Contents

Issue 43.1

Editor’s Preface

Michel FERLUS
Arem, a Vietic Language. 1-15

Hiram RING
Nominalization in Pnar. 16-23

Elizabeth HALL
Impact of Tai Lue on Muak Sa-aak phonology. 24-30

Rujiwan LAOPHAIROJ
Conceptual metaphors of Vietnamese taste terms. 31-46

Paul SIDWELL
Khmuc classification and homeland. 47-56

Mathias JENNY
Transitivity and affectedness in Mon. 57-71

J. MAYURI, Karumuri V. SUBBARAO, Martin EVERAERT and G. Uma Maheshwar RAO
Some syntactic aspects of lexical anaphors in select Munda Languages. 72-83

Stephen SELF
Another look at serial verb constructions in Khmer. 84-102

V. R. RAJASINGH
Interrogation in Muôt. 103-123

Issue 43.2

Suwilai PREMSIRAT, Kenneth GREGERSON
Fifty Years of Mon-Khmer Studies i-iv

Anh-Thur. T. NGUYÊN
Acoustic correlates of rhythmic structure of Vietnamese narrative speech. 1-7

P. K. Choudhary
Agreement in Ho 8-16


Editors’ Preface

The 5th International Conference on Austroasiatic Linguistics (ICAAL5) was held at the Australian National University (ANU) over September 4-5, 2013. The meeting was run in conjunction with the 19th Annual Himalayan Languages Symposium (HLS19), organised locally by Paul Sidwell and Gwendolyn Hyslop. The meetings were made possible by support provided by the following at ANU:

- Department of Linguistics, College of Asia and the Pacific
- Research School of Asia Pacific
- School of Culture, History and Language
- Tibetan Cultural Area Network

Some 21 papers were read over two days at the ICAAL meeting, nine of which have found their way into this special issue of MKS. To ensure academic standards, all abstracts submitted to the meeting were vetted by the Standing Committee before acceptance and presentation. After the meeting presenters were invited to submit their papers for inclusion in the special issue, and those papers which were received were subject to another round of independent review. The full list of papers read at the meetings follows:

**Plenary: K.S. Nagaraja** “Korku-Khasi, a typological study”

**K.V. Subbarao** “Relative Clauses in Munda Languages with Special Reference to the Comitative PP as Head”

**J.D. Mayuri** “Some syntactic aspects of Lexical Anaphors in Select Munda Languages”

**Indresh Thakur** “Pronominalization in Santhali”

**Supakit Buakaw** “A phonological sketch of Katiang: A new Palaung dialect of northern Shan State”

**Sujaritlak Deepadung, Ampika Rattanapitak, Supakit Buakaw** “Grammar Sketch: Dara-ang Palaung”

**Elizabeth Hall** “Sound changes based on language contact in Muak Sa-aak”

**Mayuree Thawornpat** “A community strategy for language development: The Lavua of Mae Hong Son Province, Thailand”

**Michel Ferlus** “Arem, a Vietic Language”

**Rujiwan Laophairoj** “Conceptual metaphors of Vietnamese taste terms”

**Plenary: Felix Rau** “Approaching the Family from the South-West”

**Hiram Ring** “The Pnar verbal complex”

**Keren Baker** “General phonological features and word patterning in Lyngngam”

**V. R. Rajasingh** “Interrogation in Muöt”

**Kevin Baetscher** “Patterns of Connected Speech in Mlabri”

**Paul Sidwell** “Proto-Khmuic: reconstruction and sub-grouping”

**Meng Vong** “The infix [-āmn-] in Khmer”

**Stephen Self** “Another look at multi-verb constructions in Khmer”

**Mathias Jenny** “Transitive directionals in Mon - form, function and implications for linguistic typology”

**Siripen Ungsitipoonporn** “The impact of Nyahkur revitalization”

**Murshed, Sikder** “Santali Scripts: Selection and debate in Bangladesh”

At the conclusion of ICAAL5 a business meeting was held. To ensure the continuity of the movement the meeting appointed a new Standing Committee (without formal structure of office bearers at this time). The committee members are: George van Driem, Paul Sidwell, Mathias Jenny, Felix Rau, K.S. Nagaraja, Sikder Murshed, Meng Vong, and Michel Ferlus. Details of past and future ICAAL meetings can be found online at http://icaal.org.
Pictured at the ICAAL5 business meeting (from left to right): V. R. Rajasingh, Elizabeth Hall, George van Driem, Paul Sidwell, Sikder Murshed, J. D. Mayuri, Mathias Jenny, Kevin Baetscher, Felix Rau, Stephen Self, Meng Vong, K.S. Nagaraja, Frederic Pain.

Paul Sidwell
Brian Migliazza
MKS 43.1 Editors
December 2014
Arem, a Vietic Language

Michel FERLUS
CNRS, France (retired)
<jrmferlus@orange.fr>

Abstract
The paper reports on the state of knowledge of Arem, a small endangered language of the Vietic branch of Austroasiatic. Special attention is paid to the synchronic and diachronic phonology, which display many remarkable and highly significant features. The paper is richly illustrated with lexical examples.

Keywords: Vietic languages, lexicon, phonology
ISO 639-3 codes: aem, ptk, vie

1. Introduction

The Arem people are located in the Vietnam’s Quảng Bình province, Bố-trạch district, Tân-trạch commune, Cà Ròong hamlet. Though known of neighboring populations, they were confused with Bru khùa (Katuic) by local authorities: after their presence was reported by the military in 1959, they were eventually identified as a separate ethnic and linguistic group by ethnologists.

Demographics: around 1960, there were 53 people, 30 men and 23 women (Vương Hoàng Tuyên 1963: 69). Recently, there were 102 people (Trần Trí Dõi 1999: 110). In the past, Arem lived in rock shelters or makeshift huts. They exchanged products from hunting and gathering with the neighboring populations, Vietnamese, Bru khùa or Lao. The need to find marital partners outside the group has the consequence that only a minority, about a quarter, regularly speak their language. The Arem language is in great danger of extinction.

Ethnonym: the term Arem (Vietnamese spelling of are are are areːm ːm ːm ːm) seems to be an exonym. Those concerned prefer to be named cmbrau [cmrawˀ]. At the end of the 19th century, Captain Rivière reported more than one hundred Harême families, sometimes corrupted as Hémère, on the Lao side (Mission Pavie, Géographie et Voyages, 1902; voc. pp. 285-90). Harême has been wrongly identified to Arem (Haudricourt 1966: 136) as recently revealed by Vương Hoàng Tuyên (1963). An examination of the vocabulary of numbers as preserved in Rivière’s note clearly shows that Harême should be classified in the Maleng subgroup.

<table>
<thead>
<tr>
<th>Harême</th>
<th>Maleng brô</th>
<th>Arem</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘one’</td>
<td>moɪtte</td>
<td>mʊxc</td>
</tr>
<tr>
<td>‘two’</td>
<td>harr</td>
<td>hæːl</td>
</tr>
<tr>
<td>‘three’</td>
<td>pà</td>
<td>pæː</td>
</tr>
<tr>
<td>‘four’</td>
<td>pɔne</td>
<td>pʊnˀ</td>
</tr>
<tr>
<td>‘five’</td>
<td>darm</td>
<td>dam</td>
</tr>
</tbody>
</table>

In fact, Arem is an ethnographic term without any clear linguistic relevance: it was apparently used to refer to various semi-nomadic populations moving in the Great Cordillera, between Vietnam and Laos, about the latitude of the Mụ Giạ Pass. Regarding cmbrau [cmrawˀ], the pre-syllable cm-, which is attested only in this word, suggests that this ethnonym was borrowed from another language.

Researches on Arem: the data that serves as the basis for the present text were collected during three field trips (June 1991, December 1993, and March 1996) conducted in collaboration with Professor Trần Trí Dổi of Vietnam National University in Hanoi. The Arem language is not homogeneous, the pressure of neighboring languages is exerted on the individuals rather than on the group. The speakers are at least bilingual with Vietnamese, standard as well as dialectal, and some people can also speak Khùa or Lao.

Publications and various data on Arem people and language: the first linguistic data were published by Vương Hoàng Tuyên (1963). They are sufficient to classify Arem in the Viet-Muong
group. More recently, Trần Trí Dỗi (1999) cited some vocabulary. It is heartening to mention that Arem, an endangered language, had an academic consecration through the Master of Arts of Atsushi Kasuga (1994), unpublished so far. Consistent ethnographic data should be noticed: Vương Hoàng Tuyên (1963), Đặng Nghiêm Văn, Chu Thái Sơn, Lưu Hùng (1986), and Trần Trí Dỗi (1995, 1999). We chose for our description, the speech of Arem which seemed the most conservative regarding the phonation type register.

![Map of Vietc languages showing approximate locations.](image)

**Figure 1:** Map of Vietc languages showing approximate locations.

### 2. Phonetic system

#### 2.1 Syllabic structure

The vocabulary of Arem has monosyllables and sesqui-syllables.

- **monosyllable:** C V (C)
- **sesqui-syllable:** C-C V (C)

A sesqui-syllable is composed of a main syllable C V (C) preceded by a pre-syllable C-. The main syllable is identical to a monosyllable. The pre-syllable is short and unstressed, without distinctive vowel, and its consonantal system is reduced. A pre-syllable can not exist alone. Arem is the language of the Viet-Muong group that has preserved the highest percentage of sesqui-syllable (prenasalized being considered as initials sesqui-syllables). The percentage of retention in subgroups increases from north to south.

<table>
<thead>
<tr>
<th>Vietnamese</th>
<th>monosyllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muong/mường</td>
<td>monosyllabic</td>
</tr>
<tr>
<td>Thơ</td>
<td>monosyllabic</td>
</tr>
<tr>
<td>Toum-Liha</td>
<td>monosyllabic</td>
</tr>
<tr>
<td>Pong</td>
<td>10% of sesqui-syllabic</td>
</tr>
<tr>
<td>Thavung</td>
<td>35-40%</td>
</tr>
<tr>
<td>Maleng Brô</td>
<td>35-40%</td>
</tr>
<tr>
<td>Sách-Rutc</td>
<td>35-40%</td>
</tr>
<tr>
<td>Arem</td>
<td>55-60%</td>
</tr>
</tbody>
</table>
2.2 Voice-type register system of Arem:

- Arem has two sets of vowels: Series 1 with a high-clear register phonation; Series 2, with a low-breathy register phonation (indicated by the grave accent).

- In each series, we must distinguish the syllables ending in voiced finals (# w j l m n η), and the syllables ending in voiceless finals (-p -t -c -k) and (-h).

- In the syllables with voiced finals, the series 1 is indicated by /a - a̯/, and corresponds to tones **ngang-sắc** in Vietnamese. The series 2 is indicated by /à - à̯/ and corresponds to **huyền-nặng**.

- Final -h is preserved in the both series; /ah/ corresponds to tone **hỏi**, and /âh/ to **ngã**.

- Syllables in /aC/ and /âC/ are checked by voiceless finals (-p -t -c -k); they respectively correspond to tones **sắc** and **nặng**.

- Syllables in voiced finals show a contrast between a modal tone and a glottalized tone: in series 1 /a-a̯/, and in series 2 /â-â̯/. It should be noted that [ˀ] indicates a glottal constriction of the vowel, also named creakiness, but by no means a creaky voice. I think that this phonetic feature is not a voice type register, just a weak glottal stop.

- The combination of the contrast /clear-breathy/ and /unchecked-glottal/ forms a four tones system:

<table>
<thead>
<tr>
<th></th>
<th>voiced finals (# w j l m n η)</th>
<th>unvoiced finals (h) / (p t c k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>series 1 (high-clear)</td>
<td>a</td>
<td>a̯</td>
</tr>
<tr>
<td>series 2 (low-breathy)</td>
<td>à</td>
<td>à̯</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>voiced finals</th>
<th>unvoiced finals</th>
</tr>
</thead>
<tbody>
<tr>
<td>series 1 (tones a á à)</td>
<td>ngang</td>
<td>sác</td>
</tr>
<tr>
<td>series 2 (tones à á à)</td>
<td>huyền</td>
<td>nặng</td>
</tr>
</tbody>
</table>

Because of the phenomenon of vowel splitting in two series, it is difficult to find a same vowel represented with the four tones. Only rhymes with vowel **ıa** provide enough possibilities of combinations.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>tone a :</td>
<td>atʰıəl</td>
<td>‘écaille / fish scale’</td>
</tr>
<tr>
<td>tone a ̯ :</td>
<td>akiəl̯</td>
<td>‘sur le côté / on side’</td>
</tr>
<tr>
<td>tone à :</td>
<td>cıəl</td>
<td>‘filet-épervier / cast net’</td>
</tr>
<tr>
<td>tone à ̯ :</td>
<td>pətiəl̯</td>
<td>‘milieu / middle’</td>
</tr>
</tbody>
</table>

FERLUS, Michel. 2014. Arem, a Vietic Language. Mon-Khmer Studies 43.1:1-15 (ICAAL5 special issue)
2.3 Initial consonants

Simple initials

<table>
<thead>
<tr>
<th>Initials</th>
<th>Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p^h$</td>
<td>$t^h$</td>
</tr>
<tr>
<td>$p$</td>
<td>$t$</td>
</tr>
<tr>
<td>$t^h$</td>
<td>$c$</td>
</tr>
<tr>
<td>$k$</td>
<td>$?</td>
</tr>
<tr>
<td>$b$</td>
<td>$d$</td>
</tr>
<tr>
<td>$n$</td>
<td>$ɲ$</td>
</tr>
<tr>
<td>$ŋ$</td>
<td>$ʃ$</td>
</tr>
<tr>
<td>$s$</td>
<td>$ʃ$</td>
</tr>
<tr>
<td>$v$</td>
<td>$(z)$</td>
</tr>
<tr>
<td>$l$</td>
<td>$r$</td>
</tr>
</tbody>
</table>

Clusters with $r / ɹ$

<table>
<thead>
<tr>
<th>Initials</th>
<th>Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>$pr$</td>
<td>$tr$</td>
</tr>
<tr>
<td>$br$</td>
<td></td>
</tr>
<tr>
<td>$pl$</td>
<td>$tl$</td>
</tr>
<tr>
<td>$kl$</td>
<td></td>
</tr>
</tbody>
</table>

Pre-nasalized initials

<table>
<thead>
<tr>
<th>Initials</th>
<th>Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ⁿp$</td>
<td>$ⁿt$</td>
</tr>
<tr>
<td>$ⁿm$</td>
<td>$ⁿd$</td>
</tr>
<tr>
<td>$ⁿw$</td>
<td>$ⁿl/ⁿd$</td>
</tr>
</tbody>
</table>

2.4 Finals

<table>
<thead>
<tr>
<th>Final</th>
<th>Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$t$</td>
</tr>
<tr>
<td>$c$</td>
<td>$k$</td>
</tr>
<tr>
<td>$m(ʔ)$</td>
<td>$n(ʔ)$</td>
</tr>
<tr>
<td>$w(ʔ)$</td>
<td>$ʃ(ʔ)$</td>
</tr>
<tr>
<td>$j(ʔ)$</td>
<td>$ʃ(ʔ)$</td>
</tr>
</tbody>
</table>

Final liquid -l [ɭɭ] is vocalized in [ɭɭ] by some informants.

2.5 Pre-syllables

Note: infrequent pre-syllables are placed in the brackets:

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-</td>
<td>i-</td>
</tr>
<tr>
<td>pa-</td>
<td>(pi-)</td>
</tr>
<tr>
<td>ta-</td>
<td>(tam)</td>
</tr>
<tr>
<td>ci-</td>
<td>(cm)</td>
</tr>
<tr>
<td>ka-</td>
<td>(ku-)</td>
</tr>
<tr>
<td>la-</td>
<td>(le-)</td>
</tr>
</tbody>
</table>

Pre-syllables in decreasing order, with the number of occurrences in the lexicon:

a- (240), N- (141), ka- (83), u- (79), ci- (54), pa- (44), ta- (36), i- (30), la- (30), ku- (14).
2.6 Vowel system

<table>
<thead>
<tr>
<th>breathy vowels</th>
<th>clear vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>i  i:</td>
<td>i  i:</td>
</tr>
<tr>
<td>è  è:</td>
<td>è  è:</td>
</tr>
<tr>
<td>ie  iə</td>
<td>ie  iə</td>
</tr>
<tr>
<td>ê (é):</td>
<td>ê (é):</td>
</tr>
<tr>
<td>ë  ë:</td>
<td>ë  ë:</td>
</tr>
<tr>
<td>aë  aëːːːː</td>
<td>aë  aëːː</td>
</tr>
<tr>
<td>oë  oëːːːː</td>
<td>oë  oëːː</td>
</tr>
</tbody>
</table>

Units placed between brackets are poorly represented. The vowel è is realized [é]. The difference between (u) and æ:æ has no etymological justification. The series of hight vowels /i ë i .preventei u u:/ occurs only in Vietnamese words or in Arem words whose pronunciation is influenced by Vietnamese pronunciation. This series is not in contrast with the regular series /i e u u:/.

3. Diachrony: identification of some pairs of vowels contrasting by voice type

A pair of vowels contrasting by voice type, in short a registral pair, is composed of two vowels (one in each series) that derived from one single proto vowel. It is particularly useful to identify these pairs for the reconstruction of the vowel system of Arem that existed before the formation of series. We will proceed by comparing Arem to Vietnamese and Proto Viet-Muong. In some cases one will have to distinguish between primary and secondary correspondences.

3.1. PVM æ:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>æ - a/a/æ</td>
</tr>
</tbody>
</table>

primary (regular) correspondences

<table>
<thead>
<tr>
<th>r.ka:</th>
<th>laka:</th>
<th>gâ</th>
<th>‘poulet / chicken’</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.la:</td>
<td>uòi:</td>
<td>li</td>
<td>‘feuille / leaf’</td>
</tr>
<tr>
<td>la:</td>
<td>hael</td>
<td>hai</td>
<td>‘deux / two’</td>
</tr>
<tr>
<td>saim'</td>
<td>t’jem’</td>
<td>lam</td>
<td>‘huit / eighth’</td>
</tr>
<tr>
<td>daik</td>
<td>dêk:</td>
<td>nêu</td>
<td>‘eau / water’</td>
</tr>
<tr>
<td>k.mâ:</td>
<td>miô:</td>
<td>múa</td>
<td>‘pluie / rain’</td>
</tr>
<tr>
<td>uad’</td>
<td>ciel</td>
<td>chái</td>
<td>‘filet-épervier / cast net’</td>
</tr>
<tr>
<td>las</td>
<td>liê:</td>
<td>lô</td>
<td>‘langue / tongue’</td>
</tr>
<tr>
<td>tÎan</td>
<td>tIan’</td>
<td>sian</td>
<td>‘os / bone’</td>
</tr>
</tbody>
</table>

secondary correspondences

<table>
<thead>
<tr>
<th>lac’</th>
<th>liêt</th>
<th>lat</th>
<th>‘lien / bond’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ian)</td>
<td>liê:</td>
<td>lăng</td>
<td>‘village / village’</td>
</tr>
<tr>
<td>gan’</td>
<td>ken’</td>
<td>can</td>
<td>‘peu profond / shallow’</td>
</tr>
<tr>
<td>kr.hauk</td>
<td>kahiêk</td>
<td>khac</td>
<td>‘cracher / to spit’</td>
</tr>
</tbody>
</table>
### PVM a: (before palatals -c -ɲ -j)

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>a:</td>
<td>a:</td>
<td>ñ:</td>
</tr>
<tr>
<td>t.kaç</td>
<td>takac</td>
<td>cá</td>
</tr>
<tr>
<td>lat</td>
<td>lat</td>
<td>dan</td>
</tr>
<tr>
<td>mən</td>
<td>mən</td>
<td>mnəm</td>
</tr>
<tr>
<td>C.ənəj</td>
<td>lənəj</td>
<td>lətəj</td>
</tr>
<tr>
<td>s.nət</td>
<td>nət</td>
<td>nag</td>
</tr>
<tr>
<td>ləj</td>
<td>ləj</td>
<td>ləi</td>
</tr>
</tbody>
</table>

### PVM a: (before -w)

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>a:</td>
<td>ë:</td>
<td>ë:</td>
</tr>
<tr>
<td>cəw'</td>
<td>cəw'</td>
<td>chá</td>
</tr>
<tr>
<td>k.əw</td>
<td>təw</td>
<td>dəo</td>
</tr>
<tr>
<td>ñəw</td>
<td>ñəw</td>
<td>do</td>
</tr>
<tr>
<td>(gəw')</td>
<td>kəw'</td>
<td>cəo</td>
</tr>
<tr>
<td>(əw)</td>
<td>ləw</td>
<td>ləo</td>
</tr>
<tr>
<td>(kəw')</td>
<td>ləw</td>
<td>ləo</td>
</tr>
</tbody>
</table>

### 3.2. PVM ñ:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ñ:</td>
<td>ñ / ñ:</td>
<td>ñ / ñ:</td>
</tr>
<tr>
<td>sɨ</td>
<td>t'ɨ</td>
<td>t'āy</td>
</tr>
<tr>
<td>nət</td>
<td>nət</td>
<td>ːnət</td>
</tr>
<tr>
<td>çin</td>
<td>çin</td>
<td>chin</td>
</tr>
<tr>
<td>šit</td>
<td>ŋt</td>
<td>thit</td>
</tr>
<tr>
<td>di</td>
<td>t:</td>
<td>ñ:</td>
</tr>
<tr>
<td>nət</td>
<td>nət</td>
<td>chit</td>
</tr>
<tr>
<td>çin'</td>
<td>çin'</td>
<td>chin</td>
</tr>
</tbody>
</table>

### 3.3. PVM ə:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə:</td>
<td>ː ñ / ñ:</td>
<td>ñ:</td>
</tr>
<tr>
<td>ːcəp</td>
<td>ːnəp</td>
<td>ñep</td>
</tr>
<tr>
<td>t.ər:</td>
<td>t'ər</td>
<td>chāv</td>
</tr>
<tr>
<td>k.əcet</td>
<td>kət</td>
<td>chēt</td>
</tr>
<tr>
<td>kəcet</td>
<td>kət</td>
<td>gēt</td>
</tr>
<tr>
<td>-səl</td>
<td>at'əl</td>
<td>tē (tē)</td>
</tr>
</tbody>
</table>

In the tense series, we see that PVM ə: is treated as ñ:
3.4. PVM e:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>e'</td>
<td>id</td>
<td>id</td>
</tr>
<tr>
<td>kset</td>
<td>kiet</td>
<td>ghét</td>
</tr>
<tr>
<td>dën</td>
<td>tien</td>
<td>dën</td>
</tr>
</tbody>
</table>

Data on PVM e: are insufficient.

3.5. PVM u:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>w:</td>
<td>aw / oː</td>
<td>ùː</td>
</tr>
<tr>
<td>p.ru:</td>
<td>paraw/</td>
<td>sau</td>
</tr>
<tr>
<td>cu:v</td>
<td>caw:v</td>
<td>châu</td>
</tr>
<tr>
<td>t.uː</td>
<td>tui</td>
<td>rui</td>
</tr>
<tr>
<td>C.huː</td>
<td>pahol</td>
<td>hòi</td>
</tr>
<tr>
<td>k.pur</td>
<td>aprul</td>
<td>yö</td>
</tr>
<tr>
<td>p.tuc</td>
<td>utort</td>
<td>dót</td>
</tr>
<tr>
<td>tuc</td>
<td>tuːc</td>
<td>döt</td>
</tr>
<tr>
<td>t.tum</td>
<td>lum'</td>
<td>trôm</td>
</tr>
<tr>
<td>C.ru:</td>
<td>ciru</td>
<td>sau</td>
</tr>
<tr>
<td>bu:/pu:/</td>
<td>püː</td>
<td>bû</td>
</tr>
<tr>
<td>muc/moc</td>
<td>mûc</td>
<td>mot</td>
</tr>
<tr>
<td>gu/kur</td>
<td>kui'</td>
<td>oô</td>
</tr>
<tr>
<td>C.gu'/C.ku:'</td>
<td>cik'</td>
<td>gau</td>
</tr>
</tbody>
</table>

3.6. PVM o:

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>o:</td>
<td>uɔ</td>
<td>ùɔ</td>
</tr>
<tr>
<td>pøn'</td>
<td>puɔn'</td>
<td>bón</td>
</tr>
<tr>
<td>k.lom'</td>
<td>thuɔn'</td>
<td>trôm</td>
</tr>
<tr>
<td>k.mod'</td>
<td>umuɔl'</td>
<td>mòi</td>
</tr>
<tr>
<td>pøn</td>
<td>&quot;pùøŋ</td>
<td>bông</td>
</tr>
<tr>
<td>C.kol'</td>
<td>akuɔl'</td>
<td>gòi</td>
</tr>
<tr>
<td>k.nol'</td>
<td>tuuɔl'</td>
<td>gòi</td>
</tr>
</tbody>
</table>

In these two examples PVM ū is treated as o:

<table>
<thead>
<tr>
<th>w:</th>
<th>uɔ</th>
<th>ùɔ</th>
<th>ð</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.kud'</td>
<td>lakuɔl'</td>
<td>gòi</td>
<td>‘genou / knee’</td>
</tr>
<tr>
<td>ñuj</td>
<td>ñuɔj</td>
<td>ngòi</td>
<td>‘s’asseoir / to sit down’</td>
</tr>
</tbody>
</table>
### 3.7. PVM ə

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə</td>
<td>a</td>
<td>ə</td>
</tr>
<tr>
<td>ʔə.ə’d</td>
<td>acá.ə’</td>
<td>ʔəo</td>
</tr>
<tr>
<td>ʔə.ə’d’</td>
<td>alá.ə’</td>
<td>tet</td>
</tr>
<tr>
<td>k.ən’</td>
<td>tlán</td>
<td>trə</td>
</tr>
<tr>
<td>kən</td>
<td>kəm</td>
<td>tən</td>
</tr>
<tr>
<td>t.ka’c</td>
<td>ka’c</td>
<td>got</td>
</tr>
<tr>
<td>p.ək</td>
<td>parak</td>
<td>səc</td>
</tr>
<tr>
<td>rəʔ</td>
<td>opəʔ</td>
<td>vət</td>
</tr>
<tr>
<td>ən</td>
<td>rəʔ</td>
<td>ruəl</td>
</tr>
<tr>
<td>əŋ</td>
<td>ruəŋ</td>
<td>‘rizière / ricefield’</td>
</tr>
</tbody>
</table>

In these two examples PVM ə is treated as a

<table>
<thead>
<tr>
<th>a / ə</th>
<th>a</th>
<th>ə</th>
<th>á / a(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.əŋa/əŋa’at</td>
<td>at’əŋa’at</td>
<td>‘étêindre / to extinguish’</td>
<td></td>
</tr>
<tr>
<td>por</td>
<td>pal</td>
<td>‘voler / to fly’</td>
<td></td>
</tr>
</tbody>
</table>

### 3.8. PVM a

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>á</td>
</tr>
<tr>
<td>k.car’</td>
<td>cal’</td>
<td>chay</td>
</tr>
<tr>
<td>C.pat</td>
<td>pat</td>
<td>vat</td>
</tr>
<tr>
<td>k.lam</td>
<td>lam</td>
<td>trəm</td>
</tr>
<tr>
<td>t.ka’n’</td>
<td>ʔən</td>
<td>suŋ</td>
</tr>
<tr>
<td>C.ka’m</td>
<td>ʔəm</td>
<td>gəm</td>
</tr>
<tr>
<td>t.gar’</td>
<td>ʔəŋa’</td>
<td>ngə</td>
</tr>
<tr>
<td>mat</td>
<td>mət</td>
<td>mət</td>
</tr>
<tr>
<td>k.nap’</td>
<td>mən’</td>
<td>mən’</td>
</tr>
</tbody>
</table>

### 3.9. PVM ø

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ø</td>
<td>i / ʔ</td>
<td>i</td>
</tr>
<tr>
<td>p.əŋ</td>
<td>ʔən’</td>
<td>rəŋ</td>
</tr>
<tr>
<td>pọŋ</td>
<td>pən</td>
<td>bəŋ</td>
</tr>
<tr>
<td>pọŋ’</td>
<td>pən’</td>
<td>bəŋ’</td>
</tr>
<tr>
<td>pŋ’</td>
<td>ʔəŋ’</td>
<td>dəŋ</td>
</tr>
<tr>
<td>bŋ’</td>
<td>bəŋ</td>
<td>mət</td>
</tr>
<tr>
<td>m.əŋ’</td>
<td>ʔəŋ’</td>
<td>rəŋ</td>
</tr>
<tr>
<td>m.əŋ’</td>
<td>ʔəŋ’</td>
<td>rəŋ</td>
</tr>
<tr>
<td>t.kəŋ</td>
<td>kəŋ</td>
<td>gən</td>
</tr>
<tr>
<td>k.əŋ’</td>
<td>kəŋ’</td>
<td>qət</td>
</tr>
<tr>
<td>k.əŋ’</td>
<td>kəŋ’</td>
<td>jəl’</td>
</tr>
<tr>
<td>jən’</td>
<td>jən’</td>
<td>dəj’</td>
</tr>
</tbody>
</table>

FERLUS, Michel. 2014. Arem, a Vietic Language.
Mon-Khmer Studies 43.1:1-15 (ICAAL5 special issue)
3.10. PVM u

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>ʊ</td>
<td>ʊ / ŋ</td>
</tr>
<tr>
<td>suk</td>
<td>ur’uk</td>
<td>tóc</td>
</tr>
<tr>
<td>s.?uh</td>
<td>noh</td>
<td>ŋ, tō</td>
</tr>
</tbody>
</table>

‘cheveux / hair’

‘nid / nest’

Not that data for PVM u are insufficient.

4. The influence of a Vietnamese dialect on Arem

One can note the unexpected correspondence between pairs with tones hỏi-ngã in Viet, and pairs of type a²-a² in Arem. If these correspondences were regular, Arem words should have the final -h.

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>-h</td>
<td>tone a¹</td>
<td>tone a²</td>
</tr>
<tr>
<td>bòc</td>
<td>bô’c</td>
<td>bô’c</td>
</tr>
<tr>
<td>käs</td>
<td>kô’c</td>
<td>kô’c</td>
</tr>
<tr>
<td>t’ag</td>
<td>t’ag</td>
<td>t’ag</td>
</tr>
<tr>
<td>ngã</td>
<td>ngã</td>
<td>ngã</td>
</tr>
<tr>
<td>hoń</td>
<td>hoń</td>
<td>hoń</td>
</tr>
<tr>
<td>k.çeh</td>
<td>kâçeh</td>
<td>kâçeh</td>
</tr>
<tr>
<td>k.pas</td>
<td>kîpê</td>
<td>kîpê</td>
</tr>
<tr>
<td>k.çús</td>
<td>kûçús</td>
<td>kûçús</td>
</tr>
<tr>
<td>k.çuh</td>
<td>kûçuh</td>
<td>kûçuh</td>
</tr>
</tbody>
</table>

Arem words that show these unexpected correspondences are probably borrowings from a local Vietnamese dialect, in which the expected tones corresponding regularly to tones hỏi-ngã have merged to tones sãc-nãng according to two modalities (see tables below). In type A, the expected tone hỏi merged in the tone sãc, while the expected tone ngã merged in tone nãng. This is the dialect to which Arem borrowed a part of his vocabulary. Note in passing, the existence of another dialect, the type B, in which the expected tones hỏi-ngã merged in the tone nãng.

<table>
<thead>
<tr>
<th>type A</th>
<th>type B</th>
</tr>
</thead>
<tbody>
<tr>
<td>ngang</td>
<td>sãc  (hóï)</td>
</tr>
<tr>
<td>huyén</td>
<td>nãng  (ngã)</td>
</tr>
<tr>
<td>ngang</td>
<td>sãc  (hóï)</td>
</tr>
<tr>
<td>huyén</td>
<td>nãng  (ngã)</td>
</tr>
</tbody>
</table>

The existence of type A was revealed to me by the linguist Võ Xuân Trang in 1991. Type B is the dialect of Cao Lao Hä (Ferlus 1995).

5. Correspondences of initial consonants

In order to present the comparative evolutions of initials in Arem and Vietnamese, we have classified PVM initials into five categories:

- Obstruents */p-b t-d c-j k-g/ and */ɡ/ which have undergone a double processing in Vietnamese according to their position within the word when the change occurs (Ferlus 1982, 1992), as initial of monosyllable, or as medial in sesquisyllable (see table 5.1.). Remember that in the 17th century, v was ō/ ʁe, and d was d/ dë (de Rhodes 1651, 1991).
5.1. Table of the evolutions of PVM *p-b t-d c-ʃ k-ɡ/ and *s/s/

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Vietnamese</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>monosyll.</em></td>
<td><em>sesqui-syllable</em></td>
<td></td>
</tr>
<tr>
<td>p - b</td>
<td>p</td>
<td>b [ɓ] v [v] &lt; b/bɛ [ɓ]</td>
</tr>
<tr>
<td>t - d</td>
<td>t</td>
<td>d [ɗ] d/z/i [ɗ] &lt; d/dɛ [ɗ]</td>
</tr>
<tr>
<td>c - ʃ</td>
<td>c</td>
<td>ch [ʃ] gi [z/i] &lt; ɡ/ɡi [ʃ/i]</td>
</tr>
<tr>
<td>k - ɡ</td>
<td>k</td>
<td>c/k [k] g/ɡh [ɣ] &lt; ɡ/ɡh [ɣ]</td>
</tr>
<tr>
<td>s</td>
<td>ɪ ɹ</td>
<td>t [t] r [r/z/i] &lt; r [r]</td>
</tr>
</tbody>
</table>

Labials p-b

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Vietnamese</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>p-b</em></td>
<td>p</td>
<td>b v</td>
</tr>
<tr>
<td>par</td>
<td>par</td>
<td>ba</td>
</tr>
<tr>
<td>pʊn’</td>
<td>pʊn’</td>
<td>bʊn</td>
</tr>
<tr>
<td>k.pas</td>
<td>k.pas</td>
<td>bʊn</td>
</tr>
<tr>
<td>k.pəɾ</td>
<td>k.pəɾ</td>
<td>bʊŋ</td>
</tr>
<tr>
<td>k.pəɾ’</td>
<td>k.pəɾ’</td>
<td>vʊŋ</td>
</tr>
</tbody>
</table>

Alveolars t-d

<table>
<thead>
<tr>
<th>t-d</th>
<th>t</th>
<th>d</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>tuc</td>
<td>tuc</td>
<td>dût</td>
<td>‘allumer / to light’</td>
</tr>
<tr>
<td>tag</td>
<td>tag</td>
<td>dɑn</td>
<td>‘tisser / to weave’</td>
</tr>
<tr>
<td>k.ta’</td>
<td>k.ta’</td>
<td>dăi</td>
<td>‘pierre / stone’</td>
</tr>
<tr>
<td>k.tæ</td>
<td>k.tæ</td>
<td>dăi</td>
<td>‘scrotum / scrotum’</td>
</tr>
<tr>
<td>k.ta’</td>
<td>k.ta’</td>
<td>dăi</td>
<td>‘cloison / partition’</td>
</tr>
<tr>
<td>k.ta’</td>
<td>k.ta’</td>
<td>dăi</td>
<td>‘lait / milk’</td>
</tr>
</tbody>
</table>

Palatals c-ʃ

<table>
<thead>
<tr>
<th>c-ʃ</th>
<th>c</th>
<th>ch</th>
<th>gi</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ɛɾm</td>
<td>ɛɾm</td>
<td>ɛɾm</td>
<td>‘oiseau / bird’</td>
</tr>
<tr>
<td>k.ɛɾm</td>
<td>k.ɛɾm</td>
<td>tʃʰi</td>
<td>‘incendie / fire’</td>
</tr>
<tr>
<td>ɛɾl</td>
<td>ɛɾl</td>
<td>ɛɾl</td>
<td>‘filet-épervier / cast net’</td>
</tr>
<tr>
<td>k.ɛɾl’</td>
<td>k.ɛɾl’</td>
<td>giɛŋ</td>
<td>‘emprisonner / to imprison’</td>
</tr>
<tr>
<td>k.ɛɾl’</td>
<td>k.ɛɾl’</td>
<td>giɛŋ</td>
<td>‘papier / paper’</td>
</tr>
</tbody>
</table>
### Velars k-g

<table>
<thead>
<tr>
<th>k-g</th>
<th>k</th>
<th>c / k</th>
<th>g / gh</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.kac</td>
<td>takac</td>
<td>cât</td>
<td>'sable / sand'</td>
</tr>
<tr>
<td>kon</td>
<td>kan</td>
<td>con</td>
<td>‘fils, fille / son, daughter’</td>
</tr>
<tr>
<td>ke</td>
<td>ke</td>
<td>câ</td>
<td>‘féminin / feminin’</td>
</tr>
<tr>
<td>r.ka</td>
<td>laka</td>
<td>gà</td>
<td>‘poulet / chicken’</td>
</tr>
<tr>
<td>c.kàt</td>
<td>cikàt</td>
<td>gang</td>
<td>‘empan / handsanp’</td>
</tr>
<tr>
<td>c.gr</td>
<td>cikgr</td>
<td>gày</td>
<td>‘bâton, canne / walking stick’</td>
</tr>
</tbody>
</table>

### Fricative alveolar s

<table>
<thead>
<tr>
<th>s</th>
<th>tʰ</th>
<th>t</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>sam’</td>
<td>tʰəm’</td>
<td>tám</td>
<td>‘huit / eight’</td>
</tr>
<tr>
<td>ñai</td>
<td>ñai</td>
<td>tay</td>
<td>‘oreille / ear’</td>
</tr>
<tr>
<td>p.ßaŋ’</td>
<td>ðaŋ’</td>
<td>rán</td>
<td>‘serpent / snake’</td>
</tr>
<tr>
<td>p.se’</td>
<td>ðe’</td>
<td>rài</td>
<td>‘loutre / otter’</td>
</tr>
<tr>
<td>k.ßaŋ</td>
<td>at’ang</td>
<td>ràng</td>
<td>‘dent / teeth’</td>
</tr>
</tbody>
</table>

### 5.2. Pre-glottalized ð ð f

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>ð</td>
<td>b</td>
<td>m</td>
</tr>
<tr>
<td>c.ðæc</td>
<td>cibæc</td>
<td>mút</td>
</tr>
<tr>
<td>båh</td>
<td>abåh</td>
<td>må</td>
</tr>
<tr>
<td>båj’</td>
<td>båj</td>
<td>muoi</td>
</tr>
<tr>
<td>bûk</td>
<td>bûk</td>
<td>múc</td>
</tr>
<tr>
<td>t.ßæn</td>
<td>abåñ</td>
<td>mäng</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ð</th>
<th>d</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>dæk</td>
<td>dæk</td>
<td>nươc</td>
</tr>
<tr>
<td>dâm</td>
<td>dâm</td>
<td>nâm</td>
</tr>
<tr>
<td>dìp</td>
<td>dìp</td>
<td>nèp</td>
</tr>
<tr>
<td>k.ðæc</td>
<td>kadc</td>
<td>nài</td>
</tr>
<tr>
<td>ðæ</td>
<td>ðæ</td>
<td>nhù</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ð</th>
<th>t</th>
<th>nh</th>
</tr>
</thead>
<tbody>
<tr>
<td>fìuk</td>
<td>fìk</td>
<td>nhùc</td>
</tr>
<tr>
<td>fëh</td>
<td>fë</td>
<td>nhô</td>
</tr>
</tbody>
</table>
5.3. Evolutions of others initials

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th></th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>#</td>
<td>[?]</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>h</td>
<td>[h]</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
<td>[m]</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>[n]</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>nh</td>
<td>[ɲ]</td>
</tr>
<tr>
<td>ng</td>
<td>ng</td>
<td>ng</td>
<td>[ŋ]</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>v</td>
<td>[v]</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>d</td>
<td>[z̃]</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>[r/z̃]</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>i</td>
<td>[i]</td>
</tr>
</tbody>
</table>

The Arem initials tʃ and ʃ are only interpretations of Vietnamese phonemes.

5.4. Clusters with -r-

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.r-</td>
<td>C.r/-r-</td>
<td>s-</td>
</tr>
<tr>
<td>g.ray′</td>
<td>ciray′</td>
<td>sōi</td>
</tr>
<tr>
<td>p.razk</td>
<td>parazk</td>
<td>sōc</td>
</tr>
<tr>
<td>k.rar′</td>
<td>karar′</td>
<td>sōi</td>
</tr>
<tr>
<td>s.ran′</td>
<td>ran′</td>
<td>sān</td>
</tr>
<tr>
<td>c.ru:</td>
<td>ciru:</td>
<td>sāu</td>
</tr>
<tr>
<td>p.ru:</td>
<td>paru:</td>
<td>sāu</td>
</tr>
<tr>
<td>p.rı:</td>
<td>parı:</td>
<td>say</td>
</tr>
<tr>
<td>c.rarı</td>
<td>“raı</td>
<td>suōn</td>
</tr>
</tbody>
</table>

5.5. Clusters with -l-

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.l-</td>
<td>L/-l/-Cl-</td>
<td>tr-</td>
</tr>
<tr>
<td>...</td>
<td>lizm</td>
<td>tröm</td>
</tr>
<tr>
<td>k.lon</td>
<td>ulan</td>
<td>trān</td>
</tr>
<tr>
<td>p.le:</td>
<td>ule:</td>
<td>trai</td>
</tr>
<tr>
<td>k.lu:</td>
<td>cilu:</td>
<td>trāa</td>
</tr>
<tr>
<td>p.lep</td>
<td>ıp</td>
<td>tranh</td>
</tr>
<tr>
<td>b.laj</td>
<td>tıej</td>
<td>trōi</td>
</tr>
<tr>
<td>k.laj′</td>
<td>tlaj′</td>
<td>trāi</td>
</tr>
<tr>
<td>k.lam</td>
<td>tlam</td>
<td>trām</td>
</tr>
<tr>
<td>k.lam′</td>
<td>tlam′</td>
<td>trōn</td>
</tr>
</tbody>
</table>
6. Finals consonants

6.1. Table of the evolution of final consonants

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>ng(h)</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>ng(h)</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>i/y</td>
</tr>
</tbody>
</table>

6.2. Final palatals PVM c p

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.kac</td>
<td>takac</td>
<td>cát</td>
</tr>
<tr>
<td>c.buc</td>
<td>cibuc</td>
<td>mút</td>
</tr>
<tr>
<td>n.snp</td>
<td>n&quot;ng&quot;</td>
<td>rân</td>
</tr>
<tr>
<td>m.rnp</td>
<td>mi&quot;ng&quot;</td>
<td>rân</td>
</tr>
</tbody>
</table>

6.3. Final fricatives PVM h s

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>t.pah</td>
<td>&quot;pah</td>
<td>vá</td>
</tr>
<tr>
<td>c.ah</td>
<td>aceh</td>
<td>chè</td>
</tr>
<tr>
<td>k.ah</td>
<td>abah</td>
<td>mía</td>
</tr>
<tr>
<td>l.as</td>
<td>liah</td>
<td>lụtụ</td>
</tr>
<tr>
<td>C.kas</td>
<td>akah</td>
<td>gài</td>
</tr>
</tbody>
</table>

6.4. Final approximants PVM r l

The PVM r-l distinction is preserved only in some dialects in the Maleng group. Comparison of Arem and Vietnamese can not reconstruct this distinction.

<table>
<thead>
<tr>
<th>PVM</th>
<th>Arem</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>par</td>
<td>pal</td>
<td>báv</td>
</tr>
<tr>
<td>har</td>
<td>hael</td>
<td>hài</td>
</tr>
<tr>
<td>t.kar</td>
<td>ikal</td>
<td>gày</td>
</tr>
<tr>
<td>pal</td>
<td>cial</td>
<td>chái</td>
</tr>
<tr>
<td>k.mod</td>
<td>umial</td>
<td>mói</td>
</tr>
</tbody>
</table>

FERLUS, Michel. 2014. Arem, a Vietic Language. Mon-Khmer Studies 43.1:1-15 (ICAAL5 special issue)
7. Brief conclusions

Despite its great interest for comparatism and reconstruction of Viet-Muong, the Arem has the disadvantage of being not a more homogeneous language. In addition to its own lexical funds, it contains layers of borrowings from standard Vietnamese and from a local dialect yet to be described. One can also detect borrowings from Katuic, Lao and Cham.

Not withstanding these problems and challenges, it is important to present a substantial study to the community of specialists, but we must recognize that there remains important lexical work to do.

References


Trần Trí Đôi. 1995. *Thực trạng kinh tế và văn hóa của ba nhóm người đang có nguy cơ bị biến mất* [Actual state of the economy and culture of three endangered ethnic groups]. Nhà xuất bản văn hóa dân tộc. [Arem: population pp. 71-76; no vocabulary]


Nominalization in Pnar

Hiram RING
NTU, Singapore
<hiram1@e.ntu.edu.sg>

Abstract

Pnar, an Austroasiatic (AA) language located in the state of Meghalaya in northeast India, is typologically interesting because of its range of nominalization strategies. These include derivational verbal morphology, pronominal gendered noun-class clitics that derive nouns when they attach to verbs, and a relativizer. The relativizer *wa* has intriguing similarities to the function of some nominalizers in nearby Tibeto-Burman languages (as identified by Matisoff (1972) for Lahu, and for other languages in the area by Noonan (1997); Bickel (1999); Watters (2008) among others). Unlike most TB languages, however, this relativizer is pre-verbal, more similar to constructions in other AA languages. Similar in form is the *wa* ‘with/and’ comitative coordinator that also occurs in Pnar. In this paper I review the morphemes, their syntax and interaction.

Keywords: nominalization, grammatical gender, derivation, relativisation

ISO 639-3 codes: pbv

1. Introduction

Pnar is spoken by about 400,000 people in the eastern West Jaintia Hills and East Jaintia Hills districts of Meghalaya, a state in northeast India. The district seat of Jowai in West Jaintia Hills district is acknowledged by speakers as the standard, and is thus the focus of my forthcoming grammatical description and the main source of my data on Pnar. Pnar uses a roman-based script where characters are for the most part phonemic. This script is used throughout the examples in this paper, with an additional line of IPA characters included for reference. In Pnar script, the digraph *ch* represents the affricate /ʧ/, *j* represents /ʤ/, *ṅ* the palatal nasal, and *ŋ* the velar nasal. Diphthongs ending in *i* identify the following *t* or *d* as laminal-dental (/t̪/, /d̪/), though there is no marking of syllable-initial laminal-dental sounds. Orthographic *h* following a voiceless consonant represents affrication (so *ph*, *th*, *kh* represent /pʰ/, /tʰ/, /kʰ/ respectively), while word-finally it represents the glottal stop /ʔ/ (soh /sɔʔ/ ‘fruit’), and in all other contexts represents the glottal fricative /h/. Orthographic *γ* also represents three different sounds depending on where it occurs: as the single onset of a syllable it represents the palatal approximant /j/ (γap /jap/ ‘die’), when occurring as the second constituent of the onset it represents the glottal stop (pyut /pʔut/ ‘rot’), and when occurring immediately before a nasal/trill/lateral it marks the following sound as syllabic (so *yn* represents /n̩/ ‘REF’ and *pyn* represents /pn̩/ ‘CAUS’). To distinguish between phonemic vowels, orthographic *o* represents /ɔ/, *oo* represents /o/, and the character *æ* represents /ɛ/. Other consonant and vowel characters represent their corresponding IPA symbols.

2. Defining nominalization

According to the most recent volume on nominalization in Asian languages (Yap et al., 2011), nominalization is “the process by which we derive nominal expressions” (p. 3), a definition used originally by Comrie and Thompson (1985). Some examples from English of nominalization processes are:

- destroy -> destruction (action nominal, morphological derivation)
- treat -> treatment (action nominal, morphological derivation)
- teach -> teacher (agentive nominal, morphological derivation)
- he works [v.] -> the work [n.] (event/action nominal, lexical or syntactic derivation)

The link between nominalization, relativization, and genitivization was first identified and described in Tibeto-Burman languages by Matisoff (1972), who was followed by other TB
researchers (Noonan, 1997; Bickel, 1999; Watters, 2008, to name a few). This link is also evident in other southeast Asian languages such as Burmese (Hopple, 2003). Nominalization in Austroasiatic languages have been less widely studied, though some discussion of nominalization exists for Semelai (Kruspe, 2004), Jahai (Burenhult, 2005) and Temiar (Benjamin, 1976), and in papers by Morev (2006); Parkin (1991); Costello and Khamluan (1998); Bradley (1980). A clearer typology of nominalization has been proposed by Gerner (2012), based on a review of the volume by Yap et al..

