg.  
 'he knows him'

- (48) ?ə=man-nás-e  
 Imperfect-1pl.CLC-know-3sg.  
 'he knows us'

- (49) ?ə=tan-nás-e  
 Imperfect-2pl.CLC-know-3sg.  
 'he knows you'

- (50) ?ə=yan-nás-e  
 Imperfect-3pl.CLC-know-3sg.  
 ‘he knows them’

These examples clearly reflect a Nominative-Accusative agreement system. Examples (42)-(44) whose objects are third singular inanimate (whether overt, i.e., (42) and (44), or covert i.e., (43)) also represent the same system. The only difference is that the P is *not* cross-referenced. It may be noted that in examples (45)-(50) an affix boundary separates the oblique clitic from the stem.

Now we can turn to the transitive verbs which are formed from the past stem. Sentences (51)-(53) have an overt P. In these sentences, the P hosts the A-clitics and the P itself is *not* cross-referenced. Again segmentation and glosses are provided by me.

- (51) márek=im               kušt  
 snake-1sg.CLC       killed  
 ‘I killed a snake’ (McCarus 1958: 104)
- (52) min to=m               bin-i’  
 I       you-1sg.CLC   see-past  
 ‘I saw you’ (ibid.)
- (53) du’ helkə’=m       xwárd  
 two egg-1sg.CLC   ate  
 ‘I ate two eggs’ (ibid.)

These sentences exemplify an oblique A-cross-referencing and *no* cross-referencing for P. Considering the fact that the S in intransitives with the past stem (e.g., example (41)) is cross-referenced with verbal suffixes, it is justified to call the agreement system represented by this set a Tripartite system: The S being nominative, the A being oblique, and the P showing *no*

cross-referencing.

However, there are many examples of transitive verbs formed with the past stems which encode the covert P in the verb. Examples (54)-(58) illustrate this possibility. In (54) the preposition hosts the A-clitic and the past stem of the verb encodes the suffix which cross-references the P. This suffix is the one which cross-references the S (e.g., (40) and (41)) as well as the A of the transitive verbs formed with the present stems (e.g., (44)). In (55)-(58), the imperfect aspect prefix hosts the A-clitics and the verb stem carries the suffixes which encode the P.

- (54) pe=t wút-im  
to-2sg.CLC told-1sg.  
'you told me' (McCarus 1958: 106)
- (55) ?ə=y-kúšt-it  
Imperfect-3sg.CLC-killed-2sg.  
'he was killing you' (ibid.: 68)
- (56) ?ə=y-kúšt-in  
Imperfect-3sg.CLC-killed-1pl.  
'he was killing us' (ibid.: 68)
- (57) ?ə=man-kúšt-it  
Imperfect-1pl.CLC-killed-2sg.  
'we were killing you (sg.)' (ibid.: 68)
- (58) ?ə=tan-kúšt-in  
Imperfect-2pl.CLC-killed-3pl.  
'you (pl.) were killing them' (ibid.: 68)

It should be noted that in examples (55)-(58), the A-clitic is separated from the stem by an affix boundary. Most importantly, the agreement markers realized in (54)-(58) are Oblique clitics for

the A and the Direct suffix for the P. Thus, these examples illustrate an Ergative-Absolute agreement system.

MacKenzie's (1961) meticulous survey of the Kurdish dialects of "Iraqi Kurdistan" (p. xvii) also substantiates my interpretation of the exact nature of the Suleimaniyah Type based on the agreement parameter. But before I cite the relevant examples, I quote MacKenzie's stance on the case in the dialect under study:

"In Sul . . . there are no inflective morphemes, and hence no distinction of *grammatical* gender or case." (p. 56)

As for agreement, sentences, (59) and (60) contain intransitive verbs formed with the present and past stems respectively. Sentence (61) is a transitive verb formed with the present stem. In all of them and likewise similar examples the S and A are always expressed in the personal ending of the verb. In these examples, I have followed MacKenzie's convention and have put the optional pronoun or noun subjects in the brackets. Morphemic segmentations and glosses are provided by me.

- (59) [min] a-rō-m  
 I Indicative-go-1sg.  
 'I go' (MacKenzie 1961: 106)

- (60) [min] rōīšt-im  
 I went-1sg.  
 'I went' (ibid.: 107)

- (61) [pyāw-aka] sag-aka a-kuž-ē  
 man-definite dog-definite Indicative-kill-3sg.  
 '[the man] kills the dog' (ibid.: 106)

Examples (62)-(67) also illustrate transitive verbs formed with the present stem. In these examples, the verb is the only (major)

constituent of the sentence and hence it encodes both the A and the P. The former is expressed as agreement suffixes and the latter as the oblique enclitics which is hosted by the Imperfect/Indicative prefix.

- (62) a-y-ē ḥim [a=y-ē ḥ-im]  
                           Imperfect-3sg.CLC-leave-1sg.  
                           'I shall leave it' (MacKenzie 1961: 88)

- (63) amdōzītawa [a=m-dōz-īt-awa]  
                           Imperfect-1sg.CLC-find-2sg.-postverb  
                           'thou wilt find me' (ibid.: 91)

- (64) a-y-kužē [a=y-kuž-ē]  
                           Imperfect-3sg.CLC-kill-3sg.  
                           'he kills it' (ibid.: 107)

- (65) a-m-dāt-ē [a=m-dāt-ē]  
                           Imperfect-1sg.CLC-give-3sg.  
                           'he gives me to (him)' (ibid.: 107)

- (66) amkužē [a=m-kuž-ē]  
                           Imperfect-1sg.CLC-kill-3sg.  
                           'she will kill me' (ibid.: 134)

- (67) ba das aygirīn [ba das a=y-gir-īn]  
                           with hand Imperfect-3sg.CLC-take-1pl.  
                           'we shall take it by hand' (ibid.: 134)

Examples (59)-(67) clearly manifest a Nominative-Accusative agreement system. In these examples, the S and A are identically expressed as agreement suffixes in the verb (vide all the mentioned examples) and the P if overtly present in the sentence is not cross-referenced (e.g., example (61)) or obligatorily

cross-referenced with the oblique enclitic if the verb is the only (major) constituent of the sentence (e.g., examples (62)-(67)).

Now we turn to the transitive verbs formed with the past stems. MacKenzie has called constructions containing these verbs Agential construction. In this construction the A is obligatorily cross-referenced with oblique enclitics. The P if overtly expressed is not cross-referenced. In examples (68) and (69), the P, or “Direct Affectee” as MacKenzie has named it hosts the oblique enclitic which cross-references the A, however, the P which are plural nouns are not cross-referenced. Segmentations and glosses are provided by me.

- (68) *mināl-ak-ān=ī*                      *nwān*  
       child-definite-pl.-3sg.CLC   put to bed  
       ‘she put the children to bed’ (MacKenzie 1961: 131)
- (69) *sag-ak-ān=ī*                      *kušt*  
       dog-definite-pl.-3sg.CLC   killed  
       ‘he killed the dogs’ (ibid.: 131)

Examples (70)-(72) substantiate the above mentioned observation. In these examples the P is an independent pronoun. As MacKenzie has pointed out “when an independent personal pronoun of the 1st or 2nd person appears as the Direct Affectee of the Agential construction . . . then the verb does not agree with it in person but appears in the 3rd person singular . . .” (ibid.: 75).

- (70) *min=it*                      *dī-w-a*  
       I-2sg.CLC   saw-participle-is  
       ‘thou hast seen me’ (ibid.: 75)
- (71) *ēma=t*                      *dī-w-a*  
       we-2sg.CLC   saw-participle-is  
       ‘thou hast seen us’ (ibid.: 75)

- (72) tō=yān                        nārd-uw-a  
 you (sg.)=3pl.CLC    sent-participle-is  
 ‘they have sent thee’ (MacKenzie 1961: 75)

If we take into consideration example (60) which contains an intransitive verb whose S is expressed in the verb by an agreement suffix on the one hand and examples (68)-(72) which contain transitive verbs formed with the past stem whose As are cross-referenced by oblique enclitics and their Ps are not cross-referenced on the other hand, then we can conclude that this set represents a Tripartite agreement system.

However, examples (73)-(82) which also contain transitive verbs formed from the past stem and As and Ps are not overtly present but are cross-referenced by the oblique enclitics and the verbal agreement suffixes respectively, manifest an Ergative-Absolutive agreement system. Morphemic segmentation and glosses are added by me.

- (73) dī-w=it-im  
 saw-participle-2sg.CLC-1sg.  
 ‘thou hast seen me’ (ibid.)
- (74) dī-w=it-īn  
 saw-participle-2sg.CLC-1pl.  
 ‘thou hast seen us’ (ibid.)
- (75) bō    čī        šet=tān                kird-im  
 for    what    mad-2pl.CLC    did-1sg.  
 ‘Why did you make me (out to be) mad?’ (ibid.: 109)
- (76) rā=y                        kešā-n  
 preverb-3sg.CLC    dragged-3pl.  
 ‘he dragged them’ (ibid.: 109)

- (77) lē-mān=ī sand-in  
      from-1pl.CLC-3sg.CLC took-3pl.  
      ‘he took them from us’ (MacKenzie 1961: 114)
- (78) sand-in=ī lē-mān  
      took-3pl.-3sg.CLC from-1pl.CLC  
      ‘he took them from us’ (ibid.: 114)
- (79) aw-ān=ī lē sand-īn  
      he-pl.-3sg.CLC from took-1pl.  
      ‘he took them from us’ (ibid.: 115)
- (80) lē=y sand-in-īn  
      from-3sg.CLC took-3pl.-1pl.  
      ‘he took them from us’ (ibid.: 115)
- (81) pē-mān=ī dā-n  
      to-1pl.CLC-3sg.CLC gave-3pl.  
      ‘he gave them to us’ (ibid.: 114)
- (82) dā-n=ī pē-mān  
      gave-3pl.-3sg.CLC to-1pl.CLC  
      ‘he gave them to us’ (ibid.: 114)

In examples (70)-(72) and (79) we notice that the P is the host for the A enclitics. In (75) the nonverbal constituent of the compound verb is the host for the enclitic. In (76) the preverb is the host. In (77) and (81) the prepositional phrase is the host. And in (80) the preposition itself serves as the host for the A enclitic. These observations are supplemented by examples (83)-(86) in which the negative prefix, the imperfect prefix, the past tense stem, and the past participle stem are the host for the A enclitics respectively. I have added the segmentation and the glosses.

(83) na=m-a-kird

Neg.-1sg.CLC-Imperfect-did

'I used not to do (so)' (MacKenzie 1961: 79)

(84) a=m-kird

Imperfect-1sg.CLC-did

'I used to do (so)' (ibid.)

(85) kird=im

did-1sg.CLC

'I did (so)' (ibid.)

(86) kird-uw=m-a

did-participle-1sg.CLC-is

'I have done (so)' (ibid.)

Finally, sentence (87) in which the A is overtly expressed and the past stem is the enclitic host for the A substantiates the fact that Kurdish does not obey the Wackernagel's Law.

(87) xalq-aka wut=yān

people-definite said-3pl.CLC

'the people said' (ibid.: 50)

Now, I return to examples (5)-(17) quoted from McCarus (2009), examples (18)-(25) quoted from McCarus (1958), and examples (26)-(34) quoted from MacKenzie (1961), namely the examples which I started this paper with, and propose that they along with examples (68)-(87) follow one general pattern for the A enclitic placement in Suleymanieh Kurdish: The first constituent of the verb phrase serves as the clitic host for the A. I suggest that examples (5)-(34) as well as examples (73), (74), (78), and (82) where the simple past tense verb is the only available constituent of the sentence or is the only constituent which

encodes both A and P, the past tense stem itself is the host for the A enclitic. Thus, I analyze examples such as (11)-(13), (16), (17), (23), (29), (30), (78), and (82) where the A enclitic follows the P agreement suffix as instances which signal a drift towards the stabilization of the past stem plus the P agreement suffix as a base which serves as a host for the A enclitics. The variations which we observe in (31)-(34), which permit the occurrence of the third person plural A enclitic after the past stem itself or the past stem plus the P agreement suffix, can be interpreted as transitional stages in the mentioned drift. Examples (14)-(17) which show that the first person singular whether A enclitic or the P suffix precedes any plural subject or object, are aptly explained by Haig's observation that the Speech Act Participants (SAP) act on, i.e., precede the non-SAP. Furthermore, it should be noted that the order of argument cross-referencing markers Verb-A-O in which the enclitic precedes the agreement suffix is by no means an isolated and unique phenomenon. As example (83) shows, here the A enclitic occurs in between two prefixes (i.e., Negative and Imperfect prefixes). In example (84), this enclitic occurs after the Imperfect prefix and before the past stem.

Now, I turn to another Sorani/Central Kurdish variety as it is spoken in Sanandaj in the Kurdistan Province in Iran. My informants' response to examples (5)-(17) are presented and then discussed below. In each instance, I read the meaning given for each example and asked them in separate sessions to express it in the Sanandaj dialect. Their responses to items (5)-(10) are given in (88)-(93).

- (88) di'=t=əm  
 saw-2sg.CLC-1sg.CLC  
 'I saw you (sg.)'

- (89) di'=m=o  
 saw-1sg.CLC-2sg.CLC  
 ‘you (sg.) saw me’
- (90) di'=tan=man  
 saw-2pl.CLC-1pl.CLC  
 ‘we saw you (pl.)’
- (91) di'=man=tan  
 saw-1pl.CLC-2pl.CLC  
 ‘you (pl.) saw us’
- (92) di'=yan=tan  
 saw-3pl.CLC-2pl.CLC  
 ‘you (pl.) saw them’
- (93) di'=tan=yan  
 saw-2pl.CLC-3pl.CLC  
 ‘they saw you (pl.)’

In these examples the transitive past stem is followed by the P cross-referencing enclitic and the A cross-referencing enclitic, respectively. Hence the mentioned examples illustrate an Oblique-Oblique agreement system. Examples (94)-(96) correspond to examples (11)-(13) in which the third person singular A enclitic always appears after the agreement suffix which cross-references the P. Although in Sanandaji variety too the A enclitic is final but the P cross-referencing marker is also an enclitic. Thus, examples (94)-(96) also manifest an Oblique-Oblique agreement system.

- (94) di'=m=i  
 saw-1sg.CLC-3sg.CLC  
 ‘he saw me’

- (95) di'=t=i  
 saw-2sg.CLC-3sg.CLC  
 'he saw you (sg.)'

- (96) di'=man=i  
 saw-1pl.CLC-3sg.CLC  
 'he saw us'

Finally, corresponding to examples (14)-(17), which illustrate that the first person singular agreement marker whether cross-referencing the A or the P, precedes any plural subject or object, there are examples (97)-(100) which uniformly reflect an Oblique-Oblique agreement system. In these examples too the A enclitic is always final.

- (97) di'=tan=əm  
 saw-2pl.CLC-1sg.CLC  
 'I saw you (pl.)'

- (98) di'=yan=əm  
 saw-3pl.CLC-1sg.CLC  
 'I saw them'

- (99) di'=m=tan  
 saw-1sg.CLC-2pl.CLC  
 'you (pl.) saw me'

- (100) di'=m=yan  
 saw-1sg.CLC-3pl.CLC  
 'they saw me'

Additional examples which strengthen the same conclusion are provided in (101)-(104) below. These examples correspond to examples (30) and (32)-(34), respectively.

- (101) di'=yan=i  
       saw-3pl.CLC-3sg.CLC  
       ‘he saw them’
- (102) di'=t=yan  
       saw-2sg.CLC-3pl.CLC  
       ‘they saw you (sg.)’
- (103) di'=man=yan  
       saw-1pl.CLC-3pl.CLC  
       ‘they saw us’
- (104) di'=yan=yan  
       saw-3pl.CLC-3pl.CLC  
       ‘they saw them’

Interestingly, in examples (105)-(107) below, which contain a compound verb, the P enclitic and the A enclitic are both directly attached in that order to the non-verbal constituent of the compound. Obviously, the agreement system here is also Oblique-Oblique.

- (105) æziyæt=tan=əm                    kərd  
       harm-2pl.CLC-1sg.CLC    did  
       ‘I harmed you’
- (106) xošhal=man=tan                    kərd  
       happy-1pl.CLC-2pl.CLC    did  
       ‘you (pl.) made us happy’
- (107) mučyari=tan=yan                    kərd  
       advice-2pl.CLC-3pl.CLC    did  
       ‘they advised you (pl.)’

However, in examples (108)-(110) the non-verbal part of the compound hosts the A enclitic and the verbal part (i.e., the light verb) encodes the agreement suffix which refers to the P. These examples represent an Ergative-Absolutive agreement system.

- (108) æziyət=ət            kərd-ən  
harm-2sg.CLC did-3pl.  
'you (sg.) harmed them'
- (109) xošal=ət            kərd-īn  
happy-2sg.CLC did-1pl.  
'you (sg.) made us happy'
- (110) mučyari=man        kərd-en  
advice-1pl.CLC did-2pl./3pl.  
'we advised you/them'

In all of the examples we have discussed so far the P is not overt in the sentence. In (111) and (112) the P is overtly expressed as a pronoun and a noun respectively. In these and similar examples the A is encoded through enclitic and the P is not cross-referenced. Taking into consideration the fact that S is always cross-referenced by agreement suffixes in the verb, then one can speak of this set as representing a Tripartite agreement system: S being nominative, A being oblique, and P showing no agreement.

- (111) mən    to=m                            di  
I        you (sg.)-1sg.CLC    saw  
'I saw you (sg.)'
- (112) æw-an    ow    žən-gæl=yan-æ                            di  
he/she-pl. that woman-pl.-3pl.CLC-definite    saw  
'they saw those women'

Now, the question is what the Type in the Kurdish of Sanandaj is. I reserve my answer for a little while later when I have dealt with the data from Kalhorî Kurdish, a Southern Kurdish variety spoken in Eslâmâbâd-e Gharb in the Kermanshah province. Examples (113)-(118) which correspond to examples (5)-(10) are highly revealing.

- (113) di'-m=aed  
saw-1sg.-2sg.CLC  
'I saw you (sg.)'
- (114) di'-d=aem  
saw-2sg.-1sg.CLC  
'you (sg.) saw me'
- (115) di'-men=aedan  
saw-1pl.-2pl.CLC  
'we saw you (pl.)'
- (116) di'-n=aeman  
saw-2pl./3pl.-1pl.CLC  
'you (pl.)/they saw us'
- (117) di'-n=eyan  
saw-2pl./3pl.-3pl.CLC  
'you (pl.)/they saw them'
- (118) di'-n=aedan  
saw-3pl.-2pl.CLC  
'they saw you (pl.)'

The agreement system manifested in these examples is uniformly Nominative-Accusative. In these examples the A is encoded as a verbal agreement suffix and the P is cross-referenced

by an oblique enclitic which systematically follows the agreement suffix. Examples (119) and (120) which contain an overt P substantiate this conclusion. In these examples the A is encoded by a verbal agreement suffix and the P is not cross-referenced.

- (119) me te di-m  
 I you (sg.) saw-1sg.  
 'I saw you (sg.)'

- (120) te me di-d  
 you (sg.) I saw-2sg.  
 'you (sg.) saw me'

Finally, we may notice example (121) which is the present tense counterpart of example (113). In this example too, the A is encoded as a verbal agreement suffix followed by the oblique enclitic which cross-references the covert P.

- (121) dün-em=ad  
 see-1sg.-2sg.CLC  
 'I see you (sg.)'

At this stage, I address the main point of this section, namely the identification of the exact type of the varieties of the Kurdish studied here. I start with Kalhori Kurdish. Kalhori Kurdish is uniformly Nominative-Accusative with respect to the agreement parameter. Thus, I call it a Strict Agreement Type. On the other hand, the two varieties of the Sorani Kurdish do not fall under a single categorization. In these varieties of Kurdish the sentences containing the intransitive verbs as well as the transitive verbs formed with the present stem manifest a Nominative-Accusative system. But the behavior of transitive verbs formed with the past stems is not uniform. In Sanandaji, the sentences which contain verbs formed with the past stems illustrate three patterns: (i)

Tripartite (cf. examples (111) and (112)), (ii) Oblique-Oblique (cf. examples (88)-(107)), and Ergative-Absolutive (cf. examples (108)-(110)). Taking for granted the fact that the S is always encoded as a verbal agreement suffix, then it is appropriate to propose that the Tripartite pattern is manifested in sentences in which the P is overtly present in the sentence. The Oblique-Oblique pattern is realized predominantly if the P is not overtly expressed but is cross-referenced by oblique enclitics. The Ergative-Absolutive pattern may be observed if P is not overtly expressed and the verb is a compound verb. So, to what type the Kurdish of Sanandaj belongs? I would suggest that it manifests a Split Agreement Type: Nominative-Accusative versus Non-Nominative-Accusative (which embraces the Tripartite, the Oblique-Oblique, and the Ergative-Absolutive patterns). The Sorani variety of Kurdish spoken in Suleimaniyeh could also be described as revealing a Split Agreement Type, i.e., Nominative-Accusative versus Non-Nominative-Accusative. However, it differs from the Kurdish of Sanandaj in that the former manifests two patterns with the verbs formed with the past stems. If the P is overtly expressed it is Tripartite (cf. examples (38), (39), and (51)-(53)). If the P is not overtly expressed but it is cross-referenced by an agreement suffix, then we witness an Ergative-Absolutive pattern (cf. examples (5)-(35), (54)-(58), (73)-(78), and (80)-(82)). In (79), the verb agrees with the prepositional object but not P. Thus, I conclude my characterization of Kurdish and suggest that the Sorani/Central Kurdish manifests a Split Agreement Type (namely Nominative-Accusative versus Non-Nominative Accusative) but the Kalhori Kurdish/Southern Kurdish reveals a Strict Agreement Type (namely a uniformly Nominative-Accusative Type).

Before I move on and turn to the typology of other Iranian languages based on the agreement parameter, I would like to briefly discuss the theoretical implications of my findings for our understanding of a general theory of human language.

In Figure 1, Comrie's proposed case-marking systems for S, A, and P were presented. Later, we learned that Comrie extended his case-marking system to verb-agreement. What our description of the Suleimaniyeh and Sanandaj dialects of Sorani Kurdish shows is that Comrie's type (d), the Tripartite system, and his type (e), which he has not given any name to it, that is the system whose A and P are identically marked and are distinct from S, are highly productive systems. This is an important lesson on type which we learn from the Iranian languages. The importance of this lesson becomes more transparent when we consider the more recent literature on this topic.

Croft (2003) assumes that "The Conceptual categories defined by S, A, and P are points in the conceptual space." (p. 144). He then presents four alignments between these conceptual categories which correspond to the (a-d) systems proposed in Figure 1 which I quoted from Comrie (1978). Croft describes these systems "... as semantic maps on the conceptual space . . ." (p. 145). His position on two logical types based on the case-marking is highly relevant to our present discussion. In his words:

"The fourth semantic map [i.e., system (d) in Figure 1] . . . , with all three distinct [namely the tripartite system], is quite rare, except with limited subclasses of S, A, and P. Dixon reports several languages in which tripartite marking is found on a subset of nouns, or in a subset of contexts; the only languages in which all noun phrases are reported to be consistently marked with distinct forms for S, A, and P are some Australian languages in southeast Queensland, including Wangkumara and Galali (Dixon 1994: 41). The one unattested type is the one in which one category subsumes A and P, and another S. This type would violate the Semantic Map Connectivity Hypothesis with respect to the conceptual space . . . , because there is no direct link between A and P. The only

instance of this pattern that I know of is found in the past tense only of Rushan, an Iranian language, and is restricted to pronouns and demonstratives (Payne 1980: 155; see also Dixon 1994: 39, n. 1). Moreover, the anomalous pattern is being replaced by the nominative-accusative present tense alignment by younger speakers.” (p. 145-146).

### 3.2. Talyshi

Talyshi is a North-Western Iranian language. The variety of Talyshi whose data will be discussed here is spoken in the Anbarānbālā village forty kilometers north of the city of Ardebil close to the Republic of Azerbaijan border. This variety belongs to the Northern Talysh group.

With the intransitive verbs and the transitive verbs formed with the present stem, S and A are encoded by verbal agreement suffixes. Examples (122) and (123) illustrate this point.

- (122) az umæn-æm  
 I come-1sg.  
 ‘I come’

- (123) az kitob-ə sæn-æm  
 I book-oblique buy-1sg.  
 ‘I buy the book’

These examples reflect a Nominative-Accusative system with respect to the agreement parameter.

However, in the sentences containing the transitive verbs formed with the past stem, the A is cross-referenced by the agent enclitic whose host can be the P or the verb itself. Sentences (124) and (125) support this observation. In these sentences the P is never realized by the verbal agreement suffixes.

- (124) man æv-ün=əm zənæ  
 I he/she-pl.-1sg.CLC knew  
 'I knew them'

- (125) man æv-ün zənæ=me  
 I he/she-pl. knew-1sg.CLC  
 'I knew them'

In this variety of Talyshi, the P is always overtly expressed and is never encoded in the verb. Sentence (126) provides another supporting example for this observation.

- (126) æmæ=š zæ  
 we-3sg.CLC hit  
 'he/she hit us'

Thus, I propose that in the sentences with the verbs formed with the past stem, Talyshi of Anbarānbālā manifests a Tripartite system based on the agreement parameter: S is expressed by the verbal agreement suffixes. A is cross-referenced via oblique enclitics and the P is not encoded through agreement at all. Therefore, this language has grammaticalized a Split Agreement system.

### 3.3. Davani

Davani is a South-Western Iranian language spoken in the village [dævan], locally called [dovū] in the North-East of Kazerun in the Fars province.

