

# Mon-Khmer Studies

VOLUME 43

The journal of Austroasiatic  
languages and cultures  
1964—2014 50 years of MKS

Copyright vested with the authors  
Released under Creative Commons Attribution License

Volume 43 Editors:  
Paul Sidwell  
Brian Migliazza

ISSN: 0147-5207  
Website: <http://mksjournal.org>

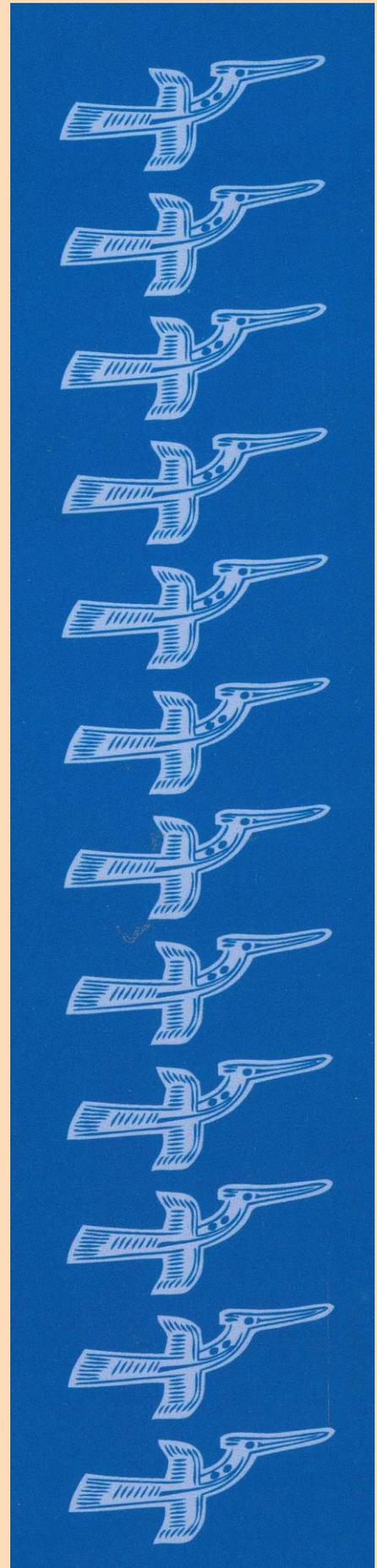
Published by:



Mahidol University (Thailand)



SIL International (USA)



## Contents

### Issue 43.1

Editor's Preface	iii
Michel FERLUS <i>Arem, a Vietic Language.</i>	1-15
Hiram RING <i>Nominalization in Pnar.</i>	16-23
Elizabeth HALL <i>Impact of Tai Lue on Muak Sa-aak phonology.</i>	24-30
Rujiwan LAOPHAIROJ <i>Conceptual metaphors of Vietnamese taste terms.</i>	31-46
Paul SIDWELL <i>Khmuic classification and homeland.</i>	47-56
Mathias JENNY <i>Transitivity and affectedness in Mon.</i>	57-71
J. MAYURI, Karumuri .V. SUBBARAO, Martin EVERAERT and G. Uma Maheshwar RAO <i>Some syntactic aspects of lexical anaphors in select Munda Languages.</i>	72-83
Stephen SELF <i>Another look at serial verb constructions in Khmer.</i>	84-102
V. R. RAJASINGH <i>Interrogation in Muöt.</i>	103-123

### Issue 43.2

Suwilai PREMSRIRAT, Kenneth GREGERSON <i>Fifty Years of Mon-Khmer Studies</i>	i-iv
Anh-Thư T. NGUYỄN <i>Acoustic correlates of rhythmic structure of Vietnamese narrative speech.</i>	1-7
P. K. Choudhary <i>Agreement in Ho</i>	8-16

## Editors' Preface

The 5<sup>th</sup> 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 19<sup>th</sup> 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 Laophaioj** “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 [-ɔmn-] 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, Bồ-trạch district, Tân-trạch commune, Cà Ròng hamlet. Though known of neighboring populations, they were confused with Bru khùa (Katuić) 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 of Arem, about a quarter, regularly speak their language. The Arem language is in great danger of extinction.