Yap et al. (2011) suggest a typology of nominalization divided into three general semantic classes or types: participant vs. event nominalization, lexical vs. clausal nominalization, and embedded vs. non-embedded nominalization. Each of these classes can be realized in Asian languages through either morphological or syntactic means. Gerner (2012) proposes a similar typology of nominalization divided somewhat differently and based on 1) morphology, 2) syntax, 3) semantics, 4) pragmatics, and 5) diachrony. Morphological processes can be further subdivided into unmarked or zero-marking, and morphological marking on the verbal or nominal complex. Gerner states: “Markers in the verbal domain are dedicated nominalizers. Markers in the nominal domain are nominalizers whose main function is to mark syntactic cases, possession, specificity and so forth” (Gerner, 2012: 804-805). Syntactic processes are where nominalization “constrains the verb phrase” (816) or “the nominalized expression assumes a syntactic function in the main clause” (816). In terms of semantics, nominalization encodes participants of the verbal expression, its nonphysical properties, or the situation it denotes (824). Pragmatic uses of nominalization can be contrastive focus, tense and aspect, modality, evidentiality, and attitudes of speakers (829). Diachronically, nominalization forms derive from forms with other functions and can make way for new functions or meanings (833). This typology allows for a clearer analysis of individual languages and their particular organization of nominalization processes, and will be used as a guide for the following description of Pnar nominalization processes. In particular, this paper will focus on morphological and syntactic processes of nominalization.

Table 1: Nominalization in Tibeto-Burman languages and in Pnar

<table>
<thead>
<tr>
<th>TB nominalization, V-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivational: [V-NMZ]NOUN</td>
</tr>
<tr>
<td>[V-NMZ]ADJ</td>
</tr>
<tr>
<td>Clausal:</td>
</tr>
<tr>
<td>[(NP)... V-NMZ]NP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pnar nominalization, V-initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivational: [NMZ-V]NOUN</td>
</tr>
<tr>
<td>[NMZ V]ADJ</td>
</tr>
<tr>
<td>Clausal:</td>
</tr>
<tr>
<td>[NMZ V... (NP)]NP</td>
</tr>
</tbody>
</table>

The data provided in the volume by Yap et al. is extremely useful to scholars of southeast Asian languages, particularly in regards to potential language contact patterns. For example, Genetti’s chapter in the volume looks at Tibeto-Burman languages, describing two basic nominalization processes and their formal properties (morphological and syntactic). Considering that speakers of Pnar share a fluid state border with several TB languages, the comparison of nominalization forms is striking. A table summarizing these processes is reproduced as Table (1), which also includes similar processes in Pnar. As can be seen here, nominalization in both TB languages and Pnar are remarkably similar functionally, with the word order (verb final in TB languages, verb-initial in Pnar) resulting in almost mirror-image formal realization.

The following sections detail morphological and syntactic nominalization processes in Pnar.
3. Pnar Nominalizers

As noted above, in Pnar morphology there are both derivational nominalizers and a clausal nominalizer. The derivational nominalizers include prefixes (§2.2.1) and pre-verbal clitics (§2.2.2). Prefixes derive full nouns, while pre-verbal clitics derive non-finite states, resultatives, action nominals and property concepts (equated here to adjectives, as they modify nouns). The clausal nominalizer wa (2.2.3) has the same form as the morpheme that derives property concepts and similarly precedes the head [modified clause]. Derivational processes cannot be negated, while clausal processes can be negated.¹

3.1 Verb root prefixes

The verb root prefix jing- is a general nominalizer that prefixes to the verb root. Evidence that this is a prefix is the fact that it is always preceded by a gender clitic, which only attach to nouns.² Example (1) is of a verbless clause where jing- derives the event nominal ‘regarding’ by prefixing to the verb ya-toh ‘have.relations’ (the verb yatoh seems to have lexicalized from ya- ‘BEN’ and toh ‘be, exist’). This is used to set up the following verbal clause where the speaker discusses the things he wants to say about the referent u=woh Lakriah. In example (2) jing- is a patient nominalizer, prefixing to ‘rule, ruling’ in order to allow the verb to function in a referential manner so it can accept the benefactive/dative case-marking of ya.

(1) i=jing-ya-toh u=ni u=woh Lakriah
   i=dʒiŋ-ja-tɔʔ u=ni u=woʔ lakriaʔ
   N=NMZ-BEN-be M=PROX M=elder Lakriah
   ‘regarding this elder Lakriah, ...’ [PP01CSE_070]

(2) daw chim kti noh u ya ka=jing-synchar sa chi-sein
   daw ʧim kti nɔʔ u ja ka=dʒiŋ-sŋfəɾ sa ʧi-sen
   IRR take hand IMM 3S.M BEN F=NMZ-rule once one-each
   ha ka=kti ka yong oo
   ha ka=kti ka jɔŋ o
   LOC F=hand 3S.F GEN 3S.M.TOP
   ‘he will immediately take hold of the ruling once again in his hand’ [BPDJ_032]

The prefix nong- derives agentive nominals in Pnar. This is an extremely productive prefix, similar to the -er suffix in English (play -> player, etc...). In example (3) it derives an agentive meaning from the verb pyllai ‘organize’ (a verb which seems to have lexicalized from pyn- ‘CAUS’ and lai ‘go’). As a prefix, this form must also be preceded by a gender clitic, similar to jing-.

(3) ka=aɲ wa da chna ki=nong-pyllai
   ka=ŋaŋ wa da ʧna ki=nɔŋ-pɬlaj
   F=rule NMZ REAL make PL=AG.NMZ-organize
   ‘the rule that was made by the organizers’ [AIJ_042]

The prefix yu- derives instrumental nominals, i.e.: ‘thing used for V-ing’. This morpheme is no longer very productive in Jowai-Pnar, being replaced by the more general nominalizer jing-. However, it can still be found in some older Pnar words, such as the word for yu-spong ‘turban’, which is still the required head covering for priests in the traditional religion, or yu-slein ‘loincloth’, again a traditional item of clothing. Speakers also reported that it could also be used for tools and implements, though most often by villagers coming to Jowai for market, or if a speaker couldn’t remember the word for a particular thing. In example (4) the morpheme yu- modifies the verb spong ‘wrap’, again being preceded by a gender clitic.

¹ This, along with word-order restrictions, is criteria for distinguishing a class of ‘adjective’ in Pnar, an issue that will not be discussed here in detail.
² Out of all my texts [172 tokens of jing-] this morpheme occurs without a preceding gender clitic only once, which may simply be an error in production.
3.2 Pre-verbal clitics

Pre-verbal derivational clitics consist of the three gender clitics (ungendered plural clitic ki= is not found to derive nominals) and the nominalizer wa. The former three clitics have the same form as those required for nouns and serve nominalizing functions when they (optionally) attach to the verb root. With the exception of u=, these clitics cannot combine with aspect or mood morphemes.

The gender clitic u derives a non-finite state when cliticized to a verb root. Example (5) shows how the verbs ‘farm’ and ‘work’ become stative when pre-cliticized by u. In example (6) u cliticizes to pyn-yap ‘kill’, following the declarative matrix verb hoi hi ‘be.ok DEC’ whose A-argument is the pronoun i ‘1PL’. Here, u=pyn-yap serves a nominal referential function as a non-finite state.

(5) biang i=pynthor u=rep u=khih
   biaŋ i=pŋʰɔɾ u=rep u=kʰiʔ
   enough N=farmland NF=farm NF=work
   ‘enough farmland to farm, to work’ [PP04SKO_044]

(6) hoi hi u=pyn-yap i ki
   hoj hi u=pp-jap i ki
   fitting DEC NF=CAUS-die 1PL 3PL
   ‘it is ok for us to kill them (animals).’ [BMPJ_036]

The gender clitic ka derives a resultative nominal when cliticized to a verb root. For comparison we have the following two examples. In example (7) the verb khih ‘work’ is used in a question, with the S-argument phi ‘2PL’ topicalized in pre-verbal position as well as given in the standard immediate post-verbal position. In example (8) from the same conversation, the verb khih is being used in referential function, and is describing an abstract notion that has actualized, i.e. the result of work.

(7) tæ phi khih phi, nong?
   te pʰi kʰiʔ pʰi nɔŋ
   NVIS 2PL work 2PL CONF
   ‘so you, you work, right?’ [AIJ_012]

(8) he-i=jooh i=pɔr man ko ka=khih
   he-i=dʒoʔ i=pɔr man ko ka=kʰiʔ
   LOC-N=same N=time happen 3S.F.TOP RES=work
   ‘at the same time it is work’ [AIJ_072]

The neutral gender clitic i derives an abstract action nominal when pre-posed to a verb. In example (9), from the same conversation as (7) above, the clitic attaches to khih ‘work’. Unlike in example (8), which refers to a resulting state, here the speaker is referring to a situation or event which is ongoing. Example (10) is similar - the clitic i attaches here to the verb bam ‘eat’ in an idiomatic expression. As compared to the function of u, this ‘eating of betel nut’ is not a non-finite state, but is rather an ongoing referential state - an abstract action with no clear grounding in actualization.

(9) i=ni hæh i=khih yong i
    i=ni hɛʔ i=kʰiʔ jɔŋ i
    N=PROX only N=work GEN 1PL
    ‘this is our only work’ [AIJ_013]
(10) **myntu da dæp u=yap tæ ong. i=bam kwai ha**

**myntu da dæp u=jap tæ ong. i=bam kwaj ha**

now REAL CPL NF=die NONV say N=eat betel.nut LOC

dwar **u=blai**
dwar **u=blaj**
door M=god

’after death it is said: ‘people eat kwai in heaven’ (lit. eat betelnut at door of God)’

[PP12BL_008]

While the clitic **u** derives non-finite states, it should also be clear that this morpheme can encode intentionality or certainty. When speakers refer to what in English is understood as future time (after now, tomorrow, etc.) they use the irrealis marker **daw**. However, **daw** (11a) also indicates uncertainty, which follows from its status as a marker of events or processes that are not actualized. When speakers want to indicate certainty regarding the actualization of an event, they use the non-finite marker **u** instead (11b).

(11a) **daw khærh i mynstæp**

**daw kʰrɛʔ i mn ̩ stɛp**
IRR prepare 1PL tomorrow

’we will prepare tomorrow’ (intention, uncertain)

(11b) **u=khærh i mynstæp**

**u=kʰrɛʔ i mn ̩ stɛp**
NF=prepare 1PL tomorrow

’we will prepare tomorrow’ (certainly)

The nominalizer **wa** generally acts as a clitic when it precedes verbs, however the degree to which it cliticizes tends to vary from speaker to speaker, and I therefore write it as a separate word. This morpheme derives property concepts from verbs, and the resulting construction follows the noun that it modifies. For example, in (12a) the verb **mane** ‘worship’ is serving as a transitive verb, while in (12b) when **wa** is pre-posed, **mane** is acting as a property modifier for **ki**, the worshippers.

(12a) **mane ki ka na ki=paid**

**ma.ne ki ka na ki=pad ̪**
worship 3PL 3S.F ABL PL=people

‘they worship her, the people..' [AIJ_161]

(12b) **ki wa mane**

**ki wa mane**
3PL NMZ worship

‘worshippers’ [AIJ_159]

Example (13a) is similar to (12b), though in this case the post-posed **ka** causes the resulting expression to be interpreted as a nominal genitive, of which **ka** is the possessor. This is typical of possessor constructions in Pnar, though often the possessor is case-marked by **yong**. Speakers said example (13a) could easily be said as (13b) with **yong** clearly marking **ka** as the possessor, but that (13a) is perfectly clear.

(13a) **ki wa mane ka**

**ki wa mane ka**
3PL NMZ worship 3S.F

‘her worshipers’ [AIJ_159]

---

3 That this is a genitive expression is clear from the pronominal form - were the 3S.F referent the A- argument of the verb **mane**, it would take the form **ko**. As it is the possessor, it takes the form **ka**.
In examples (13a-b) above, a possible analysis is that the pronoun ki is actually cliticizing to wa (which is potentially cliticized to mane), deriving a full noun of property:ki=wa=mane. However, this disregards the reference tracking function of the pronoun, which here is referring to ki=pai’d ‘the people’ of example (12a) above. Perhaps a better translation of example (12b) above is ‘those who worship’. As will be shown below, wa is serving a relativization function in (12b) similar to the function of English ‘that’ or ‘who’.

3.3 Relativizer, coordinator

The same morpheme wa that preposes verbs to form property concepts is used to mark complete clauses. The only distinction between the two morphemes involves associated morphemes. The morpheme can be preceded by a pronoun which acts as the head of the relative clause and is referential with the gender clitic attached to the full nominal head (14-15 and above). When this is the case, the verb being relativized can be negated. Alternatively, it can simply modify the noun directly (16a), in which case the verb form (in this case maya ‘love’) cannot be negated (thus serving as a derived adjective). Example (16b) illustrates how wa can relativize a full clause.

(14) ym toh ka wa bha
NEG be 3S.F NMZ be.good
‘it isn’t good’ [BMPJ_037]

(15) ki=tæ ki wa lai skur
ki=kæ ki wa laj skur
PL=NVIS 3PL NMZ go school
‘the school-goers’ [BPVM_007]

(16a) tæ kam-tæ ki=lok wa maya
te kam-te ki=lok wa maja
NVIS like-NVIS PL=friend NMZ love
‘so in that case dear/beloved friends...’ [BPDJ_044]

(16b) ha-dein wa da am ka=kur soo kpoθ...
LOC-back NMZ REAL exist F=clan four womb...
‘after the Soo Kpoh clan came into being...’ [PP05KO_001]

Examples (17 a-b) illustrate the relativization function of wa further. Both sentences are taken from a conversation regarding why another village celebrates a certain traditional festival on a different day. After a question about whether the other village has the correct date, the traditional priest being interviewed is asserting that the other village has made a mistake (17a) and follows that statement with an explanation (17b). In example (17a) wa serves to relativize the verbal construction bakla ki, which can also be translated here as a genitive. In (17b) both wa morphemes relativize clauses: one relativizes the happening (pyn-man ye-i=tu ‘for that (thing) caused to happen’) and one relativizes the lack of knowledge (ym tip ki u=keiñ ‘they don’t know to count’). Here toh acts as an equative auxiliary (copula), indicating that both wa-clauses are subordinated to

---

4 In the case of (14) the nominal being referred to by ka is outside this utterance and is understood from context.
ki, which acts as the A/S argument of both clauses and refers to ‘them’ (the ones who made the mistake).

(17a) ym toh, wa [bakla ki]
m toʔ wa bakla ki
NEG be NMZ make.mistake 3PL
‘(it’s) not, that’s their mistake’ (or ‘they make.mistake’) [PP09MW_027]

(17b) man ki wa [pyn-man ye-i=tu] tæ toh wa
man ki wa pny-man je-i=tu te toʔ wa
happen 3PL NMZ CAUS-happen DAT-N=MDIST NVIS be NMZ
[ym tip ki u=ke旖]
m tip ki u=ke旖
NEG know 3PL NF=count
‘the reason they make that (mistake) is that they don’t know how to count’ [PP09MW_028]

The morpheme wa has a homophone which functions as a noun phrase coordinator, conjoining phrases similarly to English ‘and’ or ‘with’. This use can be syntactically identified by the fact that it occurs between the two nouns it is coordinating (18a) rather than being pre-posed to a verb (as seen above and in 18b). As comitative ‘with’, it offers insight into the potential origins of the relativizer/nominalizer: property concepts could easily be interpreted as ‘NP with V[property]’.

(18a) æm ar ngut chi lok ki u=Nik wa u=Singh.
em ar nt u=nik wa u=siŋ
have two CL.HUM set friend 3PL M=Nik CONJ M=Singh
‘there were (lived) two friends (a friend set), Nik and Singh’ [KP_002]

(18b) he-i=tæ toh u=Nik toh [u [wa malik]] [u [wa yoh]]
he-i=tæ toh u=Nik toh [u wa malik] u wa yoh
LOC-N=NVIS be M=Nik be 3S.M NMZ be.boss 3S.M NMZ get
[u [wa æm]], tæ u=Singh toh [u [wa dooh]], ym toh
u wa æm te u=siŋ toʔ u wa doʔ m toʔ
3S.M NMZ have NVIS M=Singh be 3S.M NMZ be.poor NEG be
[u [wa yoh]]
3S.M NMZ get
‘then it is that Nik is a boss and has many things, while Singh is poor and doesn’t have much’ [KP_004]

4. Conclusion

To summarize, Pnar morpho-syntactic nominalization processes can be grouped into two categories: those which affix directly to verbs, and those which cliticize. Affixed forms are rather straightforward in deriving nouns, whereas cliticized forms employ multi-functional pronominal morphemes that generally attach directly to nouns to indicate gender (noun class). These gender morphemes allow Pnar speakers to categorize the nominal elements of their environment, specifically refers in terms of gendered singular and plural. The same noun gender markers derive nominals from verbs, classifying them as particular types of nominalizations salient to Pnar speakers. Pnar further exhibits a versatile clause marker that derives property concepts from verbs as well as creating relative clauses. These processes give us insight into potential pathways for the diachronic development of both gender and relativization in Pnar.
References


Impact of Tai Lue on Muak Sa-aak phonology

Elizabeth HALL
Linguistics Institute, Payap University
<ellie.hall@outlook.com>

Abstract
The Austroasiatic language Muak Sa-aak belongs to the Angkuic branch of the Palaungic subgroup. Speakers live primarily in eastern Shan State of Myanmar. This analysis is based on the variety of Wan Fai village. Although Burmese and Chinese are influential, their primary contact language is the Tai Kadai language Tai Lue. Borrowing from this language is extensive, even to the extent of replacing their numerals with Tai Lue. Although Muak Sa-aak underwent the Germanic shift typical of Angkuic languages, replacing proto-voiced initials with voiceless ones, it still retains some voiced initials. There is some evidence that language contact may have resulted in a three-tone system, where pitch would otherwise have been largely predictable. For the majority of rhotic-initial loan words, borrowing shows a direct correspondence of Muak Sa-aak /r/ with initial /h/ in Tai Lue. Some evidence from Assamese Tai languages suggests that rather than being a replacement, this might reflect a time when Tai Lue possibly still had a rhotic.

Keywords: Palaungic, Angkuic, phonology
ISO 639-3 language codes: tlq, khb, aho, aio, mya, cmn, huo, uuu, kjg, lbn

1. Introduction
Muak Sa-aak is an Austroasiatic language, belonging to the Angkuic subgroup of the Eastern branch of the Palaungic languages. They are also called Loi, or Tai Loi. This term however encompasses various languages which are not mutually comprehensible. Muak Sa-aak varieties are spoken in Eastern Shan State in Myanmar, primarily Mong Yawng Township, and also across the border in China. The estimated population total is 4,460 in Myanmar and China (Hopple 2007). This paper is based on the variety spoken in the village of Wan Fai in Eastern Shan State of Myanmar. A wordlist of 1,643 items was collected in whole or in part from three male mother-tongue speakers, aged between 42 and 55, who came from that village and had spent most of their lives there. None had any formal education.

The Muak Sa-aak are exposed to three major non-Austroasiatic languages: Tibeto-Burman Burmese, Chinese, and the Tai-Kadai language Tai Lue. Burmese, the national language of Myanmar where the majority of the Muak Sa-aak are located, has been used as a language of education in the government schools but many people in that part of Myanmar do not speak it due to lack of formal education. Like other adults in their community, none of the speakers interviewed for this research could speak Burmese. However, some Muak Sa-aak children in Wan Fai and other Muak Sa-aak villages have attended Burmese language schools. Recently many of the children have begun to attend schools with Chinese, the second contact language, as the language of instruction, especially in border areas. Speakers from Wan Fai village also report that their children watch Chinese television and that Chinese traders come through their village.

The most influential contact language is Tai Lue, classified as Tai Kadai, Kam Tai, Tai, Southwestern (Lewis, Simons and Fennig, 2013). Tai Lue is the major language of wider communication used in the Muak Sa-aak area of Shan State. Muak Sa-aak speakers do not seem to draw an ethnic or linguistic distinction between the Tai Kadai languages Shan and Lue. They will refer to both by the same ethnic group name, Shan, but they speak only Tai Lue as a second language, not Shan. In recent generations there has been intentional borrowing; one speaker in his fifties reported his parents teaching that not borrowing words from Tai Lue would be like eating food with no salt. Nevertheless, Muak Sa-aak parents still speak their mother tongue to their children and children continue to learn Muak Sa-aak first. The children may also use Tai Lue, Chinese, or Burmese at home.
Muak Sa-aak has borrowed extensively from the surrounding Tai Lue. Out of the initial wordlist of 1,643 items, 518 words or 32%, include Lue morphemes. In order to evaluate any effects of borrowing from Tai Lue on Muak Sa-aak, a brief sketch of the phonologies of both languages will be presented, followed by a comparison and analysis of loan words.

2. Comparison of Muak Sa-aak and Tai Lue phonology

The Muak Sa-aak phonology given below is based on Hall (2013). Muak Sa-aak is a tonal language. Words are sesquisyllabic, which means some words have reduced initial syllables made up of a consonant /p, pʰ, t, k, kʰ, m, s/ with a predictable vowel or the reduced vowel alone. Tone is not contrastive in reduced syllables.

There are three types of full syllables in Muak Sa-aak: open, checked, and smooth. Open and checked syllables can be short or long. In smooth syllables, when the vowel preceding a sonorant final is short, the sonorant is lengthened; when the vowel is long, the sonorant is short, so that the overall smooth syllable is always long. In full syllables, all consonants may occur in syllable-initial position. Initial consonant clusters include /pr, kr, pʰr, kʰr, pw, kw, pʰw, kʰw/. Final consonants are limited to /p, t, c, k, m, n, ɲ, ŋ, w, j, l/. All full syllables carry one of the three contrastive tones. The low tone is labeled with 1, the constricted tone is labeled 2, and the falling tone is labeled 3. Tone 1 is found on all long syllable types. Tone 2 occurs on both long and short syllables, with a high allotone in short syllables and a falling allotone in long syllables. Tone 3 occurs only on long smooth syllables.

The 21 consonant phonemes and 18 vowel phonemes are shown in Tables 1 and 2.

Table 1. Muak Sa-aak consonant phonemes

<table>
<thead>
<tr>
<th></th>
<th>pʰ p b</th>
<th>m</th>
<th>f</th>
<th>w</th>
<th>rʰ r d</th>
<th>cʰ c</th>
<th>kʰ k</th>
</tr>
</thead>
<tbody>
<tr>
<td>pʰ</td>
<td>p b</td>
<td>m</td>
<td>f</td>
<td>w</td>
<td>rʰ r d</td>
<td>cʰ c</td>
<td>kʰ k</td>
</tr>
</tbody>
</table>

Table 2. Muak Sa-aak vowel phonemes, Wan Fai variety

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back unrounded</th>
<th>Back rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td>i iː</td>
<td>u uː</td>
<td></td>
</tr>
<tr>
<td>Close-mid</td>
<td>e eː</td>
<td>r xː</td>
<td>o oː</td>
</tr>
<tr>
<td>Open</td>
<td>e</td>
<td>a aː</td>
<td>o</td>
</tr>
<tr>
<td>Diphthongs</td>
<td>ia</td>
<td></td>
<td>ua</td>
</tr>
</tbody>
</table>

A brief summary of Tai Lue phonology, based on Hudak (1996), is presented for comparison with Muak Sa-aak. Tai Lue has 22 consonant and 18 vowel phonemes, shown in Tables 3 and 4.

Table 3. Tai Lue initial consonants (adapted from Hudak 1996)

<table>
<thead>
<tr>
<th></th>
<th>p pʰ b</th>
<th>t tʰ d</th>
<th>cʰ c</th>
<th>k</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>pʰ</td>
<td>p b</td>
<td>t tʰ d</td>
<td>cʰ c</td>
<td>k</td>
<td>?</td>
</tr>
<tr>
<td>f</td>
<td>s</td>
<td>n</td>
<td>r</td>
<td>j</td>
<td>h</td>
</tr>
<tr>
<td>m</td>
<td>n</td>
<td>p</td>
<td>ɲ</td>
<td>η</td>
<td>j</td>
</tr>
<tr>
<td>w</td>
<td>l</td>
<td>r</td>
<td>j</td>
<td>η</td>
<td>j</td>
</tr>
</tbody>
</table>

The velar fricative /x/ in Tai Lue is the equivalent of the aspirated velar stop / kʰ/ in some Tai languages and in Muak Sa-aak. The phoneme /w/ is realized [v] initially. Initial consonant clusters according to Hudak include/ kw, tw, tʰw, xw, tr, tʰr/. Hudak (1996) includes the rhotic in his Tai Lue phoneme inventory, although it is only found in its written form, never spoken. Final consonants are limited to /p, t, k, m, n, ɲ, j, w/.
Hudak identifies six tones for Tai Lue with slight differences in pronunciation according to two dialects, Chiang Hung and Muong Yong. Tones 3 and 6, low and mid-falling, are glottalized. The other tones are high, mid, low and falling.

As may be seen in Tables 1-4 above, there is a great amount of similarity in phonology between Muak Sa-aak and Tai Lue, including voicing contrast of initial stops, which Angkuic languages like Muak Sa-aak are expected to have lost. Major differences between the sound inventories of Muak Sa-aak and Tai Lue include the number of contrastive tones and the presence of a rhotic. The phoneme /r/, occurring only in written forms in Tai Lue, occurs widely in Muak Sa-aak, including as a medial consonant in clusters. Tai Lue, on the other hand, forms many consonant clusters with /w/ as the medial consonant, while Muak Sa-aak has only a few of these. In the vowels, this variety of Muak Sa-aak has two diphthongs /ia/ and /ua/, which Tai Lue does not have. In addition, the lateral /l/ and the palatal consonants /c, ɲ/ are found word-finally in Muak Sa-aak, but not in Tai Lue, suggesting that words with these finals are unlikely to be borrowed from Lue. Since data on borrowings with final /c, ɲ/ into Lue are not available, it is not possible to say how these might be adapted in Lue, which has only bilabial, alveolar and velar oral and nasal stops and the glides /w, j/ in syllable-final position.

3. Phonological changes due to borrowing

Tai Lue loan words are adjusted to, but also impact, Muak Sa-aak phonology as seen in the vowel system, the rhotic, voiced initials, and tone.

3.1 Adjustment of alien vowels

The Tai Lue vowels /ɔː/ and /ɛː/ are realized as the diphthongs /ia/ and /ua/ in the Wan Fai variety of Muak Sa-aak. This suggests that the Muak Sa-aak diphthongs are phonologically the equivalents of the missing long open vowels /ɛː/ and /ɔː/ which are found in other Muak Sa-aak varieties without diphthongs, such as the variety spoken in Wan Saw village. Thus, the Wan Fai Muak Sa-aak rendering of the Tai Lue monophthongs /ɛː, ɔː/ as /ia, ua/ suggests that these diphthongs developed from an older /ɛː/ and /ɔː/. Examples of borrowed words with the adjustment to Muak Sa-aak diphthongs are given in Table 5.

| Table 5. Wan Fai Muak Sa-aak diphthongs in Tai Lue loanwords |
|-------------------|-------------------|-------------------|
| Tai Lue           | Muak Sa-aak       | gloss             |
| pɛːt²             | piat¹             | ‘eight’           |
| sip¹ sɔŋ¹          | sip² suan³        | ‘twelve’          |
| men⁴ saːp²         | mian⁴ saːp¹       | ‘cockroach’       |
| lem¹               | liam¹             | ‘sharp-pointed’   |
| sɔɔk²              | suak¹ tɔ²         | ‘elbow’           |
| hɔn⁶               | ruan²             | ‘hot’             |

The number ‘two’ forms an exception as it is not pronounced consistently. As a monosyllable the vowel /ɔː/ is retained, while in numeral compounds it is adjusted to Muak Sa-aak /ua/, as in sip² suan³ ‘twelve’.

3.2 Correspondence of /h/ and a rhotic in Muak Sa-aak

Muak Sa-aak and Tai Lue differ in phonotactics. In Tai Lue, initial /r/ is found only in the written language and always realized as /h/ in the spoken language. In Muak Sa-aak, /r/ and /h/ are contrastive, with /r/ being more common. Initial /h/ in Tai Lue words is often replaced with /r/ in

---

**Table 4. Tai Lue vowels (Hudak 1996: xxiii)**

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back unrounded</th>
<th>Back rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i iː</td>
<td>u uː</td>
<td>u uː</td>
</tr>
<tr>
<td>Mid</td>
<td>e eː</td>
<td>ϕ ϕ</td>
<td>o oː</td>
</tr>
<tr>
<td>Low</td>
<td>e eː</td>
<td>a aː</td>
<td>o oː</td>
</tr>
</tbody>
</table>
Muak Sa-aak; for example, Lue $hok^2$ ‘six’ > Muak Sa-aak $rœk^2$. This does not appear to be motivated by phonotactics or meaning. In addition to the replacement of /h/ with /r/, some apparent borrowings into Muak Sa-aak have initial consonant clusters /$p)r-/ or /$pr-/ clusters which do not occur in Tai Lue; for example, Lue $p\overset{\text{u}}{\text{ŋ}}^2$ ‘bee’ > Muak Sa-aak $p\overset{\text{r}}{\text{ŋ}}^2$. In some places where /t/ is found in Muak S-aak, /t/ is also found in standard Thai: Muak Sa-aak $ruan^2$ ‘hot’ > Thai $r\overset{\text{ɔ}}{\text{n}}$, Muak Sa-aak $r\overset{\text{x}}{\text{n}}$‘[CLSF] house’ > Thai $ruan$ ‘house’, Muak Sa-aak $r\overset{\text{x}}{\text{t}}$ ‘boat’ > Thai $rua$. In other apparent loanwords containing /t/ in Muak Sa-aak, the /t/ does not occur in Thai either.

It is useful here to examine the Tai languages of Assam, India, specifically Tai Ahom and Tai Aiton which, like Tai Lue, are Southwestern Tai languages. These have retained /t/ in places that the other Tai languages do not, including consonant clusters with /t/. Ahom was spoken by people who originally came from Shan state of Myanmar (Tabassum and Morey 2009). The language is no longer spoken as a mother tongue, but is found in manuscripts. Initial clusters with /t/ do occur in the manuscripts, but it is not clear how many of these reflect actual clusters in the spoken language. Pronunciation today in religious usage may reflect hypercorrection. Tai Aiton, however, which is closely related to Ahom, is still spoken in northeast India and the rhotic may be heard there in spoken language, both as an initial and in initial consonant clusters (Tabassum and Morey 2009). Some Tai cognates found in Muak Sa-aak borrowings have initial /t/ or clusters with /t/ which are not found in Tai Lue or Tai, whereas Ahom or Aiton contain these rhotics; see Table 6 for examples. The numeral ‘five’ is included for contrast as an example of a word which does not have /t/ in any of these languages, showing that /h/ is not systematically being replaced with /t/. There must rather have been some other mechanism responsible for the occurrence of the rhotic in Muak Sa-aak loanwords from Tai Lue.

Table 6. Muak Sa-aak loanwords with /t/ in comparison to Tai Lue, Tai Ahom, and Tai Aiton (Tabassum and Morey 2009).

<table>
<thead>
<tr>
<th>Muak Sa-aak</th>
<th>Thai</th>
<th>Tai Lue</th>
<th>Tai Ahom</th>
<th>Tai Aiton</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ro^{\cdot}caj^2$</td>
<td>$h\overset{\text{u}}{\text{a}}\overset{\cdot}{\text{c}}\overset{\cdot}{\text{a}}$</td>
<td>$h\overset{\text{o}}{\text{u}}\overset{\cdot}{\text{c}}\overset{\cdot}{\text{a}}^1$</td>
<td>$ru\ u\ chau$</td>
<td></td>
<td>heart</td>
</tr>
<tr>
<td>$r\overset{\text{œ}}{\text{k}}^2$</td>
<td>$h\overset{\text{o}}{\text{k}}$</td>
<td>$h\overset{\text{o}}{\text{k}}^2$</td>
<td>$ruk$</td>
<td></td>
<td>six</td>
</tr>
<tr>
<td>$ha^{\cdot}^2$</td>
<td>$h\overset{\cdot}{\text{a}}$</td>
<td></td>
<td>$ha$</td>
<td></td>
<td>five</td>
</tr>
<tr>
<td>$p\overset{\text{r}}{\text{ŋ}}^2$</td>
<td>$p\overset{\text{r}}{\text{ŋ}}\overset{\cdot}{\text{ŋ}}$</td>
<td>$p\overset{\text{r}}{\text{ŋ}}^1$</td>
<td>$ph\overset{\text{r}}{\text{ŋ}}\overset{\cdot}{\text{ŋ}}$</td>
<td>$ph\overset{\text{r}}{\text{ŋ}}^1$</td>
<td>bee</td>
</tr>
<tr>
<td>$pr\overset{\text{u}}{\text{t}}^2$</td>
<td></td>
<td></td>
<td></td>
<td>$ph\overset{\text{r}}{\text{t}}^2$</td>
<td>sorceror</td>
</tr>
</tbody>
</table>

The existence of the rhotic in other Tai languages suggests that Muak Sa-aak has not replaced /h/ with /t/ in some instances and kept the /h/ in others. Rather, it is likely that it has instead retained an initial /r-/ in loanwords from Tai Lue where the rhotic has subsequently been lost. Muak Sa-aak loanwords from Tai Lue containing clusters with /t/ are still found in Tai Ahom and Aiton, suggesting that Tai Lue also had those clusters when the respective cognates entered the Muak Sa-aak lexicon.

3.3 Voicing contrast of initial plosives

According to Svantesson, applying the historical Germanic shift to Angkuic languages (1991), the voicing contrast in Palaungic languages was replaced in the Angkuic languages with an aspiration contrast; that is, the proto-voiceless initials became aspirated and the proto-voiced initials became voiceless. Comparison of Muak Sa-aak initials with other Palaungic languages shows that Muak Sa-aak also underwent this so-called Germanic shift (Hall 2010). Table 6, taken from Hall (2010), compares Muak Sa-aak proto-voiceless and voiced initials with Hu and U, two other Angkuic languages. Three Mon-Khmer languages that do not belong to the Angkuic subgroup, Lamet, a Palaungic language, and Northern Khmu and Southern Khmu, Northern Mon-Khmer languages that belong to the Khmuic group, are included to demonstrate that the Germanic shift is an Angkuic innovation.
Table 7. Germanic shift in Muak Sa-aak initial consonants (adapted from Svantesson 1991)

<table>
<thead>
<tr>
<th>Muak Sa-aak</th>
<th>Hu</th>
<th>U</th>
<th>Lamet</th>
<th>Northern Kammu</th>
<th>Southern Kammu</th>
</tr>
</thead>
<tbody>
<tr>
<td>*voiceless</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pʰɤɲ³</td>
<td>phiŋ</td>
<td>phɛt</td>
<td>pijŋ</td>
<td>pijŋ</td>
<td>pijŋ</td>
</tr>
<tr>
<td>tam⁴</td>
<td>thɛŋ</td>
<td>thən</td>
<td>tɑːɲ</td>
<td>tɑːɲ</td>
<td>taːɲ</td>
</tr>
<tr>
<td>kʰəp⁵</td>
<td>khɛp</td>
<td>khap</td>
<td>kɑːp</td>
<td>kɑːp</td>
<td>kaːp</td>
</tr>
<tr>
<td>*voiced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pri²</td>
<td>pʁiʔ</td>
<td>qi</td>
<td>príʔ</td>
<td>príʔ</td>
<td>briʔ</td>
</tr>
<tr>
<td>puc² ta:k¹</td>
<td>phltąk</td>
<td>?atąʔ</td>
<td>pltą:k</td>
<td>ktą:k</td>
<td>kda:k</td>
</tr>
<tr>
<td>ka:ŋ³</td>
<td>kąŋ</td>
<td>kąː</td>
<td>---</td>
<td>ką:ŋ</td>
<td>ga:ŋ</td>
</tr>
<tr>
<td>kak²</td>
<td>kąk</td>
<td>kąk</td>
<td>kąk</td>
<td>kąk</td>
<td>--</td>
</tr>
</tbody>
</table>

Voiced initial plosives are not expected in an Angkuic language such as Muak Sa-aak, due to the established phonological correspondences of initial voiceless plosives with proto-initial voiced plosives. The preservation of a voicing contrast of bilabial and alveolar stops in Muak Sa-aak may therefore be the result of contact with Tai Lue. These voiced initials may be seen in borrowed words from Tai Lue: for example, Lue bii¹ ‘gallbladder’ > Muak Sa-aak biː³, Lue bup¹ ‘hit’ > Muak Sa-aak buːp², Lue dàː¹ ‘kick’ > Muak Sa-aak daː², Lue dɛn⁴ ‘border’ > Muak Sa-aak dian⁴.

Not all instances of voiced initials /b,d/ however, may be easily explained as borrowing from Tai Lue. The initials /b,d/ appear in some apparent Austroasiatic cognates as well. For example Muak Sa-aak bil³ ‘forget’ may be found in U pìn and has been reconstructed as proto-Mon-Khmer *[b]iir; this may be seen in Palaung bir and Praok pi[m]. Muak Sa-aak dəŋ⁴ ‘quiet’ has the proto-Mon-Khmer reconstruction *ʔok, seen also in Palaung ðɔʔ and Riang-Lang ðək. Muak Sa-aak bɔt² ‘cloud’ may be linked to the proto-Mon-Khmer reconstruction *ʔut, *ʔuut, which does not contain the voiced initial, but the realizations of ðɔt in Bo Luang Lawa and of piːt in Yunnan Khmu do have the bilabial initial (Svantesson 1988, Shorto, Cooper, Sidwell and Bauer 2006). These last two, ‘quiet’ and ‘cloud’, raise another possibility, that the voiced initials /b, d/ in Muak Sa-aak could be reflexes of earlier glottalized initials. This is an area for further research, as the data available at present is limited.

3.4 Impact of Tai Lue tone on Muak Sa-aak

Given the extensive amount of vocabulary borrowed from Tai Lue- even the numeral system has been borrowed essentially intact- it is worth asking to what extent this has affected the tones seen today.

Previously it has been suggested that tonogenesis in Muak Sa-aak was motivated primarily by final consonants and vowel length (Hall 2010). In native words, sonorant finals mostly occur with Tone 3 whereas checked syllables only occur with Tones 1 or 2. This suggests that final consonants have played an important role in tonogenesis.

This pattern is not entirely regular; in particular for nasal finals. Most of the words ending in nasals which match up to Svantesson’s (1988) Lamet list, and therefore do not appear to be loanwords, occur with falling Tone 3. However, there are a number of Muak Sa-aak words ending with nasal finals that carry low Tone 1 or constricted Tone 2. One possible explanation is that these words are largely borrowed words from Tai Lue. Words ending in palatal nasals are native Muak Sa-aak because Tai Lue does not have these. These Muak Sa-aak words always carry falling Tone 3; there are no occurrences with the other two tones.

Tai Lue also does not have the lateral approximant /l/ occurring in final position. Nearly all of those Muak Sa-aak words ending in /-l/ occur with Tone 3 as well; only a handful of exceptions in the corpus occur with Tone 1. These are: pal¹ k.tit² ‘(be) equal’, muːl¹ ‘boundary’, pul¹ pil¹ ‘crush to powder’, pul¹ ‘fall’. Even though these words are not borrowings from Tai Lue, the rarity of Tone 1 and Tone 2 sonorant finals suggests that sonorant finals naturally carry Tone 3. The falling allotone
of Tone 2 on long syllables includes a number of words with final nasals. About half of those in the corpus collected for this study are Tai Lue cognates.

Another phenomenon to take into account is the small group of long open syllables occurring with constricted Tone 2. Generally this tone is found on short open syllables, and the majority of these exceptions are borrowed from Tai Lue. Examples are given in Table 8.

Table 8. Long open syllables with constricted Tone 2

<table>
<thead>
<tr>
<th>Tai Lue</th>
<th>Muak Sa-aak</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>naː³</td>
<td>naː² naːj³</td>
<td>‘face’</td>
</tr>
<tr>
<td>kʰaː³</td>
<td>kʰaː²</td>
<td>‘slave’</td>
</tr>
<tr>
<td>ŋaː⁴</td>
<td>t.ŋaː²</td>
<td>‘sesame seed’</td>
</tr>
<tr>
<td>juː⁶</td>
<td>juː²</td>
<td>‘push’</td>
</tr>
</tbody>
</table>

Three of these four examples occur with the Tai Lue glottalized tones 3 and 6; this constriction is being carried into Muak Sa-aak. Although t.ŋaː² ‘sesame seed,’ does not follow this pattern, there still appears to be a correlation between tone of borrowed words from Tai Lue and tone in Muak Sa-aak. Tai Lue words with a glottalized tone tend to occur with constricted Tone 2 in Muak Sa-aak. The first of the items in Table 8 is a tautological hybrid form: naː² ‘face’ coming from Tai Lue, and ŋaːj³, ‘face’ or ‘eye’ from Muak Sa-aak.

To summarize, sonorant finals naturally carry Tone 3; if they carry Tone 2 they are likely to be Tai Lue loan words. Similarly, long open syllables usually carry Tone 1 or 3; those carrying the long allitone of constricted Tone 2 are mostly Tai Lue borrowings. The natural environment for Tone 2 appears to be either the short open syllable or the short checked syllable. Without these borrowings, tones on sonorant-final syllables would largely be predictable, based upon vowel length and final consonant. These predictable indigenous Muak Sa-aak tones correlate to the pitch accompanying register described for Lampang Lamet. In this variety, pitch appears to relate to the interaction of vowel length and final consonant type, or dead and live syllables (Narumol 1982). Svantesson also describes a similar distinction between final consonant types for U (1988). The reason Muak Sa-aak tones are not predictable may be found in this borrowing of long glottalized open or sonorant-final syllables, as well as the loss of some final consonants.

4. Summary and outlook

Borrowing is an interesting phenomenon in this language, as it has been heavily influenced by Tai Lue, a language of an entirely different family from which it has borrowed large amounts of vocabulary. Even the numeral system has been adopted essentially intact from Tai Lue. The effects of lexical borrowing from Tai Lue on Muak Sa-aak can be seen at several levels. The Muak Sa-aak variety under study has two diphthongs /ia/ and /ua/ and replaces Tai Lue /ɛː, ɔː/ with these diphthongs in borrowed words. Although Angkuic languages are expected to have lost their voicing contrast, Muak Sa-aak retains this contrast, which is found in Tai Lue as well. It did however, undergo the Germanic shift like other Angkuic languages. Further in-depth studies to investigate the possible motivation for retaining the old Mon-Khmer voicing contrast as well as developing aspiration in other Mon-Khmer cognates are needed. The borrowings appear to have influenced the tonal patterns of Muak Sa-aak, possibly furthering the development of tone in an otherwise largely predictable pitch system.

Despite the borrowing, the majority of lexical items and phonotactics stand out as native to Muak Sa-aak. This includes palatal final consonants and a final lateral consonant. One feature which may reflect something that has been lost in the surrounding Tai Lue but not in Muak Sa-aak is the phoneme /r/. The rhotic occurs frequently in Muak Sa-aak, both word-initially and in consonant clusters, while spoken Tai Lue does not have an /r/ in its phoneme inventory. In many words borrowed from Tai Lue, Muak Sa-aak actually replaces initial Tai Lue /h/ with the rhotic, which appears to reflect the existence of a former rhotic in these Tai Lue cognates, since rhotics are found in some Assamese Tai equivalents. Conversely, this particular variety of Muak Sa-aak shows at least one development which does not appear in Tai Lue, the development of diphthongs /ia/ and /ua/ from a former /ɛː/ and /ɔː/. These also may be seen in words originally borrowed from Tai Lue,
where they do not contain diphthongs, but in Muak Sa-aak usage the long monophthongs are replaced with a diphthong.

The exposure to the school languages could create further pressure on the Muak Sa-aak language. Together with the Chinese television programs, this could cause a shift in borrowing patterns, as the children growing up now use more Burmese and Chinese and less Tai Lue than their parents. Since the Muak Sa-aak community seems to be linguistically quite absorbent, future investigations of the lexicon and phonology promise to be of interest for the study of contact phenomena.

References


Hopple, Paulette. 2007. Notes re: the visit of Plang from Mong Yawng. ms. (Unpublished)


Conceptual metaphors of Vietnamese taste terms

Rujuwan LAOPHAIROJ
Mahidol University Thailand
<pormin_anne@hotmail.com>

Abstract
The objectives of this study were to examine the use, metaphorical meanings and metaphorical concepts of four Taste Terms in Vietnamese, namely bitter, salty, sour and sweet. The data of the taste terms were collected from dictionaries, linguistic corpora and interviews with five informants. The result on metaphorical distribution of the taste terms based on their structure reveal that the terms in Vietnamese languages have metaphorical use in two types. The first type is single taste terms with metaphorical meaning. The second type is the taste terms in combination with other words with metaphorical meanings. This type is divided into the taste terms in combination with other taste terms and the taste terms in combination with other words. Each type is different in details.

With the regard to the result on metaphorical meanings in Vietnamese, the taste terms are categorized in one semantic domain: state metaphor. The state metaphor was semantically divided into three subgroups: feelings metaphor, habits/manners metaphor and qualities metaphor. For metaphorical concepts of Vietnamese taste terms can be categorized into concept: human qualities are taste

Keywords: metaphor, semantics, taste terms
ISO 639-3 codes: vie

1. Introduction
Taste is one of the five basic perceptions of human beings, including sight, sound, taste, smell and touch. Taste is important to the learning and existence of mankind, because man has to eat on a daily basis in order to survive. When food is eaten, the tongue perceives the tastes and sends signals to the brain to process the perceptions. Moreover, humans have created lexical items to express or describe various kinds of tastes. These words are called taste terms, which vary from one language and culture to another.

Taste terms in certain languages are used not only to describe or refer to tastes but to compare other entities as well. In Vietnamese, the term chua ‘sour’ in the expression

Cô ấy giọng chua lắm.
woman 3rd voice sour very
‘That woman has a very high pitched voice.’

The word chua refers to an attribute of a high, ear-splitting voice. It can be seen that chua ‘sour, acidic and corrosive to the throat when eating something sour’ has been transferred its inherent meaning of taste to the perception of sound.

From the above examples, it is seen that the meanings of taste terms have been used comparatively to refer to other entities with different meanings. In cognitive linguistics, the process is called metaphor. This is different from literature which describes metaphor as a figurative language meant to provoke emotions and visualize images that authors or poets want to convey as well as to classify types of metaphors in the language. However, cognitive linguistics views metaphor as a concept system that humans learn from the environment by means of comparison or analogy. What they have learned reflects in the language they speak on a daily basis.
and metaphorical language expresses the concepts, thinking and beliefs of members of a particular society.

It is thus interesting to examine meanings and metaphorical concepts of the taste terms Vietnamese based on the cognitive linguistics, because the investigation would enhance better understanding of the mental and cultural systems of native speakers of the Vietnamese. Additionally, it would help us to better understand perspectives of the Vietnamese culture. More importantly, there have not been previous studies on metaphor of taste terms in the Vietnamese language.

1.1 Metaphor according to cognitive linguistics

This is a new trend of language study, viewing metaphor as a language used in daily life and reflecting the processes of thinking and perception of language users.

Lakoff and Johnson (1980) were pioneers in applying cognitive linguistics to analyze meanings. They stated that metaphor is a language form used in daily life and reflects the thinking process, beliefs and concepts of social members. Language users are usually unaware that the language they use for daily communication is metaphor. Additionally, metaphor expresses systematization and this systematization reflects conceptual metaphor that tells how language users think about and understand events and things in their culture.

Lakoff and Johnson (1980) gave an example of metaphorical language use about time of English speakers. It is found that there are several expressions about time in English that express temporal metaphor.

How do you spend your time these days?
That flat tire cost me an hour.
I’ve invested a lot of time in her.
You need to budget your time.
You don’t use your time profitably.

(Lakoff and Johnson,1980: 8)

The above sentences use words usually collocated with money (spend, cost, invested, budget, use and profitably) to collocate with time. They are metaphorical and systematic, which implies the conceptual metaphor of English speakers that TIME IS MONEY. It is further found that there are other types of metaphor in English, such as LOVE IS A JOURNEY, MORE IS UP, or ARGUMENT IS WAR. These metaphors support Lakoff’s view that human thinking processes are in the form of metaphor.