In the sentences with the intransitive verbs and the transitive verbs formed with the present stem, the verb always encodes the S and A by identical verbal agreement suffixes. Examples (127) and (128) represent this situation:

- (127) ušu me:-r-en  
 they Incomplete-come-3pl.  
 ‘they come’

- (128) ušu ketav-ačku me:-sen-en  
 they book-definite Incomplete-buy-3pl.  
 ‘they buy the book’

In (128) the P is not cross-referenced. Hence, the system revealed in these examples is Nominative-Accusative based on the agreement parameter.

In the sentences with the transitive verbs formed with the past stem, I have observed three possibilities. Most productively is the possibility in which the A is cross-referenced by an oblique enclitic and the P is not expressed through agreement suffix in the verb. This is shown in (129).

- (129) ušu=šu mu di  
 they-3pl.CLC we saw  
 ‘they saw us’

This example along with example (127) in which S is encoded via the agreement suffix manifests a Tripartite system based on the agreement parameter.

Very infrequently, I have come across examples such as (130) and (131) below in which the A is expressed via oblique enclitics and the P, which is overt in the sentence, is encoded by verbal agreement suffixes. This represents an Ergative-Absolutive system.

- (130) mæ=t deð-e  
 I-2sg.CLC saw-1sg.  
 ‘you (sg.) saw me’

- (131) to=t                                deð-u  
       you (sg.)-2sg.CLC    saw-1pl.  
       ‘you (sg.) saw us’

Finally, in compound verbs I have noticed the occurrence of two oblique enclitics both on the non-verbal constituent of the compound verb. The first of them cross-references the P and the second cross-references the A. Examples (132) and (133) substantiate this observation.

- (132) næsihæt=eš=omu                        ce  
       advice-3sg.CLC-1pl.CLC    did  
       ‘we advised him/her’
- (133) þar=emu=tu                                ze  
       call-1pl.CLC-2pl.CLC    hit  
       ‘you (pl.) called us’

The agreement system reflected in these examples is Oblique-Oblique.

Thus, I conclude that Davani has a Split Agreement system. Sentences with the present tense stem are Nominative-Accusative. Sentences with the past tense stem encompass three situations: (i) Tripartite, (ii) Ergative-Absolutive, and (iii) Oblique-Oblique.

#### 4. Conclusions

I conclude that the Iranian languages described here save Kalhorî Kurdish, would best be classified into two types: (i) Nominative-Accusative and (ii) Non-Nominative-Accusative. The latter embraces a number of possibilities including (a) Tripartite, (b) Oblique-Oblique or I here suggest Bipartite system, and (c)

Ergative-Absolutive. Thus, it will be an oversimplification to label any tense-sensitive Iranian language as Split-Ergative. I propose that this conclusion might be valid for many modern Iranian languages and that this observation for the state-of-the-art in modern Iranian languages might be equally valid for our understanding of the formation and development of modern Iranian languages diachronically. Namely a number of the Iranian languages in the past also simultaneously contained more than one system. I suppose that this is a valuable lesson that synchrony teaches us about diachrony, a perspective which I call synchronic diachrony.

These multi-system languages may stay in this position for several centuries. Thus, they should not necessarily be viewed as transitional, unstable, and temporary systems.

## References

- Anderson, S. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.
- Capell, A. 1969. *A Survey of New Guinea Languages*. Sydney: University of Sydney Press.
- Comrie, B. 1978. Ergativity. In W. Lehmann (ed.), *Syntactic Typology*. Sussex: The Harvester Press.
- Corbett, G. 2003. Agreement: The Range of the Phenomenon and the Principles of the Surrey Database of Agreement. *Transactions of the Philological Society* 101, 155-202.
- Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- \_\_\_\_\_. 2003. *Typology and Universals*. Cambridge: Cambridge University Press.
- Dixon, W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

- \_\_\_\_\_. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Fattah, M. 1997. *A Generative Grammar of Kurdish*. Ph.D Dissertation. University of Amsterdam.
- Greenberg, J. 1974. *Language Typology: A Historical and Analytic Overview*. The Hague: Mouton.
- Haig, G. 2008. *Alignment Change in Iranian Languages: A Construction Grammar Approach*. Berlin: Mouton de Gruyter.
- MacKenzie, D. 1961. *Kurdish Dialect Studies I*. London: Oxford University Press.
- McCarus, E. 1958. *A Kurdish Grammar: Descriptive Analysis of the Kurdish of Sulaimaniya, Iraq*. New York: American Council of Learned Societies.
- \_\_\_\_\_. 2009. Kurdish. In G. Windfuhr (ed.), *The Iranian Languages* 587-633. London & New York: Routledge.
- Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago, IL: University of Chicago Press.
- Payne, J. 1980. The Decay of Ergativity in Pamir Languages. *Lingua* 51, 147-186.
- Yar-Shater, E. 1969. *A Grammar of Southern Tati Dialects*. The Hague: Mouton.

Journal of Universal Language 13-1  
March 2012, 71-99

# **Spirantization and the Notion of Phonological Strength Relations in Assamese: An Optimality Theoretic Account**

**Hemanga Dutta**

*The English and Foreign Languages University (EFLU)*

## **Abstract**

The paper investigates the problem of phonological strength relations that account for the organization of speech sounds in a specific fashion in the light of spirantization process as attested in the Assamese language, an Indo-Aryan language spoken in the northeastern part of India.

In Assamese aspirated stops /p<sup>h</sup>/ and /b<sup>h</sup>/ are spirantized as [f] and [v] in word-final position, leaving the unaspirated stops intact. In word-initial and medial positions spirantization is blocked in Assamese. De-aspiration in Assamese does not apply to word final position when no following consonant occurs. In Assamese, coda is de-aspirated when it is followed by an aspirated onset. However

---

Hemanga Dutta

Department of Linguistics and Contemporary English, The English and Foreign Language University (EFLU), Hyderabad-50005, India  
Phone: 091-9550575673; Email: hemangadutta1@gmail.com

Received December 8, 2011; Revised January 20, 2012; Accepted January 31, 2012.

the fricatives /f/ and /v/ never lose their feature [+asp] despite the fact that both consonants occur in the word-final position or in coda position, being followed by aspirated onset. They turn in to /ph/ and /b<sup>h</sup>/ respectively when they are followed by obstruents. Nevertheless, the feature [+asp] is maintained in the onset position which does not undergo alternation. Distribution of Assamese aspirated phonemes at word boundary inform us that only the labial stops spirantize at the word-final position unlike coronal and velar stops which are not susceptible to the process of spirantization.

Keywords: phonological strength, spirantization, de-aspiration

## 1. Introduction

Strength is treated in pertinent literature either as a form of perceptual salience or intrinsic property embedded in a particular segment, as becomes evident in the works of Steriade's (1997) 'Licensing by cue' model, Beckman's (1998) 'Positional faithfulness view,' Kirchner's (1998) 'Integrated models of phonetics and phonology,' and Flemming's (2001) work on 'Phonetic bases of markedness.'

Strength relations can be understood to reflect on asymmetric licensing relations between units in representation. This issue of strength asymmetries in the patterning of segments can be correlated with the process of lenition, which has been described from various perspectives in domain phonological literature, such as feature spreading, sonority promotion, minimization of articulatory effort, and so on. Phonological strength can play a central role in the analysis of phonological patterning not only in the areas of language acquisition, pitch accent patterns and tonal phenomena but also in the well documented cases of segmental asymmetry. Strength relations can be understood to reflect on

asymmetric licensing relations between units in representation. This issue of strength asymmetries in the patterning of segments can be correlated with the process of lenition, which has been described from various perspectives in the domain phonological literature, such as feature spreading, sonority promotion, minimization of articulatory effort, and so on. This paper revolves around the theme of spirantization that can be subsumed under the rubric of lenition. This research work is designed to address the process of spirantization in Optimality theoretical model with the data drawn from Assamese, an Indo-Aryan language.

In Assamese aspirated stops /p<sup>h</sup>/ and /b<sup>h</sup>/ are spirantized as [f] and [v] in word-final position, leaving the unaspirated stops intact. In word-initial and medial positions spirantization is blocked in Assamese. De-aspiration in Assamese does not apply to the end of a phrase when no following consonant occurs. In Assamese, coda is de-aspirated when it is followed by an aspirated onset. However the fricatives /f/ and /v/ never lose their feature [+asp] despite the fact that both of these consonants occur in the word final position or in coda position, being followed by aspirated onset. They turn in to /p<sup>h</sup>/ and /b<sup>h</sup>/ respectively when they are followed by obstruents. Nevertheless, the feature [+asp] is maintained in the onset position which does not undergo alternation. Distribution of Assamese aspirated phonemes at word boundary informs us that only the labial stops spirantize at the word final position unlike coronal and velar stops which are not susceptible to the process of spirantization.

### **1.1. Lenition from the Perspective of Feature Spreading and Sonority Promotion**

The process of lenition in the domain of phonology has been analysed from various perspectives ranging from feature spreading under autosegmental framework to lenition as sonority promotion

treatment, but both the processes failed to offer an explicit and empirically adequate account of unified formal account of lenition or account for the contexts in which lenition normally occurs. Let us first treat the autosegmental treatment to analyse lenition which have been proposed by Selkirk (1982), Harris (1983), Mascaro (1983, 1987), Jacobs & Wetzel (1988), Cho (1990), and Lombardi (1991). However this treatment of feature spreading is confined to de-gemination, de-buccalisation, and elision as deletion or delinking of phonological material. It fails to provide an explicit account of most typical lenition context, that is the intervocalic position as it suffices to spread the relevant feature from either adjacent vowel and hence the role of the other vowel in conditioning the lenition is unexplained. Next, we can take in to consideration the other perception of lenition as sonority promotion theory which has been proposed by Foley (1977), Clements (1990), Hock (1992), and Lavoie (2001). But this notion also lacks explicit and unified phonetic grounds in explaining the process of lenition. As, for instance, we can consider the sonority scale proposed by Dell & Elmedlaoui (1985) that is stated below:

- (1) stops > voiceless fricatives > voiced fricatives > nasals > liquids > high vowels/glides > low vowels

This sonority scale will lead us to an incorrect assumption that fricatives ought to be able to lenite to nasals. Secondly, sonority account says nothing about the environment and the contexts in which lenition occurs.

## **1.2. Lenition and Articulatory Effort**

Kirchner (1998) proposed that lenition is driven by phonetic factors such as minimization of articulatory effort. Thus he proposed explicit and unified phonetic properties in characterizing

the notion of lenition, in particular the effort cost associated with a given set of articulatory gesture. Language specific lenition constraints emerge, according to Kirchner, from the effort minimization constraint, which is termed as LAZY. It interacts with some lenition blocking constraints within Optimality Theory. The conflict between Lazy and faithfulness constraints can be extended to a variety of lenition phenomena. The type of structural change that occurs in a specific language depends on lenition blocking constraints ranked below: In case of A, B occurs/A triggers B/if A is present, then B can be expected.

- (i) PRES (length) leads to de-gemination;
- (ii) PRES (voice) results in voicing;
- (iii) PRES (sonorant) leads to the reduction of an obstruent to an approximant;
- (iv) PRES (place feature) results in de-buccalization;
- (v) PRES (continuant) results in de-aspirantization in the language.

Lenition, thus assumes a unified characterization, in terms of ranking schema:

## (2) LAZY >> lenition blocking constraint

Spirantization as lenition phenomena: It is a process in which a segment turns in to features along with features associated with it. Kirchner argued that the PRES (continuant) will be dominated by LAZY; the resultant output will be spirantization in the language which is demonstrated in Table 1.

Table 1. Representation of Spirantization Where LAZY >> IDENT  
(cont)

/d/	LAZY	PRES (cont)
d	**!	
↳ ð	*	*

Table 2. Representation of Non Spirantization Where IDENT  
(cont) >> LAZY

/d/	PRES (cont)	LAZY
↳ d		**
ð	*!	

Spirantization occurs in Table 1 where LAZY dominates PRES (cont); in case of the opposite ranking, as exhibited in Table 2 spirantization is blocked. Such kinds of rankings point out that stops involve higher effort cost than continuants due to the greater distance which the articulator must travel in the former.

Prior to optimality analysis let us see how spirantization process is interpreted within the rubric of autosegmental model. I will also briefly comment on the favourable positions in which spirantization can be assigned cross-linguistically. Rhee (1998) claims that spirantization in most cases occurs in the contexts where target stop is adjacent to a [+cont] segment: prevocalic, postvocalic, or intervocalic position. Nevertheless intervocalic position flanked by both preceding and following [+cont] is found to be the most favourable environment cross-linguistically for a stop to lose its [-cont] and turns in to a segment bearing [+cont].

- (3) Examples of Intervocalic Spirantization
  - a. Kupiya (Christmas & Christmas 1975)

[ku:piya] - [ku:Φiya]	'kupia'
[su:pu] - [su:Φu]	'winnowing basket'

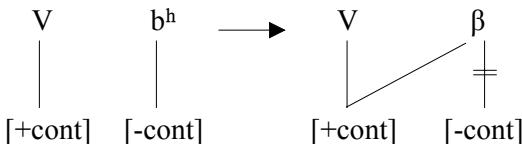
- b. Shina (Rajpurohit 1983)
- |                     |          |
|---------------------|----------|
| /babə/ - /baβo/     | 'father' |
| /darbak/ - /darβak/ | 'race'   |
- c. Florentine Italian (Kirchner 1996)
- |                           |               |
|---------------------------|---------------|
| /la tavola/ - [la əavola] | 'the table'   |
| /e dorme/ - [e ðɔrme]     | 's/he sleeps' |

The cases of prevocalic and postvocalic spirantization are also reported in phonological literature. In Boro (Bhat 1968, Bhattacharya 1977), it is found that a voiceless stop spirantizes in prevocalic position and in Nepali (Bandhu & Dahal 1971) aspirated stops spirantize in postvocalic word final position. Even Tigrinya (Kenstowicz 1982) and Tiberan Hebrew (Malone 1993) are reported to exhibit postvocalic spirantization.

- (4) a. Boro (Bhat 1968, Bhattacharya 1977)
- |   |          |
|---|----------|
| /p <sup>h</sup> ip <sup>h</sup> a/ - [Φip <sup>h</sup> a] | 'father' |
| /k <sup>h</sup> un/ - [xun]                               | 'cotton' |
- b. Nepali (Bandhu & Dahal 1971)
- |                             |          |
|-----------------------------|----------|
| /gəp <sup>h</sup> / - [gəΦ] | 'gossip' |
| /cəb <sup>h</sup> / - [cəβ] | 'dip'    |

The fact that spirantization always involves the assignment of the feature [+cont] to a stop can be addressed in analogy to the example drawn from Nepali postvocalic word final spirantization, as shown in (4b).

Figure 1. Typical Autosegmental Analysis: Assimilation to [+cont]



From the above representation in Figure 1, it is quite evident that spirantization is an instance of [+cont] assimilation. It may not be explicitly addressed in all cases but what one can conclude is that spirantization is an instance of ‘ease of articulation.’

The concept of ease of articulation for spirantization is more explicitly and formally expressed in Kirchner (1996, 1997, 1998) embedded in terms of ‘articulatory effort minimization.’ Kirchner (1998) claims that the phonetic imperative involved in the phonological interpretation of spirantization, minimizes the effort expenditure by means of reducing the magnitude of the articulatory gesture involved in consonant in question. This phonetic imperative in OT framework is classified as LAZY constraint which interacts with the faithfulness constraint to the continuancy that regulates the exactness of the correspondence between the input and the output specifications. A faithfulness constraint that militates against LAZY in the case of spirantization is IDENT (cont).

The constraints can be defined in the following fashion:

### (5) LAZY

Minimize articulatory effort.

### IDENT (cont)

Correspondent segments in input and output have identical values for continuancy.

Kirchner (1998) upholds this approach by virtue of the fact that spirants involve less articulatory effort than stops under the definition of effort as ‘a mental estimate of the neuromuscular activation levels required to achieve some set of gestures’ (Kirchner 1996: 1). Kirchner (*ibid.*) concludes that an articulatory gesture of greater displacement is more effortful than the one of lesser displacement. Hence, fricatives involve lesser articulatory effort than stops as articulatory displacement for target constriction for fricatives is lesser than that for stops. However, this view of Kirchner contradicts with the observation made by Boersma (1998) according to whom it is not fair to measure the effort merely by articulatory displacement. Boersma (*ibid.*) claims that it is rather likely that a ballistic movement for a stop, despite greater articulatory displacement, is easier articulatorily than a controlled articulatory movement found in a fricative. Underlying his belief lies is that fricative is more difficult to produce than a stop in articulation, and thus demands more effort than the stops which stands in contradiction with the assumption put forward by Kirchner. Silverman (1997: 5) also gives the view which is in consonance with Boersma when he says “fricatives are marked and presumably involve more effort to properly implement in comparison to stops.” Even the data from language acquisition contradict the views put forward by Kirchner. Stops are acquired earlier than fricatives, thereby standing as a setback to provide a positive conclusion that fricatives involve less effort than stops. In Boersma’s analysis (1998) articulatory effort is calculated with six primitives: energy, the presence of articulatory gestures, synchronization of gestures, precision, systemic effort, and coordination. Without a proper combination of all these parameters a study concerning articulatory effort expenditure, Kirchner’s approach remains elusive.

However, there are two studies conducted by Jacobs (1994) and Hahn (1998) concerning the process of spirantization bear affinities with that of Kirchner with some alternations. In his

analysis of diachronic French spirantization, Jacob proposed a consonantal markedness constraint called anti association constraint which is stated below:

(6) \*LE/[+vce, -cont]

Within LE (LE stands for lenition context), do not associate [+vce, -cont] to a margin after a peak.

The above constraint claims that a voiced stop is disfavoured in lenition context. So spirantization from voiced stop can be represented by ranking \*LE/[+vce, -cont] over IDENT (cont). Similarly Hahn (1998), in his analysis of German final spirantization, has given a constraint \*LENI (-cont)] $\sigma$ . The dominance of this constraint over IDENT (cont) is responsible for the final spirantization process in German.

(7) \*LENI (-cont)] $\sigma$ .

Stops in syllable final position are banned (i.e., do not associate [-cont] in the lenition context).

Both the processes given by Hahn and Jacob are somewhat similar with that of Kirchner although the methodology which they have employed for exhibiting the instance of spirantization is different. However, these approaches fail to make a correlation between spirantization and other closely related lenition processes such as intervocalic voicing and elision.

Prior to the phenomenon of spirantization let us consider the case of coda de-aspiration in Assamese.

## **2. Coda De-aspiration in Assamese**

In Assamese aspirated stops /p<sup>h</sup>/ and /b<sup>h</sup>/ are spirantized as [f]

and [v] in word-final position, leaving the unaspirated stops intact. Consider the following data:

- |                                       |                      |              |
|---------------------------------------|----------------------|--------------|
| (8) a. /kɒp <sup>h</sup> /            | [kɒf]                | 'phlegm'     |
| b. /map <sup>h</sup> /                | [maf]                | 'excuse'     |
| c. /sap <sup>h</sup> /                | [saf]                | 'clean'      |
| d. /baʊrp <sup>h</sup> /              | [baʊrf]              | 'ice cream'  |
| e. /lab <sup>h</sup> /                | [lav]                | 'profit'     |
| f. /xulɒb <sup>h</sup> /              | [xulɒv]              | 'cheap'      |
| g. /k <sup>h</sup> job <sup>h</sup> / | [k <sup>h</sup> iov] | 'anger'      |
| h. /nispɹɒb <sup>h</sup> /            | [nispɹɒv]            | 'lustreless' |

Sometimes the coronal stop /t<sup>h</sup>/ spirantizes to dental /t̪<sup>h</sup>/, as shown in the following data:

- |                            |                      |  |
|----------------------------|----------------------|--|
| (9) a. /zɛt <sup>h</sup> / | [zɛt̪ <sup>h</sup> ] | 'the second month of the Assamese year<br>corresponding to May-June' |
| b. /pit <sup>h</sup> /     | [pit̪ <sup>h</sup> ] | 'place'  |

However, unlike the labial and coronal stops the velar stops sometimes resist spirantization. But in the speech of some speakers (very rare) the coronal stop spirantizes to velar fricative, which is shown in (10).

- |                             |                     |           |               |
|-----------------------------|---------------------|-----------|---------------|
| (10) a. /svk <sup>h</sup> / | [svk <sup>h</sup> ] | vs. [svx] | 'style'       |
| b. /dɛk <sup>h</sup> /      | [dɛk <sup>h</sup> ] | vs. [dex] | 'to see'      |
| c. /lak <sup>h</sup> /      | [lak <sup>h</sup> ] | vs. [lax] | 'lacs'        |
| d. /jɔk <sup>h</sup> /      | [jɔk <sup>h</sup> ] | vs. [jɔx] | 'measurement' |
| e. /xɔk <sup>h</sup> /      | [xɔk <sup>h</sup> ] | vs. [xɔx] | 'happiness'   |
| f. /dɔk <sup>h</sup> /      | [dɔk <sup>h</sup> ] | vs. [dɔx] | 'sadness'     |

In word-initial and medial positions spirantization never occurs in Assamese, which is evident from the following data in (11).

- (11) a. /p<sup>h</sup>ɔli/ [p<sup>h</sup>ɔli] ‘slate’  
 b. /p<sup>h</sup>ɔlam/ [p<sup>h</sup>ɔlam] ‘colourful’  
 c. /p<sup>h</sup>ɔl/ [p<sup>h</sup>ɔl] ‘flower’  
 d. /p<sup>h</sup>ɔaṛa/ [p<sup>h</sup>ɔaṛa] ‘fountain’  
 e. /p<sup>h</sup>agɔn/ [p<sup>h</sup>agɔn] ‘name of a month’  
 f. /p<sup>h</sup>ɔrɔj/ [p<sup>h</sup>ɔrɔj] ‘cricket’  
 g. /b<sup>h</sup>ɔmɔṛa/ [b<sup>h</sup>ɔmɔṛa] ‘an insect’  
 h. /b<sup>h</sup>ɔṛi/ [b<sup>h</sup>ɔṛi] ‘foot’  
 i. /b<sup>h</sup>ɔmɔkap<sup>h</sup>ɔlia/ [b<sup>h</sup>ɔmɔkap<sup>h</sup>ɔlia] ‘very colourful’  
 j. /b<sup>h</sup>alɔk/ [b<sup>h</sup>alɔk] ‘beer’  
 k. /ɛb<sup>h</sup>ɔṛi/ [ɛb<sup>h</sup>ɔṛi] ‘a unit of measurement’  
 l. /dɔb<sup>h</sup>ag/ [dɔb<sup>h</sup>ag] ‘divide’  
 m. /gab<sup>h</sup>ɔṛɔ/ [gab<sup>h</sup>ɔṛɔ] ‘young girl’  
 n. /gɔrb<sup>h</sup>a/ [gɔrb<sup>h</sup>a] ‘womb’  
 o. /t<sup>h</sup>ɔṛia/ [t<sup>h</sup>ɔṛia] ‘ear rings’  
 p. /t<sup>h</sup>aṛi/ [t<sup>h</sup>aṛi] ‘small branch’  
 q. /t<sup>h</sup>ikɔna/ [t<sup>h</sup>ikɔna] ‘address’  
 r. /t<sup>h</sup>ɛh/ [t<sup>h</sup>ɛh] ‘offended’  
 s. /t<sup>h</sup>ik t<sup>h</sup>ak/ [t<sup>h</sup>ik t<sup>h</sup>ak] ‘ok’  
 t. /t<sup>h</sup>vga/ [t<sup>h</sup>vga] ‘a place where holy books are kept’  
 u. /at<sup>h</sup>ɔa/ [at<sup>h</sup>ɔa] ‘net’  
 v. /gat<sup>h</sup>i/ [gat<sup>h</sup>i] ‘joint’  
 w. /gat<sup>h</sup>ɔlɔ/ [gat<sup>h</sup>ɔlɔ] ‘dwarf’  
 x. /jat<sup>h</sup>i/ [jat<sup>h</sup>i] ‘a tool used in warfare’

Assamese word final spirantization of the stops /p<sup>h</sup>/ and /b<sup>h</sup>/ into /f/ and /v/ can be explained in the framework of Kirchner’s model as shown in Table 3 within OT framework:

Table 3. Spirantization of /p<sup>h</sup>/ in Assamese Where LAZY >> IDENT (cont)

/p <sup>h</sup> /	LAZY	IDENT (cont)
p <sup>h</sup>	*!	
↳ f		*

Table 4. Spirantization of /b<sup>h</sup>/ in Assamese Where LAZY >> IDENT (cont)

/b <sup>h</sup> /	LAZY	IDENT (cont)
b <sup>h</sup>	*!	
V		*

The data of spirantization in Assamese can be correlated with the phenomenon of Assamese coda de-aspiration. Consider the following data on coda de-aspiration in Assamese in (12).