Ethnonym: the term *Arem* (Vietnamese spelling of **are:m**) seems to be an exonym. Those concerned prefer to be named *cmbrâu* [**cmraw**?]. At the end of the 19th century, Captain Rivière reported more than one hundred *Harème* families, sometimes corrupted as *Hémère*, on the Lao side (*Mission Pavie, Géographie et Voyages*, 1902; voc. pp. 285-90). *Harème* 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 *Harème* should be classified in the Maleng subgroup.

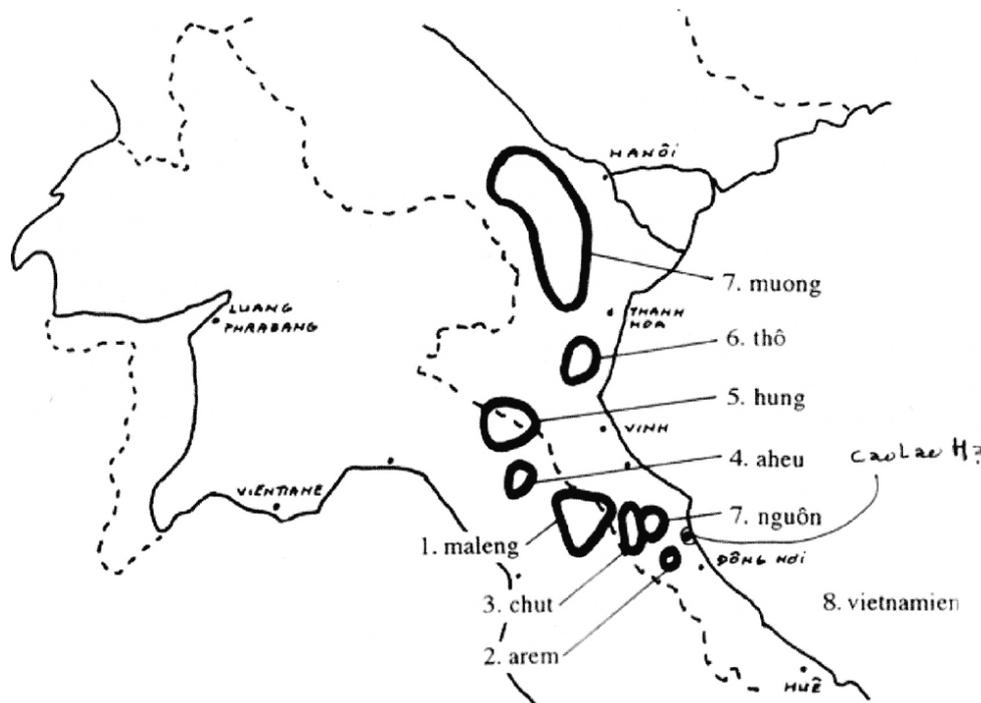
	<i>Harème</i>	Maleng brô	Arem		<i>Harème</i>	Maleng brô	Arem
'one'	<i>moitte</i>	<i>mo:c</i>	<i>mù:c</i>	'six'	<i>prao</i>	<i>paraw</i>	<i>paraw</i> <sup>?</sup>
'two'	<i>harr</i>	<i>ha:r</i>	<i>hæ:l</i>	'seven'	<i>paé</i>	<i>paj</i>	<i>pɰh</i>
'three'	<i>pà</i>	<i>pa:</i>	<i>pæ:</i>	'eight'	<i>xâme</i>	<i>sa:m</i>	<i>t<sup>h</sup>æ:m</i> <sup>?</sup>
'four'	<i>pône</i>	<i>po:n</i>	<i>puən</i> <sup>?</sup>	'nine'	<i>gîne</i>	<i>ci:n</i>	<i>ci:n</i> <sup>?</sup>
'five'	<i>damm</i>	<i>dâm</i>	<i>dam</i>	'ten'	<i>meû il</i>	<i>mə:j</i>	<i>"cùk</i>

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 *cmbrâu* [**cmraw**?], 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 Dõi 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 Dõi (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 Vạn, Chu Thái Sơn, Lưu Hùng (1986), and Trần Trí Dõi (1995, 1999). We chose for our description, the speech of Arem which seemed the most conservative regarding the phonation type register.