Lakoff further explained that metaphorical thinking processes map common aspects or qualities of source and target. The mapping is unidirectional from source to target, not the other way round. He cited an example of mapping of metaphorical concept of LOVE IS A JOURNEY. The source is JOURNEY and the target is LOVE. The mapping is illustrated as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOURNEY</td>
<td>LOVE</td>
</tr>
<tr>
<td>The travelers</td>
<td>The lovers</td>
</tr>
<tr>
<td>The vehicle</td>
<td>The love relationship itself</td>
</tr>
<tr>
<td>The journey</td>
<td>Events in relationship</td>
</tr>
<tr>
<td>The distance covered</td>
<td>The progress made</td>
</tr>
<tr>
<td>The obstacles encountered</td>
<td>The difficulties experienced</td>
</tr>
<tr>
<td>Decisions about which way to go</td>
<td>Choice about what to do</td>
</tr>
<tr>
<td>The destination of the journey</td>
<td>The goals of the relationship</td>
</tr>
</tbody>
</table>

(Kovecses,2002 cited in Chatchawadee, 2007)

The above diagram illustrates the ontological correspondences, the ‘the travelers’ as the source domain and ‘the lovers’ as the target domain. The two domains are in the one-on-one correspondence as follows.
‘The travelers’ are in correspondence with ‘The lovers’.
‘The vehicle’ is in correspondence with ‘The love relationship itself’.
‘The journey’ is in correspondence with ‘Events in relationship’.
‘The distance covered’ is in correspondence with ‘The progress made’.
‘The obstacles encountered’ is in correspondence with ‘The difficulties experienced’.
‘Decisions about which way to go’ is in correspondence with ‘Choice about what to do’.

and ‘The destination of the journey’ is in correspondence with ‘The goals of the relationship’.

Regarding semantic correspondences, Croft and Cruise (2004) categorize the correspondences into two types as follows.

1) Ontological correspondences refer to semantic correspondences from source to target on a one-to-one basis, as shown in the following example.

The ontological correspondences of the “anger” metaphor

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Container</td>
<td>- Body</td>
</tr>
<tr>
<td>- Heat of liquid</td>
<td>- Anger</td>
</tr>
<tr>
<td>- Level of heat</td>
<td>- Level of anger</td>
</tr>
</tbody>
</table>

2) Epistemic correspondences refer to the semantic correspondences from source to target on a group or collective basis. In the source, components are related to a collective basis and they correspond to the target components which are related to a collective basis as well, as shown in the following example.

The epistemic correspondences of the “anger” metaphor

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>- When liquid in a container is overheated, the pressure increases to the point of explosion.</td>
<td>- When anger is at its peak, one loses self-control.</td>
</tr>
<tr>
<td>- The explosion damages the container and harms people nearby.</td>
<td>- Loss of self control is harmful and endangers others nearby.</td>
</tr>
</tbody>
</table>

Lakoff and Johnson (1980) proposed that metaphor could be found in language of daily use and the metaphors used in daily life reflected the thinking systems, beliefs and concepts of social members. Their proposal has brought about an interest among linguists to study meanings of words. Many have investigated metaphor in language in everyday use or metaphor according to cognitive linguistics. Investigations on metaphor in Vietnamese are summarized as follows.

- Phoung Vi Vo (2003) investigated the conceptual metaphor of “love” in Vietnamese from past to present in poems and songs in contemporary and folk literature.
- Ly Toan Thang (2011) investigated the Vietnamese Expression of BODY and SOUL: A Cognitive and Cultural Linguistic Study

As for metaphorical studies of taste terms, there have not been any previous investigations. Most studies focused on linguistic forms of taste terms, their structures and meanings in order to examine the mentality and understanding of language users. Therefore, it is interesting to investigate metaphors of taste terms to understand the mentalities and concepts of Vietnamese speakers.

2. Results

The results of the analysis are presented in three topics as follows.
2.1 The metaphorical use of taste terms

According to the investigation and analysis of the four Vietnamese taste terms, namely *chua* ‘sour’, *đắng* ‘bitter’, *mặn* ‘salty’ and *ngọt* ‘sweet’, it is found that their metaphorical distribution of taste terms can be grouped in into two main groups according to their structures as follows.

2.1.1 Single word taste terms with connotative meanings
2.1.2 Taste terms compounded with other words with connotative meanings

This type was divided into two subtypes based on their occurrence as follows:

A taste term co-occurring with another taste term
A taste term in combination with other words

2.1.1 Single word taste terms with connotative meanings

There is only one taste terms *chua* that can occur as single words with metaphorical meanings in context. This is shown in the following examples.

1) *chua*
   Cô ấy chua lắm.
   woman 3rd sour very
   ‘That woman is very outrageous.’

2.1.2 Taste terms compounded with other words with connotative meanings

A) A taste term co-occurring with another taste term

There are two terms: *chua* and *đắng* that can co-occur with other taste terms with metaphorical meanings. Each taste term can precede or follow another taste term as a compound word. This is shown in the following examples.

2) *chua* ‘sour’ + *cay* ‘spicy’
   Lời châm biếm chua cay.
   word satirize sour spicy
   ‘Bitter sarcasms’
   In (2) *chua* precedes a taste term *cay* forming a compound.

3) *cay* ‘spicy’ + *chua* ‘sour’
   Lời châm biếm cay chua.
   word satirize spicy sour
   ‘Bitter sarcasms’
   In (3) *chua* follows a taste term *cay* forming a compound.

4) *đắng* ‘bitter’ + *cay* ‘spicy’
   Trên khuôn mặt cô ấy, tràn đầy những giọt nước mắt đắng cay.
   on face woman 3rd brimful some drop tear bitter spicy
   ‘Her face is covered with tears of bitterness.’
   In (4) *đắng* precedes a taste term *cay* forming a compound.

5) *cay* ‘spicy’ + *đắng* ‘bitter’
   Bao nhiêu cay đắng, bây nhiêu niềm tin.
   how much spicy bitter so much sense of confidence
   ‘So much bitterness, so much faith’
   In (5) *đắng* follows a taste term *cay* forming a compound.
B) A taste term is in combination with other words

Furthermore, the term chua, mặn and ngọt also have metaphorical meanings when they compound with other words, either before or after the compounding elements, as shown in the following examples.

6) giòng ‘voice’+ chua ‘sour’
Cô ấy giòng chua lâm.
woman 3rd voice sour very
‘That woman has a very high pitched voice.’
In (6) chua follows a noun giòng ‘voice’ forming a compound.

7) nói ‘speak’+ ngọt ‘sweet’
Nói ngon nói ngọt.
speak delicious speakl sweet
‘To use sweet words’
In (7) ngọt follows a verb nói ‘speak’ forming a compound.

8) chua ‘sour’+ xót ‘sting’
Cảnh ngộ chua xót.
Plight sour sting
‘A heart-rending plight’
In (8) chua precedes a noun xót ‘sting’ forming a compound.

9) mặn ‘salty’+ tình ‘love’
Mặn tình anh em.
Salty love older brother younger brother
‘Warm brotherhood’
In (9) mặn precedes a noun tình ‘love’ forming a compound.

10) ngọt ‘sweet’+ ngào ‘cook in syrup’
Cười cười nói nói ngọt ngào.
smile smile say say sweet cook in syrup
‘To smile and speak suavely.’
In (9) ngọt precedes a noun ngào ‘cook in syrup’ forming a compound.

2.2 Metaphorical meanings and semantic mapping of Vietnamese taste terms

The investigation of the four taste terms further reveals that they have one semantic domain: state metaphor.

Before presenting examples of the analysis of metaphorical meanings and semantic mapping of the Vietnamese taste terms, it would be better to discuss attributes of the four taste terms in order to facilitate better understanding about the analysis of their semantic mapping. This is due to the fact that one taste term has different attributes in different contexts. Therefore, some prominent attributes that are in line with the aspects in the target domain are presented. Details of the attributes of the taste terms are shown in the alphabetical order as follows.\(^1\)

\(^{1}\) [ ] indicates a semantic attribute of a word.
2.2.1 State metaphor

A state metaphor refers to linguistic forms with reference to deeds and manners. In this study, state metaphor includes feelings which are states occurring in the mind, and habits and manners which are states of actions expressed externally.

It was found from this investigation that the state metaphor was semantically divided into three domains: feelings metaphor, habits/manners metaphor, and qualities metaphor. Details are as follows.

Feelings metaphor

Feelings metaphor refers to taste terms with comparative meanings to feelings that occur in the minds of humans.

There are four taste terms: **chua**, **đắng**, **mặn** and **ngọt**, co-occurring with other taste terms and other words, which metaphorically express feelings. This is shown in the following example.

In presenting the data for analysis, the taste terms from each language were arranged in alphabetical order.

11) Nhếch mép cười chua chát.
part a corner corner.of.mouth smile sour tart
‘To smile with bitterness’

**chua chát** means bitterness. **Chua** is sour and **chát** is tart. When the two words co-occur as a single unit, it means bitterness. **Chua chát** metaphorically represents a mental corrosive feeling. The transfer of its semantic attributes can be illustrated as follows.

---

2. This and subsequent English translations of Vietnamese taste terms follow the dictionary of Nguyễn Sanh Phúc (2000).

*Mon-Khmer Studies* 42.1:31-46 (ICAAL5 special issue)
Source Domain | Target Domain
---|---
Taste term | feelings
đắng ‘bitter’ + cay ‘spicy’ | suffering
[Perceived taste] | [Perceived feelings]
[Oral perception] | [Mental perception]
[Unfavorable] | [Unpleasant]
[Corrosively hot] | [Suffering]

Diagram 3: metaphorical mapping process of đắng in đắng cay

Source Domain | Target Domain
---|---
Taste term | feelings
cay ‘spicy’ + đắng ‘bitter’ | suffering
[Perceived taste] | [Perceived feelings]
[Oral perception] | [Mental perception]
[Unfavorable] | [Unpleasant]
[Corrosively hot] | [Suffering]

Diagram 4: metaphorical mapping process of đắng in cay đắng

From the above semantic mapping, attributes of cay ‘spicy’ and its compounding element, đắng ‘bitter’ are transferred to the target domain, which is a feelings metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. Đắng transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings], from [oral perception] to [mental perception] and from [unfavorable] to [unpleasant]. Cay also transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings], from [oral perception] to [mental perception] and from [corrosively hot] to the target domain [corrosive feeling].

15) Anh chị ấy yêu nhau mặn nồng lắm.
Brother sister love together salty warm very

‘That couple loves each other warmly.’

Mặn nồng means warm and timely love. Mặn is salty and nóng is warm or hot. When the two words co-occur as a single unit, its metaphorical meaning is a timely love. The transfer of its semantic attributes can be illustrated as follows.

Source Domain | Target Domain
---|---
Taste term | feelings
mặn ‘salty’+ nóng ‘warm’ | warm and timely love
[Perceived taste] | [Perceived feelings]
[Oral perception] | [Body perception]
[Be full-flavored] | [A fairy high degree of hot][Good Feeling]
[Closely] | [Affectionate]

Diagram 5: metaphorical mapping process of mặn in mặn nóng

From the above semantic mapping, attributes of mặn ‘salty’ and its compounding element, nóng ‘warm or hot’, are transferred to the target domain, which is a feelings metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. Mặn transfers its semantic attributes from the source domain [oral perception] to the target domain [mental perception] and from [be full-flavored] to [closely]. Nóng transfers its semantic attributes from the source domain [perceived feelings] to the target domain [perceived feelings], from [body perception] to [mental perception] and from [not cold] to [good feeling].
16) Mặn tình anh em.
salty love older brother younger brother
‘warm brotherhood’

Mặn tình means intense love. Mặn is salty and tình means love. When the two words co-occur as a single unit, its metaphorical meaning is intense love. The transfer of its semantic attributes can be illustrated as follows.

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taste term</td>
<td>feelings</td>
</tr>
<tr>
<td>mặn ‘salty’ + tình ‘love’</td>
<td>intense love</td>
</tr>
<tr>
<td>[Perceived taste]</td>
<td>[Perceived feelings]</td>
</tr>
<tr>
<td>[Oral perception]</td>
<td>[Mental perception]</td>
</tr>
<tr>
<td>[Intense]</td>
<td>[Overwhelming feeling]</td>
</tr>
<tr>
<td>[Taste that is difficult to change]</td>
<td>[Feeling that is difficult to change]</td>
</tr>
</tbody>
</table>

Diagram 6: metaphorical mapping process of mặn in mặn tình

From the above semantic mapping, attributes of mặn ‘salty’ and its compounding element, tình ‘love’, are transferred to the target domain, which is a feelings metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. Mặn transfers its semantic attributes from the source domain [intense] to [overwhelming feeling] and from [taste that is difficult to change] to [feeling that is difficult to change]. Tình transfers its semantic attributes from the source domain [perceived feelings] to the target domain [perceived feelings] and from [mental perception] to [mental perception].

17) Trải qua biết bao khó khăn, cô ấy đã cảm nhận được những ngọt bùi.
through know many difficulty woman 3rd already feel get sweet crunchy
‘After having been through life’s difficulties, she has finally found happiness.’

Ngọt bùi means fun or happiness. Ngọt is sweet and bùi is crunchy. When the two words co-occur as a single unit, it metaphorically means happy or enjoyable feelings. The transfer of its semantic attributes can be illustrated as follows.

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taste term</td>
<td>feelings</td>
</tr>
<tr>
<td>ngọt ‘sweet’ + bùi ‘crunchy’</td>
<td>fun or happiness</td>
</tr>
<tr>
<td>[Perceived taste]</td>
<td>[Perceived taste]</td>
</tr>
<tr>
<td>[Savory]</td>
<td>[Favorable]</td>
</tr>
</tbody>
</table>

Diagram 7: metaphorical mapping process of ngọt in ngọt bùi

From the above semantic mapping, attributes of ngọt ‘sweet’ and its compounding element, bùi ‘crunchy’, are transferred to the target domain, which is a feelings metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. Ngọt transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings] and from [savor] to [favorable]. Bùi transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings] and from [pleasingly] to [enjoy].

Habits/manners metaphor

From the investigation, there are two taste terms, chua and ngọt, which metaphorically express habits or manners. Chua can occur alone or co-occur with other words, as shown in the following example.
18) Cô ấy chua lắm.
woman 3rd sour very
‘That woman is very outrageous.’

Chua metaphorically represent an overconfident manner of a woman or a woman dressed in outlandish color. The transfer of its semantic attributes can be illustrated as follows.

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taste term</td>
<td>habits or manners</td>
</tr>
<tr>
<td>chua ‘sour’</td>
<td>an overconfident manner of a woman or a woman dressed in outlandish color.</td>
</tr>
<tr>
<td>[Perceived taste]</td>
<td>[Perceived personality]</td>
</tr>
<tr>
<td>[Sharp taste]</td>
<td>[Strike the eyes]</td>
</tr>
<tr>
<td>[A unique taste]</td>
<td>[A unique personality]</td>
</tr>
</tbody>
</table>

Diagram 8: metaphorical mapping process of chua

From the above semantic mapping, it is found that the attribute, [perceived taste], [sharp taste] and [unique taste] of the term chua is transferred to the target domain, [perceived personality], [strike the eyes] and [unique personality] respectively, which is a habits/manners metaphor, which is a subgroup of the state metaphor category.

19) Chị Hạnh là người chanh chua trong gia đình đó.
older sister Hanh is person lemon sour in family that
‘Ha is the most outlandish person in that family.’

Chanh chua means a confident and tomboyish manner. Chanh means lemon and chua means sour. When the two words co-occur as a single unit, it metaphorically represents an overconfident manner of a woman. The transfer of its semantic attributes can be illustrated as follows.

<table>
<thead>
<tr>
<th>Source Domain</th>
<th>Target Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taste term</td>
<td>habits or manners</td>
</tr>
<tr>
<td>chanh ‘lemon’ + chua ‘sour’</td>
<td>overconfident manner</td>
</tr>
<tr>
<td>[Fruit]</td>
<td>[Perceived taste]</td>
</tr>
<tr>
<td>[Acidic taste]</td>
<td>[Perceived personality]</td>
</tr>
<tr>
<td>[A unique taste]</td>
<td>[A unique personality]</td>
</tr>
</tbody>
</table>

Diagram 9: metaphorical mapping process of chua in chanh chua

From the above semantic mapping, attributes of chanh ‘lemon’ and its compounding element, chua ‘sour’, are transferred to the target domain, which is a habits/manners metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. Chanh transfers its semantic attributes from the source domain [acidic taste] to the target domain [outlandish color]. Chua transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived personality] and from [acidic taste] to [strike the eyes].

20) Cô gái ấy là người chua ngoa.
woman 3rd is person sour rude
‘She is rude’

Chua ngoa means a confident manner. Chua means sour and ngoa means rude. When the two words co-occurs as a single unit, it metaphorically represents an overconfident manner of a woman. The transfer of its semantic attributes can be illustrated as follows.
*Mon-Khmer Studies* 42.1:31-46 (ICAAL5 special issue)
Source Domain | Target Domain
---|---
Taste term | qualities
*giọng* ‘voice’ + *chua* ‘sour’ | a high-pitched voice
[Speech production] | [Perceived taste] | [Hearing voice]
[Irritating to throat] | [Irritating to ears]

**Diagram 12:** metaphorical mapping process of *chua* in *giọng* *chua*

From the above semantic mapping, attributes of *giọng* ‘voice’ and its compounding element, *chua* ‘sour’, are transferred to the target domain, which is a qualities metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. *Giọng* transfers its semantic attributes from the source domain [speech production] to the target domain [hearing voice]. *Chua* transfers its semantic attributes from the source domain [irritating to throat] to the target domain [irritating to ears].

23) Người thầy giáo ấy luôn nói ngọt với học sinh.

‘Teacher always speak sweet words with students’

*Nói* ngọt means to speak sweetly. *Nói* means to speak and *ngọt* means sweet. When the two words co-occur as a single unit, it metaphorically means to speak sweetly or pleasantly, which is a state of speech. The transfer of its semantic attributes can be illustrated as follows.

Source Domain | Target Domain
---|---
Taste term | state of being
*nói* ‘speak’ + *ngọt* ‘sweet’ | to speak sweetly
[Speech production] | [Oral perception] | [Auditory perception]
[Tasty look] | [Savory] | [Pleasant]

**Diagram 13:** metaphorical mapping process of *ngọt* in *nói* *ngọt*

From the above semantic mapping, attributes of *nói* ‘speak’ and its compounding element, *ngọt* ‘sweet’, are transferred to the target domain, which is a qualities metaphor, a part of the state metaphor. Details of the attribute transfer are shown as follows. *Nói* transfers its semantic attributes from the source domain [speech production] to the target domain [auditory perception]. *Ngọt* transfers its semantic attributes from the source domain [oral perception] to the target domain [auditory perception] and from the [savory] to [pleasant].

24) Giọng nói ngọt ngào.

‘Sweet tone of voice’

*Ngọt ngào* means sweet tone of voice. *Ngọt* means sweet and *ngào* means to mix in syrup. When the two words co-occur as a single unit, it metaphorically represents sweet voice. The transfer of its semantic attributes can be illustrated as follows.

Source Domain | Target Domain
---|---
Taste term | qualities
*ngọt* ‘sweet’ + *ngào* ‘to mix in syrup’ | sweet
[Oral perception] | [Conformity] | [Auditory perception]
[Savory] | [Outside] | [Outside aspects]
[Outside aspects] | [Pleasant]

**Diagram 14:** metaphorical mapping process of *ngọt* in *ngọt* *ngào*

From the above semantic mapping, attributes of *ngọt* ‘sweet’ and its compounding element, *ngào* ‘to mix in syrup’, are transferred to the target domain, which is a qualities metaphor, a part of...
the state metaphor. Details of the attribute transfer are shown as follows. Ngọt transfers its semantic attributes from the source domain [oral perception] to the target domain [auditory perception] and from [savory] to [pleasant]. Ngào transfers its semantic attributes from the source domain [outside] to the target domain [outside aspects].

From the data presented, it can be seen that semantic mapping of the taste terms has been transferred from the oral perception to various kinds of perceptions, be they visual perception, auditory perception, and mental perception, which is similar to Thai taste terms.

As for Vietnamese taste terms in combination with other words, it is found that most of the compounding elements are nouns, but one word is found to be a verb, nói ‘speak’, as in (23), which is similar to Thai taste terms. However, the compounding nouns are more numerous. For the compounding elements, if their meanings are related to perceptions, these shared properties with the taste terms are transferred to the target domain as well.

2.3 Metaphorical concepts of the Vietnamese taste terms

From examining metaphors of the taste terms, it is found that, besides realizing their comparative meanings, the metaphors enable us to perceive the concepts of language users, as reflected in Lakoff:

“...Metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.”
Lakoff (1980: 3)

Metaphor that reflects the concepts of language users is called a conceptual metaphor. From the above-mentioned examples, the relationship between the taste terms, metaphor and concept can be illustrated in the following figure.

![Figure 2: Process of conceptual metaphor of taste terms](image)

From the above diagram, it can be explained that taste terms are used to describe tastes. When they are used to compare things, they are regarded as a source domain and objects being compared are regarded as a target domain. This use of comparative taste terms reflects how language users have metaphorical concepts about the terms. During the comparison, attributes of the terms will be transferred to the objects being compared.

It is found in Vietnamese that metaphorical meanings of the four Vietnamese taste terms, they can be categorized into one semantic domain: state metaphor as mentioned earlier. The state metaphor was divided into three subgroups: feelings metaphor, habits/manners metaphor, and qualities metaphor. Thus, it can be generalized in terms of taste terms as concepts that HUMAN QUALITIES ARE TASTES. As shown in following figure.

![Figure 3: Process of conceptual metaphor of Vietnamese taste terms](image)
In summary, the taste terms in Vietnamese reflect metaphorical concepts among Vietnamese speakers: human qualities are tastes.

3. Summary and conclusion

It is found from the analysis that there are two major types of metaphorical use of Vietnamese taste terms. The first type is single taste terms with metaphorical meanings. The second type is the taste terms in combination with other taste terms with metaphorical meanings. This type is also divided into the taste terms in combination with other taste terms and the taste terms in combination with other words.

First, the taste terms occur singly and chua is the only one in this category. Second, the terms occurring with other taste terms and đắng and chua are found in this category. They generally occur as the first element of compound words. Finally, the terms occurring with other words and mặn, chua and ngọt are found in this category. They generally occur as the first element of the compound words.

For metaphorical meanings, it is found that the taste terms can be categorized into one domains: state metaphor as expressed by chua, đắng, mặn and ngọt.

Regarding metaphorical concepts, the terms are metaphorically used in one domain as mentioned earlier. The metaphorical use of the terms can be categorized into one concept: human qualities are tastes.

The investigation reveals that the findings are in line with the concept posited by Lakoff in that metaphor is related to the thinking system of human beings. When man perceives a new thing, it will be compared with existing experience and reflects the new concept in a language form rather than stating that a metaphor is a form of linguistic creativity of language users. When taste terms are considered, it is clear that they could not be categorized into any forms of conventional metaphors. Nevertheless, the taste terms do have metaphorical meanings that reflect the concepts of native speakers.

On the issue that human thinking system compares abstract objects to concrete objects; the findings from this investigation indicate that the taste terms are compared with many other things. It is still not clear whether the taste terms are concrete or not, as the terms are individualistic and idiosyncratic. For instance, a fruit may be sour to one but not to another. Anyway, the taste terms are regarded as a fundamental experience of humans, because every normal person can perceive tastes of various kinds. The comparison is based on daily fundamental experience in order to make it easy to understand. This thesis is also in accordance with embodiment theory in cognitive linguistics.

References


**Websites/online resources**

http://en.wikipedia.org/wiki/Sugar_addition
http://sealang.net/vietnamese/corpus.htm
http://www.bioresearchonline.com/article.mvc/Taste-Perception-Differs-Fundamentally-from-O-0001
http://www.scientificpsychic.com/workbook/chapter2.htm
http://icaal.org/ICAAL-4.2.pdf

**Definition of terms**

*Taste terms* are words that humans have created to express or describe various kinds of tastes that they perceived. In this investigation taste terms include bitter, salty, sweet and sour.

*Conceptual metaphor* is a linguistic form derived from comparatively using a word with its inherent meaning that refers to one entity to refer to another entity and the use of such a language form exists on a daily basis, which reflects a concept system of language users.

*Metaphor of taste terms* refer to a language form derived from a comparative use of taste terms to refer to other entities.

* Semantic domain is a group of words with shared semantic properties. In other words, it is a word group that native speakers have the concept (perception and understanding) on that the words have common properties, so they are classified into the same categories.
Source domain is a representation of word properties that native speakers have the concept about (perception and understanding). Some properties can be mapping to representation of words in the target domain.

Target domain is a representation of word properties that native speakers have the concept about (perception and understanding). And some features of those words can be linked with some features in the source domain. In this way, the target domain is being understood in terms of the source domain.

Semantic mapping is a process that shows a semantic relationship of words used in comparative manners. The mapping is unidirectional: from source domain to target domain.

Acknowledgement

This research had been accomplished with kind guidance, supports and encouragement from the following institution and individual to whom I would like to express my sincere thanks here.

I would like express my gratitude to Assoc. Prof. Dr. Sophana Srichampa, my major advisor, for her advice, care and encouragement during my times of studies and conducting the research; to Prof. Dr. Somsonge Burusphat, my co-advisor, for her invaluable suggestions and unfailing moral support; to Assoc. Prof. Dr. Chatchawadee Saralamba, my co-advisor, for providing me a perspective on cognitive linguistics and inspiration to investigate this topic; to Prof. Dr. Nick Enfield, my co-advisor, for his academic insight on the topic and care during my time at the Max Planck Institute; and to Dr. Nuntana Wongthai for her advice and recommendations to improve my dissertation.

I would also like to give my special thanks to the Royal Golden Jubilee Program for providing me a grant to conduct my study at the Max Planck Institute, where I have gained invaluable experience and academic connection.

Finally, I would like to offer my very special thanks to the ten informants and individuals who had contributed to the completion of my dissertation.
Khmuic classification and homeland

Paul SIDWELL
Australian National University
<paul.sidwell@anu.edu.au>

Abstract
The paper discusses the author’s comparative investigation of Khmuic in progress. Original proposals are made concerning the classification of Khmuic languages and their original homeland and migrations that established present distributions. The arguments are made with reference to the author’s working version of his phonological and lexical reconstruction. The evolution of the consonant system is found to be uninformative in regard to the classification, so special attention is given to an apparent chain shift in the reflexes of Austroasiatic *aː; broadly schematized as *aː > *eː > *iː > *iː* that supports strongly nested family tree. However, there are also counter-examples that fail to show these developments, such that we are forced to posit various parallel correspondences. To explain these facts, it is suggested that there were several phases of Khmuic expansion historically, each radiating from a homeland in the north-west of Laos, and resulting in dialect mixing that has confused the correspondence patterns.

The pKhmuic phonology as it is reconstructed to date is quite straightforward; lacking tones, registers or an implosive contrast in the stop series. Complex initial clusters are regarded as archaic, although relatively few are yet reconstructed on the basis of lexical comparisons. The vowel correspondences are somewhat complex, but this is consistent with an apparently rich history of dialect borrowing, so absolute regularity is not expected. On the whole pKhmuic resembles the Khmu Cuang dialect phonologically.

Keywords: Khmuic, classification, reconstruction, homeland
ISO 639-3 codes: kig, khf, tyh, prb, mlf, ptt, pry, mra, kjm, pnu, puo

1. Introduction
Comparative studies on Khmuic are relatively meagre; there are no published reconstructions of pKhmuic, and published classifications are contradictory and lacking justification. This is actually a typical situation in Austroasiatic (AA) linguistics where a branch is represented by one dominant and readily accessible language (e.g.: Khasi is well known but Khasian is neglected, Khmer is well studied but Khmer dialects get less attention etc.), and quite understandable in the circumstances. It is rather striking that if one erases the imprint the Lao language, the linguistic map of Northern Laos is dominated by a single language, Khmu (also spelled Kammu, Khmu’, Khmu’), whose speaker population comprises approximately a tenth of the population of the Lao PDR, and is the second largest ethnic group after the Lao Loum. An additional handful of languages make up the remainder of Khmuic branch of Austroasiatic, some of which also spill over the borders into Thailand, China and Vietnam.

Khmu has been known to scholars since the 19th century (e.g. lexicon of Khmu features in the materials of the Garnier expedition; Garnier 1873), and Khmu comparisons played a crucial role in Haudricourt’s famous (1953, 1954) reconstruction of Vietnamese tonogenesis, so the historical importance of the branch has been appreciated. However, with much scholarly attention focussed on Khmu (e.g.: Smalley 1963, Delcros & Subra 1966, Lindell et al. 1981, Svantesson 1983, Preisig et al. 1994, Premririt 1993, 2002) at the expense of the lesser Khmuic tongues, a fuller understanding of the branch has taken time to emerge. The situation began to improve especially from the 1970s e.g.: Filbeck (1971, 1978, 2009) T’iinic, Pogibenko & Bui Khanh-Thê (1990) on Ksimgmul, Rischel (1989a,b, 1995, 2007) and Rischel & Egerod (1987) and Egerod & Rischel (1987) on Mlabri, Bui Khanh Thê (2000) on Phng/Kaniang, Ferlus (1970) and Dang Nghiem Van (1983) on Thai Hat/Odu, Maspéro (1955) on Theng.
Given the relatively small size of the branch, the availability of some reliable and lengthy lexicon, and the fact that the major contact languages (Lao, Vietnamese, Thai, Tai etc.) are well known, one would predict that a comparative reconstruction - at least of phonology and lexicon - should be a practical proposition. This is made all the more reasonable by the extent of low-level comparative work that has already been published; it focuses specifically on:

- Mlabri and T’inic sub-groups, found in the more accessible Nan Province of Northern Thailand and the adjacent Sayabouly Province of Laos (Filbeck 1978, Rischel 1989b, 2007); and

The above studies follow Haudricourt (1965), who compared Thinic, Mlabri and Khmu data, demonstrating the conservatism of Khmu and Mlabri consonants and the innovative restructuring of initial stops in the Thinic dialects. Consequently, while these works clarify a tremendous amount of Khmuic historical phonology, until now scholars have apparently not tried to synthesize this body of work and model the phonology of pKhmuic and the phonological divergences that mark the diversification of the branch. The principle explanation for this lies in the tendency for scholars to specialise in a specific language or group of related dialects; additionally the contemporary emphasis on language documentation imposes extensive burdens on field linguistics making it even less attractive to engage in comparative studies.

The problems of assembling and organising materials to support a reconstruction are real and multiple. First of all, there is actually no agreement on which languages are necessarily counted within the branch. Khang/Khao and Khabit/Phsing are treated by Diffloth (ms.) as Khmuic but this writer considers them to be Palaungic with Khmuic strata (see discussion below). Rischel, in several publications (including 1995, 2007) wonders whether Mlabri is an independent branch of Austroasiatic that was relexified with Khmuic loans, although this view seems to have no significant support. Secondly, the lexical sources that do exist vary extensively in their lexical coverage, and even when they do overlap, lexical replacement within languages has been so great that it can be very difficult to identify cognates. And finally, it is also clear that there has been substantial inter-dialectal borrowing within Khmuic, much of it from Khmu into the smaller languages, and as Rischel has variously identified, from T’inic into Mlabri. These problems are significant, but not insurmountable, and in this paper I offer a first framework of a pKhmuic reconstruction, and strategy for dealing with the problems of borrowings between closely related members of the group.

2. Classification

2.1 Defining Khmuic

The first problem is to determine the membership of the Khmuic branch. Proschan (1996) provides the following list:

<table>
<thead>
<tr>
<th>Language</th>
<th>Alternate names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kmhmu</td>
<td>(many¹)</td>
</tr>
<tr>
<td>Phong</td>
<td>Tay Phong</td>
</tr>
<tr>
<td>Thin</td>
<td>Mal, Pray</td>
</tr>
<tr>
<td>Ksing Mul</td>
<td>Puok, Pou Hok</td>
</tr>
<tr>
<td>Bit</td>
<td>Khabit</td>
</tr>
<tr>
<td>Mlabri</td>
<td>Phi Tong Luang</td>
</tr>
<tr>
<td>Theen</td>
<td>Kha Sam Liam</td>
</tr>
<tr>
<td>Iduh</td>
<td>Tay Hat</td>
</tr>
<tr>
<td>Khang</td>
<td>Mang U</td>
</tr>
</tbody>
</table>

Effectively the same list is provided by Chazée (1999), Diffloth (2003), Anderson (2006) and elsewhere. All sources agree on the following basic groups, plus a couple of doubtful languages:

¹ Note Filbeck’s spelling “Kmhmu” for the name of the language. Proschan (1997:97) lists 35 (!!) different romanized spellings attested in the literature.

• Khmu, Kmhmu’, Kammu etc. (many dialects)
• Mlabri, Yumbri (Phi Tong Luang)
• Thinic: Thin, Mal, Pray, Phay, Lua’, Lawa
• Khsingmul (Puok, Puoc, Pou Hoc)
• Pramic: Phong (Pong, Kanieng), Odu (Idu, Thai Hat)

Figure 1: Fragment of Language map (Diffloth 2001): Khmuic language light blue, Palaungic dark blue, Vietic yellow. Misclassified languages Khabit and Khang circled.

The doubtful languages are Khabit and Khang (Khao, Mang U), listed by Parkin (1991) as Khmuic. The most recent lexicostatistical study (Peiros 2004, see Figure 3) recognises both of these as Khmuic, although places them in a sub-group branching above the rest of the group. Diffloth (1982) suggested a Palaungic affiliation for Khabit, and this is confirmed below, with data from Kingsada & Kosaka (1999). More controversial is Khang (ISO 639-3 kjm, xao, not to be confused with Mang zng, spoken either side of the Vietnam-China boarder), for which Mikami (2003) is the source used here. The problem arises in the first place because both of these languages are effectively adjacent to other Khmuic speakers (see map at Figure 2) and not in contact with Palaungic, and therefore on geographical grounds we would be surprised if they were anything other than Khmuic. Throughout AA a tendency is that multiple branches don’t overlap so much, with the most marked exceptions involving Pearic and Vietic (due to the expansion and dominance of Khmer and Vietnamese as national languages).

Comparative analysis demonstrates that Khabit and Khang show significant Palaungic lexical innovations, plus a shared phonological development (raising of *a: to u/ɯ/ɔ²) that strongly hints at subgrouping

---

2 Interpretation of the sources is somewhat problematic, but it would appear that u and ɯ/ɔ are effectively indicating the same phoneme, a high unrounded non-front vowel.
Table 2: Lexical comparisons supporting Palaungic classification of Khabit, Khang (plus other forms bracketed)

<table>
<thead>
<tr>
<th>Khmuic</th>
<th>Mlabri</th>
<th>Mal</th>
<th>Phong</th>
<th>Khsing-Mul</th>
<th>Khabit</th>
<th>Khang</th>
<th>Palaung</th>
<th>Lamet</th>
<th>Khasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>'eye'</td>
<td>mat</td>
<td>mat</td>
<td>mat</td>
<td>mat</td>
<td>pŋaːj</td>
<td>ŋaj²</td>
<td>ŋaj</td>
<td>ŋaj</td>
<td>ŋaj</td>
</tr>
<tr>
<td>'fire'</td>
<td>—</td>
<td>ʔuːlh</td>
<td>ʔoːh</td>
<td>ʔoːs (halōy)</td>
<td>ʔŋal²</td>
<td>ʔŋal²</td>
<td>ʔŋor</td>
<td>ʔŋal (diŋ)</td>
<td></td>
</tr>
<tr>
<td>'sated'</td>
<td>biʔ</td>
<td>biʔ</td>
<td>piʔ</td>
<td>kʰːj (ʔkʰːj)</td>
<td>suːk</td>
<td>sɪʔ²</td>
<td>huʔ</td>
<td>saːk (hun)</td>
<td></td>
</tr>
<tr>
<td>'blood'</td>
<td>ma m</td>
<td>ma m</td>
<td>miːm</td>
<td>miːm</td>
<td>sʰnum</td>
<td>num²</td>
<td>snaːm</td>
<td>naːm</td>
<td>snaːm</td>
</tr>
<tr>
<td>'laugh'</td>
<td>kʰːras</td>
<td>--</td>
<td>khieh</td>
<td>kʰːrh</td>
<td>kʰːlō</td>
<td>ʔŋal¹</td>
<td>--</td>
<td>kʰːŋaːs</td>
<td>rkʰːe</td>
</tr>
<tr>
<td>'moon'</td>
<td>(moŋ)</td>
<td>kiʔ</td>
<td>(thuaʔ)</td>
<td>kʰː (bluːn)</td>
<td>(kʰːrɪŋ) (kʰiaŋ¹)</td>
<td>(kʰːr)</td>
<td>kʰːʔ (bnaj)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'water'</td>
<td>(?om, ʔɔːk)</td>
<td>(jраːk)</td>
<td>(?ʔk)</td>
<td>(?ʔʔʔ)</td>
<td>(ʔh:t)</td>
<td>ʔom</td>
<td>ʔom²</td>
<td>ʔom</td>
<td>ʔom</td>
</tr>
</tbody>
</table>

The Khmuic branch is readily defined by a very specific phonological innovation, the loss of pAA medial *h, which is evident in the reflexes of ‘blood’ in the table 2 (and other etyma including ‘sated’ and ‘moon’ in Table 2). Unfortunately, Khabit and Khang reflexes of the specific etyma expected to show pAA medial *h are lacking in the sources available to this writer, however they do show specific Palaungic lexical innovations. Of particular note, Khabit and Khang reflect pPalaungic *ŋal ‘fire’, *ʔŋaːj ‘eye’ and *saːk ‘sated’. Additionally Khabit and Khang reflect etyma for ‘blood’ and ‘water’ that are otherwise restricted to Palaungic and Khasian, except that ʔom ‘water’ is borrowed into Khmu (indicated by its narrow distribution in Khmuic and the short vowel). Also Khabit and Khang show the distinctive Palaungic form for ‘laugh’ with medial palatal nasal. The presence of Palaungic and Khasi-Palaung innovations in two languages which are not in contact with Palaungic, and are under strong influence form other groups, is best explained as indicating that these are actually Palaungic languages displaced by migration.

2.2 Internal classification

Studies and reference literature on Khmuic classification have been sparse and ambiguous; a typical pattern, such as listed at Table 1. A rare exception is Chazée (1999), providing the following tree diagram, citing Diffloth & Proschan as the sources, although no bibliographic reference is given.

Figure 2: Khmuic languages tree from Chazée (1999)

There is also the lexicostatistical tree offered by Peiros (2004) reproduced at Figure 3. My view is that we ought to disregard this analysis as it is not grounded in historical phonology, and it is distorted by a failure to properly identify borrowings that accounts for his placement of Khmu between Pramic (Phong/Odu) and Ksingmul. More interesting is Chazée’s (1999) tree, which does strong correspond to the analysis based on phonology - specifically vowel changes - offered here.

In the first place we would expect classification to be based upon innovations identified by a comparative reconstruction. The present author has been compiling a Khmuic comparative lexicon and developing a reconstruction of pKhmuic phonology and lexicon, and the first version of this was released online in 2013 at sealang.net/monkhmer. It must be acknowledged up front that this reconstruction is limited by the factors discussed above, which mean that most etymologies are incomplete, but it is far from an elaborate untestable hypothesis. This is because the bulk of the 750+ etymologies so far compiled are firmly grounded in deeper AA etymologies documented in published references (primarily Shorto 2006). The main consonantal correspondences3 underlying the reconstruction are set out at Table 3.

### Table 3: Khmuic consonant correspondences underlying the reconstruction of Sidwell (2013)

<table>
<thead>
<tr>
<th>Initials</th>
<th>pKhmuic</th>
<th>K-Cuang</th>
<th>Mlabri</th>
<th>Mal</th>
<th>Khsing-Mul</th>
<th>Phong</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>p</em></td>
<td>p</td>
<td>p</td>
<td>ph</td>
<td>p</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td><em>t</em></td>
<td>t</td>
<td>t</td>
<td>th</td>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td><em>c</em></td>
<td>s</td>
<td>ch</td>
<td>s</td>
<td>c</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td><em>c₁</em></td>
<td>c</td>
<td>ch</td>
<td>s</td>
<td>c</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td><em>k</em></td>
<td>k</td>
<td>k</td>
<td>kh</td>
<td>k</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td><em>ʔ</em></td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
<td>ʔ</td>
</tr>
<tr>
<td><em>b</em></td>
<td>b</td>
<td>b~b</td>
<td>p~mp</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td><em>d</em></td>
<td>d</td>
<td>d</td>
<td>t~nt</td>
<td>d</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td><em>j</em></td>
<td>j</td>
<td>j</td>
<td>c~pc</td>
<td>c</td>
<td>j, s (/C)</td>
<td>j, s (/C)</td>
</tr>
<tr>
<td><em>g</em></td>
<td>g</td>
<td>g</td>
<td>k~ŋk</td>
<td>g, k (/R)</td>
<td>g, k (/R)</td>
<td></td>
</tr>
<tr>
<td><em>m</em></td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td><em>n</em></td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td><em>ŋ</em></td>
<td>ŋ</td>
<td>ŋ</td>
<td>ŋ</td>
<td>ŋ</td>
<td>ŋ</td>
<td>ŋ</td>
</tr>
<tr>
<td><em>w</em></td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td><em>r</em></td>
<td>r</td>
<td>r</td>
<td>j</td>
<td>g, l (/p_), r (/C_)</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td><em>l</em></td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td><em>j</em></td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>z</td>
<td>j</td>
<td>j</td>
</tr>
<tr>
<td><em>s</em></td>
<td>h</td>
<td>th~ch</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td><em>h</em></td>
<td>h</td>
<td>h</td>
<td>h, Ƿ (/C)</td>
<td>h</td>
<td>h, Ƿ (/C)</td>
<td>h, Ƿ (/C)</td>
</tr>
</tbody>
</table>

3 These are the correspondences established with multiple etymologies and confirmed with external comparisons.
The proposal offered here is that the developments of pKhmuic *aː; *a are indicative of the internal classification of Khmuic. Changes in the consonants found within Khmuic are of the kind that involve devoicing and or aspiration, in other words delay in voice onset timing, and such are typical of both areally conditioned change and spontaneous drift. Within Khmuic, devoicing of stops is attested even between dialects of the same language (e.g. southern Khmu dialects show devoicing and registrogenesis while northern dialects do not: see Premsrirat 2001, 2004). However, there is a particular pattern of correspondences among reflexes of the long and short low central vowel which is taken as significant because these vowels are otherwise quite stable, and no conditioning factors are evident. In summary, there are apparently four distinct correspondences reflecting what are assumed to be only two pAA vowels. These are illustrated in the examples at Table 4.

### Table 4: Lexical comparisons illustrating pKhmuic *aː, *a, *a developments

<table>
<thead>
<tr>
<th>Finals</th>
<th>*p</th>
<th>*t</th>
<th>*k</th>
<th>*h</th>
<th>*m</th>
<th>*n</th>
<th>*r</th>
<th>*l</th>
<th>*j</th>
<th>*s</th>
<th>*h</th>
<th>*Ø</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKhmuic</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Khmu Chuang</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Mlabri</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Mal. Pray</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Khsing-Mul</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Odu</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Pong</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
</tbody>
</table>

The situation is that there appears to be two distinct correspondences each for *aː and *a, for which the notation *a; *a, *a; *a is adopted. pKhmuic *aː and *a have reflexes that are essentially unchanged across the branch, while *a; *a have phonologically marked reflexes: beyond Khmu reflexes of *aː show fronted and raised vowels, while for *a it is the Khmu reflexes that are marked, being raised and central, and in Mlabri a little raised, but otherwise little changed in the rest of the branch. The asymmetries in these changes strongly suggest that this is not a vestige of an earlier register system. Elsewhere among the vowel correspondences there is no indication of these kinds of split correspondences; admittedly some of the correspondences are

SIDWELL, Paul. 2014. Khmuic classification and homeland. Mon-Khmer Studies 43.1:47-56 (ICAAL5 special issue)
difficult to interpret, but this is due to lack of regularity (probably due to dialect borrowing) whereas the correspondences at stake here are quite regular. The interpretation of *aː; *aː; *a, *aː; offered here is as follows:

- *aː; is a regular reflex of pAA *aː; the conditioning is obscure, but all involve voiced initials and all but three so far identified have final nasals. I propose that the unchanged forms with similar environments are loans, presumably from Khmu into the other Khmuic languages, although some could also have come in from other AA contact.
- *aː; is less frequent, being only about 1/10 the frequency of *a, so it is quite marked. Also, there are sporadic examples of high vowel reflexes in other languages (e.g. Odu bil ‘thick’) which look suspiciously like loans from Khmu. However, there is no clear indication of conditioning; both voiced and voiceless initials are found, and stop and continuant finals, nor is there any evident semantic link. So it looks like a poorly understood change that originated within Khmu, and may have diffused out in some loans.

Consequently the *aː; correspondence seems to be relevant for Khmuic internal branching. If, on general phonetic grounds, we assume that there was a raising of pAA *aː; in the sequence *aː; > *ɛː > *iː, the branching/subgrouping indicated in the following figures (both the Venn diagram and family tree modified from Chazée (1999) are indicated:

### Figure 4: Venn diagram representation of *aː; developments

### Figure 5: Khmuic tree based on *aː; developments

#### 3. Homeland and migration

The proposed classification we have thusly arrived at is strongly nested, with a primary split between Khmu and a Mlabri-Pram sub-branch that diversified into the rest of the branch. However, the geographical interpretation is problematic, as we do not have an obvious centre of diversity, which would suggest a homeland location on the basis of the assumptions of dialect geography. Mlabri-Pram languages are spread over a wide area, from Northern Thailand to North Vietnam, and even the Phray-Pram sub-group below this has a similar distribution, only the very low level Pramic dialect chain has a narrower distribution in the eastern part of the Khmuic range. Khmu
also has a similarly wide range, although it is evident that the phonologically conservative dialects, and the greatest diversity of Khmu dialects overall, is quite localised to the northwest of Luang Phrabang, over more or less the area of present day Oudomxay Province (interestingly in proximity to the Palaungic languages Lamet and Khabit). These facts suggest that the Khmu language diversified and spread out of Oudomxay historically.

The next split in the tree is between Mlabri and Phray-Pram; both Mlabri and part of the Phray-Pram group (specifically the Thinic languages) are localised southwest of Luang Phrabang, while all of the non-Khmu languages in the eastern range fall within the Phray-Pram group (specifically Pramic and Khsing-Mul). This suggests a very specific migration path: the first split saw a group move directly south out of Oudomxay, over the Mekong, and into Sayabouly, where Mlabri and Thinic speakers still live today. A subsequent movement east into the region of Houaphan Province and adjacent Vietnam, and later diversification in the area, give us the Khsing-Mul and Pramic dialects.

The above scenario, however, remains incomplete. Khmu dialects are spoken as far west and even further south than Mlabri-Phray languages, and those Khmu dialects are relatively homogenous (see Lindall et al. 1981, Premsrirat 2002), indicating a fairly recent dispersal (although surely before Lao established social dominance, so some time in the first half of the 2nd Millennium is likely). This clearly suggests another phase of Khmuic expansion that saw Khmu speakers spread out over and among the various Mlabri-Pram communities after the latter had become differentiated to some extent and established in their present ranges, and offers an explanation of the problematic split correspondences *aː, *aː.

For whatever reasons, the Khmu are substantially more numerous and enjoy more prestige and status than their fellow Khmuic speakers. The linguistics and geographical facts suggest that there was a period of Khmu dominance of Northern Laos, with Khmu influence over a diversity of smaller Khmuic communities, before this gave way eventually to Lao and Vietnamese hegemony (although may still exist to some extent). Such a phase of Khmu dominance could have seen substantial relexification with Khmu words that do not show the sound changes (such as the raising of *aː) that mark the non-Khmu sub-groups.

Figure 6: Map indicating proposed Khmuic homeland (brown oblong) and migration routes: black arrows indicate initial movement of Mlabri-Pram and subsequent Phray-Pram migrations; purple arrows suggest later Khmu radiation(s).
4. Conclusion

This short paper has presented evidence for Khmuic subgrouping and homeland localisation based on the author’s emerging comparative reconstruction. The data discussed here are quite limited, but strongly suggestive of the conclusion that Khmuic originated in the area of Oudomxay, and that several phases of out-migration originated from this area, the later associated with Khmu dominance over diverse smaller communities. Moving forward, the challenge is to further build the comparative lexicon and the identification of regular correspondences and indication of borrowing. The work is extremely problematic, but the discussion present here, including the tabled correspondences, provides a framework for progress in Khmuic reconstruction which offers both linguistic and historico-cultural explanation.

References


Filbeck, David. 2009 Mal (Thin) - Thai - English Lexicon. Unpublished manuscript.

Garnier, Francis. 1873. Voyage d'exploration en Indo-Chine effectué pendant les années 1866, 1867, et 1868 par une commission française, présidée par Doudart de Lagré, et publié...sous la direction de...Francis Garnier. 2 volumes. Paris, Hachette.