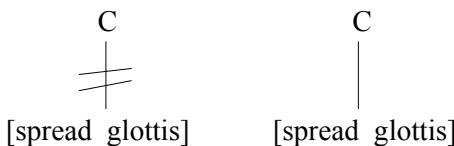
- (12) a. k<sup>v</sup>t<sup>h</sup> k<sup>h</sup>v<sup>n</sup> - k<sup>v</sup>t k<sup>h</sup>v<sup>n</sup> ‘the mat’  
 b. x<sup>v</sup>t<sup>h</sup> k<sup>h</sup>v<sup>n</sup> - x<sup>v</sup>t k<sup>h</sup>v<sup>n</sup> ‘the chariot’  
 c. ad<sup>h</sup> b<sup>h</sup>ag - ad b<sup>h</sup>ag ‘middle part’  
 d. at<sup>h</sup> k<sup>h</sup>ila - at k<sup>h</sup>ila ‘eight pieces’  
 e. kath<sup>h</sup> p<sup>h</sup>vta - kat p<sup>h</sup>vta ‘wood splitting’  
 f. bag<sup>h</sup> b<sup>h</sup>al<sup>h</sup>k - bag b<sup>h</sup>al<sup>h</sup>k ‘tigers and beers’  
 g. d<sup>v</sup>k<sup>h</sup> k<sup>h</sup>ini - d<sup>v</sup>k k<sup>h</sup>ini ‘the sadness’  
 h. b<sup>v</sup>r<sup>h</sup>f b<sup>h</sup>a<sup>j</sup>a - b<sup>v</sup>r<sup>h</sup>b<sup>h</sup> b<sup>h</sup>a<sup>j</sup>a ‘breaking of the ice’  
 i. d<sup>v</sup>k<sup>h</sup> b<sup>h</sup>ag - d<sup>v</sup>g b<sup>h</sup>ag ‘portion of the sadness’  
 j. bag<sup>h</sup> g<sup>h</sup>v<sup>j</sup> - bag g<sup>h</sup>v<sup>j</sup> ‘tigers and leopard’  
 k. b<sup>v</sup>r<sup>h</sup>f k<sup>h</sup>ini - b<sup>v</sup>r<sup>h</sup>p<sup>h</sup> k<sup>h</sup>ini ‘the ice’  
 l. kath<sup>h</sup> b<sup>h</sup>a<sup>j</sup>a - kat b<sup>h</sup>a<sup>j</sup>a ‘breaking of the wood’  
 m. k<sup>v</sup>f k<sup>h</sup>ini - k<sup>v</sup>p<sup>h</sup> k<sup>h</sup>ini ‘the phlegm’  
 n. saf b<sup>h</sup>ɔri - sab<sup>h</sup> b<sup>h</sup>ɔri ‘clean feet’  
 o. saf g<sup>h</sup>ɔri - sab<sup>h</sup> g<sup>h</sup>ɔri ‘clean house’

- p. *kv̪f b<sup>h</sup>v̪ra* - *kv̪v̪b<sup>h</sup> b<sup>h</sup>v̪ra* ‘filled with phlegm’  
q. *lak<sup>h</sup> k<sup>h</sup>elimeli* - *lak k<sup>h</sup>elimeli* ‘dozens of problems’  
r. *lav k<sup>h</sup>ini* - *lab<sup>h</sup> k<sup>h</sup>ini* ‘the profit’  
s. *xɔlv̪v̪ b<sup>h</sup>v̪uti* - *xɔlv̪v̪b<sup>h</sup> b<sup>h</sup>v̪ut* ‘admission at cheap rate’

## 2.1. Assamese Coda De-aspiration and Autosegmental Representation

This process of coda de-aspiration can be represented in the following fashion in autosegmental framework:

Figure 2. Representation of Coda De-aspiration in Autosegmental Module



Here the C connected to the node of de-linking line represents the consonant segment occurring in the coda position which is de-aspirated, but the consonant in the onset position retains its feature [spread glottis] thereby enhancing the onset coda asymmetry in phonological literature. But there are some exceptions to this pattern as exhibited by the segments /b<sup>h</sup>/ and /p<sup>h</sup>/.

What is noticeable in this context is that only the labial and velar consonants (optionally) are subject to spirantization. The coronals are never subject to spirantization in word-final position. Secondly, the process of spirantization is blocked when the word final fricatives are followed by the stops. Rather the fricatives lose their feature [+cont] and turn in to [+asp] stops. Thirdly, it is observed that coda de-aspiration in Assamese followed by another aspirated stop in the onset position is blocked in the case of

spirants.

As, for instance, consider the cases:

- (i)  $p^h \rightarrow f/-\#$  and when followed by liquids, nasals, and fricatives. However, when it is followed by obstruents it becomes  $p^h$  or  $b^h$ .

(13)	a. k <sub>v</sub> f svf - kvph svf b. saf sikvn - saf sikvn c. saf zilikil - saf zilikil d. b <sub>v</sub> rnf xarx - brvrf xarx e. arad d <sub>v</sub> rvn - arad d <sub>v</sub> rvn f. b <sub>v</sub> rnf kh <sub>v</sub> sa - brvrf kh <sub>v</sub> sa g. b <sub>v</sub> rnf lva - brvrf lva h. b <sub>v</sub> rnf nai - brvrf nai i. saf zuta - sav zuta j. svf g <sub>v</sub> ti - svb <sup>h</sup> g <sub>v</sub> ti	'phlegm and so on' 'cleanliness' 'crystal clear' 'sweeping of ice' 'falling of ice' 'to have ice cream' 'to take ice cream' 'no ice' 'clean shoes' 'spices'
------	---	--

- (ii)  $b^h \rightarrow v/-\#$  or when followed by fricatives, nasals, and liquids. But it becomes  $b^h$  when it is followed by obstruents.

(14)	a. lav sab <sup>h</sup> - lav sav b. lav za <sup>h</sup> - lav za <sup>h</sup> c. lav ximan - lav ximan d. lav h <sub>v</sub> sa - lav h <sub>v</sub> sa e. lav b <sub>v</sub> ra - lab <sup>h</sup> b <sub>v</sub> ra f. lav dija - lab <sup>h</sup> dija g. lav go <sup>h</sup> - lab <sup>h</sup> go <sup>h</sup> h. lav gh <sub>v</sub> rvn - lab <sup>h</sup> gh <sub>v</sub> rvn i. lav k <sub>v</sub> ra - lab <sup>h</sup> k <sub>v</sub> ra j. lav khini - lab <sup>h</sup> khini	'profit and so on' 'whose profit' 'that much profit' 'to make profit' 'profits' 'to give profit' 'to lose profit' 'to return profit' 'to make profit' 'the profit'
------	---	---

## 2.2. Spirantization, Coda De-aspiration, and OT Constraints

Autosegmental representation fails to show the process of spirantization in Assamese in an explicit manner. Hence it becomes imperative on our part to analyse this phenomenon with the help of some constraints within Optimality Theory which are addressed below:

- (i) Coda Condition (Coda Con): (Kager 1999)

The feature spread glottis [s.g.] is not allowed in the coda position. This constraint prohibits the feature spread glottis [s.g.] from appearing in the coda position.

- (ii) \*s.g.: (Davis & Cho 2003)

The feature spread glottis is prohibited. It can be treated as a general markedness constraint militating against the feature [s.g.].

- (iii) MAX-s.g.: (Davis & Cho 2003)

The feature spread glottis [s.g.] in the input must have a corresponding feature [s.g.] in the output. This constraint implies-maximize input segments in the output in terms of feature spread: one violation for each segment in regard to feature spread does not appear in the output. The main function of this constraint lies in the fact that it prevents segments to be deleted.

- (iv) \*OCP-s.g.: (motivated by Goldsmith 1976; ‘No identical adjacent autosegments’)

This constraint implies: avoid two instances of adjacent [s.g.]. The roots of this constraint can be traced back to the Obligatory Contour Principle, which was originally conceived for tonal

dissimilation phenomena in Mende and other African tone languages (Leben 1973), and manifested in proper fashion in Goldsmith (1976):

(15) The Obligatory Contour Principle (OCP)

“At the melodic level, adjacent identical elements are prohibited.”

In OCP the central notion revolves round ‘adjacency’ (Myers 1987, Archangeli & Pulleyblank 1994) which can be categorised in to segment adjacency and tier adjacency. It is the second aspect of adjacency that is instrumental in classical autosegmental theory, as developed by Leben (1973) and Goldsmith (1976), with the pursuit of reducing apparent action at a distance of locality. In the further pursuit of this theory additional representational assumptions became important which include feature geometric separation of feature groups (Clements 1985, Ito 1986, Sagey 1986), morphemically defined tiers (McCarthy & Prince 1986) and crucial underspecification of certain features (Archangeli & Pulleyblank 1994, Steriade 1997, Ito & Mester 1998, and many other works).

OCP is analysed in the framework of Optimality Theory too (Prince & Smolensky 1993) the main motivation of which lies in understanding the process of OCP, and of featural dissimilation in general. In OT two types of principles have emerged regarding the notion of OCP: one group of phonologists want to maintain the principle in a more or less unchanged form, with diversification in terms of special features and feature groups, as one of rankable and violable constraint that make up an OT grammar and this assumption has been reflected in the works such as the tonal area (Myers 1987) and in segmental phonology (McCarthy & Prince 1986). Another group is more interested to explore whether it is possible in OT to reduce the OCP to more fundamental notions and restrictions, thereby achieving a deeper level of explanation.

Ito & Mester (1998) claim that there is no Obligatory Contour Principle per se: Universal Grammar is not concerned about adjacent identicals qua identicals. Rather, OCP effects arise when markedness constraints are violated more than once.

OCP effects obtain when a given marked type of structure is present more than once within the same local domain.

Multiple violations of one and the same markedness constraint do not simply add up, but interact more strongly, so that a double violation within a given domain is worse than simply the sum of two individual violations. This notion of violation enhancement can get a formal expression by means of self conjunction of constraints (Tesar & Smolensky 1998).

#### (v) Local Conjunction of Constraints (LCC)

Kager (1999) claims that under Local Conjunction, two constraints are conjoined as a single composite constraint which is violated if and only if both of its components are violated within the same domain. However this constraint can be traced back to Prince & Smolensky (1993) according to whom in addition to a set of Con of universal constraints, Universal Grammar contains an operation on Con: local conjunction. In addition to reranking of constraints, local conjunction shows the way in which a line of demarcation can be drawn between individual grammars by making use of combined constraints. The chief motivation of combined constraints lies in allowing grammars to capture a particular type of constraint interaction that cannot be obtained in a theory exclusively built on direct strict domination (Prince & Smolensky 1993), but is attested in the phonologies of natural languages. We can express this constraint following Ito & Mester (1998).

## (16) a. Definition:

Local Conjunction is an operation on the constraint set forming composite constraints: Let C1 and C2 be members of the constraint set Con. Then their local conjunction C1 and C2 is also a member of Con.

## b. Interpretation:

The local conjunction C1C2 is violated if and only if both \*C1 and \*C2 are violated in some domain  $\sigma$ .

## c. Ranking (universal): C1 &amp; C2 &gt;&gt; C1

$$\text{C1} \& \text{C2} >> \text{C2}$$

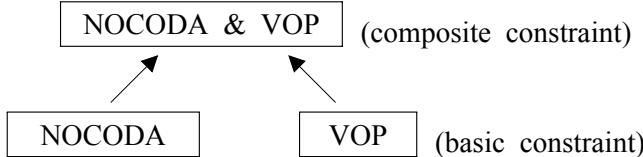
C1 & C2 is potentially active when there is some constraint C $\alpha$  (typically, but not necessarily a faithfulness constraint militating against violations of C1 and/or C2) ranked between the conjoined constraint and at least one of the two basic constraints, as shown below:

d. C1 &  $\delta$ C2 >> C $\alpha$  >> C1 or,

$$\text{C1} \& \text{C2} >> \text{C}\alpha >> \text{C2}$$

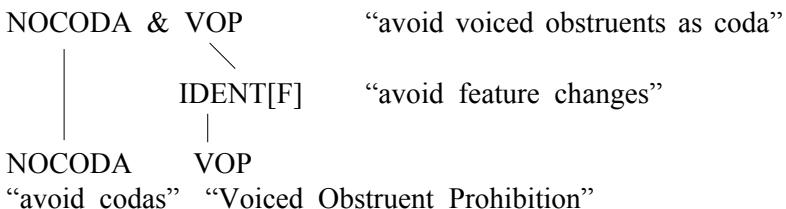
The constraint of Local Conjunction can better be analysed in the constraint based analysis of German Coda Devoicing, which implies that voiced obstruents are marked elements, and syllable codas are marked positions. The phonology of German permits both, thereby emphasizing on the faithful parsing of the input and what is ruled out is the marked in the marked position: a voiced obstruent as a coda. Following Ito & Mester (1998) below is given the constraint conjunctive analysis of German final devoicing where the two individual constraints involved are the syllable structure constraint NOCODA and the segmental markedness constraint against voiced obstruents, which is termed Voiced Obstruent Prohibition (VOP):

(17)



The additional factor which is responsible for coda devoicing, while at the same time limiting it to coda position, is the faithfulness ranking: the position of IDENT[F] below the conjoined constraint NOCODA & VOP and above the simple feature markedness constraint VOP.

Figure 3. Representation of Constraint Conjunctive Analysis of German Final Devoicing



Indeed, the chief motivation for Local Conjunction of constraints resides in chain shift (Kirchner 1996). According to Kager (1999) a chain shift implies a situation in which sounds are promoted or demoted stepwise along some scale in some context. The chain shift crucially does not result in neutralization, since each input occupies one step precisely. This can be represented in the following schemata:

(18)  $A \rightarrow B$  and  $B \rightarrow C$ , but not  $*A \rightarrow C$

### 2.3. Assamese Data on Spirantization and the OT Constraints: An Analysis

Table 5. Assamese Data on Spirantization

Final consonant	Voiceless unaspirated stop	Voiced unaspirated stop	Voiceless aspirated stop	Voiced aspirated stop	Voiceless aspirate	Voiced aspirate
f = p <sup>h</sup>	p <sup>h</sup> k	b <sup>h</sup> g	p <sup>h</sup> k <sup>h</sup>	b <sup>h</sup> g <sup>h</sup>	f s	v z
t <sup>h</sup>	t <sup>h</sup> k	d <sup>h</sup> g	t k <sup>h</sup>	d g <sup>h</sup>	t s	d z
k <sup>h</sup>	k <sup>h</sup> k	g <sup>h</sup> g	k k <sup>h</sup>	g g <sup>h</sup>	k s	g z
v = b <sup>h</sup>	b <sup>h</sup> k	b <sup>h</sup> g	b <sup>h</sup> k <sup>h</sup>	b <sup>h</sup> g <sup>h</sup>	v s	v z
d <sup>h</sup>	d <sup>h</sup> k	d <sup>h</sup> g	d k <sup>h</sup>	d g <sup>h</sup>	d s	d z
g <sup>h</sup>	g <sup>h</sup> k	g <sup>h</sup> g	g k <sup>h</sup>	g g <sup>h</sup>	g s	g z
P	p k	b g	p k <sup>h</sup>	b g <sup>h</sup>	p s	b z
t	t k	d g	t k <sup>h</sup>	d g <sup>h</sup>	t s	d z
k	k k	g g	k k <sup>h</sup>	g g <sup>h</sup>	k s	g z
b	b k	b g	b k <sup>h</sup>	b g <sup>h</sup>	b s	b z
d	d k	d g	d k <sup>h</sup>	d g <sup>h</sup>	d s	d z
g	g k	g g	g k <sup>h</sup>	g g <sup>h</sup>	g s	g z

The above table represents the distribution of Assamese aspirated phonemes at word boundary where it is observed that only the labial stops spirantize at the word final position unlike coronal and velar stops which are not susceptible to the process of spirantization. This can be represented within OT constraints following Kirchner (1998) as has been shown in section.

The fact the coda de-aspiration in fast tempo in Assamese following another aspirated stop in the following onset position can be analysed from the perspective of Obligatory Contour Principle, that is, identical segments do not occur in adjacent

position. Consider the following example:

- (19) k<sup>pt<sup>h</sup></sup> k<sup>hvn</sup> - k<sup>pt</sup> k<sup>hvn</sup> ‘the mat’

In the above example the coda is de-aspirated when it is followed by another aspirated stop in the onset position. It implies that two consecutive aspirated stops cannot occur. Here, \*OCP-s.g. is higher ranked than MAX-s.g., Coda Con and \*s.g. as evident from the following table:

- (20) /k<sup>pt<sup>h</sup></sup> k<sup>hvn</sup>/ - [k<sup>pt</sup> k<sup>hvn</sup>] ‘the mat’

Table 6. Representation of Coda De-aspiration in /k<sup>pt<sup>h</sup></sup> k<sup>hvn</sup>/  
Where \*OCP-s.g. > MAX-s.g. > Coda Con > \*s.g.

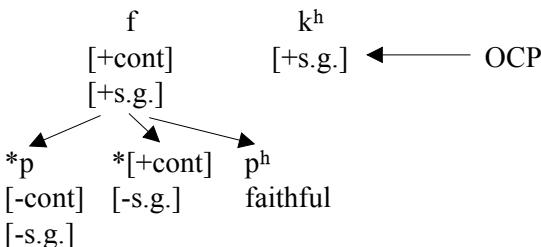
k <sup>pt<sup>h</sup></sup> k <sup>hvn</sup> --/	*OCP-s.g.	MAX-s.g.	Coda Con	*s.g.
(a)/k <sup>pt<sup>h</sup></sup> . k <sup>hvn</sup> /	*!		*	***
↳ (b) k <sup>pt</sup> . k <sup>hvn</sup>		*		**
(c) k <sup>pt<sup>h</sup></sup> . k <sup>vn</sup>		*	*!	**

Here from this table we get the winning candidate (b) which, in addition to violating lower ranked \*s.g., violates MAX-s.g. and it is higher ranked than Coda Con and \*s.g. In this table the candidate (a) violates \*OCP-s.g. which outranks all other constraints. On the other hand candidate (b) and candidate (c) violate \*s.g. twice and MAX-s.g. once yet candidate (b) appears as winner. It is because candidate (c), in addition to violating \*s.g. and MAX-s.g., violates Coda Con. So in this instance of coda de-aspiration the ranking schemata can be represented in the following way:

- (21) Constraint ranking: \*OCP-s.g. >> MAX-s.g. >> Coda Con,  
\*s.g.

But this OCP constraint is violated in case of spirants following by another aspirated stops in onset position. As, for instance, the fricatives /f/ and /v/ never lose their feature [+asp] despite the fact that both of these consonants occur in the word final position or in coda position, being followed by aspirated onset. They turn in to /ph<sup>h</sup>/ and /bh<sup>h</sup>/ respectively when they are followed by obstruents. Nevertheless, the feature [+asp] is maintained in the onset position which is never susceptible to change. An analysis is given below in the light of the example given below:

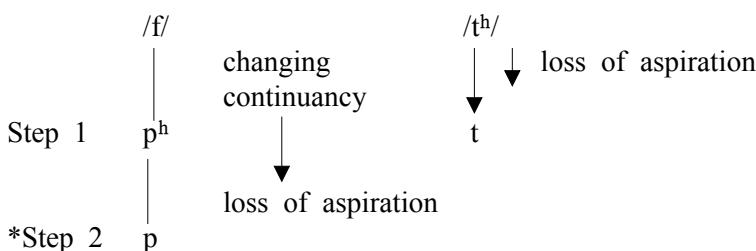
- (22) /bvəmf k<sup>h</sup>ini/ - [bvəmɒp<sup>h</sup> k<sup>h</sup>ini]



This phenomenon can be observed by using OT theoretical model by ranking constraints. In the above example it is evident that the continuancy is a feature which is not susceptible to OCP constraint whereas the stops having the features [-cont] lose their feature [s.g.] under OCP when it is followed by another aspirated stop in onset position. In the above interpretation of /f/ three possible outputs can be expected: segments having the feature ([-cont][[-s.g.]]) which is violable under faithfulness constraint. The segments having the features ([+cont][[-s.g.]]) is not phonetically plausible. It is ruled out by the phonetic constraint \*/+cont, -s.g./. So it assumes the feature [+asp] in spite of occurring in the coda position thereby being faithful to the feature continuancy in input. In contrast, in Assamese the segments having [-cont] and [+s.g.]

undergo transformation and lose their [+asp] feature when they are followed by another aspirated stop. This issue can also be analysed from the perspective of derivational steps. In the case of coda de-aspiration in Assamese the final stop undergoes only one transformation. It implies that de-aspiration occurs under the contact of the following aspirated stop. But in case of spirants in the final coda position, we do not see a single derivational step involved in the process of de-aspiration. The feature [+cont] blocks the spirants to lose the feature [+s.g.] although they are bound to lose [+cont]. It is evident from the following diagrammatic representation:

Figure 4. Representation of Steps Involved in the Transformation of Features



In order to authenticate this notion I am going to take in to consideration some OT constraints that I have discussed in the previous section. Consider the following example:

- (23) a. arad̪ d̪arad̪ - arad̪ d̪arad̪ ‘falling of ice’  
 b. kʰɔɑ kʰɔɑ - kʰɔɑ kʰɔɑ ‘to have ice cream’

Now consider the input /f kʰ/.

Table 7. OT Analysis of the Assamese Input /f k<sup>h</sup>/

Input	*[+continuant] [-continuant]	Faith[s.g.] & Faith[cont]	*OCP-s.g.	*[s.g.]	Coda Con
(a) f k <sup>h</sup>	!*		*	*	*
⌚ (b) p <sup>h</sup> k <sup>h</sup>			*		*
(c) pk <sup>h</sup>		*!		*	
(d) p <sup>h</sup> k		*!		*	*

Here, in this OT framework, the candidate (b) appears as the optimal candidate as the constraints violated by (b) are lower ranked in Assamese.

The candidate (a) violates \*[+continuant][-continuant], that is, a consonant which is [+continuant] can not be followed by another consonant which is [-continuant]. In addition it violates the constraint \*OCP-s.g., that prohibits identical adjacent segments. It violates \*[s.g.], that is a markedness constraint militating against the feature spread glottis, and Coda Con, which implies that the feature spread glottis is not allowed in the coda position. The candidate (c) violates Local Conjunction Constraint, that is Faith[s.g.] & Faith[cont], that implies that the output must be faithful in terms of feature spread glottis and continuancy to its input counterpart. This Local Conjunction is violated if both \*C1 and \*C2 are violated in a local domain. The candidate (c) violates LCC by virtue of violating not only Faith[s.g.] but also Faith[cont]. Candidate (d) is ruled out on the ground of violating Faith[s.g.] & Faith[cont] apart from \*[s.g.] and Coda Con.

The fact that spirantization of labial fricatives at word-final position in Assamese is blocked when followed by stops can be thus represented through the reranking of the constraints within OT framework.

The ranking of the constraints held responsible for this process can be depicted in the following fashion:

- (24) \*[+cont][-cont] >> Faith[s.g.] & Faith[cont] >> \*OCP-s.g.  
 >> \*s.g. >> Coda Con

### 3. Conclusions

Spirantization is treated in this paper as a process of lenition which is detected in the word-final position. Similarly to coda de-aspiration, spirantization occurs in the coda or the word-final position and thereby supports the claim of positional asymmetry or positional privilege. Spirantization is never found in word-initial and medial position, as evident from the data on Assamese. However cross linguistically word medial position is found to be the suitable place for spirantization to occur. Hence it must be argued that language-specific phonotactics is bound to play a crucial role in the patterning of speech sounds. We have demonstrated that the fricatives /f/ and /v/ never lose their feature [+s.g.] despite the fact that both consonants occur in the word-final position or in coda position, followed by aspirated onset. They turn into /p<sup>h</sup>/ and /b<sup>h</sup>/ respectively when followed by obstruents. Nevertheless, the feature [+s.g.] is maintained in the onset position which does not undergo alternation. The distribution of Assamese aspirated phonemes at word boundary reveals that only the labial stops spirantize at the word-final position unlike coronal and velar stops which are not susceptible to the process of spirantization. Hence the ranking of the constraints responsible for spirantization in Assamese is:

- \*[+cont][-cont] >> Faith[s.g.] & Faith[cont] >> \*OCP-s.g. >>  
 \*s.g. >> Coda Con

## References

- Archangeli, D. & D. Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.
- Bandhu, C. & B. Dahal. 1971. *Nepali Segmental Phonology*. Kirtipur: Tribhuvan University Press.
- Beckman, J. 1998. Positional Faithfulness. Ph.D Dissertation. University of Massachusetts, Amherst.
- Bhat, D. 1968. *Boro Vocabulary: With a Grammatical Sketch*. Poona: Deccan College, Postgraduate & Research Institute.
- Bhattacharya, P. 1977. *A Descriptive Analysis of the Boro Language*. Gauhati: Gauhati University Press.
- Boersma, P. 1998. Functional Phonology. Ph.D Dissertation. University of Amsterdam.
- Cho, Y. 1990. Parameters of Consonantal Assimilation. Ph.D Dissertation. Stanford University.
- Christmas, R. & J. Christmas. 1975. *Kuria Phonemic Summary*. Kathmandu: Tribhuvan University Press.
- Clements, G. 1985. The Geometry of Phonological Features. In C. Ewen & E. Kaisse (eds.), *Phonology Yearbook II* 225-252. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1990. The Role of the Sonority Cycle in Core Syllabification. In J. Kingston & M. Beckman (eds.), *Laboratory Phonology I: Between the Grammar and Physics of Speech* 283-333. Cambridge: Cambridge University Press.
- Davis, S. & M-H. Cho. 2003. The Distribution of Aspirated Stops and /h/ in American English and Korean: An Alignment Approach with Typological Implications. *Linguistics* 41, 607-652.
- Dell, F. & M. Elmedlaoui. 1985. Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7, 105-130.
- Flemming, E. 2001. Contrast and Perceptual Distinctiveness. In B. Hayes et al. (eds.), *The Phonetic Bases of Markedness*.