**Figure 1:** Map of Vietic 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.

Vietnamese	monosyllabic
Muong/mường	monosyllabic
Thổ	monosyllabic
Toum-Liha	monosyllabic
Pong	10% of sesqui-syllabic
Thavung	35-40%
Maleng Brô	35-40%
Sách-Rục	35-40%
Arem	55-60%



### 2.3 Initial consonants

#### Simple initials

$p^h$	$t^h$			$k^h$	
$p$	$t$	$tʃ$	$c$	$k$	$ʔ$
$b$	$d$		$ʃ$		
$m$	$n$		$ɲ$	$ŋ$	
	$s$	$ʃ$			$h$
$v$	$(z)$	$(ʒ/ʒ)$	$j$		
	$l$	$r$			

#### Clusters with r / l

$pr$	$tr$		$cr$	$kr$
$br$				
$pl$	$tl$			$kl$
	$t^hl$			

#### Pre-nasalized initials

	$^nth$			$^nk$	$^nh$
$^mp$	$^nt$	$^ntʃ$	$^nc$	$^nk$	$^nʔ$
$^mb$	$^nd$		$^nʃ$		
$^nv$	$^nl$	$^nr/^ndʒ$	$^nj$		

### 2.4 Finals

				$h$
$p$	$t$		$c$	$k$
$m(?)$	$n(?)$		$ɲ(?)$	$ŋ(?)$
$w(?)$		$l(?)$	$j(?)$	$ʔ(?)$

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

### 2.5 Pre-syllables

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

$a-$   $i-$   $u-$

$pa-$  ( $pi-$ ) ( $ba-$ ) ( $ma-$  /  $mu-$ )

$ta-$  ( $tam$ ) N- (first segment of pre-nasalized)

$ci-$  ( $cm$ )

$ka-$   $ku-$  ( $kə-$ )

$la-$  ( $le-$ ) ( $ra-$  /  $rə-$ )

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

<i>breathy vowels</i>					<i>clear vowels</i>						
<i>ì</i>	<i>ì:</i>	<i>î</i>	<i>î:</i>	<i>ù</i>	<i>ù:</i>	<i>i</i>	<i>(i:)</i>	<i>í</i>	<i>í:</i>	<i>u</i>	<i>u:</i>
						<i>ɪ</i>	<i>ɪ:</i>	<b>ɨ</b>	<b>ɨ:</b>	<i>ʊ</i>	<i>ʊ:</i>
	<i>ie</i>		<i>iẽ</i>		<i>ùo</i>	<i>ie</i>		<i>iẽ</i>		<i>uo</i>	
<i>è</i>	<i>(è:)</i>	<i>(ə̃)</i>	<i>ə̃:</i>	<i>ò</i>	<i>ò:</i>	<i>(e)</i>	<i>e:</i>	<i>ə</i>	<i>ə:</i>	<i>o</i>	<i>o:</i>
		<i>è̃</i>	<i>è̃:</i>			<i>ɛ</i>	<i>ɛ:</i>	<b>(ɐ)</b>	<b>ɐ:</b>	<i>ɔ</i>	<i>ɔ:</i>
						<i>æ</i>	<i>æ:</i>	<b>ʌ</b>	<b>ʌ:</b>	<i>a</i>	<i>a:</i>
								<i>a</i>	<i>a:</i>		

Units placed between brackets are poorly represented. The vowel **æ:** is realized [ɛa]. The difference between **(ɐ)/** and **ʌ/ʌ:** has no etymological justification. The series of high vowels /**i i: í í:** **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 /**ɪ ɪ: ɨ ɨ:** **ʊ ʊ:**/.