Transitivity and affectedness in Mon

Mathias Jenny
University of Zurich

Abstract
The present study investigates the different uses of causative/transitive directionals in Mon and the functional differences between the basic and causative forms. Dealing with a typologically rare phenomenon, this study adds to our understanding of complex verbal predicates and transitivity not only in the Southeast Asia context, but also cross-linguistically. The study is based on original data collected in Thailand and Myanmar from different varieties of Mon, supplemented by published texts such as journal articles and short stories, as well as elicited data.

Keywords: Mon, syntax, transitivity
ISO 639-3 codes: mnw, shp, sbe

1. Introduction

Many secondary verbs, including directionals, in Mon appear in two forms, basic/intransitive and causative/transitive. The causative/transitive forms are in either morphological causatives or suppletive lexemes. The choice of the form of the directional employed depends on the movement or affectedness of the participants of an expression, rather than the transitivity value of the main predicate. If the S/A argument is described as moving by the main verbal predicate, the basic form of the directional is used, if the P (or T) argument is set in motion, the causative form of the directional is obligatorily used. In transitive expressions, the basic form is used if the A rather than the P argument is set in motion, or if the setting in motion of P is backgrounded. In ditransitive expressions, the causative directional refers to the movement of the T, never the G argument. The main trigger for the choice of the directional is apparently the “affectedness of the O argument” (Hopper & Thompson 1980). This systematic distinction between basic and causative directionals, which is rare not only in Southeast Asian languages, but also globally, allows a distinction in the degree of (semantic) transitivity of an event based on the linguistic expression. It can be shown, for example, that morphological causatives in Mon have a higher degree of transitivity than periphrastic causatives, as only the former trigger the causative directionals.

2. Transitivity

The notion of transitivity covers both syntactic and semantic transitivity, and the phenomena are often treated together, though there are important differences between the two. As it is semantic transitivity that is relevant to the present study, syntactic transitivity in Mon will only briefly be outlined here, before giving a more detailed account of semantic transitivity and related features.

2.1 Syntactic transitivity

Verbs in Mon can take one, two or three arguments, that is they can be syntactically intransitive, transitive, or ditransitive. There is probably only one real ditransitive verb in Mon, namely kn ‘give’, which occurs in the pattern A V G T. With other ‘ditransitive’ predicates, such as həɓah ‘show’ and pəciʔəʔ ‘feed’, the recipient G is obligatorily marked by the oblique marker kn, which is homonymous with the verb kn ‘give’, and the structure is A V kn G T.

Transitive verbs take two arguments, A and P, which may be overtly expressed or left understood in a sentence, if their referents are known or recoverable from the linguistic or extralinguistic context. Typical transitive verbs include ciaʔ ‘eat’, chan ‘love’, and causatives like həcɛr ‘kill’. These verbs may be labeled unrestricted transitives, as they felicitously combine with an object of any semantic type, possibly resulting in non-sensical, but grammatical collocations.

Intransitive verbs are verbs that cannot take more than one argument. Their class is probably rather small in Mon, compared to intransitive verbs in European languages, as many verbs may take a direct object from a semantically restricted set of nominals. This is for example true for...
directed motion verbs (directionals), which combine directly with a noun expressing a conventionalized location, as seen in example (1).

(1) Restricted transitives
   * cao hiəʔ ‘return house’ ‘return home’
   * ʔa phja ‘go market’ ‘go to the market’
   * ceh dac ‘descend water’ ‘go down into the water’
   * məŋ phəɔ ‘stay monastery’ ‘stay at the monastery; be at school’

   These restricted transitive verbs contrast with unrestricted transitives seen above in that they grammatically combine only with an object of a closed set. Intransitive verbs as shown in (2), on the other hand, cannot combine with an object, even if it is semantically related, without an overt marker such as the oblique ko or the locative də. Besides undirected motion verbs, the class of intransitive verbs also includes expressions like mìp ‘be happy, enjoy’, toc ‘sleep’, and khjɪt ‘die’.

(2) Intransitives
   * kwac phja ‘walk market’ intended: ‘walk to the market’
   * nùm hiəʔ ‘exist house’ intended: ‘be at home’
   * mìp puə ‘happy fair’ intended: ‘enjoy the fair’
   * khjɪt kə a ‘die thirst’ intended: ‘die from thirst’

   Syntactic transitivity is always a discreet notion, each verb having a fixed transitivity value (valency) of 1, 2, or 3 arguments. Semantic transitivity, on the other hand, is more flexible, as will be seen in the following paragraph.

2.2 Semantic transitivity

A number of authors have dealt with the notion of transitivity, establishing a number of factors that make an expression more or less transitive. In their seminal paper on the topic, Hopper and Thompson (1980) list ten parameters that define transitivity:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HIGH</th>
<th>LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>2 or more</td>
<td>1</td>
</tr>
<tr>
<td>Kinesis</td>
<td>action</td>
<td>non-action</td>
</tr>
<tr>
<td>Aspect</td>
<td>telic</td>
<td>atelic</td>
</tr>
<tr>
<td>Punctuality</td>
<td>punctual</td>
<td>non-punctual</td>
</tr>
<tr>
<td>Volitionality</td>
<td>volitional</td>
<td>non-volitional</td>
</tr>
<tr>
<td>Affirmation</td>
<td>affirmative</td>
<td>negative</td>
</tr>
<tr>
<td>Mode</td>
<td>realis</td>
<td>irrealis</td>
</tr>
<tr>
<td>Agency</td>
<td>A high in potency</td>
<td>A low in potency</td>
</tr>
<tr>
<td>Affectedness of O</td>
<td>O totally affected</td>
<td>O not affected</td>
</tr>
<tr>
<td>Individuation of O</td>
<td>O highly individuated</td>
<td>non-individuated</td>
</tr>
</tbody>
</table>

   In a given expression, the value for each parameter may be either HIGH or LOW. According to Hopper & Thompson (1980), a clause is high in transitivity if it fulfills a high number of a set of factors in the HIGH column (p. 252), such as the presence of two or more participants, description of an action (rather than a state or non-action), telicity, punctuality, volitionality of the A argument, affirmation (rather than negation or questioning) of the situation described, realis mode, A argument high in potency, O argument totally affected and highly individuated. These factors involve different components of the effectiveness or intensity with which the action is transferred from the A to the O participant. “Each component of Transitivity involves a different facet of the effectiveness or intensity with which the action is transferred from one participant to the other.” (p. 252)

   It is evident from the above list that the parameters are almost purely semantic, with only the first involving a syntactic criterion. The system applied by Hopper & Thompson to establish the transitivity value of an expression also clearly shows that transitivity is values are gradual, rather
than discreet values. The more components there are with [HIGH] marks, the more the situation is transitive.

Hopper & Thompson’s definition of transitivity has been widely received and still is considered a classic on the topic. One point of criticism that has been raised is that all parameters are apparently given equal importance, though some may be more relevant to the notion of transitivity than others. Also, a number of parameters logically correlate, so that it is not clear how valid they are in establishing the transitivity of an expression.

Subsequent authors have addressed a number of these issues, including the question of syntactic vs. semantic transitivity (e.g. Kittilä 2002). Kittilä distinguishes syntactic (structural) transitivity from semantic transitivity. In the latter, agency of the instigator and direct affectedness of the patient along with an efficient transfer of energy from agent to patient must be present (Kittilä 2002:38). The state of the patient before and after the event must be distinguishable. (p. 41) Agency and affectedness are gradual, rather than binary features, that is, they form continua. Agent may be more or less agentive, and a patient more or less affected.

Kittilä (2002:39f) breaks up a transitive event into four distinct phases, namely: planning-initiation-event-result. All four phases are relevant to the transitivity value of an event/clause, and the further an event develops towards stage four, the higher is its transitivity value. As the stages are inherently ordered temporally, they also imply a hierarchical structure. Events that contain only the planning and instigation stages are less transitive than events that contain only the event and result stages. The former include intended acts that are not carried out to completion and therefore do not (fully) affect the patient, while the latter include non-intentional acts that nevertheless affect the patient. Individual languages may mark high transitivity by syntactic means, such as ergative marking on the agent (pp. 61ff). This shows that semantic transitivity may be relevant also in the syntax of a language. As will be shown below, Mon marks at least one type of high transitivity by the use of special forms of directionals.

In another approach, Næss (2007) uses three features that are involved in the notion of (prototypical) transitivity. Both participants in a transitive have + or - values for volitionality, instigation, and affectedness: [+VOL, ±INST, ±AFF]. Prototypically, transitive events show the following constellation of agent and patient.

Agent (prototypical) = [+VOL, +INST, -AFF]  
Patient (prototypical) = [-VOL, -INST, +AFF]

The agent volitionally instigates the event which affects the patient, but not the agent. The patient is neither volitional nor actively instigating the event. The ‘volitionality’ corresponds roughly to Kittilä’s ‘planning’ stage, ‘instigation’ to Kittilä’s ‘initiation’, and ‘affectedness’ to Kittilä’s ‘result’. Kittilä’s ‘event’ stage is probably included partly in both Næss’s ‘instigation’ and ‘affectedness’. Though the different analyses do not match exactly with one another, all accounts of transitivity apparently take two factors as (equally) crucial: agency/volitionality of the agent and affectedness of the patient. As will be seen below, in Mon the affectedness of the patient is the main characteristic of (high) transitivity.

2.3 ‘Transitivity harmony’

In multi-verb predicates, a number of languages exhibit what has been termed ‘transitivity harmony’ (see e.g. Valenzuela 2011). According to Valenzuela (2011:186), “Transitivity Harmony is understood as a morphosyntactic process whereby a semantically modifying verb or verbal morpheme adjusts its valency to match the transitivity value of a semantically main verb with which it combines, either in a mono-clausal or chained construction.” In Shipibo-Konibo (Pano, Amazonia; Valenzuela 2011), basically transitive phasal verbs and modal auxiliaries are detransitivized by a middle suffix when combined with an intransitive main predicate. The transitivity harmony is structural/syntactic and works also across clause boundaries in non-nuclear serialization and clause chaining. Similar structures to the ones found in Saliba also occur in Rawang (LaPolla 2010). In the case of Rawang, too, basically transitive secondary verbs are detransitivized in combination with an intransitive main verb. Examples (3) to (6) illustrate syntactic transitivity harmony in Shipibo-Konibo (from Valenzuela 2011).
JENNY, Mathias. 2014. Transitivity and affectedness in Mon. *Mon-Khmer Studies* 43.1:57-71 (ICAAL5 special issue)

(3) sani-n-ra "jobin "tsaka-xon pake-ke.
PN-ERG-EV fruit:ABS hit-P.SS.AO cause.to.fall-COMPL
’Sani hit the fruit and caused it to fall down.’ (192)

(4) jaka-jaka-kin "oin-a-ronki
sitting.INTR:MID-sitting.INTR:MID-SIM.EVENT.SS.AO see-PP2-HSY
ik-a iki westiora bimi ani jiwirema-x
be-PP2 AUX one fruit:ABS big tree:LOC:ABL-so
pake4-i jene-nko-shaman.
drop-MID-SIM.EVENT.SS.SO flowing.water-LOC-INTEN
‘While sitting there, he saw that a fruit fell from a big tree into the water.’ (192)

(5) e-a-ra ranxa-i peeokoo-ke.
1-ABS-EV dance-SIM.EVENT.SS.S begin:MID-CMPL
‘I began to dance (e.g., at a party).’

(6) e-n-ra (xeki) bana-kin pee-ke.
1-ERG-EV (corn:ABS) sow-SIM.EVENT.SS.AO begin-CMPL
‘I began to sow (it/the corn).’

In Saliba, an Oceanic language of Papua New Guinea, verbs can be combined into multiverb predicates of the nuclear serializing type (Margetts 1999:99ff). The complex verbal predicate functions as one unit with one S/A prefix and one P suffix, that is, all arguments are shared by the constituent parts of the multi-verb predicate. Sharing all arguments, all the verbs of the complex predicate must have the same transitivity value, which according to Margetts (1999:102) is due to the ‘same-subject constraint’.

Margetts (1999:58) describes “transitivity as a system of discrete morpho-syntactic features”, that is, it is syntactic transitivity that is relevant in the description of the language, presumably also to the transitivity harmony in Saliba. Margetts (1999:102) states that “generally, the transitivity status of complex verbs is determined by the initial stem of the construction and the following stems agree with it in transitivity status” and that “it is not possible for a non-initial stem to add a further argument.” If an intransitive non-initial verb is combined with a transitive initial verb, it must be transitivized by the causative marker he. This is illustrated in examples (7) to (10), all from Margetts (1999:103ff). The use of a non-causative secondary verb in these expressions is ungrammatical.

(7) je-koi-he-mwaloi-Ø. je-koi-mwaloi-Ø
3SG-hit-CAUS-dead-3SG.O ‘He hit it dead.’

(8) ku-he-sigi-sae-Ø.
2SG-CAUS-move-go.up-3SG.O ‘Move it up.’

(9) je-koi-he-beku-Ø.
3SG-hit-CAUS-fall-3SG.O ‘He made it fall down.’

(10) kaputi ku-inti-he-mwajau-Ø.
cup 2SG-pour-CAUS-full-3SG.O
‘Pour the cup full.’

If more than one intransitive verbs combine with a transitive initial verb, the causative marker is added only to the first of these, having scope over all following verbs, as seen in (11).

(11) je-sikwa-he-beku-dobi-ei-Ø.
3SG-poke-CAUS-fall-go.down-APP-3SG.O
‘He poked it down.’
According to Margetts (1999:143) “Complex verbs play an important role in the description of valence and transitivity in Saliba [...] and provide tests for word-level transitivity as well as root valence.” Although Margetts apparently takes syntactic transitivity as trigger for transitivity harmony in Saliba, all the above examples also involve (high) semantic transitivity. There are in fact exceptions to the transitivity harmony, and therefore to the same-subject constraint, as in example (12) from Margetts (1999:103). Though kita ‘see, look’ is syntactically transitive, the directional sae ‘go up’ is not transitivized in the complex predicate. This may be due to the fact that the (unexpressed) patient is not directly affected by the event, that is, the event has a low semantic transitivity value.

(12) je-kita-sae.
3SG-see-go.up
‘He looked up.’

After setting the scene for transitivity and transitivity harmony, we now turn to Mon, which exhibits a system apparently very similar to Saliba.

3. Complex verbal predicates in Mon

Multi-verb constructions are a common feature of Mon, as they are in most other languages of mainland Southeast Asia. Mon makes use mostly of multi-verb predicates of the nuclear serial type, that is, all verbs making up a complex predicate are adjacent to each other. Arguments occur before (S and A) or after (P, T, and G) the complex predicate in the following pattern: A V V; A V V P, A V V G T. The position of the individual verbs in a multi-verb predicate may vary, resulting in different readings. The directed motion verb ṭa ‘go’ in (13) has a different function from the same verb in (14), due to the different positions they occupy in the verb. While (13) can be seen as expressing two events (going and buying), (14) consists of only one event. It is evident that though all three verbs in (13) are syntactically transitive, they only share the S/A argument, not the patient. In (14), the motion verb ṭa ‘go’ functions as an orientation verb, together with the directional cao ‘return’. The transitivity value of the two directionals cao and ṭa is different from the value of the intransitive initial predicate kwac ‘walk’. In other words, secondary verbs can introduce new arguments in Mon, unlike in Saliba.

(13) ṭa ṭa rān ciaʔ kwag.
1SG go buy eat snack
‘I went to buy a snack to eat.’

(14) dč kwac cao ṭa hnaʔ.
3 walk return go house
‘He walked back home.’

If the initial verb expresses an event with a high semantic transitivity value in which the patient is the main affected entity, the non-initial verb(s) must have the same transitivity value as the initial verb. This is illustrated in (15).

(15) dč pna hæcn t klb.
3 shoot CAUS.die dog
‘He shot (and killed) a dog.’

No constraints as to transitivity apply in core serialization, where the patient occurs between two verbs, as in (16) from Jenny (2005:128).

(16) dč pæc kʃʔ pnt ṭa.
3 cut.down neck break.off go
‘He cut off the [bird’s] neck.’ (WK)

In core serialization the P of the first verb usually functions as S/A of the second verb. The verbs thus share their arguments, but the arguments may change their syntactic function. The patterns found in complex verbal predicates in Mon can be summarized as follows:

S/A V V V (V) P for nuclear serialization and
S/A V P→S/A V for core serialization.
3.1 Secondary verbs and Resultative Verb Compounds

Secondary verbs in Mon can express a wide range of functions, including modality, aspect, manner, direction, and resultative. In most cases, verbs functioning as secondary verbs also occur as main predicates, and their semantic content may restrict their applicability to certain contexts. In other words, many secondary verbs expressing grammatical functions are not fully grammaticalized.

Most secondary verbs occur in nuclear serialization, but a few modals, such as kiʔ ‘get; can’ and təh ‘hit, come into contact; know how to do, do correctly’ are found in core serialization, that is, they occur after the patient argument.2 As seen in example (16) above, also resultative verb compounds can appear in core serialization, though there are alternative expressions using nuclear serialization, as the variant of (16) in example (17) illustrates.

(17) ðec həc həŋmt na koʔ.
3 cut.down CAUS.break.off CAUS.go neck
‘He cut off the [bird’s] neck.’

In this case, all non initial verbs must agree with the initial verb in their transitivity value, that is, they are causativized. The transitivity also applies to directionals and orientation verbs, which can be used as main predicates or as secondary verbs, indicating absolute (directional verbs) or relative (orientation verbs) direction, in a complex predicate. The directionals form a closed set of verbal morphemes, consisting of the three orientation verbs òa ‘movement away from origo; go’ and kluŋ ‘movement towards origo; come’, and the directional verbs ‘movement up; ascend’, ‘movement down; descend’, ‘movement in; enter’, ‘movement out; exit’, cao ‘movement back to point of origin; return’, and czp ‘arrive’. The terminology ‘orientation verb’ (Orientierungsverben) and ‘directional verb’ (Richtungsverben) is used by Bisang (1992:67f) after Gorgoniev (1966, quoted in Bisang 1992), which together form the category of direktionale Verben (‘directional verbs’).

In addition to directionals, Mon has an open class of verbs indicating manner of motion, such as kwac ‘walk’, krip ‘run’, pn ‘fly’, etc. These manner of motion verbs, which most commonly occur as main predicates, are syntactically intransitive and do not include a direction or path and cannot combine directly with an NP expressing the goal or location of movement. Some motion verbs are syntactically bivalent and may take an object. One example is pək ‘follow, move behind someone or something’. As with other verbs of motion, direction of movement is not part of the semantics of pək, and the NP following it is interpreted as the entity behind which the A moves. Directed motion is expressed by the combination of a verb describing the manner of motion and one or two directionals. The first set of directionals consists of six verbs indicating relative direction, namely ceh ‘move down’, tm ‘move up’, ləp ‘move in’, tet ‘move out’, cao ‘move back’ and czp ‘arrive’. The directionals can further combine with one of a set of the orientation verbs.

The maximal structure of motion verbs in Mon is the following:

MANNER DIRECTION (DIRECTION) ORIENTATION

Two or more directionals can be combined in a clause, though series of more than three verbs are rarely found in spontaneous language. In example (18), two directionals without verb of manner of motion and orientation verb, while (19) shows the combination of manner, direction, and orientation.

(18) ləp czp ləŋ si.
enter arrive PN
‘They arrived inside Lounzi.’ (WW2Monland)

(19) poj təʔ khreʔ ceh òa.
1PL PL step.proudly descend go
‘We walked down proudly.’ (KM_SR)

---

2 Both kiʔ ‘get’ and təh ‘hit’ are also used in nuclear serialization with different functions.
Unlike verbs of manner of motion, the directionals and orientation verbs can take as object an unmarked NP expressing the goal of the motion. The unmarked object of têt ‘move out’ is usually understood to be the goal of the movement, with the source of the motion obligatorily marked by the ablative preposition nù. More commonly, têt ‘exit’ is combined with ṭa ‘go’ (less commonly klṣŋ ‘come’) to introduce a goal.

The orientation verb ṭa ‘go’, less commonly klṣŋ ‘come’, also has aspectual function, indicating that an event extends from the point of reference to the (relative) future, resulting in a continuous or imperfective reading. In other contexts ṭa ‘go’ indicates that the event is completed or its result is out the sphere of control of the speaker, giving a perfective reading, often implying (or implicating) a notion of irreversibility and discontinued interest of the speaker. With stative verbs, ṭa ‘go’ as secondary verb often indicates a change of state which has come about, combining aktionsart and aspect.

Mon has two complete sets of directionals and orientation verbs, one basic or intransitive, the other causative or transitive. The two sets are given in Table 1.

Table 1: Directionals and orientation verbs

<table>
<thead>
<tr>
<th>Directional</th>
<th>Basic</th>
<th>Causative</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Movement</td>
<td>Causative</td>
<td>Movement</td>
</tr>
<tr>
<td>ceh</td>
<td>‘down’</td>
<td>ṭa</td>
<td>‘away from Col’</td>
</tr>
<tr>
<td>ton</td>
<td>‘up’</td>
<td>klṣŋ</td>
<td>‘toward Col’</td>
</tr>
<tr>
<td>lùp</td>
<td>‘into’</td>
<td>nèŋ</td>
<td></td>
</tr>
<tr>
<td>têt</td>
<td>‘out’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cao</td>
<td>‘back’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cɒp</td>
<td>‘arrive’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In most cases, the causative directional verbs are transparent morphological derivations of the basic forms with the causative prefix pə-/p-. This prefix regularly merges with initial c into phj, as in phjeh ‘bring down’, and is replaced by the semi-productive prefix ha- in some colloquial varieties, as in hatet ‘take out’. The shape of həcp ‘bring to’ (for the expected, but unattested *phjip) suggests that it is of more recent origin, which is also confirmed by the fact that in Old Mon the basic form cip was used in both transitive and intransitive contexts. The causative form first appears in Middle Mon inscriptions as <bacuip>, which apparently goes back to an unattested frequentative-causative form *<phipcup>, or is built after analogous forms in other verbs. In modern Mon too, cɒp ‘arrive’ as secondary verb can be used in transitive and intransitive expressions and in this respect differs from the other directionals, as will be seen below.

The causative orientation verbs are not directly historically related to the basic forms. Venitive nèŋ ‘bring here’ goes back to Pld Mon <raï> ‘bring’, which already in Old and Middle Mon functions as venitive V2. The morphological causative of Old Mon <tlũŋ> ‘come’ (modern Mon klṣŋ) survives in Modern Mon as kəlŋ with the specialized meaning ‘welcome, receive’. In na ‘take away’, Old Mon <rin araç ‘take away, carry off’ seems to merge with the Old Mon particle <nā> ‘away’ (Shorto 1971:318). The regular reflection of the latter in Modern Mon would be ṭnèŋ. The form <rin araç> is seen as a contraction of <raï ṭar>, which came to serve as causative form of <ʔar> ‘go’ (Shorto 1971:318). In Middle Mon the shortened form <nǎ> is used as a main predicate and as V2, while in Modern Mon na only occurs as secondary verb.

Examples (20) and (21) illustrate the use of the directionals in intransitive and transitive contexts.

(20) kon.jāc kwac cao ṭa phęx.
child walk return go school.
‘The child walked back to school.’
(21) rə kok phjaọ na hnaʔ. (*kok cao ʔa)
friend call CAUS.return CAUS.go house
'The friend brought (her) back home.'

In the following section we will have a closer look at the use of intransitive and transitive forms of the directionals, as these show rather consistent patterns of transitive-intransitive alternation.

4. Intransitive and transitive directionals

4.1 Intransitive directionals

Intransitive directionals are regularly used with intransitive main predicates expressing a manner of motion. If no spatial movement is involved in the event, especially ʔa (and less frequently kləŋ) may be used to indicate aspectual values, such as change of state, as seen in (23) and (24). In this function, they often implicate completion of the event or telicity, though this can be canceled by the context. Without the addition hùʔ ʔa, example (23) would normally be interpreted that Lamaing burned down completely.

(22) krip ceh ʔa, krip tnn plon.
run descend go run ascend again
'We ran down, then we ran up again.' (KM_SR)

(23) ləmàŋ tao ʔa hùʔ ʔon.
PN burn go NEG few
'Lamaing burned down a lot.' (WW2Monland)

(24) ʔu.phə.ʔon ʔa liə ʔa lə-kəh lè.
PN PL easy go TEMP-MEDL EMPH
'Now this was when U Hpo Aung and his family got rich.' (WW2Monland)

If the main predicate does not describe a movement, intransitive directionals are possible, even if the main predicate is transitive. In this case the directional has a non-spatial value, usually aspectual. In (25), ʔa ‘go’ indicated that the talking went on for some time, in (26) marks the event of forgetting as complete and irreversible. In (27) the same directional expresses an emotional distance of the speaker to the event or its outcome, along the lines ‘nothing can be done about it’, together with a notion of irreversibility.

(25) dəh həm ʔa ʔərè bət ʔə nəfi.
3 speak go language about two hour
'He continued to talk for about two hours.'

(26) poŋao chan poj nèh ʔa dəh wət ʔa jəʔ.
matter love IPL person two 3 forget go FOC
'She has forgotten our love.' (song lyrics)

(27) ləp kləŋ dəŋ sem ʔə lə-kəh təh ʔa pəŋ.
enter come land Thai PROX MEDL hit go bomb
'In Thailand [the railway] was hit by a bomb.' (WW2Monland)

If the main predicate is transitive expresses a movement, intransitive directionals are used if the A, rather than the P argument, moves, or the movement of the agent is foregrounded. In (28), both agent and patient move, but it is the movement of the agent that is more relevant to the situation. The English ride donkeys because they want to move from place A to place B themselves, not make the donkeys move.

(28) ʔəŋkələc kəh məʔ daŋ kləŋ məʔ, daŋ kləŋ la.
English MEDL what ride come what ride come donkey
'What did the English ride coming here? They rode donkeys.' (WW2Monland)

The verb pək means ‘move behind something or someone’, and can be translated as ‘follow’ if the movement of the agent is foregrounded, as in (29). The king is not set in motion by the event described in the main predicate; therefore the use of the intransitive ʔa ‘go’ is adequate in this
context. The sentence is ambiguous as to the subject of the second part/clause. Both Tok Khæ and the king return to the palace, so either can be seen as subject of *ca*o *ʔa* ‘go back’. There is no syntactic clue that one or the other would be favored as subject.

(29) *tôk.khê  bırak* *ʔa ʔèkʰrah* *ca*o *ʔa ʔôn.*
PN follow go king return go palace
‘Tok Khæ followed the king back to the palace.’ (MCL_stories)

If the patient occurs between the main verb and the directional (in core serialization), it becomes subject of the non-initial clause and the directional appears in the intransitive form, as seen in examples (30) and (31). As (30) shows, this is also true in periphrastic causative constructions, which are structurally biclausal.

(30) *ʔǝmè ʔn kn kwa*ca*ʔa ʔphèr.*
mother give offspring walk go school
‘The mother made her child walk to school.’

(31) *dèh *həd̚ryj klo tɛt*ca*ʔa ʔn ʔn kn kloʔ.*
3 chase dog exit go ABL OBL garden
‘He chased the dog out of the garden.’

In a few cases, intransitive directionals are found where transitive forms would be expected from the context. In examples (32) and (33), the basic forms of the directional verbs *tun* ‘go up’ and *côp* ‘arrive’ are combined with the regular transitive forms of the orientation verb.

(32) *bè? *ʔəɗ kɔh həl*hè*ca*ʔan*. REF PN MEDL CAUS.free ascend CAUS.go
‘Adi let them go up there.’ (WW2Monland)

(33) *hwè? *kɔməthan nɔ? kì? pəl*ca*p ᵊ na ʔpôp. kì? tlp*
corpse corpse PROX CAUS.go get convey CAUS.go arrive get bury
dɔʔ *saoʔsàn tɔʔ kɔh kła*ja*moni hù? kì?.* LOC cemetery DIST MEDL seek man NEG get
‘I cannot find anyone who would take this body there, who would bury it in the cemetery.’ (mkp)

In summary, apart from a few unexplained exceptions, intransitive directionals are used if 1. the main predicate is intransitive, 2. the main predicate is transitive but does not express a spatial movement of the arguments, and 3. if the main predicate is transitive and expresses a spatial movement, but the agent is presented as mainly affected, rather than the patient.

4.2 Transitive directionalss

As seen above (table 1), all directionals in Mon have a basic and a derived transitive/causative form. Similar to Saliba, in Mon multi-verb predicates of the nuclear serial type share all arguments and generally have the same transitivity value. While in Saliba only the first non-initial verb is transitivized, in Mon all secondary verbs combining with a transitive main verb take the causative form individually, as seen in example (34).

(34) *dèh tek phjeə na phan.*
3 strike CAUS.descend CAUS.go glass
‘He struck the glass down.’

Another difference is that transitivity harmony in Mon is triggered purely by semantic transitivity, not syntactic transitivity, as is apparently the case in Saliba as well as Shipibo-Konibo. The situation in Mon is of rather recent origin, as a comparison with Old Mon data shows. In Old Mon causative directionals are not used as V2. To express a caused motion, a full verb like *ra* ‘bring’ can be combined with the basic directionals, as in *ra tî ’bring out’, *ra lop ’bring in’, *ra tlûn ’bring here’. If functioning as main predicate with causative meaning, directionals appear in the causative form, as *nîman plop ra* ‘invite to enter’ (Shorto 1971:312). In this example, the verb *ra* ‘bring (here)’ appears as directional. By Middle Mon, directionals show at least partial transitivity harmony in multi-verb predicates, as in *nʔà bacuip ’bring to’ and *bâk
plon phyau nʔar> ‘send back home’ (lit. ‘follow-return-CAUS.return-CAUS.go’) (Shorto 1971:318). In modern Mon, as seen above, the system of directionals is fully developed into intransitive and transitive forms. The transitive forms of he directional are used in different contexts, as shown below.

Morphological causative main verbs always trigger the transitive directionals if they occur in the same clause in nuclear serial constructions, as seen in the following examples. The normal order is A V DIR OR P. The P argument can be either the causee or the goal of the movement. Overt expression of both causee and goal is avoided in spoken Mon.

(35) ṭamè kəwac pʰjəo na kon.

mother CAUS.walk CAUS.return CAUS.go offspring

‘The mother made her child walk back.’

(36) dɛ̀ həlɛ̀ nɛ̀ kon.cao ṭəcû? mía.

3 CAUS.free CAUS.come grandchild old.man one.

‘They released the grandchild of an old man.’

(37) cəre kəh plon nɛ̀ dɛ̀ sem nəʔ.

secretary MEDL CAUS.exceed CAUS.come land Thai PROX

‘The secretary brought him across (the border) here to Thailand.’ (WW2Monland)

Sentence (38) shows the regular construction for ‘bring someone/something somewhere’, as opposed to the structure seen in example (33), where the order of directional verb and orientation verb is inverted and the basic form of the directional verb is used.

Transitive directionals are also employed if the main predicate is semantically transitive and the movement of the patient is foregrounded or the predicate describes an induced movement. As seen in example (39), the volitionality of the agent is irrelevant to the choice of directional. The first predicate kəpɛ̀ ‘gather’ is volitional, the second kɨʔ ‘get’ non-volitional, but both combine with the transitive orientation na.

(39) kəpɛ̀ na ṭəpot tɔ̀ dɛ̀ kɨʔ na həməj kəo.

gather CAUS.go stuff finish 3 get CAUS.go flowerpot

‘They gathered (and took away) stuff, and they got a flowerpot.’ (WW2Monland)

If the main verb does not in itself express an induced motion, transitive directionals can nevertheless be used, as seen in (40).

(40) dɛ̀ tɛ̀k pʰtɛ̀t nɛ̀ hə-kəh raʔ.

3 strike CAUS.exit CAUS.come ADV-MEDL FOC

‘They beat them like this (so that they left from there).’ (WW2Monland)

If the main verb has different readings according to the context, the choice of the form of the directional can disambiguate, as in example (41). The verb pɛ̀k ‘move behind someone or smething’ (consistently glossed as ‘follow’ for convenience) gets the interpretation ‘chase’ by the use of the transitive directional verb pʰjə ‘bring down’. If the intransitive cɛ̀h ‘go down’ were used, the reading would be ‘follow down’ (see section 4.3).

(41) səṭhĩ kəh kəh dɛ̀ pɛ̀k pʰjəh nɔ̀ hɔʔ.

rich.man MEDL MEDL 3 follow CAUS.descend ABL house

‘That rich man chased him out of the house.’ (WW2Monland)

---

3 According to Shorto (1971:253), <plon> is a rare variant of <plan>, ‘return; do again’. An alternative interpretation would be to see it as causative of <lon> ‘go beyond, go past’.
If the main verb does not describe a real spatial movement, but rather a metaphorical one, the transitive directionals are used if the patient is described as mainly affected by the event. Compare sentence (42) with example (25), where the intransitive orientation verb ʔa ‘go’ is used to modify the act of speaking, rather than the movement of the patient. In (42), the important part of the event is that the speech moved away from the speaker and reached the hearer, that is, the speech as patient is the mainly affected entity, though it is not overtly expressed in the sentence.

(42) ʔə ʔa kən tok həhəh ʔam pəh na.

teacher horoscope calculate show speak explain CAUS.go

‘The astrologer made the calculations and explained them to her.’ (mkp)

The choice of the venitive or andative orientation verb depends on the perspective of the speaker. In (43) the act of electing a president results in the president being conceptually closer to the sphere of interest of the speaker, therefore the venitive nəŋ ‘bring here’ is the adequate choice. In (44), on the other hand, the enumeration of objects goes on and the speaker (or writer in this case) is not involved with the enumerated objects in any way. They do not enter his sphere of interest or control, therefore the andative orientation verb na ‘take away’ is used.

(43) man pəŋəʔ pəŋəʔ kəh dəh rəj ətəm nəŋ.

CAUS.ascend CAUS.come

‘The old people chose [him].’ (WW2Monland)

(44) məʔ kon plem jəm rih riəŋ na sum nəm mənəh.

PN cry enumerate CAUS.go INCL tree jackfruit

‘In her weeping Mi Kon Plem enumerated (everything), including the jackfruit tree.’ (mkp)

With verbs of creation, both na and nəŋ can be used, depending on the real or figurative movement of the created object. In (45), the letter written is brought into the sphere of interest of the speaker. If the andative form na ‘take away’ were used in the same sentence, the interpretation would be that someone close to the speaker wrote a letter to someone further away.

(45) həka kləŋ kok, khju nəŋ ləc, khju ləc lə.

layman come call write CAUS.come text write text EMPH

‘A layman came to call, he wrote a letter for you.’ (KM_SR)

With verbs of destruction, only na can normally be used, as no movements toward center of interest is usually possible (P disappears with the event). Sentences (46) and (47) show events of destruction, in which the venitive nəŋ ‘bring here’ would be ungrammatical. Example (47) also proves that the agent does not have to be volitional, but may be an inanimate instrument rather than an animate agent.

(46) dəh pəlm nə həʔ.

3 CAUS.destroyed CAUS.go house

‘They destroyed the house.’

(47) kəʔ kəh kəpək nə taj dəh wəʔ həʔət ʔə rəʔ.

wind MEDL dash CAUS.go hut LOC filed ADV-all go FOC

‘The wind tore the hut in the field apart completely.’

As we have seen in the preceding examples, the choice of a transitive directional is always triggered by the foregrounded affectedness of the patient. We will return to other potential factors of transitivity again in section 5.

4.3 Contrastive minimal pairs

With the choice of the form of directionals being primarily based in the semantics of the clause, it is not surprising that these can be used to distinguish meanings of the same verbal lexemes. The directionals are not semantically empty grammatical morphemes, but they add to the overall meaning of the expression beyond pure directionality. There are numerous examples of minimal pairs, that is, the same verb combining with either intransitive or transitive directionals to express different notions. These notions can be different readings of the same verb, such as the specification of the semantically underspecified verb pək ‘move behind something or someone’. If
the agent is mainly affected (or moved) by the event expressed, the intransitive form is used, as in the following examples. In both examples both A and P move, but it is the motion of A that is the important fact or the aim of the act. If you drive a car, you want to move from one place to the other, the movement of the car is only a means to achieve this goal. Similarly, when following people, they necessarily move in space, but the important movement is the one by the agent.

If, on the other hand, the aim of the activity is to make the patient move from one place to another, the transitive directional is employed, as in (50).

Notice that in all three preceding examples, the general meaning of ‘A moves behind P’ is present. The different translations follow from the semantics of the secondary verbs, which foreground the movement of either the agent or patient, though in all cases both equally move in space. The following examples further illustrate the different readings arising from the use of different directionals.

Movement of agent foregrounded → intransitive DIR

(51) ᵃŋkələc kəh mii? ᵃdak klasə mii?, ᵃdak klasə la.
   English MEDL what ride come what ride come donkey
   ‘What did the English ride coming here? They rode donkeys.’ (WW2Monalnd)

Movement of Patient foregrounded → transitive DIR

(52) kii? nɛŋ Ø toə.təh wi nii pəŋə? ᵃdak nɛŋ ko θɛn klin.
   get CAUS.come then tend ABL PN ride CAUS.come OBL boat
   ‘They got him and took care of him, and they brought him over from Panga in a boat.’ (WW2Monland)

Movement of agent foregrounded → intransitive DIR

(53) kjan kii? əa stə pi pən mii?.
   PN get go fruit bael four CL
   ‘Kyan got four bael quinces (and went away).’ (KM_SR)

Movement of Patient foregrounded → transitive DIR

(54) ᵃdə kii? na həməj.kəo hənə? poj
   3 get CAUS.go flowerpot house IPL
   ‘They got (and took away) a flowerpot from our house.’ (WW2Monland)

If the main verb does not describe a concrete spatial movement, the use of transitive directionals is still possible if the patient is seen as mainly affected by the activity. The contrast of affectedness of agent vs. affectedness of patient is illustrated in examples (55) and (56). In (56) only the andative na ‘take away’ is possible, as the rice has been removed from the center of interest or sphere of control of the speaker. The result of the event, that is the affected patient, is not visible and therefore not potentially relevant to the situation anymore.

Affectedness of Agent foregrounded → intransitive DIR
JENNY, Mathias. 2014. Transitivity and affectedness in Mon. *Mon-Khmer Studies* 43.1:57-71 (ICAAL5 special issue)
obviously the third and fourth stages of an event that are relevant to high transitivity, which is regularly marked by transitivity harmony in complex verbal predicates in Mon.

Transitivity harmony only applies in monoclusal complex predicates of the nuclear serial type. Core serialization, on the other hand, is excluded from the transitivity harmony. As periphrastic causatives are biclausal, they do not participate in transitivity harmony. Compared with morphological or lexical causatives, periphrastic causatives are therefore less transitive, which is also seen in the fact that the result of a periphrastic causative expression can be canceled, as in (60) while the same is not true for morphological causatives, as in (61).

(60) dəh ʰədəŋ klø tɛt ʔə nù ʔə nù klọʔ ra,
3 chase dog exit go ABL OBL garden FOC
‘He chased the dogs out of the garden.’

chaʔ.kêh klø tɛʔ kəh hùʔ tɛt pùh.
but dog PL MEDL NEG exit NEG
‘...but the dogs didn’t go out.’

(61) dəh ʰədəŋ pʰətɛt na klø nù kə nə klọʔ ra.
3 chase CAUS.exit CAUS.go dog ABL OBL garden FOC
‘He chased the dogs out of the garden.’

*chaʔ.kêh klø tɛʔ kəh hùʔ tɛt pùh.
but dog PL MEDL NEG exit NEG
‘...but the dogs didn’t go out.’

Example (60) is less transitive than example (61), because the event is not carried out to completion, though presumably the planning and instigating stages are realized, as is the third stage of the event itself. But the event fails to produce the intended result, and the patient is not fully affected. In (61) the result is achieved and the patient fully affected, a fact that is expressed by the use of the transitive directional in a nuclear serial construction. This systematic encoding of high transitivity in Mon is of rather recent origin, and is a characteristic feature of Mon grammar. Other languages of the region, both related and unrelated, do not this consistence in distinguishing high transitivity on the one side from low transitivity and intransitivity on the other. Of the main languages influencing Mon over the past centuries, namely Thai and Burmese, Thai does not have anything similar to the Mon system, while Burmese exhibits transitivity harmony in some cases, but far less consistently than Mon. One example illustrating the phenomenon in Burmese is given in (62), where the use of the intransitive directional tʰweʔ ‘exit’ would be ungrammatical.

(62) dɪ=lu-ɡá kʰwè-dwe-go ʔeín-cè-teʔən-dè-ɡá.ne
PROX=person-SBJ dog-PL-OBJ house-front-garden-inside-ABL
tʰauʔ-tʰouʔ-qaɪʔ-te, məun-tʰouʔ-qaɪʔ-te.
CAUS.fear-CAUS.exit-follow-NFUT drive-CAUS.exit-follow-NFUT
‘This man scared the dogs away, chased them out of the frontyard.’

Though this Burmese example closely parallels the Mon sentence in (61) (both are elicited translations of the same English sentence), Burmese does not have the regular system fund in Mon and nothing can be said at the present state of research about the origin of the Mon paradigm.

Abbreviations

A agent; ABL ablative; ABS absolutive; ADV adverbial; AFF affectedness; AO agent-oriented; APP applicative; CAUS causative; COMPL completive aspect; ERG ergative; EV direct evidential; FOC focus; HRY reportative evidential; INCL inclusive; INST instigation; INTEN intensifier; LOC locative; MEDL medial demonstrative; NEG negation; O object (in Saliba); OBJ object; OBL oblique; P previous event (in Shipibo-Konibo); P patient (general); PL plural; PN proper name; PP2 past/completive participle; PROX proximal demonstrative; RED reduplication; S single argument; SBJ subject; SG singular; SIM.EVENT simultaneous event; SO subject-oriented; SS same subject; TEMP temporal adverbial; VOL volitionality
Sources:

KM_SR Conversation (two brothers, Kanni, Kayin State; audio)
MCL_stories Short stories (Sangkhlaburi, Thailand; audio)
mkp Short story (Mawlamyaing, Mon State; published in print)
WW2_Monland Conversation (elderly couple, Kawdot, Mon State; audio)

References:

Shorto, Harry L. 1971. a dictionary of the Mon inscriptions from the sixth to the sixteenth centuries. London: Oxford University Press.
Some Syntactic Aspects of Lexical Anaphors in Select Munda Languages

J. Mayuri  
Indian Institute of Technology Madras  
<mayuri.dilip@gmail.com>

Karumuri V. Subbarao  
University of Hyderabad  
kvs2811@gmail.com

Martin Everaert  
Utrecht University  
m.b.h.everaert@uu.nl

G. Uma Maheshwar Rao  
University of Hyderabad  
guraohyd@gmail.com

Abstract
This paper investigates several syntactic aspects of anaphors (reflexives and reciprocals) in Santali, Mundari and Ho (Munda languages). In this paper we intend to show that the verbal reflexivization strategy is an indigenous device and that the nominal reflexivization strategy is a form calqued from the neighbouring Indo-Aryan languages. The verbal reflexive device performs various other functions such as a detransitivizing marker, passive marker, and self-benefactive marker. The nominal reflexive is optional when the verbal reflexive is present. However, it is obligatory when the verbal reflexive is absent. Long-Distance Binding is not permitted when the anaphor occurs in a subcategorized position. Reciprocity is achieved through the infixation of a morpheme \(-pV\) in the main verb, where the vowel V in \(-pV\) harmonizes with the nucleus of the first syllable of the main verb. Some verbs have a special form when the verbal anaphor occurs.

Keywords: anaphora, pronouns, language comparison

ISO 693-3 codes: sat, unr, hoc, mni, tel

0 Introduction

Reflexives and reciprocals, generally labelled as an aphors, have been a focus of study for a considerable period of time. In this paper, we discuss several syntactic aspects of anaphors in three select languages of the Munda subfamily of the Austro-Asiatic family. The languages chosen are: Santali, Mundari and Ho. Note that the use of the term ‘anaphor’ in this narrow sense is different here from the traditional use, but follows the terminology as generally used in generative grammar (Chomsky 1981).

0.1 Occurrence of anaphors in different language families

Subbarao (2012) presents an extensive study of anaphora in Dravidian, Indo-Aryan, Tibeto-Burman and some Austro-Asiatic languages. He shows that South Asian languages have both nominal and verbal anaphors and that the occurrence of either of these by themselves or together in a sentence is a language specific property. We give three examples to illustrate this. Manipuri
A few Tibeto-Burman languages, (Ao, Tenyidie for example) have a homophonous form for a pronoun and an anaphor. Most of the Indo-Aryan languages - except Gujarati, Marathi and Sinhala - and some Tibeto-Burman languages have only the nominal form of anaphor. All Dravidian languages, except Malayalam and Toda, and many Tibeto-Burman languages and Khasi (Mon-Khmer) have both a nominal anaphor and a verbal anaphor. Most of the Munda languages have only a verbal anaphor as an indigenous device, and the nominal anaphor borrowed from the neighbouring Indo-Aryan languages.

Some of the prominent features that we will discuss are the following:

1. The indigenous device for manifesting anaphora is the **verbal device**. The detransitivizing marker -n- performs the function of a verbal reflexive. The verbal reciprocal is formed by infixing -pV- in the stem, where the vowel V in -pV- is a variable whose value is dependent on the nucleus of the initial syllable.
2. Santali, Mundari and Ho have a **simplex form** of the nominal anaphor either borrowed or calqued from the neighbouring superstrate Indo-Aryan languages such as Hindi, Sadri, and Bengali.
3. Constructions involving reflexive in a subcategorized position have a unique form of the verb stem where it occurs with a suffix -ok- in Santali and -en- in Mundari and Ho. The occurrence of such suffix is **optional**, while the occurrence of the verbal device (reflexive) is obligatory. Thus, it is the verbal device that functions as the primary device triggering reflexivity.
4. The verbal device does not occur in a non-subcategorized position. In such cases, the occurrence of the nominal reflexive is obligatory, when the verbal reflexive is absent.

The paper is organized as follows: Section 1 deals with the devices used for anaphoric reference. Section 2 provides the verbal morphology of anaphoric constructions. Section 3 lists the forms occurring as nominal anaphors. Section 4 discusses the multifarious nature of the verbal reflexive device. Section 5 focuses on the valence reducing function of the verbal reflexive -n-. Section 6 deals with special verb forms in reflexive constructions. Section 7 discusses the structural dependency between the nominal and the verbal reflexive. Section 8 concerns itself with interactions between reflexivity and Long-Distance Binding. Section 9 discusses Exceptional Case
Marking (ECM). Section 10 focuses on the issue of the ‘accusative marker’ and the nominal reflexive in the three Munda languages under consideration. Section 11 discusses the formation of reciprocal constructions and the special verb forms for reciprocals, a unique feature found in the Santali, Mundari and Ho. Section 12 is the conclusion.

1. Devices used for anaphoric reference

Santali, Mundari and Ho have both verbal and nominal anaphors. While, the verbal reflexive -n- and the verbal reciprocal –pV- (see section 11 for details) are the indigenous devices, used primarily for anaphoric reference, the nominal reflexive is borrowed or calqued from the neighbouring contact languages such as Sadri, some of the dialects of Hindi-Urdu or Bengali with which these languages have been in intense contact for a long period of time. By indigenous devices, we mean that these languages had only the verbal devices genetically, and they did not possess any nominal device. It is worth mentioning that a nominal reciprocal is absent in these languages.

In (2)-(4), -n- is the verbal reflexive and the entity in the argument position indicated in bold is a nominal reflexive. The nominal anaphor is optional, while the occurrence of the verbal device is obligatory.

Santali
2. pǝʈhǝkǝ akote-kǝ sarhao-ɛ-n-ǝ
   student-pl themselves-SAM praise-pst-vr-[+fin]
   ‘The students praised themselves.’

Ho
3. ḍobro ajek’tǝ arsi-ɛn ke-ɛn-ǝ
   Dobro himself mirror-in see-pst-vr-[+fin]
   ‘Dobro saw himself in the mirror.’

Mundari
4. hɔn-kǝ akǝte arsi-re ke-ɛn-ǝ
   child-pl themselves mirror-in-SAM see-pst-vr-[+fin]
   ‘The children saw themselves in the mirror.’

2 Verb morphology of anaphoric constructions

The morphological parsing of a verb with a verbal reflexive is as follows:

main verb + tense + (verbal reflexive) + finite marker + (subject agreement marker).

A sentence having a verb with the morpheme alignment mentioned above is given in (5) from Santali:

5. pǝʈ’uә-kǝ-kǝ sarhao-ɛ-n-ǝ
   student-pl-SAM praise-pst-vr-[+fin]
   ‘The students praised themselves.’

3 Form used as a nominal anaphor

In Santali, Mundari and Ho, the nominal reflexive is scarcely used. Hence, it was not possible to find the corresponding nominal anaphor for each of the personal pronoun as these languages use the verbal reflexive as the primary device to establish anaphoric reference. We provide the forms of the pronoun and their corresponding nominal reflexive in Tables (1)-(3). X is a variable that stands for the postposition that occurs with the nominal anaphor and the nature of the postposition depends on the structural or inherent case assigned by the predicate.