- Cambridge: Cambridge University Press.
- Foley, J. 1977. *Foundations of Theoretical Phonology*. Cambridge: Cambridge University Press.
- Goldsmith, J. 1976. Autosegmental Phonology. Ph.D Dissertation. MIT.
- Hahn, A. 1998. German Final Devoicing in Optimality Theory. Ms., University of Paderborn [ROA 241-0198].
- Harris, J. 1983. *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge, MA: MIT Press.
- Hock, H. 1992. Initial Strengthening. In W. Dressler et al. (eds.), *Phonologica 1988: Proceedings of the 6<sup>th</sup> International Phonology Meeting* 101-110. Cambridge: Cambridge University Press.
- Ito, J. 1986. Syllable Theory in Prosodic Phonology. Ph.D Dissertation. University of Massachusetts, Amherst.
- Ito, J. & R. Mester. 1998. Markedness and Word Structure: OCP Effects in Japanese. Ms., University of California, Santa Cruz.
- Jacobs, H. 1994. Lenition and Optimality Theory. Ms., Nijmegen University & Free University [ROA 127-0496].
- Jacobs, H. & L. Wetzel. 1988. *Early French Lenition: A Formal Account of an Integrated Sound Change*. London: Garland Press.
- Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
- Kenstowicz, M. 1982. Gemination and Spirantization in Tigrinya. *Studies in the Linguistic Sciences* 12.1, 103-122.
- Kirchner, R. 1996. Synchronic Chain Shifts in Optimality Theory, *Linguistic Inquiry* 27.2, 341-350.
- \_\_\_\_\_. 1997. Contrastiveness and Faithfulness. *Phonology* 14.1, 83-111.
- \_\_\_\_\_. 1998. An Effort Based Approach to Consonant Lenition. Ph.D Dissertation. University of California, Los Angeles.
- Lavoie, L. 2001. *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York: Garland Press.
- Leben, W. 1973. *Suprasegmental Phonology*. Bloomington, IN:

- Indiana University Linguistic Club.
- Lombardi, L. 1991. Laryngeal Features and Laryngeal Neutralization. Ph.D Dissertation. University of Massachusetts, Amherst.
- Malone, J. 1993. *Tiberian Hebrew Phonology*. Winona Lake, IN: Eisenbraus.
- Mascaro, J. 1983. Continuant Spreading in Basque, Catalan, and Spanish. In M. Aronoff & R. Oehrle (eds.), *Language Sound Structure* 287-298. Cambridge, MA: MIT Press.
- \_\_\_\_\_. 1987. A Reduction and Spreading Theory of Voicing and Other Sound Effects. Ms., Universitat Autònoma de Barcelona.
- McCarthy, J. & A. Prince. 1986. Prosodic Morphology. Ms., University of Massachusetts, Amherst & Brandeis University.
- Myers, T. 1987. *The Cognitive Representation of Speech*. Amsterdam & New York: North Holland Publishing Company.
- Prince, A. & P. Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University & University of Colorado.
- Rajpurohit, B. 1983. *Shina Phonetic Reader*. Mysore: Central Institute of Indian Languages.
- Rhee, S-C. 1998. Aspects of Release and Non Release in Phonology. Ph.D Dissertation. University of Illinois, Urbana-Champaign.
- Sagey, E. 1980. *The Phrase Phonology of English and French*. New York: Garland.
- \_\_\_\_\_. 1986. *The Representation of Features and Relations in Non Linear Phonology*. London: Garland.
- Selkirk, E. 1982. Syllables. In H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations* 337-383. Dordrecht: Foris Publications.
- Silverman, D. 1997. *Phasing and Recoverability*. New York & London: Garland.
- Steriade, D. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. Ms., University of California, Los Angeles.
- Tesar, B. & P. Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29, 229-268.



Journal of Universal Language 13-1  
March 2012, 101-123

## **Typological Accounts for Nominal Forms**

**Eun-Joo Kwak**

*Sejong University*

### **Abstract**

Various linguistic phenomena may serve a basis for the classification of languages and nominal forms make part of the criteria for the classification. In this study, two major approaches are closely compared for language typology based on nominal forms. Chierchia (1998a, b) assumes that morpho-syntactic features should be crucial in determining the denotations and grammatical categories of NPs. He suggests three classificatory categories with the notion of nominal mapping parameters. Contrastingly, the OT analysis of de Swart & Zwarts (2009, 2010) do not assume that morpho-syntactic features should be collapsed with countability and plurality. They propose several general constraints governing nominal forms and argue that different ordering of the constraints in optimality accounts for different nominal forms crosslinguistically. I have shown that the OT

---

Eun-Joo Kwak

Department of English Language and Literature, Sejong University  
98 Gunja-dong, Gwangjin-gu, Seoul 143-747, Korea  
Phone: +82-2-3408-3633; Email: ejkwak@sejong.ac.kr

Received September 8, 2011; Revised October 30, 2011; Accepted November 17, 2011.

analysis provides an appropriate framework to categorize languages systematically.

Keywords: countability, plurality, optimality, markedness constraint, faithfulness constraint

## 1. Introduction

Cross-linguistic observations show that languages adopt different strategies to mark the grammatical notions of countability and plurality. Specific plural morphemes may be postulated, definite and indefinite articles may give a clue, or no explicit marking is allowed. In spite of the apparent diverse patterns to mark countability and plurality, languages may be classified into a few groups depending on their morphological and syntactic features.

Given the possibility of typological classification, a deeply rooted problem is how much morphology and grammar interact each other or whether explicit morphological features determine grammatical features. Two major approaches are made to deal with this problem. One is based on the view that morpho-syntactic features determine a grammar, which is represented by Chierchia (1998a, b). Languages are classified into three categories depending on argument and predicate properties of noun phrases (henceforth NPs). The other approach made by de Swart & Zwarts (2009, 2010) accepts the view that morpho-syntactic features may not be crucial in determining a grammar, which is exemplified by an optimality-theoretic approach. Although the same set of constraints are postulated for languages, the ordering for their application may not be identical and thus languages show different morpho-syntactic patterns to mark grammatical features.

In this study, I will critically review problems with the typological approach of Chierchia (1998a, b) and show that the

optimality-based account provides a better classificatory tool for language typology. I will also argue that apparent idiosyncratic patterns for countability and plurality are attributed to unusual ranking of constraints and the development of local constraints.

## 2. A Morpho-Syntactic Approach to Language Typology

### 2.1. Chierchia (1998a, b)

One of the general assumptions in semantics has been that the countability of nouns is highly dependent on the physical nature of their references. Count nouns denote discrete objects that are countable in reality while mass nouns denote dense entities such as liquids and materials that are hard to be counted in the real world. Because only discrete objects may be grouped to make plural-numbered entities, plurality is a relevant notion to count nouns but not to mass nouns. According to this assumption, three different concepts are parallel without allowing a discrepancy: the discreteness of the real world and the countability and the plurality of nouns.

One of the challenges to this simple assumption has been made by Chierchia (1998a, b). The starting point of his analysis is division between the countability of a noun and the discreteness of its reference. He argues that entities with equal perceptual salience may belong to different nominal categories. For instance, although the object denoted by *rice* and *lentil* are not distinguished in their perceptual salience, *rice* is mass and *lentil* is count. Moreover, the same objects may be lexicalized into more than one expression which has different countability. This is exemplified by pairs of a count and a mass, e.g., *shoe/footwear*, *coin/change*, and *carpet/carpeting*.

Chierchia proposes that the solely grammatical notion of countability should be categorized by the ‘nominal mapping parameters’ of nouns: [arg] for arguments and [pred] for predicates. Since noun phrases in languages with the parameter value [+arg] denote names of kinds typed  $\langle e \rangle$ , bare NPs without determiners or number inflection may appear in argument positions. NPs in languages with [+pred] denote predicates of type  $\langle e, t \rangle$ , so bare NPs do not occur in argument positions due to the type-mismatch problem. Given the parameters, languages are classified into three different categories as follows:

- (1) a. [+arg, +pred]: English, German
- b. [-arg, +pred]: French, Italian
- c. [+arg, -pred]: Chinese, Thai

In English and German, which are classified to have [+arg, +pred], both bare NPs and NPs with a determiner may occur in argument positions. French and Italian, assigned [-arg, +pred], do not allow bare NPs in argument positions while Chinese and Thai have bare NPs in argument positions due to their features [+arg, -pred].

Chierchia’s language classification hinges on the view that morpho-syntactic properties determine the grammatical notions of countability and plurality. To substantiate this dependency, he enumerates properties of mass nouns. First, mass nouns do not have plural morphology.<sup>1</sup> Despite the similarity of their denotations, the count noun *shoe* has the plural form *shoes* but its mass

---

<sup>1</sup> Chierchia (1998a, b) provides a semantic account for the lack of plural morphology for mass nouns. He assumes that mass nouns be lexically plural, denoting kinds. Because mass nouns are already plural, there is no need to add a plural morpheme to mass nouns. Part of the evidence for the plurality of mass nouns comes from the fact that the denotation of a mass collection like furniture is not distinct from a plural NP like pieces of furniture. A problem with the plurality of mass nouns has been pointed out by Kwak (2009).

counterpart *footwear* is not pluralized. Second, mass nouns may not combine with numerals directly. Although the count noun *drop* occurs with the numeral directly as in *three drops*, the mass noun *blood* may not occur with a numeral directly like *\*three bloods*. Finally, mass nouns need a classifier or a measure phrase when counted. For example, the mass noun *rice* may be counted in forms like *three grains of rice* or *three kilos of rice*. Based on these properties, Chierchia concludes that languages with [+arg, -pred], e.g., Chinese, do not have count nouns because NPs are always accompanied by additional phrases to deliver counting readings and plural forms of NPs do not occur.

## 2.2. Problems with Morpho-Syntactic Based Typology

The seminal work of Chierchia (1998a, b) provides insightful arguments on the relations between morpho-syntactic properties and the grammatical notions of countability and plurality. However, the three general categories defined by Chierchia need a further refinement to reflect the semantic properties of nouns and diverse language variations.

According to Chierchia's classification, Chinese, having [+arg, -pred], includes only mass nouns in its lexicon, which is supported by no plural morphology and the mandatory use of a classifier for counting. However, a number of counterarguments have been made against the lack of count nouns. First, classifiers may be divided into count and mass by their semantic properties<sup>2</sup> (cf. Cheng & Sybesma 1999, Chien et al. 2003). Count classifiers provide information on how entities are partitioned in a natural

---

<sup>2</sup> According to Zhang (2007), numeral classifiers in Chinese have functions more than counting. They deliver categorization parameters such as humanness, animacy, shape, function, consistency, and size. Hence, research on classifiers should be rooted in the cognitive process of human beings (cf. Friedrich 1970, Adams & Conklin 1973, Allen 1977).

way, and mass classifiers are used to denote a unit of measure that quantifies portions of entities. Hence, count classifiers denote inherent or permanent properties of objects while mass classifiers denote temporary states of entities. Given the different semantic functions of count and mass classifiers, a noun may occur with more than one classifier. For instance, *xiangyan* ‘cigarette’ may occur with the count classifier *gen* to denote the number of cigarettes or the mass classifier *bao* ‘pack’ for the temporary grouping of cigarettes. Second, the different semantic features of count and mass classifiers are reflected in their syntax (cf. Cheng & Sybesma 1999). The modification marker *de* may be inserted between a mass classifier and a noun but it is not allowed to occur with a count classifier.<sup>3</sup>

- (2) a. san        bang        (de) rou  
three    Cl\_pound    (De) meat  
'three pounds of meat'
- b. ba        tou        (\*de) niu  
eight    Cl\_head    (\*De) cow  
'eight cows'

*Rou* ‘meat’ is a mass noun to denote a dense entity of material, and *niu* ‘cow’ is a count noun with a discrete denotation of cow. *Bang* for *rou* is a mass classifier which denotes the temporary states of meat while *tou* for *niu* is a count classifier that concerns the permanent property of the individuality of cow. This difference in countability is reflected in the optional occurrence of *de*. It may occur with *bang* but not with *tou*. Based on the semantic and syntactic properties of classifiers, Cheng & Sybesma (1999) argue

<sup>3</sup> Here is a list of abbreviations used in this article.

NP=noun phrase; DP=determiner phrase; NumP=numeral phrase; Pl=plural; Cl=classifier; Poss=possessive; ASP=aspect

that while countability distinction is marked at the level of noun in Indo-European languages, it is reflected at the level of classifier in Chinese.

Another problematic language to Chierchia's classification is Dëne Suliné (henceforth Dëne), a Northern Athapaskan language spoken in Northern Canada. Wilhelm (2008) makes an observation that nouns in Dëne occur in argument positions in their bare forms without a determiner or plural morphology. Then, Dëne should be classified as a language with [+arg, -pred] like Chinese, and all nouns in Dëne are expected to be mass nouns. However, the mandatory use of a classifier or measure phrase is not part of the grammar of Dëne.

- (3) a. solághe ejéretth'úé tili / \*solághe ejéretth'úé  
five milk container/ five milk  
'five pounds of milk/\*five milk'
- b. solághe k'ásba  
five chicken  
'five chickens'

Some nouns like *ejéretth* 'milk,' a counting reading is not available without the occurrence of a measure phrase like *tili* 'container.' However, nouns like *k'ásba* 'chicken' combine with a numeral directly. Moreover, nouns that need a measure phrase for counting denote liquids, substances, or abstract concepts while nouns which do not need a measure phrase make a group which are similar to that of count nouns in English. Hence, Whihelm argues that nouns in Dëne should be divided into count and mass nouns in spite of the lack of plural morphology and the occurrences of bare NPs in argument positions.

Korean and Japanese pose a more problem to Chierchia's classification in that the use of a plural morpheme and a classifier is optional. In Chierchia's analysis, the nominal mapping parameters determine the classificatory category of a language,

which does not leave room for the optionality of plural forms or classifier occurrences. If count nouns in plural readings may or may not occur with a plural morpheme in a given language, the mapping parameters for the language are not set to specific values. Then, the language does not belong to any of the categories, which reveals the flexibility problem of Chierchia's classification.

The three different patterns of the problematic cases to Chierchia's analysis show that the grammatical notions of countability and plurality do not hinge on the morpho-syntactic properties of NPs. An approach based on a different view is required for the typological argument of languages.

### **3. An Optimality-Based Approach to Language Typology**

#### **3.1. Optimality Theory Semantics**

For the analysis of phonological data, 'Optimality Theory' (henceforth OT) has been proposed by Prince & Smolensky (1993). Later, it turns out that the principle of OT is applicable to other areas such as syntax, semantics, and pragmatics. According to the notion of optimality, speakers are supposed to select the optimal form for a given meaning and hearers are supposed to take the optimal interpretation for a given form. Since the optimization on the side of speakers may not coincide with that of hearers, constraints in the OT framework may be in conflict. To resolve this conflict, constraints are assumed soft and violable in the OT analysis. Low-ranked constraints are ready to be violated when they are in conflict with more important or highly ranked ones.

To capture the universality of languages, the same set of constraints which are hierarchically ordered is posited for

languages. However, relative ordering among the constraints are not inherently determined. Hence, languages show diversity in the ranking and allow different nominal forms depending on which constraint is highly ordered. To see how universality and diversity are balanced in languages, let us consider what kind of constraint is necessary.

Depending on languages, bare forms are preferred over NPs marked for grammatical information such as plurality and definiteness. To account for this preference, de Swart & Zwarts (2009, 2010) propose the general markedness constraint \*FunctN.

- (4) \*FunctN: Avoid functional structure in the nominal domain.

Since \*FunctN bars the grammatical specification of NPs, a nominal structure which does not include any functional category is judged to be the most optimal. Suppose that there are four syntactic structures for NPs.

- (5) a. [NP N]  
 b. [NumP Num [NP N]]  
 c. [DP D [NP N]]  
 d. [DP D [NumP Num [NP N]]]

(5a) is the simplest structure without including any functional category, so it does not violate \*FunctN at all. (5b) and (5c) contain one functional category while (5d) occurs with two functional categories of the determiner phrase (DP) and the number phrase (NumP). Thus, (5d) is judged to be the least preferred form according to \*FunctN.

Along with the general markedness constraint, several faithfulness constraints are needed to account for optimal forms. First, nominal forms may be distinct by plurality, which is captured by FPL (cf. de Swart & Zwarts 2009, 2010).

- (6) FPL: Reference to a group of individuals must be reflected in a special plural form of the nominal.

FPL states that NPs should be morphologically distinguished by their plurality and special plural morphology is mandatory. Note that explicit morphology is not required for singulars.<sup>4</sup> Here, plural NPs are regarded as marked forms.

Second, definiteness also serves as a criterion to make different nominal forms (cf. de Swart & Zwarts 2009, 2010).

- (7) FDEF: Reference to discourse unique individuals (unique/maximal or familiar ones) requires the use of an expression of definiteness.

De Swart & Zwarts argue that a definite article plays many roles including uniqueness, maximality, or familiarity.<sup>5</sup> Hence, they propose the term ‘discourse uniqueness,’ on which the faithfulness constraint FDEF is based. Following FDEF, NPs take different forms by definiteness, and explicit marking for definite NPs is compulsory. However, the morphological specification for indefiniteness is not mandatory according to FDEF.

Finally, counting readings may be delivered only by the occurrence of a numeral in a nominal structure. However, it may be preceded or followed by a classifier or measure phrase to make counting more explicit. Hence, another faithfulness constraint is

<sup>4</sup> Typological study including Greenberg (1966) and Corbett (2000) shows that if there is only one marked form for the pair of a singular and a plural, it is always the plural that takes the marked form. FPL is proposed based on this typological observation.

<sup>5</sup> Uniqueness by a definite article is exemplified by a phrase like *the queen of the Netherlands*, which refers to the only queen in the Netherlands. On the other hand, the definiteness of *the stars* is used to denote a maximal group consisting of all the stars. *The dog* delivers familiarity, occurring in a discourse like *I saw a dog in the park, and the dog wagged its tail.*

proposed (cf. Kwak To Appear).

- (8) FCL: The occurrence of a numeral is accompanied by a classifier or measure phrase.

FCL states that the counting reading of a numeral is more specified by the occurrence of a classifier or measure phrase. Depending on languages, FCL applies to all the nouns or it is relevant to a group of nouns, e.g., mass nouns. When the application of FCL is varied by the grammatical or semantic features of nouns, the constraint may be further divided into a few local constraints which are ordered differently in their optimality.

### **3.2. Nominal Forms in English and Chinese**

Crosslinguistic studies show that languages have diverse patterns of nominal forms in their grammar while sharing some common aspects. In the morpho-syntactic approach of Chierchia (1998a, b), the common aspect is instantiated by the nominal mapping parameters and language variations are reflected in the values of the parameters. In the OT analysis, the universal aspect of languages is captured by the same set of constraints while language variations are attributed to the different ranking of the constraints.

To get the better picture of the OT analysis proposed by de Swart & Zwarts (2009, 2010), let us compare different plural forms between English and Chinese.

- (9) a. I saw a bear/bears.  
 b. Wò kànɡjiàn xiónɡ le  
     I     see       bear     ASP  
     'I see a bear/some bears/the bear(s).'

In English, singular and plural NPs are explicitly distinguished

by the occurrence of the plural morpheme ‘-s’ and that of the indefinite article.<sup>6</sup> However, since plurality is not overtly marked in Chinese, the bare NP *xióng* ‘bear’ is interpreted as either singular or plural. As discussed in the previous section, the optimal forms of NPs are determined by the interaction of the markedness constraint \*FunctN and the faithfulness constraint FPL, which are in conflict. Hence, depending on their relative ranking, one of them is decisive in determining optimal nominal forms while the other is violated. The marked form of the plural NP in (9a) shows that FPL is considered more important than \*FunctN in English. On the other hand, the unmarked form in (9b) shows that \*FunctN has a priority over FPL in determining nominal forms.

- (10) a. FPL >> \*FunctN (English)  
       b. \*FunctN >> FPL (Chinese)

FPL outranks \*FunctN in English while \*FunctN takes a higher position than FPL in Chinese.

Given the ranking of the constraints, the procedure of deriving an optimal plural form in English is represented as follows:

Table 1. Plural Marking in English

Meaning $\exists x_{\text{pl}}[\text{bear}(x) \wedge \text{see}(x)(I)]$	Form	FPL	*FunctN
	I saw bear	*	
	I saw bears		*

<sup>6</sup> De Swart & Zwarts (2009, 2010) note that some category of nouns does not show distinction as to plurality. Nouns for animals that are fished on or hunted, e.g., *carp* and *salmon*, take bare forms even in plural readings. Hence they posit a local markedness constraint \*PL<sub>FISH</sub> to block the occurrence of the plural morpheme for this category of nouns and locate it higher than FPL in the ranking. In the OT analysis, local constraints are useful to deal with idiosyncratic behaviors shown in languages.

The first column of the table shows the intended meaning, which amounts to say ‘I saw bears.’ The second column includes two possible forms to deliver this meaning, one with the bare NP *bear* and the other with the plural NP *bears*. Each of the forms violates one of the constraints. The bare NP is not marked for plurality and thus violates FPL, which is marked by the asterisk. The plural NP does not follow \*FunctN and gets an asterisk in the fourth column. Although the number of constraints the two NPs violate is equal, the bare NP violates the highly ranked constraint. Hence, the plural form is judged optimal in English, which is shown by the pointing hand (☞).

To deliver the same meaning, two nominal forms may be considered in Chinese as given in the second column of Table 2.

Table 2. Plural Marking in Chinese

Meaning $\exists x_{\text{pl}}[\text{bear}(x) \wedge \text{see}(x)(I)]$	Form	*FunctN	FPL
☞	Wò kàngjiàn xióng le I see bear ASP		*
	Wò kàngjiàn xióng_pl le I see bear_pl ASP	*	

The bare NP in the second row violates \*FPL while the plural form in the third row does not meet \*FunctN. Although the constraints they violate are identical to those of English, their ordering is reversed in Table 2. The violation of FPL is more tolerable in Chinese. Thus, the bare NP is regarded as optimal.

Nominal forms in English are distinguished not only by plurality but also by definiteness. Definite NPs are always preceded by the definite article while indefinite NPs are preceded by the indefinite article or a zero determiner. Unlike English, Chinese NPs are not marked for definiteness. For example, the bare NP *xióng* in (9b) is construed as either an indefinite NP ‘a bear/bears’ or a definite

one ‘the bear(s).’ The distinct nominal patterns between English and Chinese are again reflected in the different ordering of constraints. FDEF, which opts for NPs marked for definiteness, is ranked higher than \*FunctN in English whereas it takes a lower position than \*FunctN in Chinese.

- (11) a. {FPL, FDEF} >> \*FunctN (English)
- b. \*FunctN >> {FPL, FDEF} (Chinese)

The relative ordering between FPL and FDEF is not meaningful in both of the languages, which is represented by the bracket. Definite NPs in bare forms violate the higher constraint FDEF in English, and thus they are not regarded as optimal. However, bare NPs in definite readings satisfy the higher constraint \*FunctN in Chinese, so they are optimal forms.

Finally, the occurrence of a classifier is restricted to mass nouns in English. Hence, I argue that FCL is divided into two local constraints:  $FCL_{MASS}$  and  $FCL_{COUNT}$ . The mandatory use of a classifier for mass nouns means that  $FCL_{MASS}$  outranks its conflicting constraint \*FunctN. The lack of a classifier for count nouns means that  $FCL_{COUNT}$  is lower ranked than \*FunctN. Here is the ordering of the constraints for English.

- (12) a. {FPL, FDEF,  $FCL_{MASS}$ } >> \*FunctN >>  $FCL_{COUNT}$   
(English)
- b. FCL >> \*FunctN >> {FPL, FDEF} (Chinese)

Unlike English, all the nouns in Chinese occur with a classifier, so the general constraint FCL is operative in Chinese. Furthermore, FCL should outrank \*FunctN to make the use of a classifier obligatory.

### 3.3. Nominal Forms in Dëne and Korean

As pointed out in section 2.2., one of the problematic languages to Chierchia's analysis is Dëne, in which no number marking is allowed but the use of a classifier is different depending on the countability of nouns. Here are the data repeated from section 2.2.

- (13) a. solághe k'ásba  
five chicken  
'five chickens'
- b. solághe ejéretth'úé tili/ \*solághe ejéretth'úé  
five milk container/ five milk  
'five pounds of milk/\*five milk'

The denotation of *k'ásba* is a set of chickens, which are discrete in the real world. Although *k'ásba* is not followed by any plural morpheme, it combines with the numeral without the occurrence of a classifier. On the other hand, *ejéretth'úé* refers to milk, which is dense and hard to be counted by itself. When counted, it requires the occurrence of a classifier or measure phrase like *tili*. Direct counting only with a numeral is not allowed to this noun.

In the OT analysis, countability does not hinge on plural morphology. Suppose that nouns in Dëne are divided into count and mass nouns, following Wilhelm (2008). Then, I argue that the lack of plural morphology is attributed to the lower ranking of FPL than \*FunctN. Count nouns may be plural but their plurality is not morphologically marked due to the lower ranking of FPL. Moreover, since classifiers are mandatory only for mass nouns in counting readings, FCL is divided into two local constraints: FCL<sub>MASS</sub> and FCL<sub>COUNT</sub>. To render a classifier to follow a mass noun, FCL<sub>MASS</sub> is ranked higher than \*FunctN, which again outranks FCL<sub>COUNT</sub> to suppress the occurrence of a classifier for count nouns. As a result, the relevant ordering for the constraints should be like the following:

- (14) FCL<sub>MASS</sub> >> \*FunctN >> {FPL, FCL<sub>COUNT</sub>} (Dëne)

The last category of languages that poses a problem to Chierchia's framework includes Korean, in which a plural morpheme and classifiers are optionally used. This optionality makes it very awkward to determine the values of the nominal mapping parameters and make a typological classification of the language. Note that the explicit occurrence of the plural morpheme *-tul* makes it clear that the noun is plural. However, a bare noun without being followed by *-tul* is ambiguous between a singular and a plural.

- (15) haksayng/ haksayng-tul  
 student/ student-Pl  
 'a student or (the) students/(the) students'

The ambiguity of *haksayng* contrasts with the plural reading of *haksayngtul* in (15). To accommodate the optional use of the plural morpheme, Kwak (2011) proposes that FPL need to be coranked with \*FunctN.

- (16) {\*FunctN, FPL} (Korean)

When \*FunctN and FPL are not distinguished in their ranking, the violation of the constraints is regarded as having equal optimality. The plural NP *haksayngtul* violates \*FunctN while the bare NP *haksayng* does not satisfy FPL. Both of the NPs violate one constraint and the level of the hierarchical ranking of the constraints is not distinct. Hence, the two NPs are judged to have the same degree of optimality and used alternately.

As for definiteness, Korean has the definite article *ku* in the grammar to mark definite NPs explicitly. However, definite NPs in Korean are not always preceded by this article. The bare NPs in (15) are ambiguous between definite or indefinite readings.