### 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 **a:**

PVM	Arem	Viet
		<i>a - ủa/ớ</i>

#### *primary (regular) correspondences*

<i>r.ka:</i>	<i>lakæ:</i>		<i>gà</i>	‘poulet / chicken’
<i>s.la:?</i>	<i>ulæ:?</i>		<i>lá</i>	‘feuille / leaf’
<i>ha:r</i>	<i>hæ:l</i>		<i>hai</i>	‘deux / two’
<i>sa:m?</i>	<i>t<sup>h</sup>æ:m?</i>		<i>tám</i>	‘huit / eighth’
<i>də:k</i>	<i>dæ:k</i>		<i>nước</i>	‘eau / water’
<i>k.ma:</i>		<i>mìə</i>	<i>mưa</i>	‘pluie / rain’
<i>ʃa:l</i>		<i>ciəl</i>	<i>chài</i>	‘filet-épervier / cast net’
<i>la:s</i>		<i>liəh</i>	<i>lưỡi</i>	‘langue / tongue’
<i>tʃa:ŋ/ʃ.ʔa:ŋ</i>		<i>siəŋ</i>	<i>xương</i>	‘os / bone’

#### *secondary correspondences*

<i>la:c</i>		<i>lè:t</i>	<i>lạt</i>	‘lien / bond’
<i>(la:ŋ)</i>		<i>lè:ŋ</i>	<i>làng</i>	‘village / village’
<i>gan?</i>		<i>kè:n?</i>	<i>cạn</i>	‘peu profond / shallow’
<i>kr.ha:k</i>		<i>kah̃:k</i>	<i>khác</i>	‘cracher / to spit’

*PVM a:* (before palatals -c -ɲ -j)

PVM	Arem		Viet	
<i>a:</i>	<i>a:</i>	<i>è:</i>	<i>a - ươ</i>	
<i>t.ka:c</i>	<i>taka:c</i>		<i>cát</i>	‘sable / sand’
<i>t.sa:c</i>	<i>at<sup>h</sup>a:c</i>		<i>tát</i>	‘écoper (de l'eau) / to bail out’
<i>ta:ɲ</i>	<i>ta:ɲ</i>		<i>đan</i>	‘tisser / to weave’
<i>ma:ɲ<sup>?</sup></i>	<i>ma:ɲ<sup>?</sup></i>		<i>mượn</i>	‘emprunter / to borrow’
<i>C.la:j<sup>?</sup></i>	<i>la:j<sup>?</sup></i>		<i>lưới</i>	‘filet de pêche / fishing net’
<i>s.ɲa:j<sup>?</sup></i>	<i>ɲa:j<sup>?</sup></i>		<i>ngái</i>	‘loin de / far from’
<i>sa:j</i>	<i>t<sup>h</sup>a:j</i>		<i>tai</i>	‘oreille / ear’
<i>la:j<sup>?</sup></i>		<i>lè:j<sup>?</sup></i>	<i>lại</i>	‘revenir / to return’

*PVM a:* (before -w)

PVM	Arem		Viet	
<i>a:</i>	<i>ɛ:</i>	<i>è:</i>	<i>a</i>	
<i>ca:w<sup>?</sup></i>	<i>ce:w<sup>?</sup></i>		<i>cháo</i>	‘brouet de riz / rice gruel’
<i>k.ta:w</i>	<i>te:w</i>		<i>dao</i>	‘sabre / sword’
<i>?a:w<sup>?</sup></i>	<i>?e:w<sup>?</sup></i>		<i>áo</i>	‘veste / upper garment’
<i>(ga:w<sup>?</sup>)</i>		<i>kè:w<sup>?</sup></i>	<i>cạo</i>	‘se raser / to shave’
<i>(la:w)</i>		<i>lè:w</i>	<i>lào</i>	‘Lao (peuple) / Lao (people)’
<i>(C.la:w)</i>		<i>tè:w</i>	<i>trào</i>	‘déborder / to overflow’

**3.2. PVM i:**

PVM	Arem		Viet	
<i>i:</i>	<i>i: / i:</i>	<i>ì:</i>	<i>ay - i</i>	
<i>si:</i>	<i>t<sup>h</sup>i:</i>		<i>tay</i>	‘main / hand’
<i>p.ri:</i>	<i>pari:</i>		<i>say</i>	‘ivre / be drunk’
<i>ci:m</i>	<i>ci:m</i>		<i>chim</i>	‘oiseau / bird’
<i><sup>n</sup>si:t</i>		<i>t<sup>h</sup>i:t</i>	<i>thít</i>	‘viande / flesh’
<i>di: / ti:</i>		<i>tì:</i>	<i>đi</i>	‘aller / to go’
<i>ʰi:<sup>?</sup></i>		<i>ci:<sup>?</sup></i>	<i>chí</i>	‘sœur aînée / elder sister’
<i>ci:n<sup>?</sup></i>	<i>ci:n<sup>?</sup></i>		<i>chín</i>	‘neuf (9) / nine’
<i>ci:n<sup>?</sup></i>	<i>ci:n<sup>?</sup></i>		<i>chín</i>	‘cooked, ripe / cuit, mûr’