---

3 The preverbal constituent is the subject itself as there is no other constituent preceding the verb. Hence, the subject agreement marker (SAM) occurs to the right of the subject itself.
Table 1: Forms of nominal anaphors in Santali

<table>
<thead>
<tr>
<th>Pronoun</th>
<th>gloss</th>
<th>reflexive</th>
<th>postposition (near, about, with etc.)</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>uni</td>
<td>he/she</td>
<td>acˀ-X</td>
<td>himself/herself</td>
<td></td>
</tr>
<tr>
<td>akɔ</td>
<td>they</td>
<td>akɔ-X</td>
<td>themselves</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>a.pre-X</td>
<td>self’s</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Forms of nominal anaphors in Ho

<table>
<thead>
<tr>
<th>Pronoun</th>
<th>gloss</th>
<th>reflexive</th>
<th>postposition (near, about, with etc.)</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ae</td>
<td>he/she</td>
<td>ajekˀ-X</td>
<td>himself/herself</td>
<td></td>
</tr>
<tr>
<td>apn</td>
<td>I</td>
<td>apn-X</td>
<td>myself</td>
<td></td>
</tr>
<tr>
<td>aben</td>
<td>you</td>
<td>aben-X</td>
<td>yourself</td>
<td></td>
</tr>
<tr>
<td>akɔ</td>
<td>they</td>
<td>akɔ-X</td>
<td>themselves</td>
<td></td>
</tr>
<tr>
<td>ape</td>
<td>you (pl)</td>
<td>ape-X</td>
<td>yourselves</td>
<td></td>
</tr>
<tr>
<td>abu</td>
<td>we (incl)</td>
<td>abu-X</td>
<td>ourselves</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Forms of anaphors in Mundari

<table>
<thead>
<tr>
<th>Pronoun</th>
<th>gloss</th>
<th>reflexive</th>
<th>postposition (near, about, with etc.)</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>akɔ</td>
<td>they</td>
<td>akɔ-X</td>
<td>themselves</td>
<td></td>
</tr>
<tr>
<td>ae</td>
<td>he/she</td>
<td>ajaˀ-X</td>
<td>himself/herself</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>apna-X</td>
<td>self’s</td>
<td></td>
</tr>
</tbody>
</table>

The nominal reflexive apna (Mundari) in table 3 is borrowed from neighbouring Indo-Aryan languages.

4 Multifarious nature of the verbal reflexive device

The monomorphemic verbal reflexive -n- performs a variety of functions in the languages under discussion such as a verbal reflexive as in (6) and (8), a passive marker as in (7) and (8), an [-transitive] marker as in (9) and a self-benefactive as in (10b).

In the Santali example in (6) -n- is playing the role of a verbal reflexive marker:

6. *ophel arsi-re-j jel-ɛ-n-a*
   Ophel mirror-in-SAM see-pst-vr-[+fin]
   ‘Ophel saw himself in the mirror.’

The Ho-example in (7) illustrates -n- as a passive marker

7. *pulis-ko-te kumbu sap-ja-n-a*
   police-pl-by thief catch-pst-pass-[+fin]
   ‘The thief was caught by the police.’

Note that an example as in (8) “is ambiguous between a verbal reflexive reading and a passive reading.” (Subbarao 2012:63).

Ho

8. *pro, siŋboŋa-taɔ-re-mi ema-ke-n-a*
   you God-to-in-SAM give-pst-vr/passive-[+fin]
   ‘You gave yourself to God.’ / ‘You were given to God.’
Psychological predicates in South Asian languages are intransitive by nature and hence, they have monadic valence. Munda languages provide further evidence to substantiate this fact. An intransitive verb in these languages carries the morpheme \(-n\). In these languages it is the [transitive] marker that occurs with psych-predicates as illustrated by the Ho-example in (9):

Ho
9. \(d\)obro \(aj\)-lo \(na\-ra\-jo\-ka\-n\-a\)
   Dobro he/she-with anger-pst-[tr-]+fin]
‘Dobro is angry with him.’

There is another use for \(-n\), the [transitive] marker as self-benefactive. Take the example in (10a), the verb isin ‘cook’ in Ho is [+transitive], and it takes the [+transitive] marker \(-d\).

Ho
10a. un-\(k\)in-\(k\)in isin-akaw-\(d\)-a
   they-dual-SAM cook-pst-[+tr]-[+fin]
   ‘They (dual) cooked (something).’

In (10b), the marker \(-n\) occurs contrary to the expected occurrence of [+transitive] marker \(-d\).

Ho
10b. un-\(k\)in \(j\)oma:\(g\)-\(k\)-\(k\)in isin-akaw-\(n\)-a
   they-dual food-SAM cook-pst-self ben-[+fin]
   ‘They (dual) cooked food for themselves.’

However, \(-n\) in (10b) does not function as a valence reducer, but functions as self-benefactive. Thus, (10b) is an example where the form of the verb isin ‘cook’ indicates that it is syntactically [-transitive] but is semantically [+transitive] in nature.

5 \(-n\) functioning as an anti-causative / valence reducer

The phenomenon of valence reduction is demonstrated through the syntactic minimal pairs in (11a), (11b), (12a), (12b), (13a) and (13b), respectively. Santali (11a) contains a transitive verb and thereby, it carries a transitive marker \(-d\). However, (11b) is detransitivized into an anti-causative/medio-passive. A similar phenomenon is also observed in Ho as in (12a) and (12b) and Mundari as in (13a) and (13b). In Dravidian, Tibeto-Burman and Mon-Khmer languages too, the verbal reflexive functions as a valence reducer (see Subbarao 2012: 91 for details).

Santali
11a. baha, duw\(\varphi\)\(e\)\(i\) \(f^\i\)-\(c^2\)-ke\(d\)-a
   Baha door-SAM open-pst-[+tr]-[+fin]
   ‘Baha opened the door.’

11b. duw\(\varphi\) \(f^\i\)-\(c^2\)-\(e\)-\(n\)-(\(i\)ja)
   door open-pst-[tr]-[+fin]
   ‘The door opened.’

Ho
12a. sumi silpin kese\(k\)-ke\(-d\)-a
   Sumi door close-pst-[+tr]-[+fin]
   ‘Sumi closed the door.’

---

4 For evidence in support of this claim, see Subbarao 2012.
5 For further details see Mayuri et al 2014.
6 This construction is similar to the non-nominative subject construction in Indo-Aryan and Dravidian where the predicate is [-transitive] syntactically but is semantically [+transitive] in nature (Subbarao 2012: 177-178).
12b. \textit{silpin ajak\textsuperscript{2}-te} kesek\textsuperscript{2}-je-n-a
\begin{itemize}
\item door self-by close-pst-[\textsuperscript{tr}]-[+fin]
\end{itemize}
‘The door closed by itself.’

Mundari
13a. (en) \textit{duwɔr} nik\textsuperscript{2}-ke-d-a-jn
\begin{itemize}
\item (I) door open- pst-[\textsuperscript{tr}]-[+fin]-SAM
\end{itemize}
‘I opened the door.’

13b. \textit{duwɔr} nixo-ka-n-a
\begin{itemize}
\item door open-pst-[\textsuperscript{tr}]-[+fin]
\end{itemize}
‘The door opened.’

6 Special verb forms in reflexive constructions

As mentioned earlier, Santali, Mundari and Ho have a special verb form\textsuperscript{7} when the verb carries a verbal reflexive. Santali verb roots are marked by the verbal suffix \textsuperscript{2}\textit{ok}.\textsuperscript{8} Mundari and Ho by -\textit{en}. Thus, it is a unique feature of these languages to have distinct verb stem to denote reflexivity as in (14), (15) and (16). A list of verbs with such verbal morphology is provided in tables (4), (5) and (6) below. However, irrespective of the presence of the special form, it is the verbal reflexive -\textit{n} that manifests reflexivity\textsuperscript{9}. Moreover, the unique verbal reflexive suffix -\textit{ok}/-\textit{en} may be dropped\textsuperscript{10} without effecting the grammaticality.

Santali
14. \textit{in} arsi-re\textsuperscript{n} \textit{nel(ak)-ka-n-a}
\begin{itemize}
\item I mirror-in-SAM look(refl)-pres- vr- [+fin]
\end{itemize}
‘I am looking at myself in the mirror.’

Mundari
15. \textit{hon-kɔ} \textit{ako-te} arsi-re \textit{nel(en)-ke-n-a}
\begin{itemize}
\item child-pl they-acc mirror-in see(refl)-pst-vr- [+fin]
\end{itemize}
‘The children saw themselves in the mirror.’

Ho
16. \textit{dɔbrɔ} \textit{aj ajte} gojen/gok\textsuperscript{(en)}-ja-n-a
\begin{itemize}
\item Dobro himself kill(refl)-pst-vr- [+fin]
\end{itemize}
‘Dobro killed himself.’

The following tables provide a list of the verbs and their special forms.

\textbf{Table 4:} List of verb stems in Santali

<table>
<thead>
<tr>
<th>verb</th>
<th>gloss</th>
<th>reflexive</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{nel}</td>
<td>see</td>
<td>\textit{nel-\textit{ok}}'</td>
<td>to see oneself</td>
</tr>
<tr>
<td>\textit{ol}</td>
<td>write</td>
<td>\textit{\textit{ol-\textit{ok}}}</td>
<td>to write to oneself</td>
</tr>
<tr>
<td>\textit{em}</td>
<td>give</td>
<td>\textit{\textit{em-\textit{ok}}}</td>
<td>to give oneself</td>
</tr>
<tr>
<td>\textit{goc}</td>
<td>kill</td>
<td>\textit{\textit{guy-uk}}</td>
<td>to kill oneself</td>
</tr>
</tbody>
</table>

\textsuperscript{(Ghosh 2008:70)}

\textsuperscript{7} The special form of the verb occurs only when the verbal anaphor occurs. Thus, the special form of the verb does not occur when the verbal reflexive functions like a detransitivizer or as a “self benefactive marker”. A similar situation is found in Santali, Mundari and Ho.

\textsuperscript{8} According to Ghosh (2010: 88), \textit{\textit{ok}} in Santali functions both as a passive and a reflexive. \textit{\textit{ok}} is marked as middle voice by Neukom (2001: 64) and he also says, “the suffix -\textit{ok} loses its vowel after stems ending in a vowel, e.g., \textit{\textit{id}i} ‘take’ + -\textit{\textit{ok}} -> \textit{\textit{id}-\textit{k}’ be taken’. Santali also has a verbal suffix -\textit{ph} indicating medio-passive (Ghosh 2008: 70). The nature of the morpheme -\textit{ph} requires to be investigated further.

\textsuperscript{9} -\textit{n} and -\textit{en} are used to indicate reflexivity (Burrows 1980: 73, Deeney 1975: 12) in Ho. The verbal reflexive functions as a passive marker too in Ho. (Subbarao2012).

\textsuperscript{10} Optionality is indicated by parentheses.
Table 5: List of verb stems in Mundari

<table>
<thead>
<tr>
<th>Verb</th>
<th>Gloss</th>
<th>Reflexive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>tusiŋ</td>
<td>to put on</td>
<td>tusiŋ-en</td>
<td>to put oneself</td>
</tr>
<tr>
<td>lel</td>
<td>to see</td>
<td>lel-en</td>
<td>to see oneself</td>
</tr>
</tbody>
</table>


Table 6: List of verb stems in Ho

<table>
<thead>
<tr>
<th>Verb</th>
<th>Gloss</th>
<th>Reflexive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>em</td>
<td>to give</td>
<td>em-en</td>
<td>to give oneself</td>
</tr>
<tr>
<td>tam</td>
<td>to strike</td>
<td>tam-en</td>
<td>to strike oneself</td>
</tr>
<tr>
<td>atom</td>
<td>to move</td>
<td>atom-en</td>
<td>to move oneself</td>
</tr>
<tr>
<td>bai</td>
<td>to make</td>
<td>bai-en</td>
<td>to make oneself</td>
</tr>
<tr>
<td>abung</td>
<td>to wash</td>
<td>abung-en</td>
<td>to wash oneself</td>
</tr>
<tr>
<td>dul</td>
<td>to pour</td>
<td>dul-en</td>
<td>to pour over oneself</td>
</tr>
</tbody>
</table>

(Burrows 1980:73)

7 The structural dependency between the nominal and the verbal reflexive

In the following sections, we present evidence how the nominal reflexives occurring in a subcategorized or in a non-subcategorised position relate to presence/absence of the verbal reflexive.

7.1 The structural dependency between the nominal reflexive occurring in a subcategorized position and the verbal reflexive

In Santali, Mundari and Ho, the presence of a nominal reflexives is obligatory, when the verbal reflexive is absent as in (17a), (18a) and (19a). However, it is optional,\(^{11}\) when the verbal reflexive is present as in (17b), (18b) and (19b).

Santali

17a. hɔn-kɔ, ac’te, arsi-re, nel-ke-d-a-kɔ
child- pl themselves mirror-in see-pst-[+tr] [+fin] -SAM
‘The children saw themselves in the mirror.’

17b. hɔn-kɔ, (ac’te)h, arsi-re, nel-ke-n-a-kɔ
child- pl themselves mirror-in see-pst-[vr][-fin]-SAM
‘The children saw themselves in the mirror.’

Mundari

18a. hɔn-kɔ, akɔte, arsi-re-kɔ, lel-ke-d-a
child- pl themselves mirror-in-SAM see-pst-[+tr][-fin]
‘The children saw themselves in the mirror.’

18b. hɔn-kɔ, (akɔte)h, arsi-re-kɔ, lel-ke-n-a
child- pl themselves mirror-in-SAM see-pst-[vr][-fin]
‘The children saw themselves in the mirror.’

Ho

19a. dobroi, ajek’tei, arsi-re, nel-ke-d-a-j
Dobro himself mirror-in see-pst-[+tr][-fin]-SAM
‘Dobro saw himself in the mirror.’

\(^{11}\) The optionality of the nominal entity is depicted parenthetically.
7.2 The structural dependency between the nominal reflexive in a non-subcategorized position and the verbal reflexive

When the nominal reflexive occurs in a non-subcategorized position, the verbal reflexive is not permitted as is illustrated in (20)-(22). The non-occurrence of verbal reflexive is due to the fact that the transitivity of the predicate is unaltered and thus, the transitive marker -d/-d- takes precedence over the verbal reflexive. Since reflexivity is not manifested in the verb, co-reference has to be encoded through the presence of a nominal reflexive (in italics).

Santali

20. hopna ake-so:rer bij-kin-ci nel-ke-d-ɛɛ(j)a
   Hopna himself-near snake-dual-SAM see-pst-+[tr]-SAM-+[fin]
   ‘Hopna saw two snakes near himself.’

Mundari

21. somri aj-ak’ japa-re mijan biŋ nel-ke-d-a
   Somri herself-gen near-in one snake see-pst-+[tr]-+[fin]
   ‘Somri saw a snake beside herself.’

Ho

22. asai ajk’-japa-ɛɛ re mijak’/mɔjak’ biŋ nel-ke-d-a
   Asai himself-near one snake see-pst-+[tr]-+[fin]
   ‘Asai saw a snake near himself.’

8 Reflexivity and Long-Distance Binding

Long-Distance Binding refers to ‘the coindexation of an anaphor in a lower clause with an antecedent in a higher clause outside its minimal clausal domain. Such coindexation is an apparent violation of Principle A of the Binding principles.’ (Subbarao 2012:75), following Chomsky (1981, 1986). Factors that, among others, influence Long-Distance Binding are the following:

1. whether an anaphor occurs in a subcategorized position or not,
2. whether the anaphor is complex nominal form (polymorphic), and
3. whether the anaphor occurs in a non-finite embedded clause or not.

Factor (i) is relevant: Long-Distance Binding is not permitted when the nominal anaphor occurs in a subcategorized position as in Santali (23a) and permitted when the nominal anaphor occurs in a non-subcategorized position as in (23b). A similar restriction is obeyed in Mundari as in (24a) and (24b), and in Ho too as in (25a) and (25b).

Santali

Anaphor in a subcategorized position

23a. [montri giḍrǝ [PROj ake-te-ɛɛ] sarhawɛ metak’-de-a]
   minister child self.acc praise tell-+[tr]-+[fin]
   ‘The minister, asked the child to praise selfmj.’

Anaphor in a non-subcategorized position

23b. [baha arel [PROi] ake-aj’-lagid’ ca: benawɛ metak’-de-a]
   Baha Arel self-for tea make tell-+[tr]-+[fin]
   ‘Baha asked Arel to make some tea for selfi.’
Mundari
Anaphor in a subcategorized position
24a. [somra, suga=ke [PRO] ajak=ba:re re baqai-mente] kaje=tan-a]
   Somra Sugad-acc self-about-in praise-comp tell-pres-[+fin]
   ‘Somra asks Sugad to praise self.’

Anaphor in a non-subcategorized position
   Somra Sugad-acc self-for tea make-comp tell-pres-[+fin]
   ‘Somra asks Sugad to make tea for self.’

Ho
Anaphor in a subcategorized position
25a. [muntri k=ni-tej [PRO] ajak=tei sjahra]ici=qe-d-a]
   Minister child-pl by self praise-caus-pst-[+tr]-[+fin]
   ‘The minister, I made the children praise themselves.’

Anaphor in a non-subcategorized position
25b. [dobra, sumi=ke [PRO] aj-lagidi ca:j bajime]
   Dobro Sumi-dat tell-[+tr]-[+fin] self-for tea make
   ‘Dobro asked Sumi to make tea for self.’

These examples underscore the fact it is the subcategorized position of the anaphor that is significant and not the morphological complexity of the predicate.

9 Exceptional Case Marking (ECM)

ECM (Exceptional Case Marking) is a phenomenon that involves the case marking of the subject of the embedded clause by the verb of the matrix clause. This phenomenon is observed in all the South Asian languages, except optionally in Telugu (Dravidian), Dumi (Tibeto-Burman), Kashmiri (Indo-Aryan) (Subbarao 2012: 144-146).

The following examples are illustrative.

Santali
26a. [baha, (ac)-ac-tei adj cihr-e] monek-a]
   Baha self-self-acc a lot pretty-SAM considers-[+fin]
   ‘Baha considers herself pretty.’

Mundari
26b. [somri, (aj)-aj-tei ti:vi-re duraj-mente] ajum=ke-d-a]
   Somri self-self-acc T.V-in singing-while see-pst-[+tr]-[+fin]
   ‘Somri saw herself singing on T.V.’

Ho
26c. [asai, (ajak)-ajak-tei hosijar-e] manati/baintana-tan-a]
   Asai self-self-acc clever-SAM consider-pres-[+fin]
   ‘Asai considers herself clever.’

In (26a) in Santali, the form ac2-ac2-te occurs in a case-marked position where the matrix verb exceptionally case marks the derived object. If the postposition -te in ac2-ac2-te is treated as an emphatic, a [+transitive] verb such as monek ‘consider’ would lack a direct object in the derived structure, which is a clear violation of the Theta Criterion (Chomsky 1981). Hence, ac2-ac2 has to be treated as a reflexive and the postposition -te has to be treated as an accusative case marker. A similar structure can also be found in Mundari (26b) and (26c) above.
10 A note on the ‘accusative marker’ and the nominal reflexive in Santali, Mundari and Ho

The nominal reflexive \( ac \) in Santali, \( aj \) in Mundari and \( ajak \) in Ho are calqued on the pattern of the nominal anaphor in contact Indo-Aryan languages. The postposition \(-te\) in Santali and \(-te\) in Mundari and Ho performs several functions.

1. It functions like the instrumental postposition ‘with, by’ when it occurs with a noun such as \( ca:ku \) ‘knife’ (Santali).
2. When the postposition \(-te/-te\) follows the reflexive in a subcategorized position such as the direct object, it functions like an accusative marker. In ECM constructions too, it functions like an accusative marker as in (26).

Our analysis of treating the marker \(-te/-te\) is in contrast to the normally-held view that Santali, Mundari and Ho do not have an accusative marker. Evidence for our claim comes from the following fact. In the ECM construction, if \(-te/-te\) is treated as an emphatic marker, it is difficult to explain why either a reduplicated or a simplex form of the nominal anaphor which functions like an emphatic occurs in a position that is case-marked by the matrix verb.

Some supporting evidence comes from Psych (psychological)-predicates:

With Psych (psychological)-predicates, when the argument is coindexed with the subject, an anaphor (reflexive) occurs in a subcategorized position, it is locative case-marked in Santali (27), instrumental case-marked in Mundari (28) and Ho (29). In Mundari and Ho, it is the postposition \(-te\) that imparts the interpretation of ‘with’.

**Santali**

27. \[ upol \; a:pre-ceta:nre \; beja:r-aka-n-a-e \]

\( Upol \) self-on angry-pst-[tr] [+fin] -SAM

‘Upel was angry/ upset with herself.’

**Mundari**

28. \[ somra \; apna-te-ge \; naraj-men-a-e \]

Somra herself-with-emph anger-is+[fin] -SAM

‘Somro is angry with herself.’

**Ho**

29. \[ sumi \; ajek-te \; kurkure/ra:gu-a-ka-n-a \]

Sumi herself-with anger-pst-[tr]+[fin]

‘Sumi was angry with herself.’

11 Reciprocals

The languages under discussion (and this holds more generally for the Munda languages) do not have nominal reciprocals. It is a special feature of these languages to have the verb stem itself manifesting reciprocity through verbal reciprocal inflexion. The verb also carries \(-n\) in it which indicates that the marker \(-n\) in (30) functions as a detransitivizer/valence reducer, as illustrated in the Santali example (30) from Minegishi and Murmu (2001: 104):

**Santali**

30. \[ unkin \; -kin \; \phi \; da: -pa: -l-ka:-n-a \]

they,dual-dual hit.1-vrec-hit.2-pst-[tr]-[+fin]

‘They (two of them) were hitting each other.’

On the other hand, if the marker \(-n\) is interpreted as a detransitivizer, then, we observe that the detransitivizer and the reciprocal infix together impart reciprocity, a phenomenon not found thus far in any South Asian language that we know of.

In reciprocal verbs, the reciprocal infix \(-PV-\) is placed immediately after the first syllable of the root verb. The vowel \( V \) in \(-PV-\) indicates the position of occurrence of the vowel that harmonizes with the vowel of the first syllable of the verb stem as in tables (7)-(9) below. For
convenience, the morphological parsing of the root verb has special glossing. For example, in Ho *ti* is glossed as ‘pour.1’ which is the first part of the reciprocal verb, *pi* as ‘vrec’ and *l* as ‘pour.2’, the second part of the reciprocal verb.

**Table 7**

<table>
<thead>
<tr>
<th>verb</th>
<th>gloss</th>
<th>reciprocal</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>capa</em>da</td>
<td>to slap</td>
<td><em>ca-pa-pa</em>da</td>
<td>‘slap1-vrec-slap2’</td>
</tr>
<tr>
<td><em>til</em></td>
<td>to pour</td>
<td><em>ti-pi-l</em></td>
<td>‘pour1-vrec-pour2’</td>
</tr>
</tbody>
</table>

**Table 8**

<table>
<thead>
<tr>
<th>verb</th>
<th>gloss</th>
<th>reciprocal</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dal</em></td>
<td>‘beat’</td>
<td><em>da-pa-l</em></td>
<td>‘beat1-vrec-beat2’</td>
</tr>
<tr>
<td><em>em</em></td>
<td>‘give’</td>
<td><em>ɛ-pe-m</em></td>
<td>‘give1-vrec-give2’</td>
</tr>
</tbody>
</table>

**Table 9**

<table>
<thead>
<tr>
<th>verb</th>
<th>gloss</th>
<th>reciprocal</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ɔ</em></td>
<td>to write</td>
<td><em>ɔ-pɔ-l</em></td>
<td>‘write1-vrec-write2’</td>
</tr>
<tr>
<td><em>nel</em></td>
<td>to see</td>
<td><em>ne-pe-l</em></td>
<td>‘see1-vrec-see2’</td>
</tr>
</tbody>
</table>

Note that the detransitivizing marker *-n-* appears in reciprocal constructions, as in the Santali (31b), and likewise in Mundari (32b), and Ho (33b).

**Santali**

31a. *arel* *baha*-ɛ *nel-aka-d-e-a*

Arel Baha-SAM see-pst-[+tr]-OAM-[+fin] ‘Arel saw Baha.’

31b. *arel* a: *baha:* *ɲe-pɲ-e-l-a-ka-n-a-kin*

Arel and Baha see.1-vrec-see.2-pst-[+tr] [+fin]-SAM ‘Arel and Baha saw each other.’

**Mundari**

32a. *gangu* *asai-ke* *cɔɾɔp-ke-d-a*

Gangu Asai-acc kiss-pst-[+tr]-[+fin] ‘Gangu kissed Asai.’

32b. *gangu* *ɛndɔ* *asai* *cɔɾɔp-ɾɔp-ja-n-a*

Gangu and Asai kiss.1-vrec-kiss.2-pst-[+tr] [+fin] ‘Gangu and Asai kissed each other.’

**Ho**

33a. *somra* *somri-ke* *ica-ke-d-a*

Somra Somri-acc pinch-pst-[+tr]-[+fin] ‘Somra punched Somri.’

33b. *somra* *ar* *somri* *i-pi-ca-ja-n-a*

Somra and Somri pinch.1-vrec-pinch.2-pst-[+tr]-[+fin] ‘Somra and Somri pinched each other.’

**12 Conclusion**

In this paper, we have discussed some specific syntactic aspects of anaphors in Santali, Mundari and Ho. The indigenous devices for anaphora are the verbal reflexive and the verbal...
reciprocal. Nominal reflexives also exist, however, they are either borrowed from neighbouring Indo-Aryan languages or they are calqued forms. It is a special feature of these languages to have the reciprocal infixed in the root verb. The languages do not possess nominal reciprocals. These languages have special forms of root verbs when a verbal reflexive occurs. The nominal reflexive and a verbal reflexive have structural dependency. Further, the Long-Distance Binding is sensitive to the subcategorized or non-subcategorized position of the nominal anaphor.

Glosses and abbreviations

[+fin] : finite  
[+tr] : transitive  
[-tr] : intransitive  
AA: Austro-Asiatic  
acc: accusative  
caus: causative  
dat: dative  
DR: Dravidian  
refl: reflexive  
IA: Indo-Aryan  
incl: inclusive  

OAM : Object Agreement Marker  
pass : passive  
pl : plural  
pst: past  
SAM : Subject Agreement Marker  
self ben: self benefactive  
Sg : Singular  
TB: Tibeto-Burman  
vr: verbal reflexive  
vrec: verbal reciprocal

References


Another look at serial verb constructions in Khmer

Stephen SELF
University of Texas at Arlington
<stephen.self@mavs.uta.edu>

Abstract
Serial verb constructions (SVCs) are a widely recognized areal feature of Mainland Southeast Asia (MSEA) and Austroasiatic languages in particular. Yet discussions of SVCs in Khmer have often been limited to sporadic data in papers on SVCs in general (Schiller 1989; Clark 1989; Mikami 1981; Clark 1992; Durie 1997). The two most extensive treatments of SVCs in Khmer perse (Wilawan 1993, 1995; SakkHumphry 2005) find theory-internal grounds to deny that the language has SVCs. The more recent study of Haiman (2011) asserts that Khmer does have SVCs but includes in its discussion completive or “success verb” constructions which function differently from structures recognized cross-linguistically as SVCs. Here, I apply the “characteristic and diagnostic features of SVCs” from Kroeger (2004) and Durie (1997) to putative SVCs in Khmer, taking monoclausality as the most important characteristic of SVCs. I use the binding behavior of the bimorphemic reflexive pronoun kluæŋ to demonstrate that the sentences in question do indeed constitute single clauses comprising multiple verb phrases without coordination or subordination that share core arguments, refer to a single complex event, and must agree in polarity and tense-aspect-modality (TAM). It is concluded that Khmer does have SVCs as rigorously defined by the criteria above.

Keywords: syntax, serial verb constructions, reflexive pronouns

ISO 639-3 language codes: khm

1. Introduction
In this paper, it is argued that certain Khmer verbal constructions that have been variously accepted into and excluded from the category of serial verb constructions (SVCs) do in fact merit that label. The types of sentences under consideration can be subclassified in terms of both their semantics and syntax.

Semantically, the constructions in question coincide well with the several general meanings of SVCs outlined in Kroeger (2004:227–229) and Seuren (1990:18). Such constructions in Khmer serve to express: instrument (1a), recipient (1b), beneficiary (1c), goal of motion (1d), result of action (1e), purpose (1f), and manner (1g).

(1a) sokh jɔːk kɑmbət kaːt sac
Sokh take knife cut meat
‘Sokh cut the meat with a knife.’

(1b) kpom tɨɲ siəwphəw ʔaoj Bill
1SG buy book give Bill
‘I bought a book for Bill.’ (Spruiell 1988:252)

(1c) baək bəŋʔuəc ʔaoj kpom phɑːŋ
open window give 1SG also
‘Open the window for me, will you?’ (Huffman, Promchan & Lambert 1970:139)

I wish to gratefully acknowledge the crucial participation of four Khmer native-speaker consultants in this project. MS is a male Buddhist monk in his twenties who had lived in the United States for only eighteen months at the time I worked with him. RS is a forty-something male who has lived in the U.S since he was in his early teens. JS is RS’s wife who had just arrived in the U.S. from Cambodia a month or so prior to our first session together. RS’s mother (MR) is approximately in her seventies and has lived in the U.S. for a little more than thirty years. All speakers are completely fluent and literate in Khmer, their first language and mother tongue. Special thanks also go to the editors of Mon-Khmer Studies and the anonymous referee for able assistance and many insightful comments and suggestions.

Mon-Khmer Studies 43.1:84-102 (ICAAL5 special issue)
(1d) kɲom caol təw laːn
1SG throw ball go car
‘I throw the ball at the car.’ (Spruiell 1988:252)

(1e) kɲom wiej ckaɛ slap
1SG beat dog die
‘I struck the dog dead’ (Mikami 1981:110)

(1f) koət təw Waikiki roəm ciəmuəj jəːŋ
3SG go Waikiki dance with 1PL
‘He goes to Waikiki to dance with us.’ (Sak-Humphry 1995:181)

(1g) viə baːn mɔːk lauːp sdap nɨw kraːom pteəh
1SG PST come sneak listen be.at beneath house
‘He came and listened secretly from beneath the house.’ (Haiman 2011:217)

Syntactically, these constructions can be subclassified into four types, depending on what direct arguments or terms are shared among the constituent verbs. **Table 1** presents a summary of the four syntactic types with their basic word orders and constituents.²

**Table 1**: Khmer SVC types by syntax

<table>
<thead>
<tr>
<th>SVC Type</th>
<th>Verb Types</th>
<th>Word Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHARED AGENT</td>
<td>TR</td>
<td>A V1 T1 V2 T2</td>
</tr>
<tr>
<td>SHARED THEME</td>
<td>TR INTR</td>
<td>A V1 T=T V2</td>
</tr>
<tr>
<td>SHARED AGENT &amp; SHARED THEME</td>
<td>TR</td>
<td>A V1 V2 T</td>
</tr>
<tr>
<td>PIVOT</td>
<td>TR INTR</td>
<td>A V1 T=A V2</td>
</tr>
</tbody>
</table>

Examples of each of the basic syntactic types are provided in (2): shared agent (2a), shared theme (2b), shared agent and shared theme (2c-d), and pivot (2e).

(2a) koət jɔːk kandiəw kaːt srow
3SG take sickle cut rice.plant
‘He/she takes the sickle to cut the rice plant.’ (Sak-Humphry 1995:189)

(2b) kɲom wiej ʃaːw bajk
1SG hit glass break
‘I hit the glass and it broke.’

(2c) koːn baoh sɔmʔaːt pteəh
child sweep cleanCAUS house
‘The child sweeps the house clean.’ (Schiller 1989:408–409)

(2d) baːŋ jɔːk kasaet ?aːn
older.brother take newspaper read
‘The older brother took the newspaper to read it.’ (Wilawan 1995:61)

(2e) viə noəm kmeːŋkmeːŋ ruət ləː phnum
3SG take child-child run on mountain
‘He/she takes the children to run on the mountain.’ (Sak-Humphry 1995:191)

The interactions between these semantic and syntactic subclassification are complex, with no one-to-one correspondence between semantic and syntactic type. There are, however, some interesting correlations to note. Possible Khmer SVCs of the instrumental semantic type (1a) tend

---

² In Table 1, the term Pivot is used in the sense of Chao (1968:124–125). I am grateful to the anonymous reviewer for pointing this fact out to me. Haiman (2011:277) and Bisang (1992:438) both use the term to describe what I have here designated as “shared theme” as well. Note that this use of “pivot” is entirely separate from the use of the term “pivot” to designate the epistemic source or point of view parameter in discussions of logophoricity and long-distance reflexives (Sells 1987). Abbreviations throughout this paper conform to the Leipzing Glossing Rules with the addition of T designating “theme”.

to be of the shared agent syntactic type (2a). Beneficiary constructions (1c) tend likewise to be of
the shared agent type, since the verb ʔaoj ‘give’ has no semantic theme argument. Recipient
constructions (1b), on the other hand, tend to be of the shared agent & shared theme syntactic type,
insofar as the usual understanding of the verb ʔaoj ‘give’ in such cases is as a three-place predicate
with the agent and theme of the initial verb (V1) and its own secondary or indirect object, the
recipient. Goal-of-motion constructions (1d) tend to be of the shared theme (2b-c) or shared agent
(2a) syntactic types, while result-of-action constructions (1e) usually belong to the shared theme
category alone. Purpose constructions (1f) tend to belong to the shared agent or pivot (2e) syntactic
types, while manner constructions (1g) usually belong to the shared agent type alone.

As pointed out by a referee, possible SVC constructions in which one of the verbs belongs to a
(semantically and grammatically) restricted class (Aikhenvald’s (2006:3, 30, 35) asymmetrical-
type constructions) tend more quickly toward grammaticalization. In the case of the Khmer
examples in (1), the beneficiary/recipient, and goal semantic types (1b-d), where the second verb in
the series (V2) is almost always ʔaoj ‘give’ (recipient/beneficiary) or təw ‘go’/mɔ:k ‘come’/dɔl
‘arrive’ (goal), respectively, are more likely than the result, purpose, and manner types to represent
stages where V2 is shading off into an adposition rather than being a true lexical verb. The
instrumental type (1a), where V1 is usually either jɔ:k ‘take’ or praə ‘use’, may also be
approaching such grammaticalized status. Grammaticalization, or the lack thereof, however, is an
epiphemomenon of SVCs and not a diagnostic property for their description and definition.

The use of the term SVC in the linguistic literature has been notoriously variable. Arguments
for what constitutes a valid SVC here are made on the basis of the application of a rigorous set of
“diagnostic and characteristic features of SVCs” taken from Kroeger (2004:229–230), itself derived
in large part from Durie (1997). The most crucial of these characteristics is monoclausality. In
order to qualify as an SVC, a given construction must first satisfy the necessary condition of
comprising a single clause without coordination or subordination.

Monoclauasality is demonstrated for a subset of the possible SVCs in Khmer in (1) and (2) by
appealing to the binding behaviour of the bimorphemic reflexive pronoun kluən-æŋ, ‘body-self’. As
one of two reflexive pronouns in the language, kluən-æŋ is both clause-bounded and subject-
oriented: that is, it must take as its antecedent the subject noun phrase (NP) of its own minimal
clause and is therefore restricted to non-subject grammatical relations (i.e. object or genitive).
Together with extraction tests to rule out coordination and subordination, this bimorphemic
reflexive can thus serve as an effective delimiter of clause boundaries. The subtypes of putative
SVCs for which monoclausality is demonstrated in this way are the semantic types of instrument
(1a) and purpose (1f). These subtypes are chosen because they are less likely to involve issues of
grammaticalization and more likely to be interpretable as being biclausal than many of the other
subtypes. Thus, if monoclausality can be demonstrated even for these constructions, then, a fortiori,
it is to be expected to apply to constructions in which one of the two verbs has grammaticalized to
a greater or lesser degree. It is additionally shown that verbal constructions like those in (1) and (2)
show the remainder of the key diagnostic and characteristic features of SVCs as well, namely
direct argument sharing, reference to single (complex) events, and agreement in polarity and TAM.

The structure of the paper is as follows. Section 2 describes in detail the problem the study
seeks to address. The section begins by considering the peculiar diversity of opinion regarding the
existence of SVCs in Khmer vis-à-vis the typological and areal commonality of SVCs in Mainland
Southeast Asia (MSEA). Arguments are presented against two extensive investigations of SVCs in
Khmer in particular which have concluded, largely on theory-internal grounds, that all subsequent
verbs (V2) in putative SVCs are non-finite complements of V1. Three specific contentions support

---

3 NB: Slightly similar lists of diagnostic criteria can be found in Aikhenvald (2006:4–21) and Bisang
(2009:794). It should be noted, however, that Aikhenvald (2006:8) allows for negators in SVCs that
negate only part of the construction and do not scope over both (or all) verbs equally; the present
approach does not

4 The other reflexive pronoun, the monomorphemic kluən ‘body’, is a long-distance reflexive that may
occur in subject, object, or genitive grammatical roles within embedded clauses, coindexed with the
matrix subject (Fisher 1985; 1988). This breakdown of reflexive pronouns into bimorphemic local and
monomorphemic long-distance varieties is found in many languages throughout the world (Reinhart &
this position. First, extraction tests such as those performed by Sak-Humphry (1995) show only that the constituent verbs in a putative Khmer SVC are not “islands” for extraction by the Coordination Structure Constraint proposed by Ross (1967) and hence are not coordinate. Second, the inability to cleft and topicalize the entire V2 phrase out of these constructions suggests that it is not a complete constituent and is therefore not likely a subordinate clause. Third, tests showing that V2 cannot be independently negated and that overt coreferential subject NPs among the verbs in putative SVCs are not permitted are consistent with (indeed anticipated by) an SVC analysis. Section 2 also considers the need for terminological precision in applying the term SVC. Much of the confusion in the literature on SVCs in Khmer results from imprecise definitions of the category of SVC, lumping explicitly biclausal complement constructions in with properly monoclausal SVCs. A strict set of criteria is needed in order to be able to carve out a theoretical space in which to uniquely and exhaustively designate a given construction as an SVC and not have some other, less controversial categorical label apply equally well or even better. Section 3 lays out the proposed solution to the problem sketched in Section 2. Subsection 3.1 demonstrates the clause-bound behavior of the bimorphemic reflexive kloun-aen and applies it as a diagnostic test to demonstrate monoclausality for SVCs. Subsection 3.2 looks at the evidence of extraction tests to show that the constructions in (1) and (2) do not involve either coordination, whether overt or covert, or subordination. Subsection 3.3 demonstrates that the verbal constructions must additionally agree in polarity and TAM, thus fully satisfying the diagnostic and characteristic features of SVCs.

Evidence adduced throughout the paper includes both published data from previous grammars and articles dating from 1915 to 2011 (Maspero 1915; Huffman, Promchan & Lambert 1970; Spruill 1988; Meyer 1992; Wilawan 1995; Smyth 2008; Haiman 2011) as well as critical new data elicited from four native speakers living in Dallas, TX.

2. The Problem

SVCs are robustly represented among the languages of MSEA. In fact, they are widely regarded as an areal feature (Clark 1989; Clark 1992; Migliazza 1996; Aikhenvald & Dixon 2006). Table 2 presents just a sampling of languages across the major families of MSEA for which SVCs, by one definition or another, have been documented. As can be surmised from the table, Austroasiatic languages particularly abound in SVC behavior.5

Table 2: SVCs in MSEA

<table>
<thead>
<tr>
<th>Austroasiatic</th>
<th>Tai</th>
<th>Miao-Yao</th>
<th>Tibeto-Burman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeh (Gradin 1976)</td>
<td>Black Tai (Fippinger 1975)</td>
<td></td>
<td>Kayah Li (Solnit 1997)</td>
</tr>
<tr>
<td>Kammu (Holmer 2005)</td>
<td>Nung (Saul &amp; Wilson 1980)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semelai (Kruspe 2004)</td>
<td>Lao (Enfield 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnamese (Mikami 1981; Thompson 1984; Sophana 1997)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Yet, when we turn our attention to standard Cambodian, or Khmer, one of only two Austroasiatic languages with official status as a national language (the other being Vietnamese), the question of whether SVCs exist in the language seems vexed and open. Khmer surfaces in the literature on SVCs principally in the context of larger discussions of SVCs as a general phenomenon. Various authors cite example sentences from Khmer involving what are labelled as SVCs as though the language constituted a standard exemplar of SVC behaviour. Little to no

5 Cf. Seuren (1990:15): “The centrally relevant phenomena for SVCs are typically found in certain groups of languages in certain restricted geographical areas, notably the Kwa languages spoken in parts of West Africa, most of the Caribbean Creole languages, many East and South-East Asian languages, in particular Chinese and the Khmer group, and, it seems, some languages of Papua New Guinea, including the Creole language Tok Pisin” [emphasis mine].
discussion is usually offered. Table 3 presents a sampling from some of the available general SVC literature of the frequency of Khmer examples used to illustrate SVC behaviour.

**Table 3**: Khmer example sentences in SVC literature

- (Clark 1989): 1 Khmer example
- (Clark 1992): 4 Khmer examples
- (Durie 1997): 1 Khmer example (quoted from Jacob 1968)
- (Mikami 1981): >10 Khmer examples
- (Schiller 1989): 8 Khmer examples

Mikami (1981) stands out in this list as something of an exception. His article explicitly contrasts the behaviour of apparent SVCs in Vietnamese and Khmer. Accordingly, it provides a relative wealth of example sentences. However, Mikami’s discussion is vitiated by his imprecise application of the label “serial verb construction” to a series of complex sentences that include sentential and verb-phrase (VP) complementation and relative clause constructions. In addition, the native Khmer consultants I worked with rejected many of Mikami’s example sentences as either wholly ungrammatical or at least pragmatically odd. These faults render the article problematic as an assessment of SVCs in Khmer.

More perplexing still, two in extenso treatments of Khmer SVC behaviour per se, Wilawan (1992; 1995) and SakkHumphry (1995), outright deny that the language makes use of any construction that can be labelled an SVC while being simultaneously distinguished from a non-finite complement clause. That is, both scholars argue that since the Khmer constructions in (1) and (2) above cannot contain two overt NPs that refer to the same argument without forcing an interpretation of the whole utterance as comprising two paratactically arranged full sentences, V2 must be construed as non-finite and thus subordinate to V1 in a kind of complement clause. This conclusion is made all the more striking for the fact that the two scholars who share it also subscribe to the same theoretical orientation and both earned doctorates at the University of Hawai’i at Mānoa under the supervision of the late Stanley Starosta, originator of the grammatical theory known as Lexicase Dependency Grammar. The fact that Wilawan (1992; 1995) and SakkHumphry (1995) both apply a Lexicase analysis to the topic of SVCs in Khmer, I maintain, helps explain their conclusions on the topic of SVCs in Khmer.

As its name suggests, the Lexicase framework is a dependency grammar that has, at its heart, the dependency relation: “Every word in a sentence is the head of its own construction, and every lexical item in a sentence but one, the main verb (or non-verbal predicator), is dependent on one and only one other lexical item, its REGENT” (Starosta 1988:104). Already, one can see how this approach is likely to founder when dealing with a topic like SVCs, which have so exercised theoretical syntacticians precisely because they appear to involve multiple verbs that are syntactic sisters within a single clause. However, Starosta’s Lexicase theory also defines the concepts “finite” and “non-finite” as applied to verbs in terms of the disallowance of overt coreferential subjects within the minimal clause (Starosta 1988:68; Wilawan 1992:1243). It is likely this a priori theoretical consideration that motivates both Sakk-Humphry and Wilawan to conclude that V2 in a putative SVC must necessarily be non-finite and dependent upon V1 as regent. Wilawan (1995:55) writes:

---

6 E.g. Mikami (1981:106) example (12c): #/*knom baːn koət təw psaː ‘ʔaoj (intended meaning: ‘I got him to go to the market for me’).
7 The work of Bisang (1992; 2009) also stands out in the literature for its treatment of Khmer data in discussions of SVCs across SE Asia. However, Bisang’s discussions have paid more attention to the concept of single eventhood and parameters for its description independent of cultural factors than to a detailed analysis of monoclausality as it relates to SVCs in Khmer, such as is presented here.
8 Sakk-Humphrey (1995:201): “This paper enables me to examine the relationships between the head of the construction and its dependents in a sentence that has more than two verbs with no coordination or subordination marker (including the completive constructions) which were called by others as serial verb constructions. Thus, there is no difference between the serial verb construction and the non-finite subordination construction.” Wilawan (1995:66): “No distinct ‘serial verb construction proper’ was found to be justified. Instead, the V2 in these series is analyzed as a non-finite complement verb which is a subordination dependent [sic] of a main verb, as shown by its syntactic characteristics.”
All of the previous analyses of the SVCP [serial verb construction proper] explicitly or apparently agree that only the first verb in SVCPs allows an overt subject noun phrase. From the dependency analysis point of view, this third characteristic of SVCPs implies that only the first verb in a series is a finite verb, while the rest are non-finite.

SakkHumphry (1995:180) concurs, noting simply: “[I]f an overt subject cannot be inserted, then the V2 is a non-finite verb” and “[the verb] is non-finite if we cannot add the subject in the second position and still have a single sentence.” In a language like Khmer, where zero anaphora is the rule rather than the exception (Haiman 2011:199), this assumption can lead to significant problems. It also presupposes that what has been recognized as a key feature of SVCs cross-linguistically, viz. obligatory non-coreference of overt NPs, is an indicator of non-finiteness. Yet it is precisely because V2 cannot take an overt subject NP coreferential with the argument of the subject NP of V1 that these constructions are suspicious for being SVCs in the first place.

To make matters worse, in his recent reference grammar of Khmer, Haiman (2011) contradicts the view of both Wilawan and SakkHumphry in asserting that the language does in fact have SVCs. However, Haiman’s discussion of the phenomena suffers from the same imprecision in terminology as Mikami’s. Haiman includes in his category of SVCs resultative and so-called “success verb” constructions that behave quite differently from the types of sentences under consideration here (Haiman 2011:271).

Another look at SVCs in Khmer is clearly called for, one that avoids both the pitfalls of imprecision in usage of the label SVC to designate a particular sub-category of multi-verb construction in the language and theory-internal a priori assumptions that force us to presuppose simple answers to the very complex theoretical questions posed by SVCs to begin with.

3. The Proposed Solution

The first step in proposing a solution to the problem outlined above is to clear up precisely what is meant by the term SVC. In order to accomplish this, the “diagnostic and characteristic features of SVCs” discussed in Kroeger (2004:229–230) and given in Table 4 below are adopted.

---

9 E.g. As Wilawan (1995:61–62, n. 12) notes, sentences like (i) present a real problem for Lexicase theory. Here, the conjunction haəj ‘and’ links what is, from a Lexicase perspective, a finite clause with a non-finite clause. Assuming that coordination should conjoin only grammatically equivalent constructions (Starosta 1988:248), coordinations like this one should not be possible. As a result of her theoretical commitment, Wilawan is forced to posit a structure that she calls “non-finite coordination,” which has the unique asymmetrical property of coordinating two verbs of unequal finiteness and requiring that their relative order be fixed.

(i) kɲom tiw bɔːŋ haəj tiw psaː phɑːŋ
1SG go bank and go market also
‘I go to the bank and then [I] also go to the market.’ (Ehman & Sos 1972:73)

10 To be completely fair, there was a prior tradition in the literature of trying to fit SVCs into the complementation mold. Seuren (1990:20) refers to SVCs as an example of “pseudocomplementation”: that is, they involve embedded sentences that function as sentential complements but with the semantic roles of of “concomitant, resultative, purposive circumstance or event.” Fisher (1988:31) uses a test similar to Sakk-Humphry’s to distinguish between infinitival and sentential complements. We shall have more to say about her examples below.

11 NB: Haiman’s use of the terms “resultative” and “success” seem rather interchangeable and, as he acknowledges (Haiman 2011:271) both answer to Huffman et al.’s (1970:187–188) category of “completive verbs”.