Then, FDEF should also be coranked with its conflicting constraint \*FunctN to make the occurrence of *ku* optional. Then, the ordering of the constraints is elaborated as follows:

- (17) {\*FunctN, FPL, FDEF} (Korean)

Interestingly, the optionality of the plural morpheme does not hold in some cases. When an NP is definite, its plural reading is conveyed only through the occurrence of *-tul*. Compare the ambiguity of *haksayng* in (15) and the rigid interpretations of the NPs in (18) (cf. Song 1975, Nemoto 2005, Kwak 2010).

- (18) ku haksayng/ ku haksayng-tul  
          the student/     the student-Pl  
          'the student/the students'

*Ku haksayng* does not have a plural reading. The rigidity shown in (18) suggests that FPL in Korean needs to be subcategorized by definiteness, i.e., FPL<sub>DEF</sub> and FPL<sub>INDEF</sub>. Moreover, FPL<sub>DEF</sub> should be ranked over \*FunctN to make the occurrence of *-tul* mandatory for definite NPs.

- (19) FPL<sub>DEF</sub> >> {\*FunctN, FDEF, FPL<sub>INDEF</sub>} (Korean)

The ambiguity of indefinite NPs as to plurality is captured by the coranking of FPL<sub>INDEF</sub> and \*FunctN.

Finally, the use of a classifier is available for all nouns regardless of their grammatical or semantic features in Korean. However, it is obligatory for mass nouns.

- (20) a. sey myeng-uy haksayng(-tul)/ sey haksayng(-tul)  
          three Cl-Poss student(-Pl)/     three student(-Pl)  
          'three students'

- b. sey byeng-uy mwul/ \*sey mwul  
 three Cl-Poss water/ three water  
 'three bottles of water/\*three water'

The count noun of *haksayng* may or may not occur with the classifier, but the lack of the classifier makes the mass noun *mwul* awkward in a counting reading. Thus, FCL is also divided into two local constraints by countability, FCL<sub>COUNT</sub> namely and FCL<sub>MASS</sub>. Again, FCL<sub>MASS</sub> is ordered higher than \*FunctN. Here is the final ranking of the constraints for Korean.

- (21) {FPL<sub>DEF</sub>, FCL<sub>MASS</sub>} >> {\*FunctN, FDEF, FPL<sub>INDEF</sub>,  
 FCL<sub>COUNT</sub>} (Korean)

Dual nominal forms allowed in Korean, which are hard to be accounted for in Chierchia's analysis, are attributed to the coranking of the markedness and the faithfulness constraints. The coranking implicates that the violation of either of the constraints has the equal degree of optimality and does not bar the occurrence of violated nominal forms in Korean. Furthermore, distinct plural forms by definiteness are systematically explained by the postulation of the local constraints and their distinct ordering in optimality.

### **3.4. Comparisons between the Two Approaches**

The typological classification of languages is conducive to elucidating patterns of linguistic variations and the universality of languages. Among the criteria for the classification are nominal forms which may vary by number morphology and the occurrences of articles and classifiers. One analysis for language classification by nominal forms is provided by Chierchia (1998a, b). The underlying assumption in this analysis is that morpho-syntactic features are reflected in the grammar of a language. To

put it in another way, if morpho-syntactic patterns of nominal forms are different, the corresponding grammars are also distinct. A plural morpheme or a syntactic device to mark plurality is regarded as an explicit sign for the countability of nouns. Hence, Chierchia argues that languages may consist of only mass nouns when they are lack of morpho-syntactic features for count nouns. With the notion of the nominal mapping parameters, Chierchia makes three categories for language typology.

Chierchia's analysis is a monumental work in that typological classification is based on the semantic properties of expressions and extended by their morpho-syntactic features. However, his emphasis on morpho-syntactic features makes his framework vulnerable to cope with linguistic diversity and accommodate flexibility needed by idiosyncratic behaviors shown by some languages. First, Chierchia suggests common properties shared by mass nouns, according to which no plural morphology and the mandatory use of a classifier cannot be detached. However, these two properties need to be treated separately in some languages. This shows that plurality is a separate issue from countability, and furthermore morpho-syntactic features may not determine countability and plurality. Second, Chierchia's classification is too general to encompass all the languages. Diverse patterns of nominal forms shown by crosslinguistic data require more specified categories to handle them.

In contrast with Chierchia's analysis, the OT analysis of de Swart & Zwarts (2009, 2010) does not assume that grammatical notions of countability and plurality should coincide with typical morpho-syntactic features for them. Languages are assumed to share the same set of constraints, which reflects the generality of languages. However, the constraints may be set to different positions in the hierarchical ordering of optimality, and morpho-syntactic features triggered by lower-ranked constraints are not specified explicitly. Different specification of the features accounts for the diversity of languages. For instance, de Swart &

Zwarts (2009, 2010) provide three categories depending on the relative ordering of \*FunctN, FPL, and FDEF.

- (22) a. \*FunctN >> {FPL, FDEF}: no number morphology, no articles (Chinese)
- b. {FPL, FDEF} >> \*FunctN: number morphology, articles (English, German, French, Hebrew, Bulgarian)
- c. FPL >> \*FunctN >> FDEF: number morphology, no articles (Hindi, Polish)

When the markedness constraint \*FunctN outranks the other constraints, a language does not include any morpho-syntactic sign for countability and plurality and allows only bare NPs. This is exemplified by Chinese. On the other hand, the higher ranking of the faithfulness constraints FPL and FDEF includes number morphology and articles for countability or definiteness. Thus, bare NPs without any sign for countability, plurality, and definiteness are not allowed in this category of languages. As shown in (22c), the two faithfulness constraints may be separately ranked.

Along with the different ranking of the constraints, the OT analysis is equipped with another device to deal with linguistic diversity, namely the postulation of local constraints. A general constraint like FPL or FCL may be subcategorized by grammatical or semantic features. As discussed in section 3.2. and 3.3., FCL may be divided into  $FCL_{COUNT}$  and  $FCL_{MASS}$ , and FPL may also be divided by definiteness to deal with Korean nominal forms. The postulation of local constraints has the effect that the apparent idiosyncratic behaviors of some languages are part of the general patterns shown by other languages. This makes it possible to capture the universality and diversity of languages in a systematic way. Finally, various typological comparisons are possible depending on which constraints are selected. Not all of the constraints have to be addressed to categorize languages, but partial comparisons are also available depending on purposes.

Therefore, different levels of comparisons are accessible in the OT analysis.

#### 4. Conclusions

The typological classification of languages contributes to elucidate the universality and diversity of languages. Various linguistic phenomena serve a basis for categorizing languages and different purposes are achieved depending on which phenomena are selected for comparison. Nominal forms constitute part of the criteria to classify languages.

Two different positions may be considered for language classification by nominal forms. One is suggested by Chierchia (1998a, b), in which morpho-syntactic features are crucial to determine the denotations of NPs and the category of languages. The other is the OT analysis by de Swart & Zwarts (2009, 2010), where the grammatical notions of countability and plurality are treated separately from the morpho-syntactic features of nominal forms. Each of the approaches has its own theoretical merits and is conducive to clarifying linguistic systems underlying crosslinguistic data.

In this study, I have made close comparisons between the two approaches and argued that the underlying assumption of Chierchia's analysis is not convincing given the counterarguments raised by other studies. Furthermore, his categorization is too general to deal with the required diversity and too rigid to encompass apparent idiosyncratic behaviors of some languages. In the OT analysis, the universality of languages is well captured by postulating the same set of constraints. It still leaves room for flexibility, assuming different hierarchical ordering of the constraints and the postulation of local constraints. I have shown that the problematic phenomena in Chierchia's analysis are well explained in the OT analysis.

## References

- Adams, K. & N. Conklin. 1973. Toward a Theory of Natural Classification. *Proceedings of the 9<sup>th</sup> Regional Meeting, Chicago Linguistic Society* 1-10. Chicago, IL: Chicago Linguistic Society.
- Allen, K. 1977. Classifier. *Language* 53, 285-311.
- Cheng, L. & R. Sybesma. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30, 509-542.
- Chien, Y. et al. 2003. Chinese Children Comprehension of Count Classifiers and Mass Classifiers. *Journal of East Asian Linguistics* 12, 91-120.
- Chierchia, G. 1998a. Plurality of Mass Nouns and the Notion of Semantic Parameter. In S. Rothstein (ed.), *Events and Grammar* 53-103. Dordrecht: Kluwer.
- \_\_\_\_\_. 1998b. Reference to Kinds across Languages. *Natural Language and Semantics* 6, 339-405.
- Corbett, G. 2000. *Number*. Cambridge: Cambridge University Press.
- Friedrich, P. 1970. Shape in Grammar. *Language* 46, 370-407.
- Greenberg, J. 1966. *Language Universals with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Kwak, E. 2009. A Parallel between Countability and Atomicity. *The Journal of Language and Translation* (Now Called *The Journal of Universal Language*) 10, 113-137.
- \_\_\_\_\_. 2010. The Distributivity of Collection Terms in Korean. *The Journal of Language and Translation* (Now Called *The Journal of Universal Language*) 11, 179-205.
- \_\_\_\_\_. To Appear. Optimal Forms of Noun Phrases in Korean. *The Linguistic Association of Korea Journal* 20.1.
- Nemoto, N. 2005. On Mass Denotations of Bare Nouns in Japanese and Korean. *Linguistics* 43, 383-413.
- Prince, A. & P. Smolensky. 1993. Optimality from Neural Networks to Universal Grammar. *Science* 275, 1604-1610.
- Song, S. 1975. Rare Plural Marking and Ubiquitous Plural Marker

- in Korean. *Language Research* 11, 77-86.
- de Swart, H. & J. Zwarts. 2009. Less Form More Meaning: Why Bare Nominals Are Special. *Lingua* 119.2, 280-295.
- \_\_\_\_\_. 2010. Optimization Principles in the Typology of Number and Articles. In B. Heine & H. Narrog (eds.), *Handbook of Linguistic Analysis* 555-582. Oxford: Oxford University Press.
- Wilhelm, A. 2008. Bare Nouns and Number in Dëne Suliné. *Natural Language Semantics* 16, 39-68.
- Zhang, H. 2007. Numeral Classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16, 43-59.



Journal of Universal Language 13-1  
March 2012, 125-167

# The Representation of Korean and Other Altaic Languages in Artificial International Auxiliary Languages

**Alan Reed Libert**

*University of Newcastle, Australia*

## Abstract

Korean and other Altaic languages are generally not well represented in artificial international auxiliary languages: the best known such languages (such as Esperanto and Ido) have borrowed almost nothing from them, instead almost exclusively using Indo-European languages as sources. In this paper I will present some auxiliary languages which have taken words and/or parts of their grammar from Altaic languages, looking at which items have been borrowed and in some cases what percentage of the vocabulary they account for. The languages discussed (most of which were created relatively recently) include Ardano, Dousha, Dunia, Konya, Kosmo, Kumiko, Lingwa de Planeta, Neo Patwa, NOXILO, Olingo, Pan-kel, Sambahsa-mundialect, Sona, and Unish. In the cases of most of these languages only a small

---

Alan Reed Libert  
Department of Linguistics, University of Newcastle  
Callaghan, NSW 2308, Australia  
Phone: 61-2-49215117; Email: Alan.Libert@newcastle.edu.au

Received January 30, 2012; Revised February 26, 2012; Accepted March 6, 2012.

proportion of the total vocabulary comes from Altaic languages. Further, some of the words said to have been taken from an Altaic language originally came from an Indo-European language.

In addition, I will compare the proportion of Korean items to those taken from the other languages of the Altaic family. Overall Korean has been drawn upon less than Japanese, but (not surprisingly) more than Mongolian, Azerbaijani, and Uzbek. Most conclusions are tentative because the vocabularies of most of the auxiliary languages examined have not been fully developed and because often information about sources of words is not given.

Keywords: auxiliary language, Altaic, Korean, Japanese, Turkish, vocabulary

## 1. Introduction

Most artificial international auxiliary languages (henceforth AIALs),<sup>1</sup> and the most successful among them, have been based on one or more natural languages. However, the set of languages which have been drawn on has been rather limited, usually being from the Indo-European family, and only from a few branches of it. However, in recent years, some attempts at AIALs have made use of a broader range of languages. In this paper I shall examine the extent to which members of one non-Indo-European family, the Altaic languages, are represented in AIALs. I will also compare the representation of Korean to that of other Altaic languages.<sup>2</sup>

Artificial languages have been created for many purposes,<sup>3</sup> but,

---

<sup>1</sup> I use AIAL rather than IAL (International Auxiliary Language) because there can be, and indeed are, international auxiliary languages which are not artificial, e.g., English.

<sup>2</sup> Not all scholars consider Korean and Japanese to be part of the Altaic language family; my treatment of them together with Turkish and so on should not be interpreted as any kind of statement about classification.

as in my previous research, I am restricting my survey to those created for a serious purpose (as opposed to those created for fun or personal enjoyment), and in particular to those created to improve international communications, and those which may not have had this as their primary function, but whose designer saw it as a possible function. Such a purpose in part determines the form of artificial languages designed with it in mind, which in my view makes them more interesting for theoretical study (while a language created e.g., in connection with a work of science fiction could be much more unlike natural languages). However, although the purpose of the languages discussed here is serious, some of the actual languages may not be completely serious, at least in the sense that the designer does not really expect his language to have a chance of seeing widespread use as an AIAL. For example, Dana Nutter has designed not one but several AIALs, some of which are meant for use in particular regions of the world; it is extremely unlikely that any of them will see much use, and we might regard his AIALs as experimental creations rather than as practical solutions, or as examples of language play (but arguably of a more serious kind than e.g., science fiction languages). Nevertheless, designers of such AIALs have considered issues such as ease for potential learners and (perhaps most importantly for the purposes of the present paper) neutrality of an AIAL with respect to natural languages, and in my opinion they are worth examining.

One should bear in mind that I am reporting on the sources of words and other features as given by my sources on languages (usually the designer of the language); this does not necessarily correspond to the actual source, or at least not to the original source, as we shall see. I have not discussed every AIAL which has drawn upon the Altaic family, but I believe that I have covered the majority of them, which may give a very general

---

<sup>3</sup> See for example Gobbo (2008).

picture of the extent way in which this family is represented among AIALs.

## 2. Investigation of AIALs

In this section I shall examine individual AIALs to determine how many words and grammatical features they have taken from Altaic languages.

### 2.1. Major AIALs

As already mentioned, the most popular AIALs, e.g., Esperanto, have taken very little from languages outside of the Indo-European group. As far as I know, the only words from Altaic languages which appear in such AIALs are those for items that come from the cultures in which those languages are spoken, with the exception of Esperanto *cunamo* ‘tsunami.’ For example, Esperanto has the words *kimêio* ‘kimchi’ and *kimono* ‘kimono,’ originally from Korean and Japanese respectively. However, one might suspect that these words were not taken directly from Korean or Japanese but were borrowed from English or another European language. I only know of one Esperanto word borrowed from Japanese which could not have come via English, *hašio* ‘chopstick’ (which has not been accepted as an official word of the language, as shown by its absence from the *Akademio Vortaro* of the Akademio de Esperanto (n.d.)). Ido and Interlingua<sup>4</sup> both also have the word *kimono* ‘kimono,’ and again it may well not have come directly from Japanese (I do not know whether it has official status in the former language).

---

<sup>4</sup> By *Interlingua* I mean the language created by the International Auxiliary Language Association, not the language of the same name designed by Giuseppe Peano.

## 2.2. Ardano

Elhassi (2008: 2) states that Ardano “contains words from every natural language in the world.” This work does not contain a very large number of Ardano words, and certainly not enough so that every language is represented; further, Elhassi (*ibid.*: 10), in a section on etymology, says, “Only one example of a language the word is derived from” [sic], which could mean that if a word was taken from several languages, only one of them is mentioned (and I do not know of any other work on the language which contains more etymological information). In any case, Elhassi (*ibid.*) does not give etymological information on all the Ardano words which appear in this work, and when I give comparative figures below, it should be borne in mind that I am only comparing the number of words stated to come from particular languages, not the total number of words from those languages (which could be larger). There are some words said to be from Altaic languages: *txah* ‘tea’ from Korean, *sananal* ‘white’ from Manchu, *marul* ‘lettuce’ from Turkish, *gem* ‘boat’ from Turkish, and *tutmaco* ‘to hold’ from Azerbaijani. It will be noticed that there are no words said to be of Japanese origin, which is rather surprising; however, the prefix *o-* creates “respectful” questions; Elhassi (2008: 53) says that “The idea is taken from Japanese.” One of his examples is below:

- (1) a. Hal ti posna aiutiee min? = Can you help me?  
 b. Ohal ti posna aiutije min? = Could you help me?

Presumably if a full vocabulary for Ardano were created, there would be more words of Altaic origin. (The same holds true of other AIALs with only partly developed vocabularies to be discussed below.) If we look at the number of Ardano words said to come from some other languages, Elhassi (*ibid.*) has nine words from English (plus three from Tok Pisin and one from “Pitcairn-Norfolk” (p. 10)), one from French (and one word from

“Morisyen” (p. 36), i.e., Mauritian Creole), two from German (and two from Bavarian), one from Spanish, nine from Chinese (presumably meaning Mandarin, and not including words from Hakka and Wu Chinese), including the word for ‘Chinese,’<sup>5</sup> 17 from Arabic, and ten from Italian (plus one from Calabrese). The relatively large figures for Arabic and Italian may be due to the fact that Ardano’s designer is Libyan.

### 2.3. Dousha

Dousha dates from 2002 and was created by Almir U. Junior from Brazil. (To my knowledge no publicly available source on Dousha gives his full surname.) Almir U. (2008) says, “It is an agglutinative language, developed from the Japanese grammar and with an assorted group of vocabulary sources. It is proposed for several uses, including to be used as an auxiliary language.” According to Almir U. (n.d. a) the following Dousha cardinal numerals come from Japanese (the original Japanese word is given in parentheses): *sa* ‘three’ (*san*), *go* ‘five’ (*go*), *so* ‘six’ (*soku*), *ste* ‘seven’ (*shichi*),<sup>6</sup> *ku* ‘nine’ (*ku*), and *ju* ‘ten’ (*ju*). Some other cardinal numerals come from English or Portuguese, and no etymology is given for the words for ‘zero’ and ‘one.’ Almir U. (n.d. b) indicates that the source of the following prefixes is Japanese: *ai-/ay-* ‘posição intermediária, entre’ (‘intermediate position, between’), *ga^-/ga(t)-*<sup>7</sup> ‘com’ (comitativo) (‘with’

<sup>5</sup> The word for ‘China’ also appears to come from Chinese, though its etymology is not given. The word for ‘Japan,’ *Nipon*, appears to come from Japanese, though again this is not stated. The terms for ‘South Korea’ and ‘Turkey’ are *Yugo Corea* and *Turcyia* respectively; no etymological information on them is provided, but the second word of the former presumably comes (although perhaps indirectly) from Korean, while the latter might come from Turkish.

<sup>6</sup> It seems odd to say that *ste* was derived from *shichi*, but this is what is stated in Almir U. (n.d. a).

<sup>7</sup> To my knowledge the meaning of the symbol or diacritic <^> is not explained

(comitative)), *hiki-* ‘extracto de’ (‘extract from’), *mot-* ‘por meio de’ (instrumental) (‘by means of’ (instrumental)), and *o^-* ‘posição superior, sobre, em cima de’ (‘superior position, on, on top of’).

## 2.4. Dunia

Robertson (n.d. a) states, “Dunia draws its grammar and lexical material from the most widely spoken languages throughout the world, for example: Chinese, Hindi/Urdu, Spanish/Portuguese, English, Malay/Indonesian, Arabic, Russian, Japanese, and German. However, it does this in such a way as to permit the learner to know in advance where to expect words to come from.”

Dunia is basically an SVO language. This is not the only permitted order, but another type of allowed order requires a word borrowed from Japanese; Robertson (n.d. a) says, “If the object does not follow the verb, it is marked by the preposition *o*, e.g., ‘it is curry that they are eating,’ *o tarkari tamen komer . . .* Note that although *o* is a postposition in the source language, Japanese, Dunia always uses prepositions.” Robertson (ibid.) then comments on his choice of this word: “Although there is a word in Mandarin which would have fulfilled the same purpose (ba-3rd tone), it cannot be used because it conflicts with the word *ba*, also from Mandarin (neutral tone), which we have already allocated as the suggestive mood particle, and the Japanese equivalent is used instead.”

Dunia has other Japanese function words. Robertson (ibid.) says:

In addition to the subject-predicate sentence structure most used in Western languages, Dunia also provides for a topic-comment structure with the topic end being marked by *wa* (from Japanese—there is no word used for this in Mandarin): ‘as for my sister, she is coming on

---

in any publicly available material on Dousha.

Monday,' *woski soror wa, ta yao venir getsuyobi.*"

*San* does not seem to have exactly the same function in Dunia as in Japanese: Robertson (n.d. a) states that it "may be useful for marking personal names when placed after them."

Concerning word classes Robertson (*ibid.*) states:

There are four fundamental parts of speech in Dunia: nouns, verbs, adjectives, and 'everything else.' The 'everything else' category includes grammatical/structural particles, modal particles, time expressions, prepositions, case and sentence structure markers, conjunctions, interjections, personal and correlative pronouns, articles, numerals, and primitive adverbs. These are drawn mostly from the Chinese, Japanese, Malay/Indonesian, Arabic, and Russian languages.

Robertson later says, "The correlatives in Dunia are mostly taken from Japanese, but with some regularities imposed." (Malay/Indonesian and Russian are the source languages for conjunctions and underived adverbs respectively and the personal pronouns seem to have come from Chinese; Spanish and Hindi/Urdu are the main sources for verbs and adjectives respectively.)

Robertson (*ibid.*) makes the following remark about vocabulary: "Where a word is particularly associated with a particular place or language, there is room in Dunia for a LIMITED number of these: these might include *kuaizi* (chopsticks) from Chinese, *zeit* (oil) from Arabic, *jidosha* (automobile) from Japanese, *konyak* from French, and *kanikchak* (snow) from Inuit." One might say that at least one of his examples, the Japanese one, is somewhat odd: while chopsticks are strongly "associated with" Chinese, or at least East Asian, culture, the connection between automobiles is arguably not as strong—while a large number of the world's cars

have been made by Japanese companies, the early history of the development of automobiles is probably connected with Europe and/or the United States in many people's minds.

As far as I know, there is no publicly available Dunia dictionary. However, Robertson (n.d. b) gives a list of words (with English glosses) occurring in the Dunia version of the “Babel Story.” The words in this list said to come from Japanese are (in addition to *o* and *wa*): *dono* ‘who, which,’ *dore* ‘what,’ *dotoki* ‘when,’ *kore* ‘this,’ *kotoki* ‘now,’ *mina* ‘wholly, completely,’ *minadore* ‘everything,’ *onajidono* ‘same,’ *soko* ‘then,’ and *sore* ‘that.’

Thus while Dunia has borrowed a fair amount of material from Japanese, to my knowledge it has not made use of any items from other Altaic languages.

## 2.5. Esperadunye and Pasifika

As already noted, Dana Nutter has created several artificial languages, one of which is Esperadunye. Nutter (2012a) says that this language is “[b]asically Esperanto relexified with a mixed vocabulary from major world languages rather than just European ones, and some grammatical reforms which make the language more isolating to accom[m]odate a wider range of speaking habits.” The *Esperadunye Dictionary* (Nutter 2012b) lists a relatively small number of words and affixes in the language (93<sup>8</sup>); a (partly) Altaic etymology is attributed to two of them: *dunye* ‘world’ (from Turkish *dünya*, Hindi *duniyā*, Indonesian *dunia*, Arabic *dunyā*) and *qine* ‘person’ (from Japanese *jin*).<sup>9</sup>

Pasifika is another language designed by Dana Nutter, who says:

---

<sup>8</sup> There are two cases of three words derived from the same root being listed as separate entries, e.g., *bove* ‘ox, bovine,’ *bovine* ‘cow,’ *bovule* ‘bull,’ and *ve* ‘you,’ and *vei* ‘your’ are also listed separately.

<sup>9</sup> The letter <q> is pronounced [dʒ].

Pasifika is designed to be an auxiliary language for regional use around the Pacific Rim by drawing from the major languages of East Asia, Western Americas, Australia, Indonesia, and other parts of Oceania: Malay-Indonesian, Japanese, Korean, Vietnamese, Javanese, Mandarin, Spanish, English, Wu, and Cantonese, but also adding small influences from less popular languages such as Tagalog, Ilocano, Thai, Quechua, Aymara, Maori, and Hawaiian. The goal of Pasifika is to construct a language which is easy for speakers of all the source languages to pronounce, with a much simpler and more regular grammar, and a small but still powerful core vocabulary using words found in the diversity of the source languages (Nutter 2012c).

The *Pasifika-English Dictionary* (Nutter 2012d) gives etymological information, but only contains 16 Pasifika words, of which two are said to come from an Altaic language, *jin* ‘person’ (from Japanese *jin*) and *xima* ‘island’ (from Japanese *shimá*). Thus 12.5% of the Pasifika vocabulary on this webpage has an Altaic origin. Recall that the Esperadunye word for ‘person’ is also derived from the Japanese word *jin*. The Pasifika word for ‘world,’ *dunia*, is similar to that of Esperadunye, but only Indonesian is given as a source for it (namely the word *dunia*).