**3.3. PVM e:**

PVM	Arem		Viet	
<i>e:</i>	<i>i: / iə</i>	<i>iə</i>	<i>ê - ay</i>	
<i>de:p</i>	<i>di:p</i>		<i>nếp</i>	‘(riz) glutineux / sticky (rice)’
<i>ʰf.re:</i>	<i><sup>n</sup>ri:</i>		<i>chày</i>	‘pilon / pestle’
<i>k.cet</i>	<i>ci:t</i>		<i>chết</i>	‘mourir / to dead’
<i>kp.cet</i>	<i>ka:ci:t</i>		<i>giết</i>	‘tuer / to kill’
<i>-se:l</i>	<i>at<sup>h</sup>iəl</i>		<i>tê (tê)</i>	‘écaille, pangolin / scale, anteater’

In the tense series, we see that PVM **e:** is treated as **i:**

3.4. PVM *ɛ:*

PVM	Arem		Viet	
<i>ɛ:</i>	<i>iə</i>	<i>iə</i>	<i>e</i>	
<i>t.kɛ:t</i>	<i>kiət</i>		<i>ghét</i>	‘détester, hair / to hate’
<i>dɛ:n</i>		<i>tiən</i>	<i>đen</i>	‘chandelle / candle’

Data on PVM *ɛ:* are insufficient.

3.5. PVM *u:*

PVM	Arem		Viet	
<i>u:</i>	<i>aw / u:</i>	<i>ù:</i>	<i>au/u - ô/âu</i>	
<i>p.ru:?</i>	<i>paraw?</i>		<i>sáu</i>	‘six / six’
<i>cu:?</i>	<i>caw?</i>		<i>cháu</i>	‘petit-fils / grand son’
<i>t.mu:?</i>	<i>maw?</i>		<i>máu</i>	‘résine, sang /
<i>t.pu:j</i>	<i>pɔ:j</i>		<i>vui</i>	‘joyeux / happy’
<i>C.hu:l</i>	<i>pahɔ:l</i>		<i>hôi</i>	‘sueur / sweat’
<i>k.pu:r</i>	<i>apɔ:l</i>		<i>vôi</i>	‘chaux / lime’
<i>p.tu:c</i>	<i>utu:t</i>		<i>đốt</i>	‘nœud / node’
<i>tu:c</i>	<i>tɔ:c</i>		<i>đốt</i>	‘allumer (feu) / to lit (fire)’
<i>C.bu:c</i>	<i>cibɔ:c</i>		<i>mút</i>	‘sucrer / to suck’
<i>t.lu:m?</i>		<i>lù:m?</i>	<i>trộm</i>	‘voler / to steal’
<i>C.ru:</i>		<i>cirù:</i>	<i>sâu</i>	‘profond / deep’
<i>bu:?</i> / <i>pu:?</i>		<i>pù:?</i>	<i>bú</i>	‘téter (au sein) / to suck (breast)’
<i>mu:c/mo:c</i>		<i>mù:c</i>	<i>một</i>	‘un / one’
<i>gu:r?</i> / <i>ku:r?</i>		<i>kù:l?</i>	<i>cúi</i>	‘porc / pig’
<i>C.gu:r?</i> / <i>C.ku:r?</i>		<i>cikù:?</i>	<i>gấu</i>	‘ours / bear’