12 For specific examples of how these categories of verbal behaviour differ from putative SVCs, see section 3.2 below. Perhaps the relative imprecision of Haiman’s use of the designation SVC is to be expected given comments elsewhere in his book that seem to indicate a certain hostility to overly fine terminological distinctions in the grammatical description: “In fact, the suspicion that all of the present taxonomy of AV [auxiliary verb], MV [main verb], and SV [serial verb] is an extended example of ethnocentric grammatical pedantry is a salutary one...” (Haiman 2011:280); “Rigor has no place in Khmer grammar” (Haiman 2011:354).
**Table 4: Diagnostic and characteristic features of SVCs**

1. A prototypical SVC contains two or more morphologically independent verbs within the same clause, neither of which is an auxiliary verb.
2. There are no conjunctions or subordination intervening or separating them.
3. The serial verbs belong to the same intonation contour, with no pause separating them.
4. The entire SVC refers to a single (possibly complex) event.
5. A prototypical SVC contains only one specification for tense, aspect, modality, negation, etc. These features may be redundantly marked on more than one verb.
6. The verbs share at least one semantic argument.
7. There is no co-reference among direct arguments: a true SVC does not contain two overt NPs referring to the same argument.
8. A prototypical SVC contains only one grammatical subject.

As has been widely noted (Seuren 1990:14; Lord 1993:2; Anderson 2010:12), verb serialization is less a universal category and more a syndrome of features and phenomena. Thus a functional-contrastional approach like that which Anderson (2010:1) takes to auxiliary verb constructions is necessitated for dealing with SVCs as well: functional in that the object of study comprises a continuum of multi-verb constructions that operate over a certain range of functional domains and constructional in that we are concerned with the formal encoding of these functions in morphosyntax at the construction level. Nonetheless, in order to preserve the validity of cross-linguistic comparisons of SVCs, some kind of agreed-upon standard of definition or set of featural requirements is needed. The list in Table 4 is designed to fill just such a need.

This list comprises a highly restrictive set of criteria for judging potential SVCs. Such restrictiveness is necessary in order to ensure that whatever constructions bear the label SVC cannot simultaneously belong to any other terminological category, such as Sak-Humphry’s (1995) non-finite complement or even Seuren’s (1990) “pseudocomplementation.” In this way, the theoretical integrity of the class SVC is preserved.

No evidence will be presented here concerning the third, phonological criterion in Table 4; the reader is asked to take for granted that constructions like those in (1) and (2) are not spoken with significant pauses between the constituent verbs. The astute reader will also have noticed that the constructions contained in the category labelled “Pivot” in Table 1 and example (2) above violate feature 8 from Table 4 in that the theme of V1 is simultaneously interpreted as being the subject of V2, yielding two different grammatical subjects across the SVC as a whole. This is the only respect in which the putative Khmer SVCs surveyed here fall short of prototypicality for SVCs.\(^\text{13}\)

### 3.1 Monoclauasality in Khmer SVCs

Monoclauasality is a necessary condition for identifying a given construction as an SVC. It is not, however, a sufficient condition, as coordination of VPs is still possible within a single clause. Moreover, obligatory control can mean that a reflexive pronoun takes as its antecedent an implicit subject-controllee within its same minimal clause (such as PRO). Thus, in establishing the monoclauasality of possible SVCs in Khmer, care must be taken not only to locate probable clause boundaries, but also to rule out covert coordination of VPs and subordination of non-finite clauses with empty/null subjects.

\(^{13}\) It should be noted, however, that Anderson (2010:12) labels as “‘classic’ serializing combinations” both SVCs of the same-subject ‘take come’ > ‘bring’ type and those of the switch-subject ‘hit die’ > ‘kill’ type.

Subsection 3.1.1 below shows how the binding facts of the bimorphemic reflexive pronoun *kluənyæŋ* can be used to help locate clause boundaries in Khmer sentences. Subsection 3.1.2 then tackles the problem of ruling out covert coordination in possible SVCs by showing that suspected SVCs are not islands for extraction in the way that coordinate structures are. Subsection 3.1.3 likewise rules out interpretations of the V2 phrase as a non-finite complement by showing that it cannot be extracted and is therefore most likely not a complete constituent. Since subordinate clauses are not likely to comprise sub-minimal constituents, we can rule out the possible effects of a null subject like PRO on reflexive binding.

### 3.1.1 Locating clause boundaries

Like its bimorphemic counterpart in English, the Khmer bimorphemic reflexive *kluənyæŋ* is clause-bounded (i.e. it must find its antecedent within its own minimal clause). Unlike the English reflexive, however, Khmer *kluənyæŋ* is subject-oriented. Non-reflexive pronouns cannot take an antecedent within their same minimal clause if they function as direct arguments. As possessors, they can occur in the same minimal clause with their antecedents but, if third person (3P), are anti-subject oriented. These facts are demonstrated in (3).

(3a)  
Mora wiej Sokh niw pteah *kluənyæŋ*  
Mora hit Sokh be.at house self  
‘Mora hit Sokh at his own [i.e. Mora’s / *Sokh’s / *some other person’s] house.’

(3b)  
Mora wiej Sokh niw pteah koət  
Mora hit Sokh be.at house 3sg  
‘Mora hit Sokh at his [i.e. *Mora’s / Sokh’s / some other person’s] house.’

(3c)  
#kɲom baŋ Mora sɔmlap koət  
1SG shoot Mora kill 3SG  
#‘I shoot Mora [and] kill him [i.e. another person]’

Example (3a) shows that *kluənyæŋ* takes an antecedent within its own clause but is oriented toward the subject of the clause and not the other possible antecedent: the object, Sokh. Example (3b) demonstrates that a possessor non-reflexive pronoun can take its antecedent within the minimal clause, or it may refer outside of the clause. If the former, however, the pronoun must refer back to the object, Sokh, and not the subject, Mora, due to its anti-subject orientation. Finally, (3c) shows that if a non-reflexive pronoun is a direct argument within the clause, it must refer to some individual outside of the clause, which, in this case, gives rise to the pragmatically odd meaning that the subject of the sentence, I, shot Mora but, in so doing, killed some unnamed third party.

In (4), we observe the expected behaviour of *kluənyæŋ* in sentences comprising two clauses. The bimorphemic reflexive cannot refer outside of the downstairs clause. The non-reflexive pronoun, however, must refer outside of its same clause.

(4a)  
Mora prap Sokh thaː nìŋ sra-ləŋ *kluənyæŋ*  
Mora tell Sokh say girl love self  
‘Mora tells Sokh that the girl loves herself / *Mora / *Sokh.’

(4b)  
Mora prap Sokh thaː nìŋ sra-ləŋ koət  
Mora tell Sokh say girl love 3SG  
‘Mora tells Sokh that the girl loves him [i.e. Mora / Sokh / someone else].’

(4c)  
#kɲom twəː ʔaoj koət sɔmlap *kluənyæŋ*  
1SG make give 3SG kill self  
#‘I make him kill himself / *me.’

(4d)  
#kɲom twəː ʔaoj koət sɔmlap koət  
1SG make give 3SG kill 3SG  
#‘I make him kill him [i.e. another person].’

In (4a), *kluənyæŋ* can only take as its antecedent the girl, subject of the sentential complement; neither actor from the matrix clause is eligible. In (4b), on the other hand, the non-reflexive gender-neutral 3P pronoun *koət* can refer to either actor from the matrix clause or to some third party, but
not to the girl within its same minimal clause. Examples (4c) and (4d) demonstrate identical distinctions with a different matrix predicate and complementizer.

Turning to the behaviour of kluənyæŋ with suspected SVCs, we observe that, despite the presence of intervening verbs, the reflexive nonetheless takes as its antecedent the subject of V1 (5a). The non-reflexive pronoun must refer to someone outside of the sentence (5b).

(5a)  Sokh baɲ səməlp kluənyæŋ
Sokh shoot CAUS.die self
‘Sokh shoots and kills himself.’

(5b)  Sokh baɲ səməlp koət
Sokh shoot CAUS.die 3SG
‘Sokh shoots and kills him [i.e. someone else].’

Even with more elaborate constructions, involving additional intervening verbs, this essential behaviour of reflexive kluənyæŋ versus non-reflexive koət is maintained. For example, in sentences of the shared agent type, where each verb in the construction takes its own unique object, many Khmer speakers prefer to insert “extra” directional verbs like mɔːk ‘come’ and tiw ‘go’ before V2 (6). When asked explicitly who or what were the subjects of the inserted directional verbs, consultants responded that they interpreted the subject of V1 as supplying the subjects for both the directional verbs and V2.

(6a)  Sokh jɔːk kɑmbət (tiw) səməlp kluənyæŋ
Sokh take knife (go) CAUS.die self
‘Sokh takes the knife to (go) kill himself.’

(6b)  Sokh jɔːk kəmphləːŋ (tiw) baɲ koət
Sokh take gun (go) shoot 3SG
‘Sokh takes the gun to (go) shoot him [i.e. someone else].’

The examples in (7) present still more elaborate structures, while demonstrating the same reflexive binding facts.

(7a)  Mora prap Sokh thaː kruː-pəːt tɨw tɨn ceːk nɨw pteah kluənyæŋ
Mora tell Sokh say doctor go buy banana eat be.at house self
‘Mora tells Sokh that the doctor is going to buy bananas to eat in (his own) [i.e. the doctor’s / *Mora’s / *Sokh’s] home.’

(7b)  kɲom noəm kruː-pəːt tɨw tɨn ceːk nɨw pteah kluənyæŋ
1sg take doctor go buy banana eat be.at house self
‘I take the doctor to go buy bananas and eat them in my / *the doctor’s house.’

(7c)  Mora prap Sokh thaː kɲom noəm kruː-pəːt tɨw tɨn ceːk nɨw pteah kluənyæŋ
Mora tell Sokh say 1sg take doctor go buy banana eat
nɨw pteah kluənyæŋ be.at house self
‘Mora tells Sokh that I take the doctor to go buy bananas and eat them in my / *the doctor’s / *Mora’s / *Sokh’s house.’

In (7a), four lexical verbs intervene between the reflexive kluənyæŋ and the subject of the embedded clause, kruː-pəːt ‘doctor’. Two of these verbs have their own individual complements (tɨn ceːk ‘buy bananas’ and nɨw pteah ‘be.at house’). Nevertheless, the reflexive still takes as its unambiguous antecedent the subject of V1, doctor. In (7b), five lexical verbs intervene, three with their own complements (including noəm kruː-pəːt ‘take doctor’). Even so, the antecedent of the reflexive remains the subject of V1. Sentence (7c) does nothing more than embed (7b) inside the frame from (7a), where the 1P subject of the embedded clause continues to supply the unambiguous antecedent for the reflexive pronoun at the end of the sentence.

The evidence from the binding behavior of the bimorphemic reflexive pronoun kluənyæŋ suggests, at least initially, that the various verbs in the kinds of possible SVCs under consideration all fall within the same minimal clause. It still remains, however, to rule out definitively covert VP coordination as well as subordinated complementation structures of the sort discussed by Sak-Humphry (1995) and Wilawan (1992; 1995).
In her (1995) article, Sak-Humphry performs a series of extraction tests, demonstrating that the nominal complements of either of the verbs in a putative SVC can be fronted for topicalization or focus. Sak-Humphry explains the purpose of her tests as being to determine that the nominal complements of the verbs in a multi-verb construction are in fact nouns and the heads of their own NPs. This practice is necessitated, as she sees it, by claims attributed to unnamed analysts in the SVC literature to the effect that: 1) SVCs do not take nominal complements; 2) the V2s are not verbs at all, but adverbs; and 3) the combination of V1 and V2 is a compound verb (SakkHumphry 1995:182). Though space does not permit a repetition of all of her examples, I have given a representative sample of the tests Sak-Humphrey applies in (8) and (9). Since an NP in Khmer must be definite in order to be topicalized, the examples of topicalization in (8) all involve the addition of the demonstrative .nih ‘this’ following the fronted, topicalized NP. The focus-cleft constructions in (9) consist of the fronted focal element immediately preceeded by the focus-marking word kɨ: and followed by the relativizer daɛl.

(8a) sɪəwphəw nih kmeːŋ mɔːk psaː tɨɲ
book DEM child come market buy
‘These books the child comes to the market to buy.’ (SakkHumphry 2005:186)

(8b) sroːw nih Sokh jɔːk kandiəw kaːt
rice.plant DEM Sokh take sickle cut
‘This rice plant Sokh takes the sickle to cut.’ (SakkHumphry 2005:190)

(8c) kmeːŋ kmeːŋ nih wiə noəm ruət ləː phnum
children.children DEM 3SG take run up mountain
‘These children, he took [them] walking on the mountain.’ (Sakk-Humphry 2005:193)

(9a) kɨː sɪəwphəw daɛl kmeːŋ mɔːk psaː tɨɲ
FOC book REL child come market buy
‘It is the books that the child comes to the market to buy.’ (Sakk-Humphry 2005:187)

(9b) kɨː kandiəw daɛl Sokh jɔːk kaːt sroːw
FOC sickle REL Sokh take cut rice.plant
‘It is the sickle which he takes to cut the rice plant.’ (Sakk-Humphry 2005:191)

(9c) kɨː kmeːŋ kmeːŋ nih daɛl wiə noəm daːə ləː phnum
FOC children.children DEM REL 3SG take walk up mountain
‘It is these children whom he took walking on the mountain.’ (Sakk-Humphry 2005:193)

These tests do serve to counter the three positions on multi-verb constructions that Sak-Humphry specifically militates against. What they show even more clearly, though, is that the domains of the two verbs are not “islands” for extraction in the sense proposed by Ross (1967). If the two VPs were coordinate, their respective NP complements would not be able to be extracted, as is shown in (10).

(10a) *The girl he hit and kicked the boy. / *The girl he hit the boy and kicked.
(10b) *It is the girl he hit and kicked the boy. / *It is the girl he hit the boy and kicked.

Because such extractions are possible with the putative Khmer SVCs, we can conclude that they do not involve covert coordination.

3.1.3 Ruling out subordination

Sakk-Humphry (1995) additionally notes at various points in her discussion that what she considers to be in each case the dependent non-finite clause cannot itself be extracted for focus (1995:184) or topicalization (1995:183). Though she asserts this fact to be true, she does not actually give ungrammatical examples. The sentences in (11) supply the missing negative examples.

(11a) *tɨɲ sɪəwphəw nih kmeːŋ mɔːk psaː
buy book DEM child come market
‘To buy these books, the child comes to the market.’
SakkHumphry intends this fact to confirm that the V2s are non-finite subordinate clauses dependent on the main or regent V1s, though, I admit, the precise rationale behind such a claim remains unclear to me. SakkHumphry’s tests are actually similar to the facts Roberts (2012:219–220) cites in support of the conclusion that certain non-finite clauses in English multi-verb constructions are not subordinate. In making this claim, Roberts is summarizing an argument from Van Valin & LaPolla (1997:469–471) that certain non-finite complement clauses cannot be passivized or focused (12-13) in the way that other expressly subordinated clauses can (14). Presumably, the tacit assumption behind Robert’s claim is that the clausal complements in question occupy the object position in their matrix clauses and must, therefore, have some of the properties of regular nominal (i.e. direct object) complements. Similarly, in Khmer we can show that certain expressly subordinated clauses can be focused, as in (15). Since we cannot similarly topicalize or focus V2 phrases from putative SVCs, as shown in (11) above, then, following Robert’s logic, we might be tempted to conclude that these facts provide some evidence against considering V2 part of a subordinate clause.

(12a) Mary stopped crying.
(12b) *Crying was stopped by Mary.
(12c) *It was crying that Mary stopped.
(13a) Sam helped run the tournament.
(13b) *Run the tournament was helped by Sam.
(13c) *It was run the tournament that Sam helped.
(14) To do our best is demanded of us all.

Indeed, what the extraction tests in (11) show most clearly is that the V2 phrases which could not be topicalized or clefted must not constitute complete constituents, because only complete constituents (i.e. full NPs, PPs, VPs, etc.) are usually available for topicalization and clefting (Kroeger 2004:27–29). Thus, the V2 phrases most likely cannot constitute complete clauses unto themselves.

To further underscore this conclusion, we can contrast this behavior with that of a true non-finite complement clause for a control predicate like cɔŋ ‘want’. Predicates like ‘want’ can take open complement constructions that look superficially much like potential SVCs in Khmer. As shown in example (16), when the subject of the complement is the same as the subject of cɔŋ itself, an overt coreferential subject NP is not allowed. On the other hand, predicates like ‘think’ and ‘say’ that take a sentential complement following the complementizer thaː ‘say’ permit subsequent verbs to appear with or without such overt coreferential subjects.

(16) Sina, prap thaː (wiə i) cɔŋ (*wiə) tiw psa:
Sina say COMP 3SG want 3SG go market
’Sina says that he wants to go to the market.’ (Fisher 1988:31)

Fisher (1988:31) uses these facts as evidence to support a claim that same-subject uses of cɔŋ take a necessarily infinitival complement with null subject PRO.
The verb *cɔŋ* ‘want’ may also occur with sentential complements and a complementizer (the verb *ʔaoj* ‘give’), but only when the subject of the complement is distinct from that of *cɔŋ* (switch-subject). Since this construction is immediately recognizable for the biclausal structure that it is and is thus not superficially similar to potential SVCs, it need not concern us here.

As is demonstrated in (17), the surface form of a same-subject sentence with *cɔŋ* appears at first much like the shared agent-shared theme type of potential SVCs given in Table 1 and (2c) above.

(17) kɲom cɔŋ niʔjiej piəsaː khmaɛ
1SG want speak language Khmer
‘I want to speak Khmer.’

Its behavior with reflexive pronouns is also similar, as shown in (18).

(18) kɲom cɔŋ sɔmlap kluənyæŋ
1SG want CAUS.die self
‘I want to kill myself.’

However, tests attempting to extract the V2 phrase from these sentences produce different results from those obtained for other apparent SVCs.

(19a) ?kiː sɔmlap kluənyæŋ daɛl kɲom cɔŋ
COP CAUS.die self REL 1SG want
Intended meaning: ‘It is to kill myself that I want.’

(19b) kaː daɛl kɲom cɔŋ kiː sɔmlap kluənyæŋ
NMLZ REL 1SG want COP CAUS.die self
‘What I want is to kill myself.’

Whereas the focus cleft in (19a) was generally judged as dubious by consultants, the pseudo-cleft in (19b), where the focused element occurs at the end, still preceded by the focus-marker *kiː*, proved acceptable. This evidence would seem to suggest that, in fact, the V2 phrase constitutes a complete constituent here (VP2) with an implicit subject controlled by the subject of *cɔŋ* itself. This behavior stands in stark contrast to that of potential SVCs, which do not allow extraction of the V2 phrase because it is a sub-minimal constituent and, therefore, not a subordinate clause.

### 3.2 Ruling out auxiliary verbs

Anderson (2010:11–12) observes that SVCs have provided one of the major historical sources for auxiliary verb formation. Through a process of grammaticalization, either V1 or V2 in an SVC becomes specialized and develops functional semantics. While Anderson (2010:13) argues that there is no hard-and-fast line of demarcation between a serialized verb that is in the process of grammaticalizing and an auxiliary verb (indeed a period of ambiguity is to be expected over the course of the transition), he inclines toward an auxiliary verb analysis when the functional semantics become the default interpretation. Kuteva (2001:1–2) similarly notes that once the lexical meaning of a verb has been entirely supplanted by a grammatical-functional semantics, the process of grammaticalization to an auxiliary verb is complete.

The resultative or “success” verb constructions Haiman (2011:271) includes within his category of SVC are potential candidates for being auxiliary verbs.14 These two constructions differ from the potential SVCs under consideration in two chief respects. First, resultative and success verb constructions in Khmer permit, and indeed require, independent negation of the second (i.e. resultative/success) verb (V2) (20a-b) (Huffman 1967:171). While negation of the success verb *baːn* ‘be able’ effectively scopes over the entire sequence, since negating one’s ability to do X necessarily entails that one does not do X, other cases make it clear that the interposed negator often scopes only over V2 (20c-d). In (20c), the subject is definitely looking for his wallet; the negation comes into play only in denying the success of the search. In (20d), the subject is listening...

---

14 The referee points out, however, that resultative and success verb constructions in Khmer are subject to lexical restrictions and are thus not as productive as either auxiliary verbs or SVCs.
but simply unable to hear. Each of these sentences can also be expanded into an overtly biclausal structure by interposing the conjunction bontae ‘but’ between V1 and the negator-V2 pair (20e-g). This move makes the conative semantics of V1 even more explicit.

(20a) kɲom niʔjiej piəsaː khmaɛ min baːn te
    1SG speak language Khmer NEG be.ble NEG
‘I cannot speak Khmer.’

(20b) *kɲom min niʔjiej piəsaː khmaɛ baːn te
    1SG NEG speak language Khmer be.able NEG

(20c) kɲom rɔk ka:bup min khəɲ
    think look.for wallet NEG see
‘I can’t find (lit. look for, don’t find) [my] wallet.’ (adapted from Bisang 2009:800)

(20d) kɲom sdap min luː te
    1SG listen NEG hear NEG
‘I can’t hear [you] (lit. listen, not hear).’ (Smyth 2008:107)

(20e) kɲom niʔjiej piəsaː khmaɛ bontaɛ min baːn lʔɔ:
    1SG speak language Khmer but NEG be.ble good
‘I [try to] speak Khmer, but cannot [speak it] well.’

(20f) kɲom rɔk ka:bup bontaɛ min khəɲ
    think look.for wallet but NEG see
‘I looked for [my] wallet, but didn’t find [it].’

(20g) kɲom sdap bontaɛ min luː:
    1SG listen but NEG hear
‘I [try to] listen but don’t hear.’

The second key distinction between resultative and success verb constructions and SVCs in Khmer is that the semantics of resultative and success verb structures are often not compositional (21), while SVCs not only often have entirely compositional semantics but also usually require iconic word order, such that the order of actions described matches the order of operations in the external world.15

(21a) məːl min khəɲ
    look NEG see
‘did not see’ (Maspero 1915:408)

(21b) sdap min luː:
    listen NEG hear
‘did not hear’ (Haiman 2011:291)

(21c) kɨt min khəɲ
    think NEG see
‘did not figure out’ (Huffman, Promchan & Lambert 1970:187)

(21d) hoːp kaət
    eat be.born
‘managed to eat’ (Haiman 2011:71)

Perhaps a better example of an auxiliary verb for the present purposes would be trəw. This verb can still be used on its own with the full lexical meaning of ‘to hit’, but it also functions as an auxiliary verb in two constructions that bear superficial similarity to potential SVCs. As an auxiliary, trəw may serve to indicate epistemic/deontic modality much as ‘must’ does in English. Alternatively, it may be used in passive-like structures, where it indicates that the subject has undergone the action of the subsequent V2. In this latter use, trəw may either take an infinitival open complement VP2 without overt subject, or it may take an entire sentential complement without complementizer. In this latter case, VP2 will have its own overt subject NP, expressing the

agent of the action. The agent of the action may also be expressed as the object of the preposition *daoj* ‘by’. Both of these auxiliary verb uses of *trəw* may occur together with potential SVCs.

(22a) *nisət trəw mɔːk saːlaː rion*
student must come school study
‘The student must come to school and study.’

(22b) *nisət (min) trəw (*min) mɔːk saːlaː rion te*
student NEG must NEG come school study NEG
‘The student must not come to school and study.’

(23a) *kɑmbət trəw jɔːk kaːt sac (*daoj bɔːŋ)*
knife undergo take cut meat (by older.brother)
‘The knife was taken to cut the meat by older brother.’

(23b) *sac trəw keː jɔːk kɑmbət kaːt*
meat undergo 3PL take knife cut
‘The meat was cut by them with a knife.’

Example (22a) demonstrates the modal use of auxiliary *trəw* with a potential SVC in V2 position. Example (22b) shows that, as we shall see for SVCs as well, modal *trəw* cannot be left outside of the bracketing negation: that is, both the auxiliary and any subsequent verb must agree in polarity. Thus, example (22) serves to illustrate the surface similarity between uses of modal *trəw* and the SVC structures under consideration.

Example (23) shows the passive-marking use of *trəw*. In (23a), the agent is expressed as an obligatory prepositional adjunct; in (23b), on the other hand, the agent remains in situ within the sentential complement. This distinctive behavior, coupled with the more-or-less completely grammatical/functional meaning of *trəw* in these sentences clearly sets it and similar auxiliary verbs apart from the kinds of potential SVCs surveyed above. Nevertheless, in keeping with the functional-constructional approach taken here, each potential candidate for auxiliary verb versus serialized verb status will have to be evaluated on a case-by-case basis.

### 3.3 Polarity agreement across SVCs

Along with her extensive extraction tests, Sak-Humphry (2005) also performs a negation test on the multi-verb constructions she examines in order to determine whether V2 can be negated separately from V1 (24). The Khmer negative involves an obligatory preverbal adverb (**min**, ʔɑt, *pum*) and an optional sentence-final bracketing element (*te*). On the basis of the fact that V2 cannot be separately negated, Sak-Humphry concludes that it must be non-finite and dependent upon V1 as a main verb, which must bear any negation for the construction as a whole (2005:182).

(24a) *wiə ɔŋkuj min jum kraːom daːəmchəː teː*
3SG sit NEG cry under tree NEG
**‘She sits not crying under the tree.’** (Sak-Humphry 2005:182)

(24b) *wiə min ɔŋkuj jum kraːom daːəmchəː teː*
3SG NEG sit cry under tree NEG
‘She does not sit crying under the tree.’ (Sak-Humphry 2005:181)

The behavior of the negative in (24) is quite similar to that observed in (22b) for the modal auxiliary verb *trəw*. It is also similar to the way the negative interacts with the control predicate *cɔŋ*, as shown in (25) below. Recall that it has previously been argued that this verb takes an open complement with null subject PRO whenever the controllee and controller have the same referent.

(25) *kɲom min cɔŋ (*min) sɔmlap kluənyæŋ teː*
1SG NEG want NEG CAUS.die self NEG
‘I don’t want to kill myself.’

These negation facts, however, are also consistent with (indeed anticipated by) an interpretation of these structures as SVCs. Kroeger (2004:230) writes: “One clear indication that the two serialized verbs express a single event is that we cannot negate one verb while still asserting the truth of the other.” As was seen in (21) above, the separate negatability of resultative and success verbs is one of the two chief factors differentiating those structures from the putative
SVCs in Khmer. The fact that V1 and V2 cannot be separately negated in example (24) can therefore be taken as an indication that the sentence in fact contains a true SVC.

3.4 TAM agreement across SVCs

It is well known from the extensive literature on SVCs that tense and/or aspect marking usually must be uniform across both verbs in the construction (Kroeger 2004:230; Durie 1997:291). This characteristic holds true for the Khmer multi-verb constructions under consideration as well. In Khmer, future tense is marked by what Haiman (2011:263) calls the “dedicated auxiliary verb” nɨŋ. It occurs directly in front of the verb it inflects and, in a multi-verb construction, can occur on V1 alone or on both V1 and V2; irrespective of how many times nɨŋ shows up, however, the tense value for both verbs is necessarily the same: future. By contrast, the perfective aspect marker haəj, which is itself a grammaticalized verb meaning ‘to finish’, follows the verb it inflects and can only occur at the end of both verbs in a multi-verb construction. In fact, haəj usually occurs at the end of its clause or sentence, which means that if the verb is transitive or has some oblique complement, the aspect marker will follow that element. The examples in (26a–e) below demonstrate the interaction of probable SVCs with these two tense/aspect markers. Example (26f) uses the auxiliary verb trəw ‘must’ to demonstrate that the various verbs in a potential Khmer SVC must also agree in modality.

(26a) kɲom nɨŋ jɔː:k kəmbɔːt (nɨŋ) kaːt sac
 1SG    FUT    take knife (FUT)    cut    meat
  ‘I will take the knife and cut meat.’

(26b) kɲom jɔː:k kəmbɔːt kaːt sac haəj
 1SG    take knife cut meat PFV
  ‘I took the knife and cut meat.’

(26c) kɲom trəw jɔː:k kəmbɔːt kaːt sac
 1SG    must take knife cut meat
  ‘I must take the knife and cut meat.’

(26d) *kɲom jɔː:k kəmbɔːt haəj kaːt sac
 1SG    take knife PFV cut meat
  *‘I took the knife [and] cut meat [now].’

(26e) *kɲom jɔː:k kəmbɔːt nɨŋ kaːt sac
 1SG    take knife FUT cut meat
  *‘I took the knife, will cut meat.’

(26f) *kɲom jɔː:k kəmbɔːt trəw kaːt sac
 1SG    take knife must cut meat
  Intended: ‘I must take the knife and cut meat.’

Just as with the argument from negation above, the fact that the various verbs in these potential SVCs must agree in tense/aspect and modality could be taken as evidence for either a non-finite complement clause analysis or an SVC analysis. When viewed in the context of the arguments for monoclau sality presented above, however, these facts clearly favor an SVC analysis for the present data.

3.5 Single-event interpretation of SVCs

In many ways, the oft-cited requirement that the various verbs in an SVC express actions that are construed in the minds of speakers as component parts of a single, complex event is the least satisfactory of the characteristic and diagnostic features of SVCs. This criterion is unsatisfactory for two main reasons. First, the notion of “single event” is extremely difficult, if not impossible, to define; attempts to draw a clear line of demarcation between component parts of a single complex event and separate, sequential actions involve inevitable arbitrariness (Senft 2004:53–54; Anderson 2010:12). Secondly, as Durie (1997:326–330) and Jarkey (1991:169–170) as pointed out by the reviewer, independent uses of haəj as a lexical verb are rare in Khmer, though Haiman (2011:186) reports it does occur in the idiom kɲom təw haəj (literally ‘I go finish’), meaning something like I can do everything or You know me.

SELF, Stephen. 2014. Another look at serial verb constructions in Khmer. Mon-Khmer Studies 43.1:84-102 (ICAAL5 special issue)
have both noted, different languages impose different requirements on what can constitute a single, complex event. That is, the concept of “single event” is culturally bound. Jarkey’s example from White Hmong illustrates this point particularly well. She discusses the three example sentences given in (27) below.

(27a) nws dhia tshov qeej
3SG dance blow bamboo.pipes
‘He dances playing the pipes.’

(27b) *nws dhia mloog nkauj
3SG dance listen song
‘He dances and listens to music.’

(27c) nws dhia thiab mloog nkauj
3SG dance and listen  song
‘He dances (while) listening to music.’ (Jarkey 1991:169–170)

Speakers of White Hmong interpret (27a) as a perfectly acceptable SVC. (27b), on the other hand, they reject as ungrammatical, despite the fact that the underlying grammatical structure is identical to (27a) in every way. The only acceptable way to express the concept aimed for in (27b) is by using a conjoined clause, as in (27c). To speakers of White Hmong, the actions of listening to music and dancing are not culturally compatible with a single-event interpretation. Playing the bamboo pipes and dancing, however, are conceived of as two component parts of a natural, complex single event and are, therefore, acceptable inside an SVC.

While acknowledging the dangerous slipperiness of the concept, then, we can nevertheless present a striking example from Khmer of a reflex of the kind of single-event interpretation usually associated with SVCs. The SVC example in (28a) below was felt by native speakers to depict normal, expected student behavior with regards to showing obedience to a teacher. No specific instance of an actual command is presupposed, whence the translation with a bare plural in English: ‘commands’. The sentence merely describes the expected cultural norm of student obedience to their teacher. The coordinate structure in (28b), on the other hand, emphasizes a temporal discontinuity between the two conjuncts. Consultants explained that (28b) would presuppose a situation in which the teacher had issued a specific order, which the students first heard and then subsequently followed.

(28a) nisət sdap tam bɔŋkoəp kru:ybɔŋriən
student listen follow command teacher-CAUS.learn
‘The students obey [their] teacher’s commands.’ [SVC]

(28b) nisət sdap haəj-nɨŋ tam bɔŋkoəp kru:ybɔŋriən
student listen and follow command teacher-CAUS.learn
‘The students listened and then obeyed [their] teacher’s command.’ [not an SVC]

Though the acceptability of a given sequence of verbs as depicting a “single event” is often culture-specific and hard to define, the contrast portrayed in (28) is precisely what we would expect to find when dealing with an SVC.

4. Conclusion

The main contribution of the present paper has been to put the examination of SVCs in Khmer on firmer methodological footing and to further our understanding of both Khmer grammar and the general parameters and wider typology of SVCs. It has been argued that Khmer is a typical example of languages of both its family and larger linguistic area in having a robust class of SVCs that express instrument, beneficiary, recipient, goal, purpose, manner of motion, and result. It has been shown that previous explorations of potential SVCs in Khmer by Wilawan (1992; 1995) and

In this way, the conjunction haəj-nɨŋ, which intriguingly contains the word nɨŋ that can also function as a future-tense auxiliary verb, appears to the have the same capacity to determine the temporal parameters of single eventhood as the time adverbials, temporal clauses, and tense marking that Bohnemeyer et al. (2007) note are constitutive of the “macro-event property” they use to diagnose whether apparently separate sub-events are packaged under the same “macro-event” or not (cf. Bisang 2009:803–805).
Sak-Humphry (1995), which concluded that Khmer SCVs were biclausal structures with non-finite complement clauses, were responding more to the specific, theory-internal concerns and assumptions of Lexicase Dependency Grammar than to the independent facts of the language. Once the binding facts of the clause-bounded bimorphemic reflexive pronoun kluənyæŋ are combined with data showing that the nominal complements of both V1 and V2 can be extracted for topicalization and focus, while the V2 phrase cannot be similarly extracted, the overwhelming conclusion is that potential SCVs of the type surveyed here are, in fact, monoclausal. Combining this insight with the additional facts that such constructions must have single values for polarity and TAM across all of their constituent verbs and do not tolerate overt coreferential NPs for direct arguments additionally strengthens the case for an SVC analysis. Finally, while the criterion is itself somewhat nebulous, there is nonetheless evidence from native speaker intuition that these SCVs encode single complex events. It is thus demonstrated that the structures in question display all of the “characteristic and diagnostic features of SCVs” detailed in Kroeger (2004), among the most rigorous and restrictive descriptions of what constitutes an SVC in the available literature.

One issue for further study will be to probe the upper extent of verb serialization in the language. Khmer is known to permit concatenations of up to 10 separate verbs. The question is whether the language makes a distinction between “compact” SCVs of the sort detailed here and much longer (and possibly less coherent) “narrative” SCVs similar to those which Pawley (2008) has described for the Papuan language Kalam.

References


See, e.g., Ehrman and Sos (1972:25).

A similar distinction is found in Van Staden and Reesink (2002; Foley 2004:132) between “component” and “narrative” serialization.


Interrogation in Muöt

V. R. Rajasingh
Central Institute of Indian Languages, Mysore, India
<rajasingh.1958@gmail.com>

Abstract
Interrogation is a semantic process of eliciting information by way of questioning. Muöt is one of the six varieties of Nicobarese languages. It is spoken by the ethnic Nicobarese who inhabit the three Central Nicobar Islands, namely, Nancowry, Katchal and Kamorta of the Nicobar Archipelago, India. In Muöt, interrogation is found to be carried out with interrogative sentences which are of two kinds. Those interrogative sentences which employ interrogative words as markers of interrogation constitute one type and those which employ intonation, the supra segmental feature, as marker of interrogation constitute the other. Identifying the markers of interrogation and providing a descriptive account on the process of interrogation are the foci of this paper. In order to place the findings on a strong theoretical footing, the paper will have a survey on the process of interrogation as exemplified in the extant works on the language. The paper will also, try to make necessary departures from them reiterating its relevance to the contemporary trend of documentation and description of minor languages. The paper is data bound. The data for the purpose are drawn from the Andaman Commissioned Project data base collected from the Nancowry Island between September and December of 2004 just before the killer tsunami.

Keywords: Interrogation
ISO 639-3 codes: ncb

1. Muöt

Muöt is one of the Nicobarese languages of the ethnic Nicobarese of Nicobar Archipelago, India.\(^1\) The Nicobar Archipelago is a chain of twenty two islands, with thirteen of them inhabited, lying North to South in the Bay of Bengal.\(^2\) The language is spoken by the Nicobarese presently inhabiting the three Islands, namely, Nancowry, Katchal and Kamorta of the archipelago.\(^3\) In the North, these islands are bound by islands of Teressa, Bompoka and Isle of Man, while in the South by that of Miroe. And, as with other members of the archipelago, their eastern border is covered by Thailand and Malaysia, while the western by peninsular India and Sri Lanka. Longitudinally, the three islands are between 93°22 and 93°34°50 and latitudinally between 7°56 and 8°08. As per 2001 census, the total number of people who speak the language stands as 5826 spreading over a geographical area of 515.8 sq. kms.

The data made use of for the paper have been drawn from the data collected as part of the Andaman Commissioned Project, a collaborative program entered into by the Union Territory Administration of the Andaman and Nicobar islands with the Central Institute of Indian Languages, Mysore. The objective of the collaboration is to bring out a Linguistic Description of Muöt so as to enable the Union Territory Administration chalk out programs for the educational and economic progress of the ethnic community. The Nancowry Island with an ethnic human population of 881 over a geographical area of 66.9 sq. kms which is said to be the seat of local administration for the

---

1 It is also known otherwise as Nancowry or Central Nicobarese.
2 The thirteen inhabited islands are Car Nicobar, Chowra, Teressa, Bompoka, Nancowry, Katchal, Kamorta, Trinet, Tiliong Chong, Konul, Pulomilo, Little Nicobar and Great Nicobar. Among them Tiliong Chong is devoid of ethnic inhabitants.
3 Till 2004, just prior to the tsunami, the speakers of the language were spread across four islands, the fourth one being the Trinet. After the tsunamiic devastation, the Indian Administration had to declare the island as inhospitable and the surviving inhabitants thereof have been settled down in the neighboring Kamorta Island. The Administration has named their new habitation in Kamorta as Vikas Nagar.
islands of Nancowry, Katchal, Kamorta and Trinket during the colonial era was selected as the field. Mr. Mark Paul, a native of this island, aged 60 years with the educational qualification of Higher Secondary School Examination passed became the informant. The CIIL questionnaire comprising a word list of 4202 words and a sentence list of 1555 sentences was made use of as the tool for data collection. The data have been collected between September 19th and December 26th of 2004, independently by the author both by observation and elicitation besides recording them in magnetic tapes. During the entire period of field work, the author had to stay with the speakers day and night having him immersed into their language and culture. Presumably, the launching of linguistic description of Muöt is conceived of as part of a larger objective of bringing out descriptive accounts on all the Nicobarese languages. Hence, attempts have already been initiated for collecting data from three more languages also with the author visiting the area of Takahaṅilāhngō (Great Nicobarese) and his colleague Winston Cruz, the areas of Sanēnyō (Chowra) and Lamòngsĕ (Kondul). All the data thus collected are the property of Central Institute of Indian Languages, Mysore and they are marked as Andaman Commissioned Project database in order to differentiate them from the others.

Map: Nicobar Archipelago with Muöt area circled.

---

4 He, and only he, was authorized by the Tribal Council of the Island to work as informant.
5 To determine the number of Nicobarese languages, a survey was conducted as part of this collaborative program between August 11th and October 24th of 2002, by the author along with his colleague Winston Cruz, covering all the twelve islands inhabited by the ethnic Nicobarese. The yet to be published report of the survey enables to fix the number of languages tentatively as six, the other five being, Pū (Car Nicobarese), Sanēnyō (Chowra), Larō (Teressa), Lamòngsĕ (Kondul) and Takahaṅilāhngō (Great Nicobarese). The survey excludes Shompen.
The language has been classed as a member of the Austroasiatic family through the Mon-Khmer sub-family (Lewis 2009). The attestation of Mon-Khmer specific characteristics at the phonological, morphological and syntactic levels of the language seems to substantiate such an affiliation. At the phonological level, the language is found to attest, among others, ɯ ‘high back unrounded vowel’ in its vowel inventory. At the morphological level, it is found to have all roots as monosyllabic ones and also found to attest, among others, <an> ‘resultative infix’ in its affixal morphology. At the syntactic level, it is predominantly found to be of VOS pattern with serial verb construction.

2. Phonology of modern Muöt

The sound system of the language is found to consist only of segmental phonemes. They are in the form of consonants, simple vowels and complex vowels. The phonemic inventory identifies sixteen consonants, nine simple vowels and seven complex vowels.

2.1. Consonants

The sixteen consonants and their phonetic properties can be inferred from table-1.

Table 1

<table>
<thead>
<tr>
<th>Plosive</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>m (m)</td>
<td>n (n)</td>
<td>ē (ň)</td>
<td>ŕ (ń)</td>
<td>η (ng)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>l (l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f (f)</td>
<td>s (s)</td>
<td>Ĉ (č)</td>
<td>ħ (ḫ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>v (v)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2. Simple vowel

The nine simple vowels and their phonetic description can be inferred from table-2.

Table 2

<table>
<thead>
<tr>
<th>High</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rounded</td>
<td>Unrounded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-mid</td>
<td>i (i, ī)</td>
<td>u (u, ū)</td>
<td>uu (eu, eū)</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>ē (ĕ, ē)</td>
<td>o (o, ō)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-mid</td>
<td>ē (e, ē)</td>
<td>ě (ě, ě)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>a (a, ā)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All these function as nucleus of root and as well as affixal syllables.

2.3. Nasalized simple vowel

Except o, all the other eight simple vowels are found to attest their nasalized counterparts. They are ĩ (iṅ, īṅ), ũ (uṅ, ūṅ), ĉ (eṅ, eūṅ), ě (ěṅ, ěṅ), ɑ (aṅ, āṅ). They all are found to occur as nucleus of root syllables only.

---

6 Stress is perceived, but not found to be phonemic.
7 What are given in round brackets against consonants are their equivalents in Muöt orthography.
8 What are given in round brackets against vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.
2.4. Complex vowels

Complex vowels are found to be of diphthongs and the number of them identified stands as seven. They are iə (iö, īö), uɑ (ua, uā), uə (uö, ūö), eɑ (ea, eā), eə (eö, ēö), oə (oö, ōö). Among them, uɑ, uə, uə̃ and oə are found beginning with back vowels, while iə, eə and eɑ with front vowels. As is seen, they all begin with higher vowels and move towards lower ones. Like nasalized simple vowels, they also are found to occur as nucleus of root syllables only.

2.5. Nasalized complex vowel

Except eə and oə, all the other five complex vowels are found to attest their nasalized counterparts. They are iə̃ (iöṅ, īöṅ), uã (uaṅ, uāṅ), uə̃ (uöṅ, ūöṅ), ɯə̃ (euöṅ, eūöṅ), eã (eaṅ, eāṅ). Like nasalized simple vowels, they are also found to occur as nucleus of root syllables only.

2.6. Phonotactics

Among the sixteen consonants, t ‘alveolar plosive’ is found to show variation in its usage. It becomes t̪ ‘dental plosive’ at the syllable initial position (compare, for example, the syllable initial t̪ of the monosyllabic word t̪oˑp ‘drink’ in the sentence 39 with the syllable final t of the monosyllabic word nɔˑt ‘pig’ in the sentence 37).

All the nine simple vowels function as nucleus of either closed or open syllables. They are found to be stressed or unstressed, and when stressed they are of tensed. The tenseness extends to full length in the case of open syllables (see, for example, the tensed vowels uː and eː of the respective open syllables ɯː and eː of the disyllabic words ɯːtː ‘stay’ and eːtː ‘plural’ in sentences 25 and 36 respectively) and to half length in the case of closed syllables (see, for example, the half tensed vowels uː and eː of the closed syllables of the monosyllabic words kæŋ ‘get up’ and ciːn ‘what’ in sentences 31 and 24 respectively). The vowels are found stressed and tensed invariably when they happen to be nucleus of root syllables (compare, for example, the stressed tensed vowels iˑ and eˑ of the respective root syllables of the monosyllabic words ciˑn ‘what’ and neˑn ‘past’ in sentences 24 and 39 respectively with the unstressed lax vowels i and e of the respective affixal syllables PsiP and Pse of the word juɑˑŋsise ‘progressive’ in sentence 19a).

In open syllables, ə ‘mid central vowel’ becomes ɑ ‘low back unrounded vowel’ when the syllables happen to be either of prefixal ones in word initial position (compare, for example, ɑ of t̪ɑP in the word t̪ɑŋãˑŋɛ ‘there’ in sentence 19a with ə of Pŋə in the word kɑjiˑŋə ‘go’ in sentence 30) or of suffixal ones in word medial position (compare, for example, ɑ of PŋɑP in the word xɑliːŋɑse ‘prepare’ in sentence 34 with ə of Pt̪ə in the word t̪əˑŋt̪ə ‘reach’ in sentence 32).

When nasalized, ɑ ‘low back unrounded vowel’ is found to become a ‘low front unrounded vowel’ (compare, for example, ɑ of maˑt ‘sociative’ in sentence 25 with aː of ʔəˑcəʔ ‘arrow’ in sentence 22).

Like simple vowels, complex vowels also are found functioning as nucleus of either open or closed syllables, and are found stressed and tensed to full length in open syllables (see, for example, the diphthong uɑː of the open syllable of the monosyllabic word juɑˑ ‘what’ of section 4.3.1) or to half length in closed syllable (see, for example, the diphthong uɑː of the closed syllable juən of the trisyllabic word juəŋsise ‘progressive’ in the sentence 19a). But, the stress and the tenseness are found to occur either with the initial vowel sounds (see, for example, the diphthong oə of the monosyllabic word koˑn ‘child’ in the sentence 36) or with the final ones (see, for example, the diphthong oo of the monosyllabic word too k ‘toddy’ in sentence 39).

What are given in round brackets against nasalized simple vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

What are given in round brackets against complex vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

What are given in round brackets against nasalized complex vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

Hence, instead of marking them individually, the present paper employs the marker for length as symbolic of both.
3. Interrogation defined

Crystal (2008) in his definition of the word interrogative, states it as ‘a term used in the grammatical classification of sentence types and usually seen in contrast to declarative; it refers to verb form or sentence/clause type typically used in the expression of question’ (p251). As characteristics that are responsible for the typical use of the aforesaid forms in the expression of question, he mentions two: inversion of word order and use of interrogative word. He substantiates the capability of inverted word order in expressing question, with the help of the sentence,

1. Is he coming?

As tokens of interrogative word, three forms, namely, which, why and who are listed sub-categorizing them into interrogative adjective, interrogative adverb and interrogative pronoun respectively (ibid.). Later, while defining the word, question, the author states it as ‘a term used in the classification of sentence functions, typically used to elicit information or a response, and defined sometimes on grammatical, and sometimes on semantic or sociolinguistic grounds’ (p400). Taking English as the sample, he mentions three kinds of questions, namely, sentence with inversion of subject and first verb, sentence that commence with question word and sentence that ends with question tag. As respective illustrations for the three types he lists the following:

2. Is he coming?
3. Where is he?
4. He is going, is not he?

Further, with the passing reference ‘some would include the use of sentences with rising intonation to be a class of question’ (ibid.) the author spells out the possibility of having one more question.

4. Previous sources of information on interrogation

The speakers of the language have been in persistent contacts with traders, administrators, missionaries, researchers and members of various expeditions from time immemorial. Accomplishments of their objectives have resulted in documenting the language in the form of vocabularies, dictionaries, translations, grammatical descriptions etc. Among them, the present paper makes use of the Vocabulary of Dialects spoken in Nicobar and Andaman Isles by De Röepstorff (1875), A Dictionary of the Nancowry Dialect of the Nicobarese Language by De Röepstorff (1884) and A Dictionary of the Central Nicobarese Language by Man (1889) as classical sources of information on interrogation.

4.1. Treatment of interrogation by De Röepstorff (1875)

De Röepstorff (1875) in his vocabulary on Nancowry dialect doesn’t make any formal reference concerning interrogation or interrogative marker or interrogative sentence. However, he lists a few words which correspond to what the present paper views as interrogative marker. They are found in the vocabulary along with their respective glosses in English. For the sake of easy reference they have been reproduced below.

<table>
<thead>
<tr>
<th>Röepstorff (1875)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>katóm</td>
<td>Katōm</td>
<td>kä̂öm</td>
<td>‘how many’</td>
<td>(p64)</td>
</tr>
<tr>
<td>kin-kkee-en?</td>
<td>Chūn</td>
<td>c’n</td>
<td>‘what’</td>
<td>(p109)</td>
</tr>
<tr>
<td>kahēn</td>
<td>Kahēṅ</td>
<td>kaxē:</td>
<td>‘when’</td>
<td>(ibid.)</td>
</tr>
<tr>
<td>jōa</td>
<td>Chū</td>
<td>cuː</td>
<td>‘where’</td>
<td>(ibid.)</td>
</tr>
<tr>
<td>kyouja?</td>
<td>Tai chua</td>
<td>jaj cuː:</td>
<td>‘why’</td>
<td>(ibid.)</td>
</tr>
</tbody>
</table>

13 Stands for the orthography made use of by De Röepstorff (1875).
4.2. Treatment of interrogation by De Röepstorff (1884)

De Röepstorff (1884) in his introduction to the grammar of Nancowry dialect of the Nicobarese language seems to list what the present paper views as marker of interrogation under two titles, namely, the pronoun and the adverb. Under the former, he lists them as a sub category of pronoun called interrogative pronoun and under the latter, as a sub category of adverb called interrogative.