## 2.6. Euransi

Euransi was designed by Libor Sztemon of the Czech Republic. He was working on the language from at least 1998 to 2002 (he died in 2002). Sztemon (n.d.) contains a table giving the sources of 42 Euransi words, along with their “original meaning” (i.e., the meaning in the source language), including the following from Altaic languages (I only give the original meaning when it is different from the meaning in Euransi): *bahce* (from Turkish

*bahçe* ‘garden’), *kaze* ‘wind’ (N) (from Japanese *kaze* ‘wind, cold’), *bayrami* ‘holiday’ (from Turkmen *bayram* ‘festival’), *yorgân* ‘chasm’ (from Azeri *yorgvan*). Other words in this table are stated to come from English, Dutch, Latin, Hebrew, Tajik, Chechen, Indonesian, Swahili, and so on. (In fact each word comes from a different language, except that one word is said to come from Arabic and another from Egyptian Arabic; also one word comes from Hindi and one from Urdu.) Thus it can be seen that Sztemon drew upon quite a range of languages, but there is no evidence that Korean was used.

## 2.7. Jigwa

Jigwa was an (apparently abandoned) attempt to create an AIAL which was not “Euro-centric” (Jigwadx Jungdwei n.d.). The currently available sources on Jigwa do not mention or list very many of its words, but Jigwadx Jungdwei (*ibid.*) gives several of the cardinal numerals, including *sam* ‘three,’ which is “from Korean and Cantonese, cognate with Mandarin *san1* and Japanese *san*.” Jigwadx Jungdwei (*ibid.*) says, “The demonstratives, all borrowed from Asian languages, are *na* (what? which?), *jx* (this), *so* (that), *mai* (each, every), and *mu* (no, not any).” Of these *na*, *so*, and *mai* appear to come from Japanese, and *mu* from Korean and/or Japanese.

The document Jigwa (Jigwadx Jungtim 1995), which is no longer available online,<sup>10</sup> has “etymological information” on the origins of most words, although, given the fact that in most cases several source words are mentioned, it is not clear what this information means—is the Jigwa word a synthesis of the words cited, or have some of the words cited simply reinforced the idea

---

<sup>10</sup> Even the Jigwa materials mentioned in the previous paragraph, which I called “currently available,” are not obtainable from the webpage where they were first posted.

of borrowing a particular word (because they sound like it)? The Jigwa words in relation to which words from one or more Altaic languages are listed below, along with the “etymological information”:

- (2) a. du ‘two’—English: two, dual; Italian: due; Korean: tul; Nepali: dui
- b. dung ‘east(ern)’—Cantonese: dung; Mandarin: dong<sup>1</sup>; Korean: tong(-jjok); Vietnamese: Đồng
- c. chx<sup>11</sup> ‘seven’—Mandarin: qil; Cantonese: chat; Japanese: shichi
- d. dou ‘bean, pea’—Mandarin: dou4; Vietnamese: dāu; Japanese: tō
- e. gwa ‘language’—Italian: linguaggio; Japanese: go; Cantonese: wa
- f. jung ‘middle, center’—Korean: chung-ang; Cantonese: jung; Mandarin: zhong<sup>1</sup>
- g. nam ‘south(ern)’—Korean & Vietnamese: nam; Mandarin: nan<sup>2</sup>; Cantonese: naam
- h. nen ‘year’—Mandarin: nian<sup>2</sup>; Cantonese: nin; Japanese: nen; Korean: nyon
- i. nung ‘agriculture’—Cantonese: nung; Vietnamese: nong; Korean: nong
- j. pal ‘eight’—Korean: pal; Mandarin: bal; Cantonese: baat
- k. sam ‘three’—Korean & Thai: sam; Cantonese: saam; Japanese: san, sam-
- l. wan ‘10,000’—Mandarin: wan4; Cantonese: maan; Japanese: man

Since such information is only given on less than one third of the Jigwa words in the “wordlist” in Jigwadx Jungtim (1995), it is impossible to know the extent to which Korean and Japanese

---

<sup>11</sup> The letter <x> stands for schwa.

have been drawn upon in creating the vocabulary (which in any case is presumably not to be the full lexicon of the language). However, based on the words given above, it appears that both Korean and Japanese are fairly well represented in the Jigwa lexicon (12 out of 24 words for which “etymological information” is given involve Korean and/or Japanese), or at least they have had an influence (in case the Korean and Japanese words only reinforced the choice of a word). This figure may be misleading, as some of the words in the “wordlist” for which information about sources is not given seem to have been taken (only) from Mandarin, e.g., *bei* ‘northern.’ On the other hand *ba* ‘place, location’ and *zen* ‘whole, all entire’ appear to have a Japanese origin; at least the former does not seem to involve Mandarin at all.

## 2.8. Konya and Ilomi

Sulky (2005) says the following about “influences” on his language Konya:

Q: What are Konya’s influences?

A: There is a sizable handful. But they are only influences. For example, the marking of major clauses is something we see in Japanese, so Japanese could be considered an influence; but even if Japanese didn’t exist, I would still have used the concept in Konya, because it’s a good idea in the overall framework of the language.

English is an influence on the vocabulary, but then, so are the Romance languages and many others. The heavy use of compounding is something we see in Mandarin and other Sino-Tibetan languages, as well as in English. The syllable structure is influenced by Indonesian and Japanese. The limited phoneme set is

influenced by a now-dormant constructed language initially called Kalim. The self-segregating morphology is influenced by Loglan and its step-child Lojban.

He then cites the artificial language Ladekwa as “a major influence,” and brings up another artificial language, Ceqli, so it may have contributed to his ideas for Konya. The Konya dictionary (Sulky n.d.) does not contain any etymological information, so I cannot say whether and to what extent words were borrowed from Altaic languages. In any case, the language apparently remains incomplete.

A later language designed by Sulky is Ilomi; to at least some extent it can be seen as a development of Konya. Sulky’s (2006a) remarks about the “influences” for this language are quite similar to those just quoted, in fact the first paragraph is identical (with “Ilomi” substituted for “Konya”), as is the second paragraph, except that “Malayo-Polynesian languages” is added after “the Romance languages.” The dictionary of Ilomi (Sulky 2006b) does give what appear to be sources for some words, two of which are said to have Altaic etymologies (as well as one proper noun); the entries for these words are (with minor modifications):

- (3) a. *afito* ‘person (not implicitly adult), human, people’  
[Japanese *hito*]
- b. *amanka* ‘cartoon, comic’  
[Japanese *manga*, Korean *manhwा*]

## 2.9. Kosmo

Kosmo was designed by G. Schröder, who apparently never published anything on his language, and dates from approximately 1949. According to Monnerot-Dumaine (1960: 176) it “[a]ppartient à la même catégorie que l’Esperanto, mais n’est pas un espérantide” (‘belongs to the same category as Esperanto, but is

not an Esperantido [a language derived from Esperanto]'). I cannot say anything about the sources of the vocabulary of this language, as to my knowledge there is no publicly available information on it, but the system of case marking was taken from Japanese; Monnerot-Dumaine (1960: 54) says:

Peu de langues ont emprunté la déclinaison analytique du japonais. Celui-ci marque le cas au moyen de particules détachées: *ga* (ou *wa*) pour le nominative, *no* pour le génitif, *ni* pour le datif, *wo* pour l'accusatif. Ces particules suivent le substantif. Schröder s'en inspire pour son Kosmo qui marque le nominative par *ga*, le génitif par *no*, le datif par *ni*, l'accusatif par *vo*. Les particules *ga* et *vo* ne s'emploient que dans le mode emphatique, ou lorsqu'il est utile de préciser. Toutefois, le Kosmo laisse le choix pour le génitif, entre la préposition romane *de* et la postposition japonaise. Ex.: *le livre de la tante* peut se dire *la libra de la tanta* ou *la tanta no libra*. Schröder estime la seconde forme plus harmonieuse. Pour le datif, même choix entre la préposition *a* et postposition *ni*.

('Few languages have borrowed the analytic declension of Japanese. It marks the cases by means of detached particles: *ga* (or *wa*) for the nominative, *no* for the genitive, *ni* for the dative, *wo* for the accusative. These particles follow the noun. Schröder was inspired by it for his Kosmo, which marks the nominative with *ga*, the genitive with *no*, the dative with *ni*, the accusative with *vo*. The particles *ga* and *vo* are only used in the emphatic mode, or when it is useful to clarify. However, Kosmo allows the choice with respect to the genitive between the Romance preposition *de* and the Japanese postposition. E.g.,: 'the aunt's book' can be said *la libra de la tanta* or *la tanta no libra*. Schröder considers the second form

more harmonious. With respect to the dative, [there is] the same choice between the preposition *a* and the postposition *ni*.'

## 2.10. Kumiko

Kumiko was created by Ming Tang, who was 17 in 2010 and a "Chinese immigrant living in Canada" (Tang 2010). Tang (2011a) says that "Kumiko is the newest installation of logical languages," i.e., it is a language of the type of Loglan and Lojban. While its main purpose was not to be an AIAL, Tang (2011b) states, "Kumiko is a candidate for the next international auxiliary language." According to Tang (2011a: 2), "Kumiko is a phonetic language, and the syllable structure is based on Japanese, which is clear and easy to pronounce." On the same page Tang says, "Kumiko's main source languages for its vocabulary are Chinese, Japanese, and English." He then (ibid.: 3) gives more information about what Kumiko has taken from various sources:

- (4) Kumiko attempts to incorporate the advantages [of] many other languages
  - a. Japanese
    - Vocabulary
    - Sounds
    - Syllable structure
    - Loan word system
    - Mixing of writing systems
  - b. Korean
    - Featural alphabet

The other languages mentioned here are Chinese, English, Lojban, Ceqli (another artificial language), "Mathematical programming language," Perl 6, APL, and J (the last three of these are computer languages). Tang (ibid.) mentions katakana in the section entitled

“Alternative Writing Systems [for Kumiko]”; in fact it is the only system brought up there.

Below is a list of the 86 Kumiko words which come from Japanese, according to Tang (2011c):<sup>12</sup>

- (5) *aki* ‘autumn,’ *ama* ‘sweet,’ *ame* ‘rain,’ *ao* ‘cyan,’ *aR* ‘exist,’ *aso* ‘play,’ *au* ‘meet,’ *cuQ* ‘soil,’ *dai* ‘large,’ *daki* ‘hug, embrace,’ *eR* ‘obtain,’ *Fyu* ‘winter,’ *gacu* ‘month,’ *hae* ‘fly (insect),’ *haJ* ‘first,’ *hana* ‘flower,’ *haR* ‘spring,’ *ika* ‘squid,’ *iku* ‘go,’ *iro* ‘color,’ *iu* ‘say,’ *kami* ‘paper,’ *kani* ‘crab,’ *kao* ‘face,’ *kara* ‘spicy,’ *kasa* ‘umbrella, parasol,’ *kiku* ‘listen,’ *kiR* ‘wear,’ *ku-* ‘long time,’ *kumi* ‘grouped,’ *kuQ* ‘mouth,’ *kuro* ‘black,’ *kyo-* ‘capital city,’ *macu* ‘wait,’ *mi-* ‘beautiful,’ *mi+* ‘ear,’ *miku* ‘future,’ *mise* ‘store,’ *Mne* ‘chest,’ *mo* ‘peach,’ *mocu* ‘hold’ (V), *moe* ‘moe’ (anime term), *MX* ‘insect,’ *nacu* ‘summer,’ *naW* ‘rope,’ *neko* ‘cat,’ *neM* ‘rest,’ *neN* ‘year,’ *niN* ‘person,’ *noM* ‘drink,’ *nue* ‘nue,’<sup>13</sup> *omo* ‘think,’ *oni* ‘oni,’<sup>14</sup> *qa-* ‘tea,’ *qo-* ‘butterfly,’ *Qsa* ‘small,’ *re-* ‘example,’ *sake* ‘sake (wine),’ *saku* ‘cherry blossom,’ *sasu* ‘point’ (V), *seka* ‘world,’ *sora* ‘sky,’ *su-* ‘number,’ *sui* ‘water,’ *suki* ‘like,’ *suru* ‘do,’ *sute* ‘throw,’ *suW* ‘sit,’ *suX* ‘sushi,’ *tabe* ‘eat,’ *tacu* ‘stand,’ *taki* ‘waterfall,’ *tama* ‘egg,’ *teR* ‘come out,’ *tomo* ‘friend,’ *uqo* ‘outer space,’ *usa* ‘rabbit,’ *uta* ‘song,’ *XB* ‘news,’ *Xma* ‘island,’ *Xnu* ‘die,’ *Xo* ‘salt,’ *Xro* ‘white,’ *Xzu* ‘quiet,’ *Yma*

<sup>12</sup> Some of the orthographic symbols of Kumiko require explanation: <-> indicates that the final vowel of a word is to be pronounced twice (e.g., *mi-* is pronounced [mi]), <+> indicates that the last final syllable of a word is pronounced twice (e.g., *mi+* is pronounced [mimi]). Most capital letters stand for sequences of sounds, e.g., <F> stands for [fu]; <N> stands for a syllabic nasal which can be pronounced in various places of articulation. The letters <q> stands for [tɛ] or [tʃ].

<sup>13</sup> I do not know what this gloss means, or whether it is an error.

<sup>14</sup> Once again I do not understand the meaning of this gloss.

‘mountain,’ *yu-* ‘soup’

This dictionary also has words said to be from Chinese, English, Latin, and several other languages (though no other Altaic ones). The total number of words in it is 377, and thus the words with a Japanese etymology make up approximately 23% of them.

## 2.11. Lingwa de Planeta (LdP, Lidepla)

According to Ivanov (2011):

Lingwa de Planeta is a neutral international auxiliary language based on the most widely spoken languages of the world, including English, German, French, Spanish, Portuguese, Chinese, Russian, Hindi, Arabic, and Persian. We attach much importance to psychology: elements of one’s own native language in a constructed language evoke positive emotions and increase greatly one’s motivation to learn it. And we hope that the bulk of the planet’s population may perceive Lidepla not only as neutral, but as a common language with a certain smack of their native tongues.

The *LdP-English Dictionary* (Ivanov 2009) has words built from 1,500 roots (according to Ivanov 2011), a fairly large number of which said to be from an Altaic language, although in many cases the Altaic language is clearly not the original source of the language (although Ivanov does not say this). (On the other hand one might imagine that *tsunami* was taken from English, not directly from Japanese, though again Ivanov does not state this.) Ivanov (*ibid.*) does say, “Word etymologies are given sparingly, and usually only one source language is indicated. However you should bear in mind that we have often adopted roots that are

widely spread across languages. For example, the word ‘*kitaba*’ is originally from Arabic, but the root is also present in Turkish, Hindi, and many other languages.”<sup>15</sup> Below is a list of words said to be constructed from Altaic roots,<sup>16</sup> with glosses and explanations from Ivanov (2009), occasionally modified:

- (6) a. Azeri (3): *divan* ‘divan (sofa),’ *gitara* ‘guitar,’ *vulkan* ‘volcano’
- b. Crimean Tatar (1): *bagaja* ‘luggage’
- c. Turkish (69): *aida* ‘let’s go, come along,’<sup>17</sup> *aksham* ‘evening,’ *albatross* ‘albatros,’ *algoritma* ‘algorithm,’ *banka* ‘bank (financial institution),’ *barut* ‘gunpowder,’ *bavul* ‘suitcase,’ *bitum* ‘bitumen,’ *buton* ‘button (fastener; device meant to be pressed; bud),’ *chapa* ‘hoe, mattock,’ *depo* ‘depot, store, storehouse,’ *dialoga* ‘dialogue,’ *dilim* ‘slice, clove,’ *dolar* ‘dollar,’ *ekran* ‘screen (viewing area),’ *ekwator* ‘equator,’ *etiket* ‘etiquette,’ *fenomen* ‘phenomenon,’ *fitil* ‘wick, fuse,’ *fok* ‘seal (animal),’ *gargari* ‘gargle,’ *garson* ‘waiter,’ *general* ‘general (commander),’ *gol* ‘goal (sports),’ *grip* ‘flu, grippie, influenza,’ *gubra* ‘dung, manure, muck,’ *gunah* ‘sin’ (N), *harem* ‘harem,’ *idris* ‘bird cherry,’ *juri* ‘jury,’ *kamar* ‘squid,’ *kamion* ‘lorry, truck,’ *kapak* ‘lid (cover),’ *karpus* ‘watermelon,’ *kazan* ‘boiler,’ *leleka* ‘stork,’ *maimun* ‘monkey,’ *mat* ‘checkmate, mate’ (N), *ochak* ‘hearth, fireplace; (fig.) hotbed, seat, breeding ground, focus, center,’ *omlet* ‘omelette,’ *orak* ‘sickle,’ *palto* ‘coat, overcoat,’ *paragraf* ‘paragraph,’ *parola* ‘password,’ *patlajan* ‘eggplant, aubergine,’ *radyum* ‘radium,’ *rejim*

---

<sup>15</sup> There are many words for which no etymological information at all is given.

<sup>16</sup> Ivanov (2009) lists words, not roots, and so the list below also contains words (but only headwords).

<sup>17</sup> Ivanov (*ibid.*) gives the etymology of this word as “Turkish, Slavic.”

- ‘regime, regimen,’ *rejisor* ‘director (in a show or film),’ *rekor* ‘record (unsurpassed measurement),’ *shorta* ‘shorts,’ *sosis* ‘sausage,’ *soya* ‘soya,’ *tapa* ‘spigot, tap, plug, stopper,’ *tavan* ‘ceiling,’ *tel* ‘wire,’ *tirpan* ‘scythe,’ *topale* ‘lame,’ *topuk* ‘heel (part of the foot),’ *tuman* ‘fog; mist,’ *turk* ‘Turkish (relating to ethnicity, language, or country); Turkish (person of Turkish ethnicity); Turkish language’ (N, A),<sup>18</sup> *tuvalet* ‘toilet (grooming or dressing),’ *uf* (an exclamation expressing tiredness or relief, appeasement), *uniforma* ‘uniform (distinctive outfit),’ *uranyum* ‘uranium,’ *ya<sub>1</sub>* (exclamatory particle) ‘yes’ (e.g., *oo ya!* ‘oh yes!’, *ya<sub>2</sub>* (grammatical particle) ‘you see, really’ (emphatic particle), *yash* ‘age (stage/time or duration of life),’ *yok* ‘there is no, not available, no,’ *yoksuni* ‘deprive, bereave,’ *yurta* ‘yurta, yurt, ger’
- d. Uzbek (3): *akwarium* ‘aquarium,’ *surma* ‘antimony, stibium,’ *tuhmat* ‘slander, calumny’
- e. Mongolian (1): *roman* ‘novel’
- f. Korean (2): *onlain* ‘online’, *tampon* ‘tampon’<sup>19</sup>
- g. Japanese (11): *bandana* ‘bandana,’ *bekon* ‘bacon,’ *fuki* ‘blow’ (V), *kiba* ‘fang,’ *mango* ‘mango,’ *nihon* ‘Japanese’ (N, A),<sup>20</sup> *pay* ‘pie,’ *tana* ‘shelf,’ *tawa* ‘tower,’ *tsunami* ‘tsunami,’ *yuma* ‘humor’

The vast majority of these words are not particular to an Altaic culture; an exception is *harem*.<sup>21</sup> Given the already noted fact that

<sup>18</sup> The word for ‘Turkey’ is *Turkiye*; its etymology is not given in Ivanov (2009).

<sup>19</sup> The words *Chosen* ‘North Korea’ and *Hanguk* ‘South Korea’ were almost certainly taken from Korean, but Ivanov (*ibid.*) does not state this.

<sup>20</sup> The word *Nipon* ‘Japan’ presumably comes from Japanese, but its origin is not given in Ivanov (*ibid.*).

<sup>21</sup> Another LdP dictionary, Ivanov & Lysenko (2007), has far more headwords (3,733 as of January 28, 2012), but the same number of words stated to be from Azeri, Crimean Tatar, Turkish, Uzbek, Mongolian, Korean, and Japanese.

many of those supposedly Altaic words are obvious (and relatively recent) borrowings into the source Altaic language (e.g., *akwarium*, *omlet*, *onlain*, *roman*), the linguistic representation of Altaic languages in LdP is considerably less than one might think when simply looking at the numbers of words from Turkish, and so on in LdP. One might wonder why Ivanov did not simply state that *roman* was taken from French, *bekon* from English, and so on. It might have been that he wanted to give a fairly high level of representation to Altaic (and other non-Indo-European) languages, but found it easier to choose some words which he was already familiar with due to their originally English, French, and so on origin. Of course it would also be easier for learners who speak English and French, but the result is an AIAL which is not as “neutral” as might be claimed. On the other hand there are more truly Altaic words, or at least words which are not recent borrowings from Western languages,<sup>22</sup> than in Esperanto, Ido, or Interlingua, including some for concepts or items not particular to one or more Altaic cultures, e.g., *yash* ‘age.’

---

However, what appears to be an earlier version of this dictionary (also © 2007, but downloaded on April 28, 2010) has some differences with respect to etymological information, attributing a (partly) Altaic origin to some words which are not said to be from an Altaic language in the current version. (There is a significant difference between the two versions: although they are both supposed to be “multilingual,” the current version only has LdP and English.) For example, in the earlier version the origin of *chay* ‘tea’ is given Turkish and Hindi, while in the current version only Hindi is mentioned. In the earlier version *chinara* ‘plane tree’ is said to be from Turkish, but in the current version Armenian is given as the source. According to the earlier version of this dictionary *adres* ‘address’ came from Turkish; in the current version no etymological information on it is given. On the other hand we also find the opposite type of situation: in the current version *divan* ‘divan (sofa)’ is said to be from Azeri, while in the earlier version there is no etymological information about this word; according to the current version *palto* ‘coat, overcoat’ came from Turkish, while in the earlier version (where it is glossed as ‘overcoat, paletot’) a Russian origin is attributed to it.

<sup>22</sup> For example, Turkish has borrowed many words from Arabic and Persian.

## 2.12. Neo Patwa

Wilkinson (2010a) states that “Neo Patwa is a pidgin-like international language that is used by people around the world to communicate when they do not share a common language.” As for the vocabulary, “the words [of New Patwa] appear to be chosen from real languages, such as English, Chinese, Hindi, Swahili, Spanish, Arabic, Russian, Indonesian, Korean, and Japanese. So learning Neo Patwa means learning words that are already in use by people from different cultures around the world.” (ibid.).

Below are the words to which the *Neo Patwa-English Dictionary* Wilkinson (2010b) attributes an Altaic origin:

- (7) a. Turkish (1): *kanun* ‘law, rule, regulation’<sup>23</sup>
- b. Mongolian (1): *altai* ‘gold, metal’<sup>24</sup>
- c. Korean (7): *acim* ‘morning,’ *anda* ‘sit, at, located,’ *babo* ‘fool, foolish, stupid,’ *ban* ‘evening, afternoon,’ *dalida* ‘run, flow, happen, current,’ *pal* ‘eight,’ *ulda* ‘cry, weep’
- d. Japanese (13): *cika* ‘near, close, approach, soon,’ *genki* ‘health, healthy, vigorous,’ *kabe* ‘wall, obstacle, hinder,’ *kaki* ‘write, scratch,’ *kona* ‘powder, flour,’ *mada* ‘still, yet, again,’ *mo* ‘also, too, neither, and, even,’<sup>25</sup> *nige* ‘escape, avoid,’ *oci* ‘drop, fall,’ *samwi* ‘cold,’ *si* ‘four,’ *tate* ‘stand, vertical, position,’ *tomo* ‘accompany, with, friend, together, mutual, mix’
- e. Korean and Japanese (1): *kaci* ‘price, value’

There is an older *Neo Patwa-English Dictionary*, Wilkinson (n.d.), which is much longer (approximately 24 2/3 pages,

<sup>23</sup> Wilkinson (2010b: 3) gives the origin of this word as “Turkish, etc.”

<sup>24</sup> Wilkinson (ibid.: 1) says “from mountains,” i.e., from the toponym *the Altai mountains*.

<sup>25</sup> Wilkinson (ibid.: 4) gives the origin of this word as “Bislama [sic]/Japanese.”

compared to the approximately 6 1/3 pages of Wilkinson (2010b) and contains many more words. The words in it to which Wilkinson attributes an Altaic etymology are:

- (8) a. Turkish (11): *balina* ‘whale,’ *bati* ‘west,’ *cati* ‘roof, top, lid,’ *kanca* ‘hook,’ *kanun* ‘law, rule, regulation,’<sup>26</sup> *mide* ‘stomach,’ *sakal* ‘beard, moustache,’ *sane* ‘stage,’<sup>27</sup> *somon* ‘salmon,’ *sosisu* ‘sausage,’ *yesil* ‘green’
- b. Korean (15): *acim* ‘morning,’ *anda* ‘sit, at, located,’ *babo* ‘fool, foolish, stupid,’ *ban* ‘evening, afternoon,’ *gwi* ‘ear,’ *hyomo* ‘yeast, ferment,’ *misu* ‘smile,’ *mogi* ‘mosquito,’ *nuwe* ‘brain,’ *pada* ‘dig,’ *pal* ‘eight,’ *paseli* ‘parsley,’<sup>28</sup> *ulda* ‘cry, weep,’ *usan* ‘umbrella,’ *yunipom* ‘uniform (clothes)’
- c. Japanese (46): *ali-dudu* ‘ant,’ *aspara* ‘asparagus,’<sup>29</sup> *budo* ‘grape,’ *cika* ‘near, close, approach,’ *cinami* ‘incidentally, by the way,’ *coco* ‘butterfly,’ *ebi* ‘shrimp,’ *egaki* ‘draw, describe,’ *fuki* ‘blow, puff,’ *genki* ‘health, healthy, vigorous,’ *gili* ‘in-law,’ *gulam* ‘gram,’ *haci* ‘bee,’ *ilasto* ‘picture,’ *ima* ‘now, present,’ *kame* ‘turtle,’ *kami* ‘bite,’ *kamome* ‘seagull,’ *kasumi* ‘mist, fog, steam,’ *kataci* ‘shape, form, appearance,’ *kicune* ‘fox,’ *kime* ‘decide, choose,’ *kona* ‘powder, flour,’ *kuxi* ‘comb,’ *mada* ‘still, yet,’ *mate* ‘wait,’ *mo* ‘also, too, neither, and,’<sup>30</sup> *momo* ‘peach,’ *muda* ‘waste,’ *nami* ‘wave,’ *nige* ‘escape, avoid,’ *numa* ‘swamp,’ *oci* ‘drop, fall,’ *sakura* ‘cherry,’ *sango* ‘coral,’ *sase* ‘cause, make, compel,’ *selori* ‘celery,’<sup>31</sup> *sumi*

<sup>26</sup> Wilkinson (n.d.: 10) again gives the origin of this word as “Turkish, etc.”

<sup>27</sup> Wilkinson (ibid.: 2) gives the origin of this word as “Japanese-English.”

<sup>28</sup> Wilkinson (ibid.: 18) gives the origin of this word as “Korean-English.”

<sup>29</sup> Wilkinson (ibid.: 20) gives the origin of this word as “Persian/Turkish.”

<sup>30</sup> Wilkinson (ibid.: 15) again gives the origin of this word as “Bisllama [sic]/Japanese.”

<sup>31</sup> Wilkinson (ibid.: 20) gives the origin of this word as “Japanese-English.”

‘corner, angle,’ *taikun* ‘business leader,’ *tate* ‘stand, vertical,’ *tawa* ‘tower,’ *tobi* ‘fly, jump,’ *tomo* ‘accompany, with, together, mutual,’ *toyo* ‘fertile, bountiful, abundant,’ *xi* ‘four,’ *yagi* ‘goat’  
d. Korean and Japanese (1): *kaci* ‘price, value’

Wilkinson (n.d.: 20) also has the word *sasa* ‘sudden, suddenly,’ the origin of which he states as “Hindi. Close to Japanese.”

There are many words from Altaic sources in Wilkinson (n.d.) which are not in Wilkinson (2010b). On the other hand, although it is much shorter, the latter has some words not present in the former: from Mongolian *altai*, from Korean *dalida*, from Japanese *kabe*, *kaki*, and *samwi*. Wilkinson (n.d.) had Neo Patwa words with the meanings of these words, but they were different words, e.g., *xita* ‘winter’ (from Arabic), *sona* ‘gold’ (from Hindi), and *metal* ‘metal’ (from English). In the case of the last two words the later dictionary has given two meanings to one word, while in the earlier dictionary there was a different word for each meaning. (Wilkinson (2010b) has no other word for ‘metal.’)

Aside from the differences in items included in the two dictionaries, the glosses for some items are different, e.g., in Wilkinson (2010b) *cika* (from Japanese) has one additional meaning in its gloss (i.e., which was not in the gloss for this word in Wilkinson (n.d.), ‘soon,’ *mada* has the additional meaning ‘again,’ and *mo* has the additional meaning ‘even’ (both also from Japanese)). Wilkinson (n.d.) has words for ‘again’ and ‘even,’ *lagi* (from Singlish-Malay) and *ance* (from Italian) respectively, while these words are not in Wilkinson (2010b), nor are any other words which only mean ‘again’ or ‘even.’ This is thus one way in which one could reduce the size of a vocabulary of an artificial language, by assigning additional meanings to a word already present, and getting rid of words which only have those meanings. On the other hand, in Wilkinson (n.d.) there is no word with the meaning ‘soon,’ so a new meaning has been added to the language. Perhaps

the most notable addition to meaning involves *tomo*, which in Wilkinson (2010b) has the new meanings ‘friend’ and ‘mix,’ along with its older meanings of ‘with,’ and so on. There are no other words for ‘friend’ or ‘mix’ in Wilkinson (*ibid.*), while the words for ‘friend’ and ‘mix’ in Wilkinson (n.d.) are *ami* (from French) and *mixen* (from German; it also means ‘combine, blend’).

In the case of the word for ‘four,’ the pronunciation and spelling have changed, from *xi* to *si*.

This is presumably due to the fact that the letter <x>, which in the earlier version of the language stood for [ʃ], was removed from the 2010 version, although the sound [ʃ] still exists in the language as a possible pronunciation of <c> (while <s> stands for [s] in both versions).

If we look at how many Neo Patwa words in Wilkinson (2010b) are said to come from other languages, we find 20 words from Mandarin, two words from Vietnamese, 21 words for which French is given as the source, or one of the sources, also 21 words for which English is the source or a source, and 5 words from German. These figures, as well as those for words from Altaic languages, do not include compounds containing a word from one or another of these languages.

## 2.13. NOXILO

Sentaro (2012) states:

NOXILO (pronounced ‘NOSHILO’) is a truly equal and easy-to-use international auxiliary language, which was made in Japan for Asians, Europeans, Africans, and North/South Americans. Unlike most other international auxiliary languages, NOXILO allows most users (American, Chinese, French, German, Indian, Korean, Japanese, Malaysian, Russian, Spanish, Thai, and so on) to write and speak in the word order of their native

tongue or in a similar word order.

There is also a considerable degree of freedom with respect to the vocabulary:

NOXILO vocabulary consists of *two* types of words, which are *Basic Words* (BW) and *International Standard Words* (ISW). Basic Words are mandatory to use. Except for the BWs, NOXILO users can freely choose proper words from ISWs, or any words from their mother language such as English, French, Spanish, Russian, Arabic, Chinese, Korean, Japanese, Swedish, Thai (Sentaro 2012).

As far as I know the sources of most NOXILO words are not given in any publicly available materials on the language in English, but there is etymological information on some words. The cardinal numerals *NI* ‘two’ and *MAN* ‘10,000’ come from Japanese, while *SAM* ‘three’ and *OK* ‘100,000,000’ are from Korean.<sup>32</sup> NOXILO had two other cardinal numerals with a Korean provenance, *QO* ‘one trillion’ and *KyOn* ‘10,000,000,000,000,000,’ but Sentaro (2011b) says that they were discarded (and words from Greek and Chinese respectively used instead) as “한국어가 너무 많다는 비판을 고려했습니다” (‘criticism that there was too much Korean [in NOXILO] was taken into consideration’). The word for ‘decimal point,’ *TEn*, is of Japanese origin.

There are two ways of saying ‘thank you,’ one of which, *KAMSA*, has a Korean etymology (the other is from Swahili). Another word of Korean origin is *XI* ‘hour.’ The word *KAn* ‘for’

---

<sup>32</sup> There is a NOXILO alphabet, but one may write NOXILO in capital letters of the Roman alphabet, with two lower case letters: <y> indicates that the preceding consonant is palatalized and <n> stands for a nasal: bilabial, alveolar, or velar, depending on its environment.

comes from Japanese; it appears to be a postposition and to mark extent of time, e.g., *25NE KAn* ‘for 25 days’ (Sentaro 2011a). One of the words for ‘autumn,’ *AKI*, was also borrowed from Japanese. It is among the words which “are originally Friendship Nouns (now part of ISWs)” (*ibid.*).<sup>33</sup>

## 2.14. Olingo

Olingo was designed by R. Stewart Jaque, who writes (1944: 24) that this language “is basically Neo-Latin and Anglo-Saxon with roots and words selected from all of the major languages of both the Western and Eastern Hemispheres.” He then (*ibid.*: 24-25) gives some Olingo words from various languages which “show the worldwide language coverage of the Olingo vocabulary” (*ibid.*: 25). Four of these words are from Japanese: *qimono* ‘kimono,’ *quchigo* ‘destroyer (ship),’ *sampo* ‘trench,’ and *seqajo* ‘patrol.’ No words from other Altaic languages are given there, though there are words from Arabic, Chinese, English, French, German, Greek, Italian, Latin, Malay, Portuguese, Russian, Sanskrit, and Spanish, and one word from Hawaiian. Jaque (*ibid.*) contains an Olingo-English Vocabulary and an English-Olingo Vocabulary, each with more than 1,000 words, but since no etymological information is given it is impossible to state the percentage of words from Japanese.

Globaquo is a revision of Olingo. I only have secondary sources for this language; Monnerot-Dumaine (1960: 172) describes its vocabulary as follows: “Racines romano-anglaises peu déformées et quelques racines arabes, chinoises, japonaises, malaises, russes, sanskrites” (‘Slightly changed Romance-English roots and some Arabic, Chinese, Japanese, Malay, Russian, [and] Sanskrit roots’).

---

<sup>33</sup> To my knowledge, in the currently available English materials on NOXILo there are no explanations of what “Friendship Nouns” or “Friendship Words” are.

## 2.15. Pan-kel

Pan-kel is an artificial language of the mixed type, i.e., made up of both *a priori* material and material taken from one or more natural languages. Its designer was Max Wald of Gross-Beeren (near Berlin). There were several editions of his book presenting the language, the first one published in 1906. (I have used the fourth edition (1909) here.)

One of Wald's concerns was with the brevity of words, and he drew upon a wide range of languages for his vocabulary, as the following quotation shows (Wald 1909: 11):

Für pan-kel sind aus *allen* Kultursprachen die kürzesten Wörter mit möglichster Vermeidung von Verstümmelung entnommen worden. Weil das Englische am verbreitesten ist, so muss auch der Wortschatz hauptsächlich dem Englischen entliehen werden, danach kommt dann das Deutsche und das Französische in Betracht. Doch finden Vertreter anderer Völker auch eine Menge Wörter in pan-kel, die völlig mit ihrer Muttersprache übereinstimmen . . . So sind ausser allen Sprachen Europas, z.B. noch Arabisch, Chinesisch, Japanisch, Malaiisch, und Persisch berücksichtigt.

('For Pan-kel the shortest words from *all* civilized languages have been taken with the greatest possible avoidance of mutilation. Because English is the most widespread, the vocabulary must be borrowed mainly from English; after it German and French come into consideration. However, representatives of all peoples will find a quantity of words in Pan-kel which completely correspond to [words in their] mother language . . . Thus, e.g., Arabic, Chinese, Japanese, Malay, and Persian are considered, in addition to all the languages of Europe.')

Wald does not give etymological information in the “Dictionary” in his (1909) work (although the first part of it is “Wörter, die in Deutsche und pan-kel gleich sind” (“Words which are the same in German and Pan-kel’)), so it is difficult to determine precisely how many words came from Japanese or other languages. Couturat & Leau (1907: 21-22) cite one Turkish word in Pan-kel, *at* ‘horse.’

## 2.16. Sambahsa-mundialect (Sambahsa)

The designer of Sambahsa-mundialect (or Sambahsa) is Olivier Simon. The oldest webpage presenting it dates from 2007. Unlike some of the languages discussed in this paper, Sambahsa has an extensive vocabulary and a large amount of learning and reference material. Simon (2010a) says the following about where the words of his language Sambahsa-mundialect came from:

The first source of inspiration for the vocabulary has been reconstructed Indo-European, but words from dialectal areas covering languages from different linguistic families (ex: sprachbünde) play a prominent role too. The technical and scientific vocabulary comes from the Greco-Latin roots common to most European languages. Generally, Sambahsa-mundialect tries to include words common to at least two different languages. The chosen languages are in general European ones (and above all English and French), but the spectrum stretches from Iceland to Japan, therefore including words from Turkish, Arabic, Persian, Swahili, Sanskrit, Indonesian, and Chinese.

There is not much information on the origin of non-Indo-European-based Sambahsa words, but there is a webpage entitled *Origine du vocabulaire sambahsa* (‘Origin of the Sambahsa Vocabulary’) (Simon 2010b). It only gives data on 703 words, but

it might give some idea of the extent to which different languages were drawn upon. Simon (2010b) says, “L’origine retenue a été l’étymologie la plus ancienne connue, ou du moins celle à laquelle le mot en question avait acquis le même sens qu’en Sambahsa” (“The origin which was kept [i.e., given on this page] was the oldest etymology known, or at least that in which the word in question had acquired the same meaning as in Sambahsa’). The percentages for different sources are: Indo-European 44.28%, Latin 15%, Germanic 9.5%, French 6.21%, Greek 4.64%, Romance 3.95%, Arabic 3.42%, English 1.45%, Slavic 1.28%, Italian 1%, Persian 0.92%, German 0.78%, and Chinese 0.71%. It can be seen that no Altaic language is even mentioned here; however, in the list of Sambahsa words on this webpage there are three words to which a Turkish etymology is attributed: *cofie* ‘coffee,’ *murdar* ‘dirty,’ and *khap* ‘tablet.’ Turkish thus accounts for 0.43% of the words in this list and no other Altaic languages are represented at all.

## 2.17. Sona

Kenneth Searight, the designer of Sona speaks of its “*a priori* basis” (1935: 21). However, it appears to be *a posteriori*, at least to some extent, as he also says (*ibid.*: 11), “There is no reason why we should not borrow, assimilate, and methodize many of the radical elements of many languages, such as I.G. [Italian German]: PA (feeding), C. [Chinese]: TA (great), A. [Arabic]: RU (go), J. [Japanese]: TE (hand), T. [Turkish]: SU (water), and so on, provided they fit in to the general scheme.” That is, not only does Sona draw upon natural languages, but it is one of the earlier AIALs to take material from Altaic languages. Searight (1935) does not give etymological information on most of the roots of Sona, so it is impossible to say what percentage of its vocabulary is Altaic.

The influence of Japanese can also be seen in the grammar of

Sona. Consider the following remarks by Searight (1935: 10-11), which also show his attitude toward using natural languages as sources:

Any national or dead language is unsuitable for our purpose if adopted complete, while at the same time recognizing the necessity for borrowing, assimilating, and even improving upon their choicest gifts. For instance, we should adopt the logical word-order of English and Chinese (almost identical) in preference to those of German or Japanese, while from the latter we may conveniently borrow the structure wherein one verb only in a clause, or series of clauses, shows the time of action. In Japanese it is the last verb; but we improve upon the idea by making it first, thereby anticipating the sense. We find that Italian, in company with Chinese and Japanese, often suppresses the personal pronouns without loss of clarity.

Searight (*ibid.*: 35) discusses an “article” of Sona which is of Japanese origin:

Sona has a special form of Article called the Honorific, borrowed from J. This is the vowel *o*; cf. J. *o mimoso san*, *o kodomo shu*. It is used before names, words of address, and verbs as an expression of politeness. Thus we have: *o ra*, ‘(honorable) man, gentleman,’ *o hara* ‘sir,’ *o tu jiko* ‘your (honour’s) children,’ *o toru* ‘please pass,’ *o min* ‘(please) come in!’

I do not think that it would generally be regarded as an article, but in any case it is an interesting example of a borrowing of an item from Japanese into an AIAL.

## 2.18. Unish

Unish is, as far as I know, the only AIAL created in Korea and the only one created by Koreans. Choo (2001: 8) says the following on which languages Unish has used as sources:

'Unish' . . . is an efficient composition of 16 languages: 13 natural languages spoken by a population [of] more than 70 million people, the root of European languages, Greek and Latin, and the most prominent internation[al] language, Esperanto . . . English accounts for 85 percent [of the vocabulary] as a result of the fact that English has more borrowed words than any other language. European languages such as Spanish, Italian, and German account for 50-60 percent. Among oriental languages, Chinese, Korean, Japanese, and Hindi account for 5-15 percent of Unish.

This adds up to over 100%, which "is due to the fact that a considerable amount of Unish words has [sic] overlapping common roots" (*ibid.*).

To my knowledge no currently available material on the language gives details on the etymology of particular words, but Alexander (2005) "hazarded a guess at the origin of a number of words—probably wrongly in some cases." He continues, "Many are obviously Chinese, Japanese, Korean, and so on." In his list there is only one word to which he attributes a Korean etymology, *buzai* 'absence' (and in fact there is another Unish word for 'absence,' *absens*), and only one to which he attributes a Japanese origin, *ana* 'hole.' Kupsala (2005) mentions another Unish word which (apparently) comes from Japanese: *kao* 'face.'

I know of no features of the grammar of Unish which clearly came from Korean (or any other Altaic language), and in fact Unish grammar closely resembles English grammar in some

respects.

## 2.19. Other AIALs

I shall briefly mention some other AIALs which apparently took material from one or more languages but about which I have very little information.

Albani & Buonarroti (1994: 269) say that Gagne-Monopanglotte (or Monopanglotte), whose designer was Paulin Gagne, “è composta di parole prese da venti lingue (tedesco, inglese, arabo, cinese, danese, spagnolo, francese, greco, ebraico, indostano, italiano, malgascio, persiano, polacco, portoghese, russo, sanscrito, svedese, turco) proporzionalmente all’importanza dei popoli che le hanno parlate o le parlano” (‘is composed of words taken from twenty languages (German, English, Arabic, Chinese, Danish, Spanish, French, Greek, Hebrew, Hindustani, Italian, Latin, Malagasy, Persian, Polish, Russian, Sanskrit, Swedish, Turkish) in proportion to the importance of the peoples who spoke or speak them’). Gagne (1843), a work presenting this language, says nothing explicit about Turkish (or about most of the languages listed by Albani & Buonarroti), though he does say (p. 5) that “La langue universelle sera formée de la réunion ou mélange radical des principales langues mères, mortes ou vivantes” (‘The universal language will be formed by the collection or mixture of roots of the main mother languages, dead or living’). Perhaps in his later works on his language, none of which I have, he gives the full list of the languages from which he took material.<sup>34</sup>

---

<sup>34</sup> Gagne wrote a book entitled *Monopanglosse, langue universelle*, which was published in 1858. Note that the title of this book begins *Monopanglosse*, and Monnerot-Dumaine (1960: 184) and Duličenko (1990: 81) refer to Gagne’s language by this name, rather than as *Monopanglotte*, perhaps because they may not have been aware of Gagne’s earlier work (both give *Monopanglosse* a date of 1858 and do not mention Gagne (1843)). One might assume that despite the minor change of name Gagne thought of *Monopanglotte* and

According to Monnerot-Dumaine (1960: 182), Charles Menet's *Langue Universelle* (from 1886) had “[r]acines monosyllabiques empruntées à l'arabe, au turc, au Volapük etc.” ('monosyllabic roots borrowed from Arabic, from Turkish, from Volapük, and so on.')

Pei (1968: 137) says, “There is unconfirmed mention of a New Chino-Japanese which blends the two great Far Eastern Languages.” I do not know anything further about this possibly existing language.

*Veltparl* was designed by Wilhelm von Arnim, whose first book on it was published in Silesia in 1896. Couturat & Leau (1903: 204) mention various sources of its vocabulary [in addition to Latin, Greek, Volapük, and so on]:

On en emprunte même aux langues non-aryennes: non seulement des mots devenus internationaux comme *algèbre*, *gong*, *islam*, *pacha*, *caravane*, . . . mais des mots magyars comme *kert* ('jardin') et *tys* ('feu'), des mots hindoustani comme *seb* ('pomme'), chinois comme *tael* ('D. thaler'), japonais comme *tok* ('horloge'), annamites comme *tam* ('huit').

('One even borrows [words] from the non-Aryan languages: not only words which have become international like *algèbre*, *gong*, *islam*, *pacha*, *caravane*, . . . but Hungarian words like *kert* ('garden') and *tys* ('fire'), Hindustani words like *seb* ('apple'), Chinese words like *tael* (German 'thaler'), Japanese words like *tok* ('clock'), annamite words like *tam* ('eight').

I do not have the primary sources on this language and so I cannot say how many words of Japanese origin are in *Veltparl*.

Unolok was the creation of Hector Wilshire (1957, Brisbane).

Monnerot-Dumaine (1960: 198) says that its words come from various languages, including Japanese, as well as English, German, French, Italian, Spanish, and Russian.

Yazu (also called Ya-zu and Ya-za Huo) is the only AIAL that I know of whose designer, namely Tsegmedin Bold, was Mongolian. My only sources for this language are several webpages on the Internet, and none of them gives much information on it. Wen (2000) says, “Dek jarojn antauxe mi ricevis libron de Yazu. Yazu estas lingvo kreita de mongolo Bold. Gxi havas pli da radikoj el aziaj lingvoj kiel la cxina, mongola, japana, korea ktp. Sed ecx en Azio gxi ne havas uzantojn!” (“Ten years ago I received a book about Yazu. Yazu is a language created by a Mongolian, Bold. It has more roots from Asian languages like Chinese, Mongolian, Japanese, Korean, and so on. But even in Asia it does not have users!”). Don Harlow, as quoted in anonymous (2005), states, “‘Ya-zu’ or ‘Yazu’ is a proposed pan-Asian language invented by the Mongolian Ts. Bold.”

Abdullayev (2010) writes about волланджо (Vollanjo), which was designed by Niyameddin Kebirov, an Azerbaijani “teacher-linguist” (“преподаватель-языковед”). This language has Azerbaijani, Korean, and Japanese among its sources, along with Latin, English, French, German, Russian, Chinese, and “Indian” (“индийский”).

### 3. Discussion

I have already given some figures on the percentages of the vocabulary of some AIALs accounted for by words taken from Altaic languages. The following table shows the number of words/roots of three other languages (with relatively large vocabularies) said to come from Altaic languages, and the percentages (in parentheses) that they form of the total number of words/roots in a lexicon of the AIAL.

Table 1. Numbers and Percentages of Words from Altaic Languages in Three AIALs

	Ardano	LdP	Neo Patwa Wilkinson (2010b)
Total words/roots	457 <sup>35</sup>	1,500	356 <sup>36</sup>
Azerbaijani (Azeri)	1 (.2%)	3 (.2%)	–
Crimean Tatar	–	1 (< .1%)	–
Turkish	2 (.4%)	69 (4.6%)	1 <sup>37</sup>
Uzbek	–	3 (.2%)	–
Mongolian	–	1 (< .1%)	1 (.3%)
Manchu	1 (.2%)	–	–
Korean	1 (.2%)	2 (.1%)	7.5 <sup>38</sup>
Japanese	–	11 (.7%)	13.5 (3.8%)
Total Altaic words/roots	5 (1%)	90 (6%)	23 (6.5%)

If we consider this table, and the representation of the Altaic languages among AIALs in general, it is clear that it is rather low in most languages, not even considering the very many AIALs which have no or almost no words from Altaic languages. Jigwa and Kumiko are notable exceptions to this: Kumiko has taken a relatively large proportion of both its grammar and lexicon from Japanese, while Japanese and Korean have at least influenced the choice of a large percentage of the (admittedly small number of) Jigwa words whose etymology has been explicitly given. The

<sup>35</sup> This only includes words for which etymological information is given.

<sup>36</sup> This (admittedly approximate) figure does not include compounds. It was arrived at by multiplying the number of Neo Patwa words on one page (p. 2), multiplying it by 5, then adding the number of words on pp. 1 and 7 and subtracting the number of compounds.

<sup>37</sup> Recall that the origin of this one word, *kanun* ‘law’ was given as “Turkish, etc.”

<sup>38</sup> The .5 in this figure and the one below are due to the fact that the origin of *kaci* ‘price, value’ is said to be “Korean and Japanese.”

AIALs which do have some representation of Altaic languages do not give equal representation to all of these languages, which is no surprise. There are some Altaic languages which have not, to my knowledge, been used as a source by any AIALs, e.g., Evenki and Chuvash.<sup>39</sup> Some other Altaic languages are represented, but only in a minimal way and only in a small number of AIALs; these include Crimean Tatar and Manchu. Compared to all these languages, Japanese is a relatively large contributor, and if an AIAL draws upon only a single Altaic language, one might predict that it will probably be Japanese. Korean and Turkish stand between Japanese on the one hand and Azerbaijani and Uzbek, on the other; the latter languages are poorly represented in AIALs. Words from Korean and Turkish are present in some AIALs, but usually they are not as numerous as Japanese words. LdP is an exception to this; it has far more Turkish roots than roots from either Japanese or Korean. Ardano, or rather the part of the Ardano vocabulary for which information on sources has been given, is also an exception—it has words from Korean and Turkish, but not from Japanese. However, if Ardano indeed includes (or will include) words from all the world's languages, then there is (or will be) at least one Japanese word in its lexicon.

Also, while some AIALs have borrowed elements of, or been influenced by, Japanese grammar, this is not obviously the case for Korean or Turkish. (Recall however, that Kumiko “attempts to incorporate” the “Featural alphabet” (Tang 2011a: 3) of Korean.) Even Unish, an AIAL created by Koreans, has no clear instances of grammatical borrowing from Korean.

These facts are probably due to the relative prominence of languages, either in the western world or in general, and this

---

<sup>39</sup> Even the relatively well known Altaic languages Kyrgyz and Kazakh have little representation in AIALs; to my knowledge the only auxiliary language which draws on them is Jalpi Türk Tili, which is meant as a sort of pan-Turkic language. It thus not surprising that it uses these languages as sources, as well as Azerbaijani, Turkish, Turkmen, Uyghur, and Uzbek.

prominence depends on political, economic, cultural, and historical factors. The vast majority of AIAL designers have come from the western world, and in this cultural-political sphere the Altaic people who have had the most impact in recent centuries are the Japanese, followed by the Turks and Koreans (while until recently Uzbeks and Azerbaijanis have barely been noticed by the west). One might then wonder whether, if Korea becomes more politically salient in the coming centuries, the Korean language will be used more as a source for AIALs. This might be the case, but the overall representation of Korean, and of Japanese and Turkish, in AIALs will probably not change much in the future. This is due to the dominance of Esperanto, which, as we have seen, has very little material from Altaic languages—it is very unlikely that any present or future AIAL which does have more substantial representation of Altaic languages will take the place of Esperanto as the most used AIAL (nor is it probable that the positions of Ido, Interlingua, and so on as the next most popular AIALs, will be taken over by e.g., Ardano). The same remarks apply to most other non-Indo-European languages as well: for example, Uralic languages have little representation in AIALs and it is unlikely that any AIALs which do have a substantial number of Uralic items will replace Esperanto.