4.2.1. Interrogative pronoun

As interrogative pronouns, he lists four forms and they have been rendered with their glosses in English. They are reproduced below for the sake of easy reference.

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiī</td>
<td>Chī</td>
<td>ciː</td>
<td>‘who’</td>
</tr>
<tr>
<td>Tiīn</td>
<td>Chīn</td>
<td>ciˑn</td>
<td>‘what’</td>
</tr>
<tr>
<td>Tiang tiī</td>
<td>Chōng chī</td>
<td>cə ˆ ciː</td>
<td>‘whose’</td>
</tr>
<tr>
<td>Katōm</td>
<td>Katōm</td>
<td>kə̌o m</td>
<td>‘how many, how much’</td>
</tr>
</tbody>
</table>

(14) Stands for the orthography made use of by De Röepstorff (1884).

Of these, Tiī ‘who’, has been said of as being used with human noun and by virtue of it been termed as personal interrogative pronoun. On Tiīn ‘what’, the author is of the view that the form has been used with human nouns as well as with non-human ones. With respect to Tiang tiī ‘whose’ and Katōm ‘how many, how much’ it has been said that the former is made use of to express possessive relationship while the latter the quantity. The author provides, what are reproduced below as 5 - 7, as sample sentences to illustrate the usage of Tiīn ‘what’ for eliciting information on human as well as non-human nouns and as 8 - 9 as that to illustrate the possessive and quantifying functions of Tiang tiī ‘whose’ and Katōm ‘how many, how much’ respectively.

5. Tiīn paiū

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Gloss</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiīn</td>
<td>paiū</td>
<td></td>
<td></td>
<td>‘Who (what man) is there?’</td>
</tr>
</tbody>
</table>

6. Tiīn io me

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Gloss</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiīn</td>
<td>io me</td>
<td></td>
<td></td>
<td>‘What do you want?’</td>
</tr>
</tbody>
</table>

7. Tiīn wētié

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Gloss</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiīn</td>
<td>wē tié</td>
<td></td>
<td></td>
<td>‘What shall I do?’</td>
</tr>
</tbody>
</table>

(15) Copula verb is found to have elided.
8. **Tiang tī**

| Röepstorff (1884) | Tiang | tī | Modern Muöt | Chōng | chī | IPA | cə ə | ci: | Gloss | poss | who
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Translation</strong></td>
<td>‘Who is it?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. **Katōm kamehewe kāktā**

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Katōm</th>
<th>kamehewe</th>
<th>Kā</th>
<th>kāt</th>
<th>Modern Muöt</th>
<th>Katōm</th>
<th>kamahēnvō</th>
<th>kōk</th>
<th>kōt</th>
<th>IPA</th>
<th>kət o m</th>
<th>kəmxə və</th>
<th>kəʔ</th>
<th>kət</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Translation</strong></td>
<td>‘In how many months will you be here?’ (ibid.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen, in Tiīn paiū ‘who (what man) is there?’ the form Tiīn is made use of to elicit information concerning human noun and as seen in Tiīn io me ‘what do you want?’ as well as in Tiīn wētię ‘what shall I do?’ it has been made use of to elicit information concerning non-human noun. Also, as seen in 8 and 9, Tiang tī ‘whose?’ and Katōm ‘how many, how much’ are made use of to convey possessive and quantifying meanings respectively. The insights obtained from the present paper enable to note that the lexical equivalent of Tiī ‘who?’ seems to be used for eliciting information concerning non-human nouns also in the present day language.

4.2.2. Interrogative

As interrogatives, the author lists three forms and they are found with their respective English glosses. For the sake of easy reference they have been reproduced below.

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiū</td>
<td>Chū</td>
<td>cu:</td>
<td>‘where’</td>
<td>(p14)</td>
</tr>
<tr>
<td>Kahæ</td>
<td>Kahæn</td>
<td>kəxɛ̃ː</td>
<td>‘when’</td>
<td>(ppxx-xxi)</td>
</tr>
<tr>
<td>Katōm</td>
<td>Katōm</td>
<td>kət o m</td>
<td>‘how many’</td>
<td>(ibid.)</td>
</tr>
</tbody>
</table>

Further, he speaks about another interrogative form kā and is of the view that it is often used as an independent interrogative particle. To illustrate such a proposition two sentences have been provided and among them one has already been reproduced as 9 and the other is reproduced below as 10.

10. **Léat kā ina kalāh oknōk omtōm**

<table>
<thead>
<tr>
<th>Röepstorff (1884)</th>
<th>Léat</th>
<th>kā</th>
<th>ina</th>
<th>kalāh</th>
<th>oknōk</th>
<th>omtōm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Muöt</td>
<td>Leāt</td>
<td>kōk</td>
<td>inān</td>
<td>kalāh</td>
<td>hungōnk</td>
<td>umtūm</td>
</tr>
<tr>
<td>IPA</td>
<td>leət</td>
<td>kəʔ</td>
<td>?ināː</td>
<td>kulə x</td>
<td>xunŋ̂ k</td>
<td>?umŋ̂ m</td>
</tr>
<tr>
<td>Gloss</td>
<td>perf</td>
<td>q</td>
<td>2du</td>
<td>taste</td>
<td>food</td>
<td>all</td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td>‘Have you (two) tasted all the food?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The insights obtained from the present paper enable to consider the form kā in the sentences 9 and 10 not as an independent interrogative particle, but as a demonstrative particle.

---

16 The subject and copula verb of the sentence are found to have elided.

17 The subject of the sentence and the marker for future tense are found to have elided.
4.3. Treatment of interrogation by Man (1889)

Man (1889) in his notes on the grammar of the dialect spoken in the central portion of the Nicobar Islands, deals with what the present paper views as interrogative marker and interrogative sentence under three titles, namely, pronouns, adverbs, and interrogative sentences. And under pronouns, he treats them as a sub-category of pronoun called interrogative pronoun and under adverbs as four sub-categories of adverbs called, interrogative adverbs of time and change; interrogative adverbs of place; interrogative adverbs of quantity, number and degree; and interrogative adverbs of manner and cause.

4.3.1. Interrogative pronoun

As interrogative pronouns, the author presents a list of nineteen forms along with their respective English glosses. They are,

<table>
<thead>
<tr>
<th>Man (1889)18</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chī</td>
<td>Chī</td>
<td>ciː</td>
<td>‘who’</td>
</tr>
<tr>
<td>Ten chī</td>
<td>Tin chī</td>
<td>tì ɲ ciː</td>
<td>‘whom’</td>
</tr>
<tr>
<td>Tai chī</td>
<td>Tai chī</td>
<td>t-deals</td>
<td>‘by whom’</td>
</tr>
<tr>
<td>Lamòngto-chī,</td>
<td>Lamòngto-tinchī</td>
<td>lampa ɲ ciː,</td>
<td>‘from whom’</td>
</tr>
<tr>
<td>Chamang-ta-chī</td>
<td>Chamōöngtö chī</td>
<td>camapa ɲ ciː</td>
<td>‘whose’</td>
</tr>
<tr>
<td>Chū19</td>
<td>Chū</td>
<td>cuː</td>
<td>‘where’</td>
</tr>
<tr>
<td>Chūa, Chūan,</td>
<td>Chūa, Chūan</td>
<td>cuː n,</td>
<td>‘what’</td>
</tr>
<tr>
<td>Chin, Kā, Ka, Kan</td>
<td>Chin, Kā, Ka, Kan</td>
<td>cuː n, ciː n,</td>
<td>‘what’</td>
</tr>
<tr>
<td>Chun</td>
<td>Chī</td>
<td>ciː</td>
<td>‘which’</td>
</tr>
<tr>
<td>Chūn-shī</td>
<td>Chūnši</td>
<td>cuː nside</td>
<td>‘why’</td>
</tr>
<tr>
<td>Kā-shīn</td>
<td>Kasí</td>
<td>kasiː</td>
<td>‘like what’</td>
</tr>
<tr>
<td>Ka-rîshe</td>
<td>Karîse</td>
<td>kaɾiːse</td>
<td>‘how much’</td>
</tr>
<tr>
<td>Karâm, Katöm</td>
<td>Karâm, Katöm</td>
<td>kapa m,</td>
<td>‘how many’</td>
</tr>
<tr>
<td>Chin-leāng-dîo</td>
<td>Chin leāng rîöv</td>
<td>ciː n leaŋ ɹiːəv</td>
<td>‘what else’</td>
</tr>
</tbody>
</table>

(PPXVI-XXVII)

As evidence for their usage in utterances, he provides two sentences which are reproduced below as 11 and 12.

11. Chamangta chī

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Chamangta chī</th>
<th>en enh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Muöt</td>
<td>Chamoöngtö chī</td>
<td>nek enh</td>
</tr>
<tr>
<td>IPA</td>
<td>camapa ɲ ciː</td>
<td>ne ɲ ciː</td>
</tr>
<tr>
<td>Gloss</td>
<td>whose</td>
<td>this20</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Whose is this?’</td>
<td></td>
</tr>
</tbody>
</table>

18 Stands for the orthography made use of by Man (1889).
19 The form is not found listed in the notes, but found listed in the dictionary part (P122).
20 See footnote 15.
12. Lamòngto ten chī en eṅh

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muôt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kâhē</td>
<td>Kahēn</td>
<td>kaxē:</td>
<td>‘when (of future only)’</td>
</tr>
<tr>
<td>Kâhē-ta-shii</td>
<td>Kahēn tô sī</td>
<td>kaxē: ḷasi</td>
<td>‘when (of a coming event)’</td>
</tr>
<tr>
<td>Kâhē-ta-tai</td>
<td>Kahēn tôtai</td>
<td>kaxē: ḷa(tp̪)j</td>
<td>‘when (future, of making, giving, &amp;c.)’</td>
</tr>
<tr>
<td>Kâhē-ngalâh</td>
<td>Kahēn ngō lah</td>
<td>kaxē: ḷa(pl̪)x</td>
<td>‘when (future, of travelling)’</td>
</tr>
<tr>
<td>Kâhē-nıga</td>
<td>Kahēn ngō</td>
<td>kaxē:ŋ</td>
<td>‘when (of past time, in reference to death, sickness)’</td>
</tr>
<tr>
<td>Kâhē-ngashī</td>
<td>Kahēn ngō sī</td>
<td>kaxē:ŋasi</td>
<td>‘when (of some past event)’</td>
</tr>
<tr>
<td>Kâhē-ngatai</td>
<td>Kahēn ngō tai</td>
<td>kaxē:ŋa(tp̪)j</td>
<td>‘when (of past time, in reference to making, giving, &amp;c.)’</td>
</tr>
</tbody>
</table>

4.3.2. Interrogative adverb of time and change

As interrogative adverbs of time and change, the author lists nineteen forms and they are given along with their glosses in English. The forms are:

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muôt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hē-chûa</td>
<td>Hĕn chūa</td>
<td>x̂:cua:</td>
<td>‘when (at what time)’</td>
</tr>
<tr>
<td>Kâ-shanî-tashe</td>
<td>Ka sanî tô se</td>
<td>kasanî ḷase</td>
<td>‘how long (time)’</td>
</tr>
<tr>
<td>Kâ-shanî-latōh</td>
<td>Kasaنانî lō tôh</td>
<td>kasanî ˈlat̪ɔˑx</td>
<td>‘how old’</td>
</tr>
<tr>
<td>Kâ-rūa-hanga-heng</td>
<td>Kareûōt hôngō hēng</td>
<td>καϻɔ txαŋαxεŋ</td>
<td>‘how long ago (today)’</td>
</tr>
<tr>
<td>Kâ-rūa-hanga-yan</td>
<td>Kareûōt hôngō yōn</td>
<td>καϻɔ txαŋαjɔn</td>
<td>‘how long ago (in the past)’</td>
</tr>
<tr>
<td>Karâm-shuâ</td>
<td>Karâm tô suâ</td>
<td>καϻα:ngasua:</td>
<td>‘how many times’</td>
</tr>
<tr>
<td>Kâ-inōa-ato</td>
<td>Ka inōǎns ô tô</td>
<td>καʔinɔǎsɔŋ</td>
<td>‘soon’</td>
</tr>
<tr>
<td>Kâ-rūala-heng</td>
<td>Ka reûōlō hēng</td>
<td>καϻɔ ˈlaxeɲ</td>
<td>‘what time (of today) is it?’</td>
</tr>
<tr>
<td>Kâ-rūala-hatôm</td>
<td>Ka reûōlō hatôm</td>
<td>καϻɔ ˈlaxaːtp̪ɔm</td>
<td>‘what time (of night) is it?’</td>
</tr>
</tbody>
</table>

(PPXXII-XXXIII)

Of these, except hē-chûa ‘when (at what time)’ all the remaining eighteen are seen occurring with the interrogative marker ka- or kâ-.

4.3.3. Interrogative adverb of place

As interrogative adverb of place, he lists two forms which occur with ka-. They are rendered with their glosses in English and are reproduced below for easy reference:

See footnote 15.
The insights obtained from the present paper enable us to view the form Ka-ra-múa-hanga ‘how near’ as the agentivized counterpart of Ka-rúa-hanga ‘how far’ and as such regarded not as conveying meaning opposite to the non-agentivized form but as conveying parallel meaning associated with an object.

4.3.4. Interrogative adverb of quantity, number, and degree

As interrogative adverbs of quantity, number, and degree the author gives a list of twenty forms, all beginning with ka-. They have been listed along with their glosses in English and they have been reproduced below for easy reference.

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ka-rī</td>
<td>Ka rī</td>
<td>kariː</td>
<td>‘how big’</td>
</tr>
<tr>
<td>Ka-ra-mī</td>
<td>Karamī</td>
<td>kəzamiː</td>
<td>‘how small’</td>
</tr>
<tr>
<td>Ka-rī-fāp</td>
<td>Karī fāp</td>
<td>kəiː fɑˑp</td>
<td>‘how fat; how wide (of ship)’</td>
</tr>
<tr>
<td>Ka-ra-mī-fāp</td>
<td>Karamī fāp</td>
<td>kəzamiː fɑˑp</td>
<td>‘how narrow (of ship)’</td>
</tr>
<tr>
<td>Karī-tāk</td>
<td>Karī tak</td>
<td>kəiː t̪ɑˑk</td>
<td>‘how wide (of plank)’</td>
</tr>
<tr>
<td>Karamī-tāk</td>
<td>Karamī tak</td>
<td>kəzamiː t̪ɑˑk</td>
<td>‘how narrow (of plank)’</td>
</tr>
<tr>
<td>Ka-chin-yāwa</td>
<td>Kachin yāvō</td>
<td>kəcinjəˑʋə</td>
<td>‘how deep’</td>
</tr>
<tr>
<td>Ka-rūat</td>
<td>Kareūot</td>
<td>kəzʊːt</td>
<td>‘how long’</td>
</tr>
<tr>
<td>Kara-mūat</td>
<td>Karamītʊt</td>
<td>kəzəmʊːt</td>
<td>‘how short (inanim.)’</td>
</tr>
<tr>
<td>Ka-rūala</td>
<td>Kareūolō</td>
<td>kəzʊːlə</td>
<td>‘how high’</td>
</tr>
<tr>
<td>Ka-rūala-kōi</td>
<td>Kareūolō kui</td>
<td>kəzəmʊːlək</td>
<td>‘how tall’</td>
</tr>
<tr>
<td>Kara-mūala-kōi</td>
<td>Karamīolō kui</td>
<td>kəzəmʊːlək</td>
<td>‘how short (anim.)’</td>
</tr>
<tr>
<td>Karūa-hanga-tai</td>
<td>Kareūo hōngō tai</td>
<td>kəzəxɑŋət̪ɑˑj</td>
<td>‘how far off (of object shot, speared, &amp;c)’</td>
</tr>
<tr>
<td>Karī-tare-she,</td>
<td>Karī tō re se,</td>
<td>kəiːt̪əˈʔeːsə,</td>
<td>‘how much more remains’</td>
</tr>
<tr>
<td>Karī-tet-she</td>
<td>Karī tɛt se</td>
<td>kəiːtətsete</td>
<td></td>
</tr>
<tr>
<td>Karī-hata-she</td>
<td>Karī hō tō se</td>
<td>kəiːxəʧəse</td>
<td>‘how much has been paid, delivered, &amp;c’</td>
</tr>
<tr>
<td>Katōm-tare</td>
<td>Katōm tō re</td>
<td>kət̪əˈʔeːsə</td>
<td>‘how many more’</td>
</tr>
<tr>
<td>Kā-yan</td>
<td>Kayōn</td>
<td>kəʔ n</td>
<td>‘how, in what state of health’</td>
</tr>
<tr>
<td>Kā-shin-mush-tai</td>
<td>Kasin mūs tai</td>
<td>kəsinmuˑst̪əj</td>
<td>‘in what style’</td>
</tr>
<tr>
<td>Ka-shī</td>
<td>Kasī</td>
<td>kəsiː</td>
<td>‘what kind, what sort’</td>
</tr>
</tbody>
</table>

To illustrate the usage of these forms in utterances, he has listed a sentence and the same is reproduced below as 13.

13. Kā yan ka an ta-linheň

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kā</td>
<td>yōn</td>
<td>ko̞k n̂</td>
<td>tō linheň</td>
</tr>
<tr>
<td>Modern Muöt</td>
<td>Kā</td>
<td>jō n</td>
<td>ko̞ʔ ?a n̂</td>
</tr>
<tr>
<td>IPA</td>
<td>kə</td>
<td>jɔˑn</td>
<td>ko̞ʔ ?a n</td>
</tr>
<tr>
<td>Gloss</td>
<td>how</td>
<td>health</td>
<td>today22</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How is he today?’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

22 See footnote 15.
The insights obtained from the present paper enable to view the forms Ka-ra-mī ‘how small’, Ka-ra-mī-fāp ‘how narrow (of ship)’, Karamī-tāk ‘how narrow (of plank)’, Karamī-muāt ‘how short (inan.)’ and Karamī-muāla-kōi ‘how short (anim.)’ as agentivized counterparts of Ka-rī ‘how big’, Ka-rī-fāp ‘how fat; how wide (of ship)’, Kari-tāk ‘how wide (of plank)’, Ka-rūāt ‘how long’ and Kā-ruāla-kōi ‘how tall’ respectively. And, as such they are regarded not as conveying meanings opposite to the non-agentivized forms but as conveying parallel meanings associated with the concerned objects.

4.3.5. Interrogative adverb of manner and cause

The author lists, as interrogative adverbs of manner and cause, eight forms all seem to be variants of a single entity. They are,

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chūan-shī,</td>
<td>Chūan si,</td>
<td>cuɑˑn ssi,</td>
<td>‘why’</td>
</tr>
<tr>
<td>Chūan-lāng-ngāshī,</td>
<td>Chūan lōng ngō ssi,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chūan-lāng-ŋitō,</td>
<td>Chūan lōng ngō t̪o,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chūan-hān-ngāshī,</td>
<td>Chūan hān ngō ssi,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chūan-hān-ŋitō,</td>
<td>Chūan hān ngō t̪o,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chūan-wī,</td>
<td>Chūan vī,</td>
<td>cuɑˑn ʋiˑʔ,</td>
<td></td>
</tr>
<tr>
<td>Chūan-tai-chūa</td>
<td>Chūan tai chūa</td>
<td>cuɑˑn t̪aˑjcuɑː</td>
<td></td>
</tr>
</tbody>
</table>

And to illustrate their usage in utterances he gives, what is reproduced below as 14, as the sample sentence.

14. Chūan wī meṅ ta watshī meṅ met dalngatō

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tāv mëṅ kōḵ aṅn</td>
<td>Chūan vīḵ mëṅ t̪o vāt si mëṅ m’it rāl ngō t̪o</td>
<td>cuɑˑn vm ɛ̃ː t̪ɑ ʋɑˑt si mɛ̃ː m’ɹɑˑlŋɑt̪ɔ</td>
<td>‘Is he your younger brother?’</td>
</tr>
</tbody>
</table>

4.3.6. Interrogative sentence

Under the heading interrogative sentence the author seems to mention two types of interrogation as if in conformity with what has been arrived at in the present paper. One is interrogation with sentence final rising intonation and the other, interrogation with interrogative word. Regarding the former, he makes a passing reference as ‘in many cases interrogation is sufficiently indicated by the tone of the voice’ (p lv). But, in respect of the latter, a relatively detailed discussion has been presented with the help of a host of interrogative sentences which includes sentences which are found to have as their markers of interrogation, sentence final rising intonation also besides interrogative pronouns and interrogative adverbs. They can respectively be inferred from the following three sentences which are reproduced below as 15, 16 and 17.

15. Tāu meṅ kā an

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Modern Muöt</th>
<th>IPA</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tāv mëṅ kōḵ kā an</td>
<td>Chūan vī kā an</td>
<td>cuɑˑn vm kəˑʔ kãˑn</td>
<td>‘Is he your younger brother?’</td>
</tr>
</tbody>
</table>

23 For a similar passing reference, see (p xi) of Introductory Remarks.
24 See footnote 15.

Rajasingh, V. R. 2014. Interrogation in Muöt.
Mon-Khmer Studies 43.1:103-123 (ICAAL5 special issue)
16.  Chī yō haiyūan

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Chī</th>
<th>yō</th>
<th>haiyūan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Muöt</td>
<td>Chī</td>
<td>yōk</td>
<td>hayeūōn</td>
</tr>
<tr>
<td>IPA</td>
<td>ciː</td>
<td>jɔʔ</td>
<td>xajwː n</td>
</tr>
<tr>
<td>Gloss</td>
<td>who</td>
<td>future</td>
<td>hunt pigs</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Who is going to hunt pigs?’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17.  Kāhē tashe meṅ dāk

<table>
<thead>
<tr>
<th>Man (1889)</th>
<th>Kāhē tashe</th>
<th>meṅ</th>
<th>dāk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modern Muöt</td>
<td>Kāhēn tō se</td>
<td>meṅ</td>
<td>reuk</td>
</tr>
<tr>
<td>IPA</td>
<td>kaxɛ̃ːt̪ɑse</td>
<td>mɛ̃ː</td>
<td>.getUserDictionary().getSymbols().get(27)</td>
</tr>
<tr>
<td>Gloss</td>
<td>when</td>
<td>you</td>
<td>come</td>
</tr>
<tr>
<td>Translation</td>
<td>‘When did you come?’ (pplv-lvi)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen, sentences 15, 16 and 17 above are found to manifest interrogation by means of sentence final rising intonation, interrogative pronoun and interrogative adverb respectively.

5. Interrogation in Modern Muöt

Of the four means of interrogation seen in section 3, namely, interrogation by means of sentence with inversion of subject and first verb, interrogation by means of sentence with question word, interrogation by means of sentence with question tag and interrogation by means of sentence with rising final intonation, Muöt, the language under discussion seems to make use of only two. They are sentences with rising final intonation and sentences with initial interrogative word.

5.1. Sentence with rising final intonation

They are declarative sentences with normal word order and become interrogative ones by taking with them the supra-segmental feature, rising final intonation. Endowed with this marker of interrogation, they are poised to elicit information concerning objects and actions. The following two pairs of sentences 18a-b\(^{25}\) and 19a-b\(^{26}\) can be made use of for understanding such a function.\(^ {27}\)

18a. Ööṅ in kōön mēṅ inkānö in Merī

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>ööṅ</th>
<th>in</th>
<th>kōön mēṅ inkānö</th>
<th>in</th>
<th>Merī</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>?ʒ.</td>
<td>ʔiˑn</td>
<td>kɔˑʔn mɛː</td>
<td>ʔinkɑˑnə</td>
<td>ʔiˑn</td>
</tr>
<tr>
<td>Gloss</td>
<td>cop</td>
<td>prox</td>
<td>your daughter</td>
<td>prox</td>
<td>Mary</td>
</tr>
<tr>
<td>Translation</td>
<td>Mary is your daughter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18b. Ööṅ in kōön mēṅ inkānö in Merī?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>ööṅ</th>
<th>in</th>
<th>kōön mēṅ inkānö</th>
<th>in</th>
<th>Merī</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>?ʒ.</td>
<td>ʔiˑn</td>
<td>kɔˑʔn mɛː</td>
<td>ʔinkɑˑnə</td>
<td>ʔiˑn</td>
</tr>
<tr>
<td>Gloss</td>
<td>cop</td>
<td>prox</td>
<td>your daughter</td>
<td>prox</td>
<td>Mary</td>
</tr>
<tr>
<td>Translation</td>
<td>Is Mary your daughter?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

19a. Yuāngsise uksök tōngāṅinge in Sipā nen

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>yuāngsise</th>
<th>uksök</th>
<th>tōngāṅinge</th>
<th>in</th>
<th>Sipā</th>
<th>nen</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ʔuŋsɪˑse</td>
<td>ʔuksʊˑk</td>
<td>ʔʊŋʊˑn</td>
<td>ʔiˑn</td>
<td>sipːə</td>
<td>neˑn</td>
</tr>
<tr>
<td>Gloss</td>
<td>prog</td>
<td>stand</td>
<td>there</td>
<td>prox</td>
<td>sheeba</td>
<td>pst</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Sheeba was standing there’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{25}\) The sentences are unmarked for present tense.

\(^{26}\) It is because of the free word-order the marker for past tense occur sentence finally.

\(^{27}\) The sentences 18a-b can also be found occurring with the elision of copula verb.
19b. Yuångsise uksök töngåñnge in Sipâ nen?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>yuångsise</th>
<th>uksök</th>
<th>töngåñnge</th>
<th>in</th>
<th>Sipâ</th>
<th>nen</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>juɑˑŋsise</td>
<td>ʔuksəˑk</td>
<td>ʔaŋãˑŋɛ</td>
<td>ʔiˑn</td>
<td>sipɔˑ</td>
<td>neˑn</td>
</tr>
<tr>
<td>Gloss</td>
<td>prog</td>
<td>stand</td>
<td>there</td>
<td>prox</td>
<td>sheeba</td>
<td>pst</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Was Sheeba standing there?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen above, sentences 18b and 19b are interrogative ones. They seem to have come into existence from their respective declarative counterparts 18a and 19a with the annexing of rising final intonation. And, as being the source for eliciting information, sentence 18b seems to elicit information regarding kinship of human nouns while 19b that regarding the act performed by human nouns. In view of the answers interrogative sentences of this type elicit, they can be termed as yes or no questions.

5.2. Sentence with interrogative words

They are sentences having interrogative words as markers of interrogation. Two kinds of interrogative words are identified in the language and they seem to occur in the sentence initial position. They are termed here as chP-interrogative words and ka- interrogative words.

5.2.1. Sentence with chP-interrogative word

They are interrogative sentences having interrogative words that begin with the digraph, ch-. Sentences with five numbers of such interrogative words are identified in the language. They are sentences with chī ‘who’, sentences with chîn ‘what’, sentences with chuān ‘what’, sentences with chū ‘where’ and sentences with chuānsi ‘why’. Having these forms as markers of interrogation, sentences of the type, elicit information on objects and actions. And, as constituents of sentences, these interrogative words seem to perform varying grammatical functions such as subject, object and adverb besides functioning as interrogative pronoun.

5.2.1.1. Sentence with chī

They are sentences that begin with the interrogative word chī ‘who?’ and seek to elicit information about human as well as non-human nouns. The sentences 20 and 21 listed below can be taken for the illustration of the fact.

20. Chī ööṅ köḵ öṅn inkônyö?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>chī</th>
<th>ööṅ</th>
<th>köḵ</th>
<th>öṅn inkônyö</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ciː</td>
<td>ʔə̃ː</td>
<td>kəˑʔ</td>
<td>ʔə̃ˑn ʔinkɔˑɲə</td>
</tr>
<tr>
<td>Gloss</td>
<td>who</td>
<td>cop</td>
<td>dist3</td>
<td>he₂⁹</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Who is he?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

21. Chī ööṅ tö ki āṅn e in nyī öṅn inkônyö?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>chī</th>
<th>ööṅ</th>
<th>tö ki āṅn e</th>
<th>in</th>
<th>nyī öṅn inkônyö</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ciː</td>
<td>ʔə̃ː</td>
<td>ʔa kiʔəˑnɛ</td>
<td>ʔiˑn</td>
<td>niˑ: ʔiˑn ʔinkɔˑɲə</td>
</tr>
<tr>
<td>Gloss</td>
<td>which</td>
<td>cop</td>
<td>among these</td>
<td>prox</td>
<td>his house²⁰</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Which is his house?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen above, sentence 20 seeks to elicit information on human noun and 21 on non-human. And while doing so, chī ‘who?’ in both cases is found to be interrogative pronoun. At the same time it is found to be grammatical object in 20 and grammatical subject in 21. As pronoun, it gets

---

28 ch represents voiceless palatal consonant in Muöt orthography.
29 The sentence can also be found occurring with the elision of copula verb.
30 See footnote 29.
inflected for case relationship, here the possessive, and the grammatical function of such case inflected form as object can be seen from the sentence 22 listed below.

22. Āṅ chōḵ chē ōōṅ uſe kōk e?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>āṅ chōḵ</th>
<th>chē</th>
<th>ōōṅ</th>
<th>uſe kōk e</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ŭãːcəʔ</td>
<td>ciː</td>
<td>ŭə̃ː</td>
<td>ŭufeː kəˑʔɛ</td>
</tr>
<tr>
<td>Gloss</td>
<td>arrow</td>
<td>q</td>
<td>cop</td>
<td>those³²</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Whose arrows are those?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.1.2. Sentence with chīn

They are sentences that begin with the interrogative word chīn ‘who?’ and seek to elicit information concerning nouns of human as well as of non-human nature. Such a phenomenon can be inferred from the sentences listed below as 23 and 24.

23. Chīn ōōṅ kōk ōṅn inkānö?

<table>
<thead>
<tr>
<th>Modern Muót</th>
<th>chīn</th>
<th>ōōṅ</th>
<th>kōk</th>
<th>ōṅn inkānö</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ciˑn</td>
<td>ŭə̃ː</td>
<td>kəˑʔ</td>
<td>ŭə̃ˑn ʔinkɑˑnə</td>
</tr>
<tr>
<td>Gloss</td>
<td>who</td>
<td>cop</td>
<td>dist₃</td>
<td>she³³</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Who is she?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24. Chīn ōōṅ in leāng mēṅ?

<table>
<thead>
<tr>
<th>Modern Muót</th>
<th>chīn</th>
<th>ōōṅ</th>
<th>in</th>
<th>leāng mēṅ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>ciˑn</td>
<td>ŭə̃ː</td>
<td>ŭiˑn</td>
<td>leaˑŋ mɛ̃ː</td>
</tr>
<tr>
<td>Gloss</td>
<td>who</td>
<td>cop</td>
<td>prox</td>
<td>your name³⁴</td>
</tr>
<tr>
<td>Translation</td>
<td>‘What is your name?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen above, sentence 23 seeks to elicit information concerning human noun and 24 that concerning non-human. And, while doing so, in both the sentences, chīn ‘who?’ is found to function as interrogative pronoun. At the same time, it is found to function as grammatical object in 23 and grammatical subject in 24 as well. As pronoun, it gets inflected for case relationship, here the sociative, and the grammatical function of such case inflected form as object can be inferred from the sentence 25 below.

25. Mat chī yòḵ nō kateū in Mark tō öt Luang?

<table>
<thead>
<tr>
<th>Modern Muôt</th>
<th>mat</th>
<th>chī</th>
<th>yòḵ</th>
<th>nō</th>
<th>kateū</th>
<th>in</th>
<th>Mark</th>
<th>tō öt Luang</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>mɑˑt</td>
<td>ciː</td>
<td>jɔˑʔ</td>
<td>nə</td>
<td>kaˑt̪ɯː</td>
<td>ŭin</td>
<td>mɑˑk</td>
<td>t̪ə ʔəˑt luɑˑŋ</td>
</tr>
<tr>
<td>Gloss</td>
<td>soc</td>
<td>q</td>
<td>fut</td>
<td>svs</td>
<td>stay</td>
<td>prox</td>
<td>Mark</td>
<td>in kondul</td>
</tr>
<tr>
<td>Translation</td>
<td>‘With who will Mark stay in Kondul?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.1.3. Sentence with chuān

They are sentences that begin with the interrogative word chuān ‘what?’ and seek to elicit information on non-human nouns, and also on actions. Sentences 26 and 27 given below can be taken as illustrations.

---
³¹ The possessive case is found to be unmarked in the language. Mere juxtaposing of the possessed and the possessor is found to yield the possessive meaning.
³² The sentence is unmarked for present tense and can also be found occurring with the elision of copula verb.
³³ See footnote 29.
³⁴ See footnote 29.
26. Chuān yòḵ topre in mën?

<table>
<thead>
<tr>
<th>Modern Muôt</th>
<th>chuān</th>
<th>yòḵ</th>
<th>topre</th>
<th>in</th>
<th>mën</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>cuɑˑn</td>
<td>jɔˑʔ</td>
<td>ʔɔˑpre</td>
<td>ʔiˑn</td>
<td>mɛ̃ːʔ</td>
</tr>
<tr>
<td>Gloss</td>
<td>what</td>
<td>will</td>
<td>drink</td>
<td>prox</td>
<td>you</td>
</tr>
<tr>
<td>Translation</td>
<td>‘What will you drink?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27. Chuān yuāngsise vīḵ āṅn ufē?

<table>
<thead>
<tr>
<th>Modern Muôt</th>
<th>Chuān</th>
<th>yuāngsise</th>
<th>vīḵ</th>
<th>āṅn</th>
<th>ufē</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>cuɑˑn</td>
<td>juɑˑŋsise</td>
<td>v̟iˑʔ</td>
<td>ʔāˑn</td>
<td>ʔufeː</td>
</tr>
<tr>
<td>Gloss</td>
<td>what</td>
<td>do</td>
<td>dist₁</td>
<td>they³⁵</td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>‘What are they doing?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the above, sentence 26 seeks to elicit information concerning non-human noun, while 27 that concerning action. In both the instances, Chuān ‘what?’ seem to function as grammatical object. In addition, in 26 it functions as interrogative pronoun also.

5.2.1.4. Sentence with chū

They are sentences that begin with the interrogative word chū ‘where?’ and seek to elicit information on the location of objects and actions. The following sentence 28 can be taken as illustration.

28. Chū ôt köḵ nyi mën?

<table>
<thead>
<tr>
<th>Modern Muôt</th>
<th>chū</th>
<th>ôt</th>
<th>köḵ</th>
<th>nyi mën</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>cuː</td>
<td>ʔɔˑt</td>
<td>kəˑʔ</td>
<td>ɲiː mɛ̃ː</td>
</tr>
<tr>
<td>Gloss</td>
<td>where</td>
<td>exis</td>
<td>dist₁</td>
<td>your house³⁶</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Where is your house?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen above, sentence 28 seeks to elicit information on the location of object and action. And, while doing so, chū ‘where?’ is found to functions as grammatical adverb.

5.2.1.5. Sentence with chuānsi

They are sentences that begin with the interrogative word chuānsi ‘why?’ and seek to elicit the reason for carrying out an action. The sentence 29 given below can be taken to illustrate the fact.

29. Chuānsi nö chīm köḵ kinyôṅm?

<table>
<thead>
<tr>
<th>Modern Muôt</th>
<th>chuānsi</th>
<th>nö</th>
<th>chīm</th>
<th>köḵ</th>
<th>kinyôṅm</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>cuɑˑnsi</td>
<td>ɲə</td>
<td>ciˑm</td>
<td>ʔɔˑʔ</td>
<td>ʔiɲɔ̃ˑm?</td>
</tr>
<tr>
<td>Gloss</td>
<td>why</td>
<td>svs</td>
<td>cry</td>
<td>dist₃</td>
<td>baby³⁷</td>
</tr>
<tr>
<td>Translation</td>
<td>‘Why does the baby cry?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen, sentence 29 above seeks to elicit the reason for the baby to cry and while doing so, the interrogative word chuānsi ‘why?’ is found to functions as an adverb.

5.2.2. Sentence with ka-interrogative word

They are interrogative sentences having interrogative words that begin with the syllable ka-. Sentences with eight numbers of such interrogative words are identified in the language. They are sentences with kahèṅ ‘when (in terms of period of time)?’, sentences with karām hōngō ‘when (in

---

³⁵ See footnote 25.
³⁶ The sentence is unmarked for present tense and can also be found occurring with the elision of existential verb.
³⁷ See footnote 25.
terms of hour)?’, sentences with kasī ‘how (manner)?’, sentences with katāi ‘how (means)?’, sentences with karāmkui ‘how many (human)?’, sentences with karīsē ‘how many?, how much?’, sentences with karīhöt ‘how much (liquid)?’ and sentences with kareūöt ‘how long?’ Having these forms as markers of interrogation, sentences of the type, elicit information on objects and actions. And, as constituents of sentences, these interrogative words seem to perform varying grammatical functions such as subject, object, adjective and adverbs.

5.2.2.1. Sentence with kahëṅ

They are sentences that begin with the interrogative word ka-hëṅ ‘when (in terms of period of time)?’ and seek to elicit information concerning the period of time at which incidence of an action takes place. The following sentence 30 can be taken as illustration of the function.

30. Kahëṅ mën yök kayëṅgö nö hayeūön in mën?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>kahëṅ</th>
<th>mën</th>
<th>yök</th>
<th>kayëṅgö</th>
<th>nö</th>
<th>hayeūön</th>
<th>in</th>
<th>mën</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑx̂ː</td>
<td>ṁː</td>
<td>jɔʔ</td>
<td>kɑjìŋə</td>
<td>ṅː</td>
<td>xajɯəˑn</td>
<td>?in</td>
<td>ṁː</td>
</tr>
<tr>
<td>Gloss</td>
<td>when</td>
<td>svs</td>
<td>will</td>
<td>go</td>
<td>purp</td>
<td>pig hunting</td>
<td>prox</td>
<td>you</td>
</tr>
<tr>
<td>Translation</td>
<td>‘When will you go for pig hunting?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen, sentence 30 seeks to elicit information concerning the period of time at which leaving for pig hunting takes place. While doing so, the interrogative word kahëṅ ‘when (in terms of period of time)?’ is found functioning as temporal adverb.

5.2.2.2. Sentence with karām höngö

They are sentences that begin with the interrogative word karām höngö ‘when (in terms of hour)?’ and seek to elicit information concerning the hour of time at which incidence of an action takes place. The sentence 31 listed below can be taken to illustrate such a function.

31. Karāmhöngö kööngö in mën tö òal hakī möök?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>Karāmhöngö</th>
<th>kööng</th>
<th>in</th>
<th>mën</th>
<th>tö òal hakī</th>
<th>möök</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑɹɑˑmxɑŋə</td>
<td>kəˑŋ</td>
<td>?in</td>
<td>ṁː</td>
<td>jɔʔ ?ɔəˑl xɑkiː</td>
<td>məˑk</td>
</tr>
<tr>
<td>Gloss</td>
<td>when</td>
<td>get up</td>
<td>prox</td>
<td>you</td>
<td>in the morning</td>
<td>will</td>
</tr>
<tr>
<td>Translation</td>
<td>‘When will you get up in the morning?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen above, sentence 31 seeks to elicit information concerning the hour of time at which getting up from sleep takes place and in the process, the interrogative word karāmhöngö ‘when (in terms of hour)?’ is found functioning as temporal adverb.

5.2.2.3. Sentence with kasī

They are sentences that begin with the interrogative word kasī ‘how (manner)?’ and seek to elicit information concerning the manner of incidence of an action. The sentence 32 given below can be taken for the illustration of such a function.

32. Kasī chōṅ rōh chōṅ tōng tō kōk matāi mën in chōṅ?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>kasī</th>
<th>chōṅ</th>
<th>rōh</th>
<th>chōṅ</th>
<th>tōng tō</th>
<th>kōk</th>
<th>matāi mën</th>
<th>in</th>
<th>chōṅ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑsiː</td>
<td>c̩ː</td>
<td>tɔ̃ x</td>
<td>c̩ː</td>
<td>tɔŋŋə</td>
<td>k̩ː</td>
<td>mɔtɑˑj ṁː</td>
<td>?in</td>
<td>c̩ː</td>
</tr>
<tr>
<td>Gloss</td>
<td>how</td>
<td>svs</td>
<td>can</td>
<td>svs</td>
<td>reach</td>
<td>dist3</td>
<td>your island</td>
<td>prox</td>
<td>I</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How can I reach your island?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

38 The sentence can also be found occurring with the elision of serial verb subject.
39 Because of free word-order, the future marker is occurring sentence finally. Again, it is because of the free word-order the temporal adverbial phrase jɔʔ ?ɔəˑl xɑkiː ‘in the morning’ occur after the subject ṁː ‘you’ of the sentence.
40 See footnote 38.
As inferred, sentence 32 seeks to elicit information concerning the manner, such as might be sailing, flying etc., by which the action of reaching the island can be realized. In the process, the interrogative word kasī ‘how (manner)?’ seems to function as adverb of manner. It is found that, the same kasī ‘how (manner)?’ can also be made use of in other instances for eliciting information concerning health of animate beings. Such a function of can be inferred from the sentence 33 listed below.

33. Kasī őt in mën?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>kasī</th>
<th>őt</th>
<th>in</th>
<th>mën</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑsɪ</td>
<td>ʔɔˑt</td>
<td>ʔɪ n</td>
<td>mẽː</td>
</tr>
<tr>
<td>Gloss</td>
<td>how</td>
<td>exis</td>
<td>prox</td>
<td>you⁴¹</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How are you?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.2.4. Sentence with katāi

They are sentences that begin with the interrogative word katāi ‘how (means)?’ and seek to elicit information concerning the means of incidence of an action. The sentence listed below as 34 can be taken as illustration.

34. Katāi halīngöse köḵ insūön larōm in mën?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>katāi</th>
<th>halīngöse</th>
<th>köḵ</th>
<th>Insūön larōm</th>
<th>in</th>
<th>mën</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑt̪ɑˑj</td>
<td>xɑliːŋɑse</td>
<td>kəˑʔ</td>
<td>ʔɪnsuˑən lɑɹoˑm</td>
<td>ʔɪ n</td>
<td>mẽː</td>
</tr>
<tr>
<td>Gloss</td>
<td>how</td>
<td>prepare</td>
<td>dist₃</td>
<td>pandanus</td>
<td>bread</td>
<td>prox</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How do you prepare pandanus bread?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As inferred, sentence 34 seeks to elicit information concerning various processes, such as might be bringing raw pandanus fruits by canoe, dressing them with knife, cooking them in pot, removing the dough with a metal piece etc., involved in the preparation of pandanus bread. While doing so, the interrogative word kɑt̪ɑˑj ‘how (means)?’ grammatically functions as manner adverb.

5.2.2.5. Sentence with karāmkui

They are sentences that begin with the interrogative word karāmkui ‘how many (human)?’ and seek to elicit information concerning the quantity of human nouns. The sentence given below as 35 can be taken as illustration for such a function.

35. Karāmkui őt tāi in mën ufē köḵ kōön?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>karāmkui</th>
<th>őt</th>
<th>tāi</th>
<th>in</th>
<th>mën</th>
<th>ufē</th>
<th>köḵ</th>
<th>kōön</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑɹɑˑmkuj</td>
<td>ʔɔˑt</td>
<td>ʔɑˑj</td>
<td>ʔɪ n</td>
<td>mẽː</td>
<td>ʔufēː</td>
<td>kɔˑʔ</td>
<td>kɔˑən</td>
</tr>
<tr>
<td>Gloss</td>
<td>how many</td>
<td>exis</td>
<td>dat</td>
<td>prox</td>
<td>2sg</td>
<td>pl</td>
<td>dist₃</td>
<td>child⁴³</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How many children do you have?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen, sentence 35 above seeks to elicit information concerning the number of children the person has. In the process, the interrogative word karāmkui ‘how many (human)?’ is found to function as grammatical adjective.

5.2.2.6. Sentence with karīsĕ

They are sentences that begin with the interrogative word karīsĕ ‘how many?, how much?’ and seek to elicit information concerning quantity of human or non-human or mass nouns. The capability of such sentences in eliciting of human noun can be illustrated by the sentence 36 given below.

See footnote 29.
See footnote 25.
See footnote 36.
36. Karīsĕ ôt tāi in mēn ufē kōḍ kōon?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>karīsĕ</th>
<th>ôt</th>
<th>tāi</th>
<th>in</th>
<th>mēn</th>
<th>ufē</th>
<th>kōḍ</th>
<th>kōon</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kari:se</td>
<td>ʔɔˑt</td>
<td>t̪ɑˑj</td>
<td>ʔiˑn</td>
<td>m̪eˑː</td>
<td>ʔufeˑː</td>
<td>kɔˑʔ</td>
<td>kɔˑon</td>
</tr>
<tr>
<td>Gloss</td>
<td>how many</td>
<td>exis</td>
<td>dat</td>
<td>prox</td>
<td>2sg</td>
<td>pl</td>
<td>dist3</td>
<td>child44</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How many children do you have?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sentences listed below as 37 and 38 can be taken as respective illustrations for the capability of such sentences in eliciting quantity of non-human noun and mass noun.

37. Karīsĕ ôt ki āṅn n̪ôt tō āṅne?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>karīsĕ</th>
<th>ôt</th>
<th>ki</th>
<th>āṅn</th>
<th>n̪ôt</th>
<th>tō āṅne</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kai̱:se</td>
<td>ʔɔˑt</td>
<td>ki</td>
<td>ʔaˑn</td>
<td>n̪oˑt</td>
<td>ʔaˑʔaˑn</td>
</tr>
<tr>
<td>Gloss</td>
<td>how many</td>
<td>exis</td>
<td>pl</td>
<td>dist1</td>
<td>pig</td>
<td>there45</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How many pigs are there?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

38. Karīsĕ mēn umkōmö in hinyuāh in mēn?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>karīsĕ</th>
<th>mēn</th>
<th>umkōmö</th>
<th>in</th>
<th>hinyuāh</th>
<th>in</th>
<th>mēn</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kai̱:se</td>
<td>m̪eˑː</td>
<td>ʔumkoˑmə</td>
<td>ʔin</td>
<td>xinjuˑx</td>
<td>ʔin</td>
<td>m̪eˑː</td>
</tr>
<tr>
<td>Gloss</td>
<td>how much</td>
<td>svs</td>
<td>receive</td>
<td>prox</td>
<td>salary</td>
<td>prox</td>
<td>you46</td>
</tr>
<tr>
<td>Translation</td>
<td>How much salary you receive?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While seeking to elicit the quantity, the interrogative word karīsĕ ‘how many?, how much? in the sentences 37 and 38 is found to function as grammatical adverb.

5.2.2.7. Sentence with karīhöt

They are sentences that begin with the interrogative word karīhöt ‘how much (liquid)?’ and seek to elicit information concerning quantity of liquid nouns. The sentence 39 given below would illustrate such a function.

39. Karīhöt mēn nēn top in tōök minyeūi in mēn?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>karīhöt</th>
<th>mēn</th>
<th>nēn</th>
<th>top</th>
<th>in</th>
<th>tōök</th>
<th>minyeūi</th>
<th>in</th>
<th>mēn</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kai̱:xət</td>
<td>m̪eˑː</td>
<td>neˑn</td>
<td>t̪oˑp</td>
<td>ʔin</td>
<td>t̪oˑʔk</td>
<td>minjuˑj</td>
<td>ʔin</td>
<td>m̪eˑː</td>
</tr>
<tr>
<td>Gloss</td>
<td>how much</td>
<td>svs</td>
<td>pst</td>
<td>drink</td>
<td>prox</td>
<td>toddy</td>
<td>yesterday</td>
<td>prox</td>
<td>you47</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How much toddy did you drink yesterday?’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen above, sentence 39 seeks to elicit the quantity of toddy, a liquid noun and while doing so, the interrogative word karīhöt ‘how much (liquid)?’ functions as grammatical adverb.

5.2.2.8. Sentence with kareūöt

They are sentences that begin with the interrogative word kareūöt ‘how long?’ and seek to elicit information concerning the length of nouns. The following sentence 40 can be taken for illustrating such a function.

44 See footnote 36.
45 The sentence is unmarked for present tense and can also be found occurring with the elision of existential verb. Again, it is because of the free word-order, the locational adverb t̪ɑˑʔaˑn ‘there’ occur sentence finally.
46 The sentence is unmarked for present tense and can also be found occurring with the elision of serial verb subject.
47 See footnote 38.
40. Kareuöt nek òal matāi nek ĭnh?