One might ask whether any particular Altaic words were particularly popular sources for AIAL words (leaving aside the obvious possibilities, i.e., words for items or actions specific to a culture). That is, do any Altaic words appear as sources in several AIALs? The word for ‘sausage’ in both LdP and Neo Patwa is said to come from Turkish, and a word for ‘blow’ in both languages comes from Japanese, and has the form *fuki* in both. The word for ‘tower’ in both LdP and Neo Patwa is *tawa* and in both cases Japanese is given as the source language. Such occurrences could be coincidences, or perhaps the designer of one AIAL was influenced by another AIAL—i.e., one might think that such words were borrowed into one AIAL from another, rather

than being taken directly from Japanese or some other natural language. To fully answer the question posed at the beginning of this paragraph we might need more etymological information on AIALs than is currently available; in fact this is true for the analyses presented in this paper in general—a clearer picture of the extent of representation of Altaic languages in AIALs would emerge if designers gave more detailed information on the sources of words.

#### 4. Conclusions

The representation of Altaic languages in AIALs is generally small, even in most of those AIALs which (claim to) strive for a less Eurocentric vocabulary. However, there may be a trend towards the creation of AIALs using a broader base of source languages, since most of the languages which are discussed in this paper, i.e., which do take some material from Altaic languages, are relatively new—if this trend continues we may eventually see more Altaic (and other non-Indo-European) material in AIALs.<sup>40</sup> This does not mean that Altaic words and grammatical features will necessarily become much more prominent among AIALs overall, since it appears unlikely that any of these more recently designed auxiliary languages will displace the most successful AIALs such as Esperanto.

In addition, as we have seen with respect to LdP (note also the Neo Patwa words *somon* ‘salmon’ and *sosisu* ‘sausage’), not all supposedly Altaic words in AIALs are really Altaic; rather some are fairly recent borrowings from a Western language (or perhaps Esperanto), and do not have any Altaic “feel” to them. Thus, using

---

<sup>40</sup> It would be interesting to do the same kind of research as in this paper on words from other non-Indo-European language families in AIALs, e.g., from Uralic and Dravidian.

a wider range of source languages for an AIAL does not guarantee greater internationality, unless one avoids such borrowings.

Taking all of this into consideration, one would say that Altaic languages have not played a major role in AIALs, and are unlikely to do so in the near future. The same is true of most other non-Indo-European languages.

## References<sup>41</sup>

- Abdullayev, K. 2010. Азербайджанский учитель составил новый искусственный язык. Available at URL <[http://ru.apa.az/news\\_%D0%90%D0%B7%D0%B5%D1%80%D0%B1%D0%B0%D0%B9%D0%B4%D0%B6%D0%B0%D0%BD%D1%81%D0%BA%D0%B8%D0%B9\\_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D1%8C\\_%D1%81%D0%BE%D1%81%D1%82%D0%B0%D0%BB%D1%82%D0%B2%D0%BB%D0%88%D0%BB\\_%D0%BD%D0%BE%D0%B0%D2%D1%8B%D0%BD%D0%B9\\_154364.html](http://ru.apa.az/news_%D0%90%D0%B7%D0%B5%D1%80%D0%B1%D0%B0%D0%B9%D0%B4%D0%B6%D0%B0%D0%BD%D1%81%D0%BA%D0%B8%D0%B9_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D1%8C_%D1%81%D0%BE%D1%81%D1%82%D0%B0%D0%BB%D1%82%D0%B2%D0%BB%D0%88%D0%BB_%D0%BD%D0%BE%D0%B0%D2%D1%8B%D0%BD%D0%B9_154364.html)>.

Akademio de Esperanto. (n.d.) Akademia Vortaro. Available at URL <[http://akademio-de-esperanto.org/akademia\\_vortaro/](http://akademio-de-esperanto.org/akademia_vortaro/)>.

Albani, P. & B. Buonarroti. 1994. *Aga Magéra Difúra*. Bologna: Zanichelli.

Alexander, A. 2005. Unish Wordlist (message posted to the Yahoo group Auxlang). Available at URL <<http://tech.groups.yahoo.com/group/auxlang/message/35946?var=1>>.

Anonymous. 2005. International Auxiliary Languages: Yazu. Available at URL <<http://ial.wikia.com/wiki/Yazu>>.

Choo, M-G. 2001. The Need for Unish, a Universal Language, and its Principles of Development. *Journal of Universal Language* 2.1, 3-14.

Couturat, L. & L. Leau. 1903. *Histoire de la langue universelle*.

<sup>41</sup> The authors of some of these sources are not given in the sources themselves, but I have been able to discover them in one way or another.

- Paris: Hachette.
- \_\_\_\_\_. 1907. *Les nouvelles langues internationales*. Paris: Hachette.
- Duličenko, A. 1990. *Международные вспомогательные языки*. Tallinn: Valgud.
- Elhassi, Z. 2008. Complete Ardano Course. Available at URL <<http://groups.yahoo.com/group/ardano/files/Ardano%20courses/>>.
- Gagne, P. 1843. *La Gagne-monopanglotte, ou la langue unique et universelle*. Paris: Ildefonse Rousset.
- Gobbo, F. 2008. Planning Languages and Language Planning: The Contribution of Interlinguistics to Cross-Cultural Communication. In F. Boers et al. (eds.), *Multilingualism and Applied Comparative Linguistics Volume II* 24-42. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Ivanov, D. 2009. The LdP-English Dictionary. Available at URL <<http://lingwadeplaneta.info/files/lpen.shtml>>.
- \_\_\_\_\_. 2011. Lingwa de Planeta. Available at URL <<http://lingwadeplaneta.info/en/index.shtml>>.
- Ivanov, D. & A. Lysenko. 2007. Lidepla-English-Lidepla. Available at URL <<http://lingwadeplaneta.info/glossword/index.php>>.
- Jaque, R. 1944. *One Language*. Santa Barbara, CA: J. F. Rowny Press.
- Jigwadx Jungdwei.<sup>42</sup> (n.d.) Jigwa Update. *Journal of Planned Languages*.<sup>43</sup>
- Jigwadx Jungtim.<sup>44</sup> 1995. Jigwa. Formerly Available on the Internet, URL Unknown.
- Kupsala, R. 2005. Re: Unish Wordlist (message posted to the

---

<sup>42</sup> This means Jigwa Central Team.

<sup>43</sup> This article appeared in the *Journal of Planned Languages* and is currently available in a PDF file (with two other articles from the same journal) at <<http://www.oocities.org/athens/5383/jigwa.pdf>>. However, nothing in the file indicates when this particular article was published, though it was probably in the mid-1990s, and definitely after 1993.

<sup>44</sup> This is not a personal name, but a name of the group of people who created Jigwa.

- Yahoo group Auxlang). Available at URL <<http://tech.groups.yahoo.com/group/auxlang/message/35959>>.
- Monnerot-Dumaine, M. 1960. *Précis d'interlinguistique*. Paris: Librairie Maloine.
- Nutter, D. 2012a. Esperadunye. Available at URL <<http://conlang.dana.nutter.net/index.php/Esperadunye>>.
- \_\_\_\_\_. 2012b. Esperadunye Dictionary. Available at URL <[http://conlang.dana.nutter.net/index.php/Esperadunye\\_Dictionary](http://conlang.dana.nutter.net/index.php/Esperadunye_Dictionary)>.
- \_\_\_\_\_. 2012c. Pasifika. Available at URL <<http://conlang.dana.nutter.net/index.php/Pasifika>>.
- \_\_\_\_\_. 2012d. Pasifika-English Dictionary. Available at URL <[http://conlang.dana.nutter.net/index.php/Pasifika-English\\_Dictionary](http://conlang.dana.nutter.net/index.php/Pasifika-English_Dictionary)>.
- Pei, M. 1968. *One Language for the Word*. New York: Biblo & Tannen.
- Robertson, E. (n.d. a) Dunia: A World Auxiliary Language with a World Outlook. Available at URL <<http://www.langmaker.com/outpost/dunia.htm>>.
- \_\_\_\_\_. (n.d. b) Dunia: The Babel Text. Available at URL <<http://www.langmaker.com/babel/dunia.htm>>.
- Searight, K. 1935. *Sona*. London: Kegan Paul, Trench, Trubner & Co.
- Sentaro, M. 2011a. NOXILO Webpage 2. Available at URL <<http://www2s.biglobe.ne.jp/~noxilo/sub1b.htm>>.
- \_\_\_\_\_. 2011b. 문법정리. Available at URL <<http://www2s.biglobe.ne.jp/~noxilo/Korean.htm>>.
- \_\_\_\_\_. 2012. NOXILO Webpage 1. Available at URL <<http://www2s.biglobe.ne.jp/~noxilo/sub1a.htm>>.
- Simon, O. 2010a. FrontPage: Welcome to the Sambahsa-mundialect Wiki! Available at URL <<http://sambahsa.pbworks.com/w/page/10183084/FrontPage%7CThe>>.
- \_\_\_\_\_. 2010b. Origine du vocabulaire sambahsa. Available at URL <<http://sambahsa.pbworks.com/w/page/21576068/origine%20du%20vocabulaire%20sambahsa>>.

- Sulky, L. 2005. The Konya Language: Q&A. Available at URL <<http://larrysulky.webs.com/konya/konya-qa.html>>.
- \_\_\_\_\_. 2006a. The Ilomi Language—Version 2: Q&A. Available at URL <<http://larrysulky.webs.com/ilomi/ilomi-qa.html>>.
- \_\_\_\_\_. 2006b. The Ilomi Language—Version 2: Vocabulary. Available at URL <<http://larrysulky.webs.com/ilomi/ilomi-lex.html>>.
- \_\_\_\_\_. (n.d.) The Konya Language: Vocabulary. Available at URL <<http://larrysulky.webs.com/konya/konya-lex.html>>.
- Sztemon, L. (n.d.) Talvezi e Zevone e Euransi. Available at URL <<http://ls78.sweb.cz/talvezi.htm>>.
- Tang, M. 2010. SHiNKiROU. Available at URL <<http://shinkirou.org/>>.
- \_\_\_\_\_. 2011a. Introduction to Kumiko. Available at URL <<http://shinkirou.org/public/downloads/Introduction%20to%20Kumiko.pdf>>.
- \_\_\_\_\_. 2011b. Kumiko Wiki: Main Page. Available at URL <[http://shinkirou.org/kumiko/wiki/index.php?title=Main\\_Page](http://shinkirou.org/kumiko/wiki/index.php?title=Main_Page)>.
- \_\_\_\_\_. 2011c. Kumiko: Vocabulary. Available at URL <<http://shinkirou.org/public/downloads/dict.pdf>>.
- U., Almir, Jr. 2008. International Auxiliary Languages: Dousha. Available at URL <<http://ial.wikia.com/wiki/Dousha>>.<sup>45</sup>
- \_\_\_\_\_. (n.d. a) Números: Numba. Available at URL <<http://almirus.atspace.com/doushanet/langdesc/numeral.htm>>.
- \_\_\_\_\_. (n.d. b) Afixos: Toldala. Available at URL <<http://almirus.atspace.com/doushanet/langdesc/afixos.htm>>.

---

<sup>45</sup> The 2008 “revision” of this page was apparently done by someone other than Almir U. However, the original version of the page was created by Almir U., who was also responsible for some of the revisions (though not all of them), and I have listed him as the author of this page. The original version of the page, and the first three revisions, were in Portuguese, while more recent revisions are in English; it seems that Almir U. was not the person responsible for the first revision in English. Recall that Almir U.’s full last name is not given in material on Dousha.

- Wald, M. 1909. *Pan-kel, leichteste Kurzsprache für den Weltverkehr.* Gross-Beeren: Apparently Self-Published.
- Wen, J. 2000. Esperanto Vortaro: Internacieco kaj facileco. Available at URL <<http://www.angelfire.com/ny2/ts/wenjingen.html>>.
- Wilkinson, J. 2006. Guide-Oldest: A Guide to Neo Patwa. Available at URL <<http://patwa.pbworks.com/w/page/14800484/Guide-oldest?rev=1164948112>>.
- \_\_\_\_\_. 2010a. Neo Patwa: FrontPage [sic]. Available at URL <<http://patwa.pbworks.com/w/page/14800479/FrontPage>>.
- \_\_\_\_\_. 2010b. Neo Patwa: English Dictionary 2010. Available at URL <<http://patwa.pbwiki.com/f/P-E.pdf>>.
- \_\_\_\_\_. (n.d.) Neo Patwa: English Dictionary. Available at URL <<http://patwa.pbwiki.com/f/N-E-dict.pdf>>.

## INSTITUTE POLICIES

The Language Research Institute at Sejong University was founded for theoretical and practical research on various modes of language usage and to provide a forum to discussions on the nature of language. In order to accomplish this goal, the Institute undertakes the following functions:

### **1. Publication of *Journal of Universal Language***

The institute publishes biannually as an international research journal, *Journal of Universal Language* to develop the communication skills among different language speaking countries and help them to communicate efficiently. Followed by the initial issue of the journal in May 1996, the second issue was published in May 2001. Since then the journal has been published biannually.

### **2. Holding Seminars**

To offer a chance for the exchange of ideas between researchers of various viewpoints, we occasionally hold seminars on language related areas ranging from linguistics, artificial language, communication & culture in language usage, typology, and universality in language.

### **3. Exchange Programme**

We contract with other scholars and academic societies inside and outside of Korea and exchange databases and academic resources.

## **ETHICS GUIDELINES FOR THE INSTITUTE**

**— Established on May 31, 2007**

1. These guidelines aim at regulating issues that could occur with publication of the *Journal of Universal Language*. The followings are strict regulations that apply to authors, editorial board and reviewers.
2. Authors have rights concerning their own research and work, and the works are recognized as their own achievements. Upon acceptance of an article by the Journal, the authors will be asked to transfer copyright of the article to the publisher. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microfilm or any other reproduction of similar nature, and translations.
3. Submission of a manuscript will be held to imply that it contains unpublished, original material that it is not being considered for publication elsewhere. Upon acceptance of a paper, the author will be asked to assign copyright (on certain conditions) to Sejong University.
4. The Editorial Board members are appointed to review papers based on the specialty and research activities. Each paper is reviewed three times. And if there are no adequate editors in a certain field on the Editorial Board, we offer a commission to other experts. Proofs will be sent to authors for correction and should be returned to the editors within three days of receipt.
5. The Editorial Board will review and evaluate the journal in strict course.

6. Reviewers are all required to appreciate all of the submitted papers and the authors' ideas, and also be fair and balanced when evaluating the submitted papers.
7. In the instance of a violation of guideline no. 6, we have an Ethics Committee. It consists of 5 Board members, and a chairperson is elected from among the Committee. The Ethics Committee members are appointed by the executive officers of the Language Research Institute. The term of office is 2 years and Committee members can be reappointed for the next term. Decisions will be made upon a majority vote of the enrolled Committee members.
8. Upon determination, disciplinary action will be immediately taken against those found to be erring in their obligations, as per the above guidelines, by the executive officers of the Language Research Institute. Disciplinary action will take the following course:
  - (1) Verbal or documented warning
  - (2) Prohibition of submission of papers for an agreed upon duration
  - (3) Disqualification of position as a Editorial Board member or editor
9. General issues which are not specified here follow standard customs.
10. These Ethics Guidelines hold from September 1, 2007.

## PUBLICATION POLICIES

1. *The Journal of Universal Language* is published biannually (in spring and fall) in English by the Language Research Institute of Sejong University.
2. The spring and autumn editions are published respectively on March 31 and September 30.
3. Manuscripts submissions are accepted year round.
4. The topic of a paper contributed for the journal should relate to any language. The journal is structured to include:
  - (1) work on artificial language, communication & culture in language usage, typology, and universality in language;
  - (2) interdisciplinary contributions written with applied linguists in mind;
  - (3) brief notes, comments or observation concerning official languages;
  - (4) reaction/replies to recent papers;
  - (5) book reviews of important linguistic titles;
  - (6) letters to the editors.
5. Submission of a paper to the journal will be taken to imply that it represents original work not previously published, that is not being considered elsewhere for publication, and that if accepted for publication it will not be published anywhere without the consent of the Editor.
6. Upon acceptance of an article by the journal, authors will be asked to transfer copyright of the article to the publisher.

7. Additional Policies:

- (1) Authors will be properly notified of acceptance or rejection within 60 days of the submission of manuscripts.
- (2) Manuscripts, once submitted, shall not be returned.
- (3) A detailed “Guidelines for Submitting Manuscripts” and “Publication Policies” shall be provided in every issue of the journal.

## **EDITORIAL COMMITTEE GUIDELINES**

1. The Editorial Committee consists of members who have made excellent achievements in various language research fields.
2. The Editorial Committee decides on the publication, organization, and style of the contributed papers.
3. The Editorial Committee chooses three reviewers for each paper from the editorial members.
4. If a paper is found to be acceptable with partial revisions, the author must make the appropriate revisions, and resubmit the paper to the Editor. If he is unwilling to comply with the Editor's recommendations, the author must present a letter to the Editorial Board stating the reasons for noncompliance.
5. If a paper requires reexamination after revision, the author must revise and resubmit it for a final review.

## **REVIEW POLICIES**

1. A manuscript can be rejected out-of-hand if the English is so poor that it cannot be understood or if there are unusually large sections in an un-translated language other than English.
2. Reviewers evaluate a manuscript according to the following four categories:
  - (1) Article to be published without any revisions
  - (2) Article to require minor revisions before publication
  - (3) Article to be reexamined after major revisions
  - (4) Article to be rejected
3. Each submitted manuscript shall be subject to peer review by three reviewers from a relevant field of study. Reviewers shall be appointed by the Editorial Committee. Publication of manuscript is determined by approval of two or more reviewers. If two or more reviewers approve a manuscript as is or with minor revisions, we examine its scope to see if it fits the theme of the journal—the universality and typology of language. From that, the Editor-in-chief and the Editor finally decide whether it is to be published or not, and gives final notification to the author.

## **GUIDELINES FOR SUBMITTING MANUSCRIPTS**

Contributions are welcomed from all countries. When submitting a manuscript for consideration, we ask that authors consider a few basic guidelines. They are designed simply to allow our reviewers to work on the manuscript with minimum complications.

There is no limit on article length, but manuscripts longer than 40 pages (including footnotes and references) are not encouraged. The manuscript will be sent out for review with the author's name(s) removed. An abstract of approximately 200 words should immediately follow the title. On a separate sheet, the author should provide the entire title, name, affiliation, full mailing address, and an e-mail address if available.

### **I. Authors' Names**

1. If there are three or more authors, the corresponding author should be mentioned first, followed by the others.
2. Authors' names should be written in full, not as initials.

### **II. Headings**

1. Authors should use quadruple space between sections.
2. Authors should use the following section numbers for headings:  
1., 1.1., 1.1.1.,... 2., 2.1., 2.1.1....
3. Section numbers should start with I (not 0). If the article has a short introduction (one to three paragraphs), authors should not give it a number or a title. A longer introduction may have a number as well as a title.
4. Authors should use include titles for sections and subsections, following the capitalization conventions for English listed in *The*

*Chicago Manual of Style* (16th ed., 2010, University of Chicago Press, Chicago, Illinois). Type the numbers and titles in roman type.

### **III. Punctuation**

1. Authors should indent quotations with quotation marks.
2. Authors should not hyphenate words containing prefixes unless a misreading will result; hyphenate if the stem begins with a capital letter: Proto-Athabaskan.
3. Authors should indicate ellipsis with three periods, close set, with a blank space before and after.

### **IV. Abbreviations**

Abbreviations should be explained at first occurrence.

### **V. Notes**

1. All notes to the body of the text should be numbered serially throughout the manuscript.
2. Reference should be made for each footnote in the text with a raised numeral following the relevant passage, not enclosed in parenthesis.
3. Authors should type all notes to the body of the text as footnotes, double spaced.
4. Authors should place any acknowledgement footnote at the end of the title of the paper, keyed with an asterisk.
5. Each footnote begins with as a raised reference number (no punctuation).

## **VI. Numbered Examples**

1. The author should type each numbered item on a separate indented line with the number in parenthesis; indent after the number; use lowercase letters to group sets of related items:
  - (1) a. Jack and Jill ran up the hill.
  - b. Jack and Jill ran up the bill.
  - c. \*Jack and Jill ran the hill up.
  - d. Jack and Jill ran the bill up.
2. Authors should refer to numbered items as (1), (1a), (1a, c), (1a-d) in the text.

## **VII. Glosses and Translations of Examples**

Examples not in English must be translated or glossed appropriately. Sometimes, both a translation and a word-for-word or morpheme-by-morpheme gloss are appropriate.

1. Authors should place the translation or gloss of an example sentence or phrase on a new line below the example.
2. Authors should align word-for-word or morpheme-by-morpheme glosses of example phrases or sentences with the beginning of each original word.
3. Authors should observe the following conventions in morpheme-by-morpheme glosses:
  - 3.1. Place a hyphen between morphemes within words in the original, and a corresponding hyphen in the gloss.

fog-okfel	probal-ni	olvas-ni
will-1sg	try-inf	read-inf

- 3.2. Abbreviate glosses for grammatical categories and list the abbreviations in a note.

### **VIII. Cross-Referencing**

1. In the text, a reference identified by means of an author's name should be followed by the date of the reference in parenthesis and page number(s) where appropriate.
2. When there are more than three authors, only the first author's name should be mentioned, followed by 'et al.'
3. In the case that an author cited has had two or more works published during the same year, the reference, both in the text and in the reference list, should be identified by a lowercase letter such as 'a' and 'b' after the date to distinguish the works.

Farmer & Harnish 1987  
 Dowty 1980: 32  
 Hausser 2001a, 2001b  
 Couturat et al. 1903

### **IX. References**

Authors should provide a full bibliography at the end of the manuscript with the heading References. In addition, authors should adhere to the following guidelines:

1. Arrange the entries alphabetically by surnames of authors.
2. List multiple works by the same author in ascending chronological order.
3. Use suffixed letters a, b, c, etc. to distinguish more than one item published by a single author in the same year.
4. If more than one article is cited from one book, list the book as a separate entry under the editor's name, with cross-references to the book in the entries for each article.

5. Replace given names with initials: Lehiste, I., but Lehiste, Ise.
6. Be sure that each entry contains the following elements in the order given:

(First) author's surname, given name(s) or initial(s). Year of publication. Full title and subtitle of the work. For a journal article: full name of the journal and volume number (roman type), inclusive page numbers for the entire article. For an article in a book: In [full name(s) of editor(s)] (eds.), title of the book, inclusive page numbers. For books: the edition, volume or part number (if applicable) and series title (if any). Place of publication: Publisher.

7. Use the following examples as guides:

Poser, W. 1978. Impersonal Passives and the Unaccusative Hypothesis. *Berkeley Linguistics Society* 4.1, 157-189.  
Yip, M. 1991. Coronals, Consonant Clusters, and the Coda Condition. In C. Paradis & J. Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence* 61-78. San Diego, CA: Academic Press.

8. In some languages, such as Finnish, French, and Estonian, capitalize only the first letter of the initial word in the title, unlike English.

## X. Tables and Figures

1. Tables should begin with a table number and heading. There should be a horizontal line above and below the column headings, and a third line at the bottom of the table. Column headings and column entries should align on the left.

Table 3. Present Tense Endings in German

Person	Singular	Plural
1	-e	-en
2	-st	-t(ihr)/-en(sie)
3	-t	-en

2. It is the author's responsibility to provide a camera-ready copy for all figures. Figures include illustrations such as graphs, charts, photographs, and line drawings. Figures do not include standard tree structures, functional structures, tableaux, and so on, which are normally handled by the typesetter. Figures should be accompanied by a separately typed list of figure numbers and captions.

Journal of Universal Language  
March 2003, 33-75

## Languages and Universals

**David Odden**  
*Ohio State University*

### Abstract

A central question [...] how do we identify them; why do they exist?

Keywords: observational universals ...

### 1. Types of Universals

One example of a universal property is the Coordinate Structure Constraint [...] a number of other languages such as Spanish and German.

- (1) a. *English*    \*Who<sub>j</sub> do you see Mike and \_\_\_\_<sub>j</sub>?  
              b. *Spanish*    \*Quien<sub>j</sub> tu ves Miguel y \_\_\_\_<sub>j</sub>?

### 34 Languages and Universals

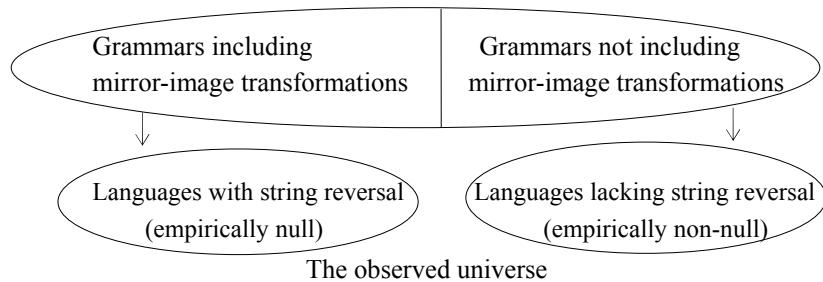
c. German    \*'Wem<sub>i</sub> siehst du Karl und \_\_\_?<sub>j</sub>'?

The principle is proposed [...] and as to what they are statements about.

#### 1.1. Three Concepts of Universals

The strongest position on linguistic universals is [...] languages which form yes-no questions in other ways.

##### (2) The Logical Universe of Architectural Principles



The connection between formal theory and observation about languages is [...] for a predictive theory which actually states what a “possible language” is.

#### 2. The Statistical Foundation of Universals

Before trying to [...] as we will discuss in the third part of this section.

## 2.1. Improbable Events

As Greenberg has pointed out [...] there is a very good reason to be skeptical about any such results.

## 4. Conclusions

The central problem in establishing universals of language is [...] the appearance of correctness because the counterexamplifying languages died out undocumented.

## References

- Archangeli, D. 1985. Yokuts Harmony: Evidence for Coplanar Representation in Nonlinear Phonology. *Linguistic Inquiry* 16, 335-372.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.