<table>
<thead>
<tr>
<th>Modern Muöt</th>
<th>kareuöt</th>
<th>nek</th>
<th>òal matāi</th>
<th>nek ĭnh</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA</td>
<td>kɑɹɯəˑt</td>
<td>neˑʔ</td>
<td>?oˑʔ maˑj</td>
<td>neˑʔ ʔɛˑx?</td>
</tr>
<tr>
<td>Gloss</td>
<td>how long</td>
<td>prox</td>
<td>village</td>
<td>this⁴⁸</td>
</tr>
<tr>
<td>Translation</td>
<td>‘How long is this village?’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen above, sentence 40 seeks to elicit the length of a village and during the process the interrogative word kareuöt ‘how long?’ seems to function as grammatical adverb. The table-3 below will provide a cursory look at the markers of interrogation dealt with so far.

Table 3

<table>
<thead>
<tr>
<th>Sentence final rising intonation</th>
<th>Interrogative word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ch-interrogative word</td>
</tr>
<tr>
<td></td>
<td>Chī</td>
</tr>
<tr>
<td></td>
<td>Chūn</td>
</tr>
<tr>
<td></td>
<td>Chuān</td>
</tr>
<tr>
<td></td>
<td>Chūnsi</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. Finding

1. All the three previous works reviewed here, Röepstorff (1875), Röepstorff (1884) and Man (1889) seem to make use of interrogative words for the purpose of interrogation. But Man (1889), in addition, mentions about the use of tone of the voice also for the purpose (cf. section 4.3.6). The present day language is found to carry out the process of interrogation with interrogative words as well as with sentence final intonation.

2. All the three works, Röepstorff (1875), Röepstorff (1884) and Man (1889) seem to make use of both ch- and ka- interrogative words for the purpose as in the present day Muöt, but with the following distinctions:

- Among the ch- interrogative words, Tīi ‘who?’ of Röepstorff (1884) and chī ‘who’ of (Man 1889) are said to be used for eliciting information on human nouns (cf. sections 4.2.1 & 4.3.1). Whereas, in the present day language the form seems to be used for eliciting information on non-human nouns also (cf. section 5.2.1.1).

- Among the ka- interrogative words, all the three works are found to have the form katōm ‘how many, how much’ for eliciting information concerning quantity. Whereas, the present day language seems to make use of the two forms, karāmkui ‘how many (human)?’ and karīsē ‘how many, how much’ for the purpose (cf. sections 5.2.2.5 & 5.2.2.6).

- In addition to ka- interrogative words, both Röepstorff (1884) and (Man 1889) mention the use of another form kā as an independent interrogative marker (cf. sections 4.2.2 & 4.3.1). But, Man (1889) goes a step further in recognizing two more forms, kā and kan as variants of kā (cf. section 4.3.1). The present day language doesn’t seem to have such independent interrogative markers.

⁴⁸ The sentence is unmarked for present tense and can also be found occurring with existential verb.
3. In all, the discrepancies observed in the data made use of for the paper from the three classical works can be related to the graphalogical, phonological, morphological, syntactic and semantic levels of linguistic investigation. The use of Ti and æ by Röepstorff (1884) to represent palatal plosive and low-mid front vowel respectively (see, for example, Tiū ‘where’ and Kahæ ‘when’ of section 4.2.2) can be taken as instances of graphalogical discrepancy. Likewise, the lacuna being observed in Röepstorff (1875) and (1884) in recognizing nasalization (see, for example, Kahæ ‘when’ of sections 4.1 and 4.2.2) and high back unrounded vowel (see, for example, tié ‘I’ of sentence 7) can be mentioned as instances of discrepancy at the phonological level. The use of the lexical forms kin-kee-en, joa and kyouja to convey the interrogative meaning ‘what’, ‘where’ and ‘why’ respectively (see section 4.1) by Röepstorff (1875), and the use of the forms kâ, ka and kan as independent interrogative particle to convey the meaning ‘what’ (see section 4.3.1) by Man (1889) can be alluded to the morphological discrepancy. The use of the sentence Chūan wī meṅ ta watshī meṅ met dalngatô ‘Why do you behave in such a way? Aren’t you ashamed of yourself?’ (see, sentence 14) by Man (1889) having the verb wī ‘behave’ of the main sentence occurring within the serial verb construction, i.e., between chūan ‘why (serial verb)’ and meṅ ‘serial verb subject’ (see, for example, sentences 25, 29, 32 and 38 where in the main verbs follow the serial verb subject) can be taken as an instance of syntactic discrepancy. The use of the agentivized forms such as Ka-ra-miūa-hange (see sections 4.3.3 and 4.3.4) to convey the semantic opposites of their non-agentivized counterparts by Man (1889) can be regarded as instances of discrepancy observed at the semantic level.

References


Abbreviations

2sg: Second person singular
cf.: Compare with
cop: Copula verb
dat: Dative case
dist₁: Distal demonstrative 1
dist₃: Distal demonstrative 3
2du: Second person dual
exis: Existential verb
fut: Future
ibid.: In the same source
p: page
perf: perfect
pp: pages
pl: Plural
pst: Past
poss: Possessive case
prog: Progressive
prox: Proximate demonstrative
purp: purposive
q: Interrogative marker
sfi: Sentence final intonation
soc: Sociative case
svs: Serial verb subject
Fifty Years of Mon-Khmer Studies

Suwilai Premsrirat
Mahidol University

It has been a great honour for me to have participated in the development of MKS Journal, and like all of my colleagues I am extremely proud to celebrate its first 50 years as a distinguished source of research on the South East Asian linguistic region.

MKS Journal was first published in 1964 under the joint sponsorship of SIL and the Linguistic Circle of Vietnam as an outlet for the work of young linguists working on various aspects of the languages of Vietnam (MKS I). Then MKS II – MKS XIV were produced by editorial staff in a variety of places including Vietnam, USA, Hawaii, and eventually in Thailand where the Research Institute of Languages and Cultures for Rural Development (now the Research Institute for Languages and Cultures of Asia), Mahidol University took over responsibility for the editing process, and University of Hawaii Press for the publishing of editions XV – XIX. SIL Dallas Academic Publications were responsible for publishing editions MKS XX to MKS 40. From MKS 41 on, the journal has been a fully online publication with each article becoming available to readers as soon as it is approved by the editorial board. Currently, MKS Journal is under the management of an active group of young MKS linguists led by Paul Sidwell, Brian Migliazza, Sophana Srirchampa and others. The MKS Journal expects to expand its coverage of languages of the Austroasiatic Language Family, and apart from remaining an outlet for the documentation of MKS languages of Mainland SEA as well as South Asia and the languages of Southwest China, we also expect to publish more work on language documentation and revitalization by both the speakers themselves, and linguists who, increasingly, recognize the treasure of these languages.

Kenneth Gregerson
SIL International

In a Memoriam for David D. Thomas (MKS 36) I wrote:

“It may be difficult to appreciate now in the twenty first century just how little was known about this Southeast Asian family of languages when Dave [Thomas] and his wife Dorothy (Dot) arrived in Vietnam in the 1950’s to begin their research. Respected scholars were still, following Pater Wilhelm Schmidt, classifying Chamic languages as Mon-Khmer—an issue laid to rest by Richard Pittman in 1959. The sub-groupings of Mon-Khmer languages were vague and had little empirical basis. Thomas, acknowledging the great French scholarly tradition in Indochina and celebrating especially the ground-breaking work of Haudricourt, set about with his colleagues both to study in detail and to classify the many Montagnard groups in the region. Dave, along with Prof. Nguyen Dinh Hoa, formed the Linguistic Circle of Saigon, which in turn launched the journal Mon-Khmer Studies in 1964. This unique journal was a ‘labor of love’ which has over the years had several homes (now at Mahidol University in Bangkok), but would surely have died on the vine without Dave’s single-handed persistence.”

1. A Beginning: Volume 1 of Mon-Khmer Studies

Ken Smith who was “present at creation” in the early years of MKS, has taken in hand his copy of Volume 1, now worn and discolored by fifty years of Southeast Asian sunlight and monsoons. He summarizes the content of that first issue:

Suwilai PREMSRIRAT; Kenneth GREGERSON. 2014. Fifty years of Mon-Khmer Studies. 
Mon-Khmer Studies (Notes, Reviews, Data-Papers) 43.2:i-iv
Suwilai PREMSRIRAT; Kenneth GREGERSON. 2014. Fifty years of Mon-Khmer Studies. Mon-Khmer Studies (Notes, Reviews, Data-Papers) 43.2:i-iv
3. An All-embracing Outlook

As reflected in Volume 1, MKS was never restricted just to Mon-Khmer. In his article (Vol 1: 149-163) Thomas provided a historical thumbnail sketch framing the MK comparative enterprise within a larger Austroasiatic scheme of things (touching even on Schmidt’s Austric as well as speculations by Hevesy and Coedes of a Finno-Ugric connection for Munda). These large-scale questions continue to the present—though Finno-Ugric has not been brought up recently! As the MKSJ journal website states:

Scope of the Journal: MON-KHMER STUDIES specializes in Austroasiatic (including Mon-Khmer and Munda) linguistics. For fifty years the MKSJ has provided a scholarly forum for high-quality articles relevant to the study of Southeast Asian languages and cultures. We welcome submissions that advance the study of any SEA language family; topics may include linguistic description, cultural description, comparison, bibliography, historical development, sociolinguistics, stylistics, orthography, and paleography.”

This breadth of inquiry was hinted at the first volume, in which papers included John Miller’s Word Classes in Brou, a descriptive tour of the basic grammar of MK language cast in a Tagmemic framework. In this same issue John Banker writes on Transformational Paradigms of Bahnar Clauses, employing a then contemporary approach due to Henry Hiz, one Thomas’s professors at the University of Pennsylvania.

This open market place of ideas and tools of explanation continues today with, for example, Edmondson, Gregerson and Sidwell applying to the North Bahnaric clade some uses of Bayesian probability that have shown promise in Proto-Indo-European linguistic classification (MKS Special Issue 3:33-37, 2011).

I am indebted to my colleague, Dick Watson, who wrote on Pacoh Phonemes in MKS 1, for pointing out that each ensuing volume likewise carried its own significance, e.g.:

MKS 2 was dedicated by Dick Pittman to two of our colleagues, Gaspar Makil and Elwood Jacobsen, who were killed during the war while going about their work. MKS 3 published on Atjehnese, a language outside of Vietnam, by Vaughn Collins. MKS 4 saw the journals sponsorship shift to Dr Hoa’s Center for Vietnamese Studies at Carbondale, Illinois.

……

And so it goes, each volume has stories to tell and they are not always linguistic ones.

4. A Circle of Friends

I remember the Linguistic Circle of Saigon in the 60’s as an informal gathering of Vietnamese professors, students, and an eclectic assortment of the rest of us from wherever—a cordial and convivial gathering to talk about and hear of all things linguistic. Sometimes an SIL member would report on a discovery he/she had made in their village fieldwork. I remember interpreting for a Vietnamese speaker who had gone to a linguistic conference abroad which revealed the latest developments in Chomskian grammar—what is the word for ‘generative’ in Vietnamese?

While MKS represents the contributions of a “circle of (mostly) friends”, who occasionally differ on this or that linguistic point, for the most part we have been able, despite various entrenched positions, to maintain after a half a century a collegial easiness and camaraderie. And happily, like a pebble tossed in the water, the ripple effect has widened the circle to include colleagues around the world.

On a final note, MKS Vol 1 lists the seven first contributors, three women and four men, not one of whom was a PhD! It started out as a kind of ‘Mom and Pop’ affair. They described what they observed with the linguistics they knew—an on-the-job training program. In a sense, they were a metaphor for the state of MK and Austroasiatic studies at that time. In the fifty years that have ensued, they have matured and so has the entire field, but it is my hope that we will never
grow out of the experience of getting our hands ‘dirty’ with data and having the ‘light bulbs go off in our heads’ as we see how each new piece of the puzzle falls wondrously into place.

Regrettably, space fails to acknowledge the many authors and supporters who have contributed so much to the success of our Journal. But here’s to the next 50 years of the Mon-Khmer Studies Journal and everyone for whom it remains a ‘labor of love.’
Acoustic correlates of rhythmic structure of Vietnamese narrative speech

Anh-Thúr T. Nguyễn
Mountain Creek, Sunshine coast, Australia

Abstract
This paper reports a study on the acoustic realization of the rhythmic structure of Vietnamese narrative prose speech. Eight speakers of Saigon dialect read part (one page) of a short story. Acoustic measurements including duration and intensity were taken for every syllable of the excerpt. The syllables were labelled into four types: s is for a monosyllable that stands alone by itself; s₀ is for the first syllable of a three-word phrase/chunk in which it is the modifier of a bisyllabic word (e.g., cai̯ ˈ kiến, thạt ˈ âu yêm); s₁ and s₂ are the first and second syllables of a bisyllabic word/chunk (e.g., ˈ kiến, ˈ âu yêm). The one-syllable, two-syllable and three-syllable units/chunks were also labelled utterance final and utterance non-final. The results showed that for both utterance medial and utterance final chunks, the monosyllable significantly had longer duration and stronger intensity than the other syllable types. Within bisyllabic words, the second syllables had longer duration and stronger intensity than the first syllables. Within three-syllable phrases/chunks, the first syllable of a three-word phrase/chunk was not significantly different from the first syllable of the bisyllabic word and the second syllables of the bisyllabic words/chunks also had longer duration and stronger intensity than the preceding first syllables. This result suggests an iambic pattern of acoustic prominence of bisyllabic and trisyllabic words/phrases in narrative speech.

Key words: Acoustic correlates, rhythmic structure, narrative speech.
ISO 639-3 language codes: vie

1. Introduction
Vietnamese is a contour tone language and has no system of culminative word stress; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson, 1965; Nguyễn Đăng Liêm, 1970). Duration, intensity, full tonal realisation of accented syllables have been observed to be important parameters for describing stress in Vietnamese (Đỗ, 1986; Chaudhary, 1983; Hoàng & Hoàng, 1975; Gsell, 1980). Regarding the stress patterning in utterances, it is generally agreed by some researchers that there is an alternating pattern of strong and weak syllables. Thompson (1965) stated that the majority of the syllables have medium stress. In a sequence of syllables, alternating ones are slightly louder (but not in a distinctive manner): “each pause group has at least one heavy stress and weak stresses are fairly frequent in rapid passages, rarer in carefully speech” (p. 50). Jones and Huỳnh (1960) stated that “normally the stresses in a Vietnamese utterance are conditioned by the junctures,” and regarded the fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress. Consistent with Jones and Huỳnh’s observation, it is remarked by Cao (2003) that due to the demarcative function of stress/accent in Vietnamese, native listeners tended to hear a juncture after a stressed syllable even though there is no such pause in reality as examined by spectrograms. In a recent study, Schiering, Bickel and Hildebrandt (2010) remarked that “Vietnamese provides ample evidence for a genuine stress domain that is preferably disyllabic and maximally trisyllabic. Within this domain, stress is realised on the final syllable in the default case. Crucially, this domain is computed irrespective of the morphosyntactic status of its constituent syllables, i.e. stress phonology does not distinguish between a word-level and a phrasal-level of prosodic structure. Metrically, polysyllabic words are thus indistinguishable from other combinations of syllables. Since the most complex structures which are referenced by the rules for iambic rhythm are phrasal, stress may most adequately be attributed to the prosodic domain of the Phonological Phrase.”(p.673).

In recent studies on more carefully phonetically controlled and specialized sets of Vietnamese disyllabic compounds and reduplications, Nguyen and Ingram (2007a, b) have found that there was at least a phonetic tendency for the right hand element of a disyllabic compound...
word to be more prosodically prominent by a number of relevant phonetic measures: greater tonal F0 range, higher intensity, greater duration of the second syllable, and formant measurements indicative of more centralized vowel nuclei (vowel reduction) on the first syllable. Nguyen (2010) investigated the rhythmic patterns in Vietnamese polysyllabic words by examining the rhythmic patterns and their acoustic correlates in polysyllabic reduplicative words (2, 3, 4, 5, 6- syllable pseudo-words). The results showed that there is a tendency of syllable coupling indicated mainly by syllable duration pattern and supported by the native listeners’ perception results, suggesting that polysyllabic words in Vietnamese tend to be parsed into bi-syllabic iambic feet with a rightward or retrograde rhythmic pattern. In a recent study, Nguyen (2013) examined the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poems made up of three-word, five-word, six-word, seven-word, and eight-word lines. The acoustic analysis showed that the duration and intensity results mirror each other in indicating a strong iambic pattern of prominence, supporting the literature that a line of folk verse with even number of syllables tends to have a series of iambics and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn, 1985). Nevertheless, the perception results showed that listeners’ parsing patterns, though to some extent reflect the acoustic patterns, do not strongly correlate with the acoustic results. This study is a follow-up of the results found in Nguyen (2010) and Nguyen (2013) that polysyllabic units of speech in Vietnamese tend to be parsed into bi-syllabic iambic pattern as indicated by the examination of duration and intensity patterns of syllables in narrative prose speech.

2. Experiment

2.1. Linguistic materials

In order to pursue the aim of the study, part (one page length) of a short story titled Tôi đi học (I went to school) by Thanh Tịnh was used. The excerpts consisted of 372 syllables in total. The excerpt is in the appendix.

2.2. Subjects

Eight speakers of the Sài Gòn dialect (4 males, 4 females) who came from Hồ Chí Minh city participated in the study. They were either visitors or newly arrived immigrants to Australia and had been in Australia from 2 weeks to 4 years. Their age ranged from 38 to 70 years. Their education levels ranged from high school to higher degrees.

2.3. Procedure

Subjects were given the short story in print to practice reading before the recording. They were asked to read the excerpts in a natural narrative manner. Recordings were made in a quiet room using sound recording and editing computer software PRAAT (Boersma and Weenink, 2007) at 22050 Hz sampling rate.

2.4. Measurement

The acoustic parameters measured included syllable duration (ms) and syllable intensity (dB). Temporal variations of F0 and tonal shapes are obvious components but will not be treated here due to the nature of the linguistic material. That is, the examination of tones requires a comparison and contrast of constant segmental compositions of the target linguistic materials which cannot be met by the nature of the short story. Peak intensity (dB) in syllables and syllable duration (ms) were measured manually via Praat (Boersma and Weenink, 2007).

The one-syllable, two-syllable and three-syllable units/chunks of speech were segmented by the researcher on the basis of auditory signal and spectrogram: cues for segmentation included final lengthening, pause, F0 reset and auditory rhythmic perceptual cues. An example of the segmented chunks was presented in Appendix 2. The syllables were labelled in a Praat Textgrid into four types by the researcher based on auditory signal and spectrogram: s is for a monosyllable that stands alone by itself (e.g., vâ [and], lại [again], ãy[that], tôi [I]); s0 is for the first syllable of a three-word phrase/chunk in which it is the modifier of a bisyllabic word (e.g., cái ý kiến, thất âu yếm); s1 and s2 are the first and second syllables of a bisyllabic word/chunk (e.g., ý kiến, âu yếm). In addition, utterances were also identified based on the punctuations in the text (periods and commas) together
with the silent pauses and pitch reset in the speech signal. The one-syllable, two-syllable and three-syllable units/chunks were also labelled utterance final and utterance non-final. The utterance has been proposed as the largest unit in the prosodic hierarchy: It is the largest span of application of phonological rules (Selkirk, 1978, 1980; Nespor & Vogel, 1986; Hayes, 1989) and its boundaries are sometimes said to be the location of non-hesitation pauses (Hayes, 1989). This unit often corresponds to a single syntactic sentence, but can include two or more sentences joined into a single higher-level sentence (Selkirk, 1978). It is noted that all of the utterances in the narrative were found to be followed by silent pauses and/or with pitch resets.

2.5. Analysis

There were in total 2616 syllables (327 syllables per excerpt x 8 speakers). A mixed (fixed and random) effects analysis of variance (ANOVA) model, using the restricted maximum likelihood method (REML) to estimate variance components was used to statistically analyse the data. The dependent variables were syllable duration and intensity. The fixed effects were syllable positions (s, s0, s1, s2), segmented chunks (one-syllable, two-syllable, and three-syllable chunks) and chunk positions (utterance final and utterance non-final). The random effect was speakers and items. Tukey post-hoc tests were carried out to determine the significant differences among levels of the main effects when necessary.

The use of REML overcomes the potentially serious deficiency of the ANOVA-based methods which assumed that data are sampled from a random population and normally distributed. REML also avoids bias arising from maximum likelihood estimators in which all fixed effects are known without errors, consequently tend to downwardly bias estimates of variance components. Moreover, REML can handle unbalanced data. The data analysis was carried out using SAS program.

2.6. Results

2.6.1. Prosodic units

Majority of the narrative speech was segmented as two-syllable units (60.4%), while three-syllable units accounted for 23.4% and one-syllable unit only 16.2%.

2.6.2. Duration

The three-way mixed effect ANOVA (syllable positions x segmented chunks x chunk positions) on syllable duration showed a significant effect for the main factor syllable positions: F(5, 2931) = 9.12, p < .0001, chunk positions F(1, 2931) = 69.77, p<.0001, but no significant effects for segmented chunks F (2, 2931) = 0.03, p = 0.5 ns. The interactions syllable positions x segmented chunks x chunk positions: F(1, 2931) = 0.9, p =.09 ns and other interactions were not significant.

A post-hoc Tukey test on the significant differences among levels of the main factor of syllable positions (figure 1 below) showed that for both utterance medial and utterance final positions, the monosyllables significantly had longer duration than the other syllable types. Within bisyllabic words/chunks, the second syllables had longer duration than the first syllables. Within three-syllable phrases/chunks in non-final position the first syllable (s0) of a three-word phrase was not significantly different from the first syllable (s1) and second (s2) of the bisyllabic words and within the bisyllabic words the second syllable (s2) was significantly longer than the first syllable (s1). In utterance final position, the first syllable (s0) of a three-word phrase was significantly different from the first syllable (s1) and second (s2) of the bisyllabic words and within the bisyllabic words the second syllable (s2) was significantly longer than the first syllable (s1). Generally, the results are very robust for both utterance final and non-final positions. Utterance final syllables are also shown to be marginally longer than those at utterance non-final position, indicating a final lengthening effect.
*Mon-Khmer Studies* 43.2:1-7

**Figure 1:** Mean duration (ms). $s$, $s_0$, $s_1$ and $s_2$ are syllable positions in segmented chunks.  
1syll, 2syll, and 3syll: number of syllables in chunks.  
Non final and final are positions of chunks in utterances. The symbol * means $p < .01$

### 2.6.3. Intensity

The three-way mixed effect ANOVA (syllable positions x segmented chunks x chunk positions) on syllable intensity showed a significant effect for the main factors syllable positions: $F(5, 2931) = 6.16$, $p < .0001$, and chunk positions $F(1, 2931) = 68.5$, $p = .0001$, but no significant effect for segmented chunks $F(2, 2931) = 0.84$, $p=0.4$ ns. The interaction syllable positions x segmented chunks x chunk positions: $F(1, 2931) = 1.31$, $p =0.2$ ns. and other interactions were not significant.

A post-hoc Tukey test on the significant differences among levels of the main factor of syllable positions (figure 2 below) showed that in both utterance non-final and final positions, the monosyllable significantly had stronger intensity than the two other syllable types ($s_0$ and $s_1$). Within bisyllabic words/chunks, the second syllables had stronger intensity than the first syllables. Within three-syllable phrases/chunks, the first syllable ($s_0$) of a three-word phrase was not significantly different from the first syllable ($s_1$) of the bisyllabic words and the second syllable ($s_2$) also had stronger intensity than the first syllable ($s_1$). The results also show that intensity value of the utterance–final syllables is lower than that of utterance non-final ones, indicating an intensity declination at utterance final.
2.7. Discussion and conclusion

In summary the duration and intensity results mirror each other in indicating an iambic pattern \((s2>s1)\) of acoustic prominence of bisyllabic words/phrases/chunks in narrative speech, consistent with Nguyen and Ingram (2007a, b) that in a bisyllabic unit of speech, the second element is more acoustically prominent than the first element. Generally, the patterns of acoustic results are consistent with the researcher’s segmentation in indicating clear units of one–syllable, two-syllable or three-syllable chunks, particularly the final syllable of the speech unit always had the most acoustic prominence. The results mirror those found in Nguyen (2010) and Nguyen (2013) that polysyllabic units of speech in Vietnamese tend to be parsed into bi-syllabic \((s2>s1)\) or tri-syllabic \((s2>s0>s1)\) iambic feet and when syllables stood by themselves, they tended to be lengthened to fill the bi-syllabic foot template. This result also reflects observations by Thompson (1965) and Jones and Huỳnh (1960) that “fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress”. The results also support Schiering, Bickel and Hildebrandt (2010) remarks that “Vietnamese provides ample evidence for a genuine stress domain that is preferably disyllabic and maximally trisyllable. Within this domain, stress is realised on the final syllable in the default case. Crucially, this domain is computed irrespective of the morphosyntactic status of its constituent syllables, i.e. stress phonology does not distinguish between a word-level and a phrasal-level of prosodic structure. Metrically, polysyllabic words are thus indistinguishable from other combinations of syllables. Since the most complex structures which are referenced by the rules for iambic rhythm are phrasal, stress may most adequately be attributed to the prosodic domain of the Phonological Phrase.”(p.673). It is wondered whether the results of this study is extended to spontaneous speech which needs to be investigated in future studies.

References


Appendix 1: the first excerpt of a short story

TÔI ĐIFEST, a short story by Thanh Tịnh

Hằng năm cứ vào cuối thu, lá ngói dương rụng nhiều và trên không có những đám mây bồng bềnh, lòng tôi lại nao nức những ký niệm hoang mang của buổi tuổi thơ.

Tôi không thể nào quên được những cảnh giác trong sáng ấy nay nọ trong lòng tôi như mấy cánh hoa tươi mỉm cười dưới cái buổi nắng rực rỡ.

Những yếu tố ấy tôi chưa lẩn nắc gì lên đầu, vì hồi ấy tôi không hề biết gì và ngày nay tôi không nhớ hết. Nhưng mỗi lần thấy mây em nó tự nhiên nở rộ lên, lòng tôi lại tùng bừng rộ rã.


Tôi không lôi qua sở thích điều như thường Quy và không ra dòng nhỏ hồi thơ mộng Sơ nứa.

Trong chiếc áo vải đụ dể tôi cảm thấy mình trung trực và đứng đảm.

Đọc đường tôi thấy mây cửu nhỏ trắc bàng tôi, áo quản tựm tắt, nhị nhịm gọi tên nhau hay trao sánh vẻ cho nhau xem mà tôi thêm. Hai quén vị mới dường ở trên tay tôi đa bất đầu thấy nặng. Tôi bấm tay ghi thật chặt, những một quen vị cùng chỉ ra và chính đầu chửi xuống đất. Tôi xúc

lên và năm lại cần thân. Máy cấu di trước o sách vor thiết nhiều lại kèm cả bút trước nữa. Nhưng máy cấu không để lỗi về khó khăn gì hết.

Tôi muốn thử sức mình nền nhìn mẹ tôi:
- Mẹ đưa bút trước cho con cảm.
Mẹ tôi cúi đầu nhìn tôi với cặp mắt thật yêu:
- Thôi để mẹ nắm cũng được.
Tôi có ngay cái ý kiến vừa non nốt vừa ngày thơ này: chắc chỉ người thạo mới cảm nổi bút trước.

Appendix 2
An example of the segmented chunks
Tôi/ không thế nào/ quên được/ những cảm giác/ trong sáng/ ấy/ nay nói/ trong lòng tôi/ như/ máy cảnh/ hoa tươi/ mỉm cười/ giữa bầu trời/ quang đãng/.
Agreement in Ho

P. K. Choudhary
CIIL, Mysore

Abstract
The aim of this paper is to show the conditions of agreement marking in Ho, an Austroasiatic language. Pronominal clitics are attached as an agreement marker on the preverbal word, the verb and may be before or after the finite marker. The agreement morpheme allows Subject, Object or both to be dropped. There is no inflectional agreement marker in Ho. The agglutinating nature of the morpheme makes North-Munda languages in general and Ho in particular, morphologically very rich. Negation too, can host agreement clitic. The function of these clitic is twofold; one is to carry the load of the NP and another is to allow that NP to be dropped.

Key words: Agreement, Pro-drop, Animacy

ISO 639-3 language codes: hoc, mai

1. Introduction
Ho language is very rich in agreement. It shows that lexical case marking may not block the agreement between the NP and the verb like Hindi. According to Subbarao (2001), a noun phrase whether lexically case marked or not, agrees with the verb if the functional head agreement is “active”. Agreement is not morphologically manifested and is “covert” if the functional head agreement is “weak”.

This paper is an attempt to define the term Agreement and it deals with the types of agreement with subject, object and possessor of an object in Ho. This paper also discusses the pro-drop parameter to show that there are languages such as Ho in particular and North-Munda languages in general where the subject clitic attached to the first material preceding the verb, which sometimes may be an object NP, and then it agrees with the verb. A subject may or may not be overtly realised.

2. Definition of Agreement
There are so many definitions of the term Agreement. According to Crystal’s dictionary of linguistics and phonetics (1995:13), agreement is “the formal relationship between elements whereby a form of one word requires a corresponding form of another”. Steel (1978:610) also defines agreement as “the term agreement commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another, for example, adjectives may take some formal indication of the number and gender of the noun they modify”. Keenan (1978), Lehman (1982) and Haegeman (1994) have also tried to define the term Agreement. Haegeman (1994) defines agreement as “a formal requirement for indicating specific syntactic properties between constituents”. She also mentions that agreement plays a major role in “Binding and Case Theory”, which are components of Government and Binding theory.

All definitions, in fact, focus on one important point, that is, the covariance of matching of features between separate elements, such as a Noun, subject NP, a Verb and an Adjective. The term Concord has been used as synonymous with agreement, with no clear distinction between the two. Nevertheless, it seems that in recent generative linguistics, Agreement has resurfaced with a new range of applicability.

In the 1990s, agreement has become the focal point of study. Linguists have proposed different ways of examining agreement with a keen interest to find out the universal principles, which govern agreement patterns with reference to parametric variations. More recently agreement

---

1 Mundari is a main language of Kherwari group of North-Munda along with Santhali language of Austroasiatic family. Grierson (1906), in Linguistic Survey of India mentioned it as a close dialect of Ho and Bhumij. Anderson (2001) and Diffloth (2005) put it under North-Munda group of Kherwarian in their classification.
became the focal point in the *Minimalist Program*. Chomsky (1981) treats agreement with superscripts in part because it fails to serve as an antecedent for the Binding Theory. Chomsky (1995a) has explored the idea that functional categories like Complement(C) and Inflection(I) are regular in terms of X-bar theory and constitute heads of phrases. Because of this analysis the subject NP is the specifier of IP. Koopman (1987) asserts that all agreement relations are Spec-Head agreement. This proposal certainly makes sense for subject-verb agreement.

3. Subject-Verb Agreement

The Agreement is overtly realized between an NP and a Verb; the most frequent pattern of this type is Subject-Verb agreement. The verb changes its form according to person, number and gender of the Subject. Some of these features may not be overtly present in all languages. Since Ho is an agglutinating language there are no such morphological modifications, but, its agreement feature occurs as an affix to the verbal root, though the place of occurrence of agreement is not fixed. Ho is also a pro-drop language, where agreement enables the subject to be dropped.

1. \( aɲ \_ u\_i \_ -\_ŋ \_ i \_ \_ jom \_ -ke \_ -d \_ -a \)
   I mango 1SG eat PST TR FM
   ‘I ate a mango’.

   In the above sentence subject pronoun is \( aɲ \). The object NP \textit{uli} ‘mango’ agrees with the subject NP and the subject pronominal clitic \(-ɲ\) occurs to the right of the object NP \textit{uli} as velar nasal \( ŋ \).

2. \( pro\_i \_ u\_i \_ -\_ŋ \_ jom \_ -ke \_ -d \_ -a \)
   mango 1SG eat PST TR FM
   ‘I ate a mango’.

   The subject can be dropped as in (2). Some more examples of subject pro-drop are given below:

3. \( (am)\_i \_ ma:qι \_ -m\_i \_ jom \_ -ke \_ -d \_ -a \)
   you meal 2SG eat PST TR FM
   ‘You have eaten meals’; ‘You ate meals’

4. \( (ini)\_i \_ laq\_ -ai\_i \_ jom \_ -a \)
   he bread 3SG eat FM
   ‘He eats bread.’

5. \( (inku)\_i \_ laq\_ -ko\_i \_ jom \_ -e \_ -a \)
   they bread PL eat 3SG FM
   ‘They eat bread.’

6. \( (ako)\_i \_ pa:iti \_ -ko\_i \_ na:m \_ -ke \_ -d \_ -a \)
   they work 3PL get PST TR FM
   ‘They got a job.’

   In sentences (3) - (6), subject pronouns can be dropped; because the subject agreement marker is coded on the object NP. For example, in sentence (6), \textit{ako} ‘they’, an NP, is subject of the sentence and it can be dropped, because the agreement marker is present on object NP \textit{pa:iti}.

4. Person Agreement and Honorific Agreement

Languages make three way distinction in person namely first person (speaker), second person (hearer) and third person (listener). There are languages called gender sensitive like Hindi where gender plays an important role in agreement marking. There are languages like Maithili where honorific marker plays an important role in agreement marking. As far as Ho is concerned, it makes a distinction of inclusive or exclusive pronoun; Inclusive includes the hearer whereas exclusive excludes the hearer. Ho also makes distinction in honorific and non-honorific pronoun.

7. \( ako\_i \_ laq\_ -ko\_i \_ jom \_ -e \_ -a \)
   they bread 3PL eat 3SG FM
   ‘They eat bread.’
In sentence (7), ako is third person non-honorific plural pronoun and it agrees with the object. In sentence (5), we have inku, which is honorific third person plural pronoun. Thus there is distinction in pronouns in third person plural, but there is no difference in agreement marker. Even in third person singular pronouns we have a distinction in Ho. For example:

8. \( \text{ini(NH) seno} -ja -n -a \)
   \( \text{he go PST INTR FM} \)
   ‘He had gone/He went’.

9. \( \text{in-kin(H) seno} ja -n -a \)
   \( \text{he go PST INTR FM} \)
   ‘They (two) had gone.’

In example (8) ini is non-honorific in Ho, but in-kin has also been used as an honorific marker along with a dual marker. However, the verb does not exhibit any agreement. In Ho too honorific marker plays a role in agreement marking as shown in example (5) and (6).

In Ho there also exists an exclusive and inclusive distinction in first person pronouns. In sentence (10) ale is exclusive pronoun and in sentence (11) abu is inclusive pronoun of first person plural. i.e.:

10. \( \text{ale an} -lagte vo:ta -ke -ɖ -a \)
   \( \text{we I for vote PST TR FM} \)
   *(inclu) ‘We voted for me.’ (Literal meaning)

11. \( \text{alei an} -le_i sa:la: -ke -a \)
   \( \text{we I 2PL elect PST FM} \)
   *(exclu) ‘We elected me’. (Literal meaning)

If we put ape ‘you’ instead of ale ‘we’ the meaning will be ‘you elected me’ and sounds fine in (11) above.

5. Number Agreement

Ho has three way number distinctions i.e. singular, dual and plural. Number agreement correspond the number of entities, which are referred to in a NP through an agreement marker. As we have shown above there are separate agreement markers for each pronoun. We have given details of pronouns and their agreement suffixes in Anaphors and Pronouns in Mundari in Choudhary (2005).

6. Gender Agreement.

In this type of agreement, the verb agrees with the gender of the subject NP. There is no gender agreement in Ho. Munda languages are not gender sensitive like Hindi. e.g.:

12. \( \text{repo ini} -e koṭari -tan -a \)
   Repo he POSS criticise PRS FM
   ‘Repo criticises him’

13. \( \text{na:masi ini} -e koṭari -tan -a \)
   Namsi he POSS criticise PRS FM
   ‘Namsi criticises him’

14. \( \text{ini na:masi} -ke koṭari -tan -a \)
   he Namsi ACC criticise PRS FM
   ‘He criticises Namsi.’

Hence, we can say that there is no gender agreement in Ho.

7. Subject-verb agreement in intransitive verbs

In Ho, if the verb is intransitive it carries subject agreement provided there is no other constituent available to host the agreement. For example:

CHOUGHARY, P. K. 2014. Agreement in Ho. 
_Mon-Khmer Studies_ 43.2:8-16
15. \( repo_i \) ja.pi\( dik \) -a -e\( i \)  
   Repo sleep FM 3SG  
   ‘Repo sleeps.’

The reason for agreement on the verb (after the finite marker) is the absence of a direct object or an adverb. In another word, since there is no pre-verbal word in (15), verb has to carry the agreement clitic.

8. Subject-Verb agreement in transitive verbs

According to Chomsky (1995), in the Nominative-Accusative type of languages with a transitive verb, the AGRs (the subject verb agreement) is “active” in which the subject of a transitive verb agrees with the functional head AgrS of the INFL. Consider (16) from Ho.

16. kowa\( a_i \) -hon kita:b -e\( i \) pa\( ra \)w -e -a  
   boy man book 3SG read 3SG FM  
   ‘A boy reads a book’.

17. ini\( i \) la\( d_j \) -ai\( i \) jom -e\( i \) -a  
   he bread 3SG eat 3SG FM  
   ‘He eats bread.’

In sentences (16) and (17), subject controls the agreement. Hence, we can drop subject, but we cannot drop object. Since objects are not only carrying subject agreement markers, but also there is no object agreement in the verb. The object is in the third person.

9. Dative subject construction in relation to agreement

It is a common phenomenon in most of the Indo-Aryan languages that predicates expressing psychological feeling; possession and duty etc. mark their subject with a dative or genitive case marker. In a set of languages of South-Asia verb agrees with the object. The appearance of subject agreement is blocked in such cases and the dative marker on the indirect object of a di-transitive verb in perfective aspect along with the presence of an ergative marker on the subject necessitates the verb to agree with its direct object in languages such as Hindi. In Ho, however there is no ergative marker. The subject agreement marker does not occur on the DO or IO if it is dative case marked.

18. repo\( i \) na.masi -ke kita:b -e\( i \) ema:q -i -/e  
   Repo Namasi DAT book 3SG give PST FM  
   ‘Repo gave a book to Namsi.’

19. (an\( j \)) (am\( j \) -ke) mia\( d \)’ uli -\( n_j \) em -le -\( d \) -me\( j \) -a  
   I you DAT one mango 1SG give PST TR 2SG FM  
   ‘I gave a mango to you.’

In sentence (18) the subject agreement marker occurs on direct object. There is a lexical case marker on the indirect object; it does not have agreement with the verb. But in (19) except subject agreement marker on the DO, a dative case marked IO is also occurring on the verb as an agreement marker. In Ho, there are some sentences, which exhibit agreement when an NP is even lexically case-marked as in (19) above, where am ‘you’ is lexically case marked but it agrees with the verb. Thus there is no blockage of agreement even an NP is lexically case marked as shown above.

10. Object-Verb Agreement

In Ho subject agreement marker generally occurs to the right of the preverbal word and object agreement marker occurs either to the left of the tense marker or to the right of the tense marker.

20. an\( j \) hapnam -ko\( i \) -\( n_j \) nel -le -\( d \) -ko\( i \) -a  
   I girl PL 1SG see PST TR 3SG FM  
   ‘I had seen girls’.
21. `ale, am –le, dula:r-te –m, –a
   we you 2PL love PRS 2SG FM
   ‘We love you.’

22. ini, laɖ –e, jom –a
   he bread 3SG eat FM
   ‘He eats bread’.

23. inku, laɖ –ko, jom –a
   they bread 3PL eat FM
   ‘They eat bread’.

In sentence (20), the pronoun aɲ ‘I’ and in (21), am can be dropped, because the verb carries the agreements of these nouns. In sentences (22) and (23), subject pronouns can be dropped, but the objects cannot be dropped since objects are carrying agreement markers of the subjects.

11. Agreement in Small Clause and ECM constructions

Radford (1998) has defined Small Clause (SC) as a clause which neither has a finite verb nor an infinitival too. It is usually abbreviated as SC and its structure can be elaborated as “NP XP / XP NP” where XP can be AP or NP etc. Let us see the agreement in such cases in Ho.

24. aɲ –re aʈkarai –min –a [(aɲ i) sugraː-ɲ i]
   I PP believe be FM I good-1SG
   ‘[I believe [myself to be good]]’

25. ini [aːj-aːj-ge sugaraː] aʈkarain ke –ɖ –a
   he himself-EMP good believe PST TR FM
   ‘He believed [himself to be smart]’

In sentence (24), there is an agreement marker on the embedded verb but in (25), there are no agreement markers either on matrix verb or embedded verb.

12. Adjective Agreement

Adjectives are used in connection with a noun or pronoun to produce a description about the person, thing or group referred to. Adjectives may either be used predicatively or attributively. The adjective “beautiful” in “a beautiful girl” is used attributively because it is placed in front of the noun it qualifies. The same adjective in “the girl is beautiful” is used as predicate because it is placed after a linking verb. Adjectives used in this way are called the complements of the linking verb. In Ho there is no agreement with adjectives.

26. en sepeɖ bugin -a
    that boy good FM
    ‘That boy is good’.

27. bugin sepeɖ
    good boy
    ‘Good boy’.

In sentences (26) and (27), there is no agreement on the NP sepeɖ ‘boy’ and adjective bugin in either case. Hence, there is no adjective agreement in Ho.

13. Demonstrative Agreement

Demonstratives are generally used to refer to people or things in a definite way. There is no agreement with demonstratives in Ho as in (26) above illustrates the point.

14. Possessive Agreement

In some languages such as Maithili, Hmar and Kurmali there is a possessor agreement. In Ho, there is no agreement with possessor as in example (28) and (29) below:
   Soma your hand wash PST TR FM 3SG
   ‘Soma washed your hands.’

29. **soːmaː**: apaː: ti jiː: abuŋ ke -d -a -i
   Soma my hand wash PST TR FM 3SG
   ‘Soma washed my hands’.

**Maithili (mai)**

30. **siːtaː**: Ḟomᵊ -or haːth Ḟho -l -okᵊ
   Sita I GEN hand wash PST 1SG
   ‘Sita washed my hands’.

   In (30), verb agrees with the possessor Ḟomᵊ, thus we have an example of possessive agreement in Maithili but there is no such agreement in example (28) and (29).

15. **Long-Distance Agreement**

   There are instances of long distance agreement in some South-Asian languages such as Hindi, Punjabi, Kashmiri, Maithili, Mizo, Hmar and Telugu (Subbarao, 2001). Let us consider the case of Ho.

31. **soːmaː;** adana -a [S₂ cilke saːikal -ko hare -eᵊ -a]
   Soma know FM how bicycle ACC drive 3SG FM
   ‘Soma knows [how to ride a bicycle.]’

32. **soːmaː**: sanan -tan -a [[S₂ amᵊ gupaiː-mᵊ -e]
   Soma wants PRS FM you fight 2SG 3SG
   ‘Soma wants [you to fight.]’ (Literally: Soma wants that I and you fight)

33. **soːmaː;** sanan -tan -a [S₂ aɲᵊ nikul -enᵊ ol -eᵊ -a]
   Soma want PRS FM I letter 1SG write 3SG FM
   ‘Soma wants [me to write a letter.]’ (Literally Soma wants that I write a letter)

   In sentences (31) - (33), we have evidence of long distance agreement. Soma is in 3rd person and it is the matrix subject. The third person agreement marker -e occurs with embedded verb. Hence, we can say that Ho has long-distance agreement.

16. **Agreement in Relative clauses**

   Ho has a relative-correlative construction as well as externally headed relative clauses such as in English. It has no internally headed relative clause. However, Denney (2002:89) claims that when Ho properly spoken there are no relative clauses, and what we would express by the use of a relative clause in English or Hindi is expressed in Ho by the use of a participle.

34. en uli [S₂ okona aɲᵊ: jiː: jom -ke -d -a]
   that mango which my friend eat PST TR FM
   ena sojaː -ka -n -a
   that useless PST INTR FM
   ‘[That mango which my friend had eaten] was useless.]

35. en tebal maran -gi -a [S₂ okonaː -re amᵊ jom -amᵊ]
   that table big EMP FM which PP you eat 2SG
   rike -ta -d -a
   keep PST TR FM
   ‘The table [on which you have kept the food] is big.’

   In sentence (34), we have a relative-correlative construction and in (35), the embedded relative occurs to the right of the matrix VP just as in English relative clause construction. The difference between the two is also found in its agreement. In (35), we have the second person agreement marker -am ‘you’ in the preverbal constituent of the embedded sentence, whereas there is no agreement marker in (34).
17. Agreement and the role of [+Animate]

Ho is very sensitive with regard to the feature [+Animate]. The inanimate NP has the object agreement marker -e/-i in the verb in [-past] tense. The 3rd person plural marker -ko is used as an agreement marker with inanimate plural nouns.

36. pro uri -ŋ kumbu -i -tan -a
cow 1SG steal 3SG PRS PROG FM
‘I am stealing the cow’.

37. aːje am -a setaː -ko -e taːm -ko -a
he you POSS dog PL 3SG hit PL FM
‘He hit your dogs’.

When tense is [-past], [+animate] direct object agreement marker occurs to the left of the tense marker as in (36), but when tense is [+past] direct object agreement appear after the tense marker. We find the same distinction with regard to [+human]. The occurrence of subject agreement marker to the left of the subject depends on the feature [+human] of the subject. However, in direct object agreement cases, there is no distinction between the two (human and nonhuman) subjects as shown in example (37) above.

18. Agreement in di-transitive verbs

Example (38) exhibits subject as well as direct object agreement but there is no indirect object agreement as it has been lexically case marked. Some time speakers of Ho may prefer marked order, where IO occurs initially in such cases.

38. am æn -taː hapnum -koj -m j kul -le -q -koj -a
you I for girl PL 2SG send PST TR PL FM
‘You sent girls to me’.

39. æn (am -ke) miːaɖ’ uli -ŋ æm -le -q -mej -a
I you DAT one mango 1SG give PST TR 2SG FM
‘I had given a mango to you’.

Thus in example (38), subject is marked on DO. And the IO is lexically case marked and hence, cannot trigger any agreement. However, subject and the DO can trigger agreement in the verb. In example (39), the lexically case marked IO can occur as an agreement clitic on the verb. Therefore, it is not clear that the lexically case marked IO can occur as agreement clitic or not.

19. Agreement Hierarchy

Direct object and Indirect object agreement marker cannot occur together in Ho. There seems to be a grammatical hierarchy in agreement marking among the subject, direct object and indirect object.

Sub Agr> IO Agr> DO Agr

40. am æn -taː hapnum -koj -m j kul -le -q -koj -a
you I for girl PL 2SG send PST TR 3PL FM
‘You sent girls to me’.

41. remo æ aːj -taː re miɖo banduk -e j moːgaːw -le -q -a
Remo he near one gun 3SG put PST TR FM
‘Remo put the gun near him’.

42. apuŋ æn -ke am æ e j emaːd -mi j -a
my father I ACC you 3SG give 2SG FM
‘My father gave me to you’.

In example (40), the IO is lexically case marked and it cannot occur as an agreement marker anywhere else in the sentence. Hence, only the subject agreement marker and DO agreement marker can occur. In example (41), only subject agreement marker can occur and in (42), subject as well as IO agreement marker occurs with the verb.
20. Conclusion

Ho is a Nominative/Accusative type of language in terms of verb agreement. The verb obligatorily agrees with the subject NP in terms of person and number. The subject is marked by the pronominal clitic either on the verb phrase or somewhere else. The agreement clitic attached with the verb, the word preceding it or may be on the verb as an affix. The pronominal agreement clitic have the same shape as personal pronouns, except that the third person is marked by -e, -kin and -ko respectively in the singular, dual and plural; when an animate noun stands as the subject NP, it agrees with the verb by its clitic form.

The marking of pronominal object plays an important role in Ho. The animate objects are marked in the verb in the form of infixed pronominal clitic. Direct object marked by pronominal clitic as an affix; and the indirect object, attached after the finite marker.

The Ho language developed a strict distribution of position to distinguish between Subject and Object suffixes, and in this way manages to use the suffixes for both Subject and Object. The syntactic function and position of these person affixes is not constant throughout the North-Munda languages. Basically, the function and position of these person affixes differs strongly throughout this family. As far as the function is concerned, the suffixes either mark for Subject or for Object (or for both). As far the position, the suffixes either occur as suffixes to the main predicate or to the immediately preverbal word. Lexical case marker may or may not block the agreement as it is shown in example (39) and (42) above; it is not constant throughout the Ho language.

Abbreviations:

1= first person  FM = finite marker  PP = post positions
2 = second person  GEN = genitive  PROG = progressive
3= third person  INTR = intransitive  PRS = present tense
ACC= accusative  NEG = negative  PST = past tense
DAT = dative  PL = plural  SG = singular
EMP = emphatic  POSS = possessive  TR = transitive

References