3.6. PVM *o:*

PVM	Arem		Viet	
<i>o:</i>	<i>uə</i>	<i>ùə</i>	<i>ô</i>	
<i>po:n?</i>	<i>puən?</i>		<i>bốn</i>	‘quatre / four’
<i>k.lo:n?</i>	<i>tluən?</i>		<i>trốn</i>	‘se cacher / to hide’
<i>k.mo:l?</i>	<i>umùəl?</i>		<i>mối</i>	‘termite / white ant’
<i>po:ŋ</i>		<i>ṃpùəŋ</i>	<i>bông</i>	‘fleur, floraison / blossom’
<i>C.ko:l?</i>	<i>akuəl?</i>		<i>gối</i>	‘oreiller, appui-tête / pillow’
<i>k.no:l?</i>	<i>tanuəl?</i>		<i>gối</i>	‘oreiller, appui-tête / pillow’

In these two examples PVM *u:* is treated as *o:*

<i>u:</i>	<i>uə</i>	<i>ùə</i>	<i>ô</i>	
<i>t.ku:l?</i>	<i>lakuəl?</i>		<i>gối</i>	‘genou / knee’
<i>ŋu:j</i>		<i>ŋùəj</i>	<i>ngôi</i>	‘s’asseoir / to sit down’

## 3.7. PVM ɔ:

PVM	Arem		Viet	
	<i>a:</i>	<i>ùə</i>	<i>o - ua/uô</i>	
<i>ʔa.cə:ʔ</i>	<i>aca:ʔ</i>		<i>chó</i>	‘chien / dog’
<i>ʔa.lə:ʔ</i>	<i>ala:ʔ</i>		<i>lúa</i>	‘paddy / unhusked rice’
<i>k.ləjʔ</i>	<i>tlajʔ</i>		<i>trói</i>	‘ligoter / to tie up’
<i>kən</i>	<i>ka:n</i>		<i>con</i>	‘fils, fille / son, daughter’
<i>t.kə:c</i>	<i>ka:c</i>		<i>gót</i>	‘tailler, ciseler / to carve, chisel’
<i>p.rək</i>	<i>para:k</i>		<i>sóc</i>	‘écureuil / squirrel’
<i>m.rəj</i>		<i>urùəj</i>	<i>ruôj</i>	‘mouche / a fly’
<i>C.pət</i>		<i>apùət</i>	<i>vuót</i>	‘lisser / to smooth’
<i>rə:c</i>		<i>rùəc</i>	<i>ruôt</i>	‘intestins / intestines’
<i>rəŋʔ</i>		<i>rùəŋʔ</i>	<i>ruông</i>	‘rizière / ricefield’

## 3.8. PVM a

PVM	Arem		Viet	
	<i>a</i>	<i>è</i>	<i>ă / a(y)</i>	
<i>k.carʔ</i>	<i>calʔ</i>		<i>cháy</i>	‘incendie / fire’
<i>C.pat</i>	<i>pat</i>		<i>vất</i>	‘presser / to squeeze’
<i>k.lam</i>	<i>tlam</i>		<i>trăm</i>	‘cent / hundred’
<i>s.ranʔ</i>	<i>ranʔ</i>		<i>săn</i>	‘manioc / cassava’
<i>t.baŋ</i>	<i>abaŋ</i>		<i>măng</i>	‘pousses (bambou) / bamboo shoots’
<i>C.kamʔ</i>		<i>kəmʔ</i>	<i>gặm</i>	‘mordre / to bite’
<i>t.ŋarʔ</i>		<i>aŋəlʔ</i>	<i>ngáy</i>	‘ronfler / to snore’
<i>mat</i>		<i>mèt</i>	<i>mắt</i>	‘œil / eye’
<i>k.məŋʔ</i>		<i>mènʔ</i>	<i>mặn</i>	‘salé / salted’

In these two examples PVM ə is treated as a

<i>a / ə</i>	<i>a</i>	<i>è</i>	
<i>p.sat/-sət</i>	<i>at<sup>h</sup>at</i>		‘éteindre / to extinguish’
<i>pər</i>	<i>pal</i>		‘voler / to fly’

## 3.9. PVM ə

PVM	Arem		Viet	
	<i>i / I / ɨ</i>	<i>ì / ì</i>	<i>ã / â</i>	
<i>p.səŋʔ</i>	<i>ut<sup>h</sup>iŋʔ</i>		<i>rắn</i>	‘serpent / snake’
<i>pəŋʔ</i>	<i>pìŋʔ</i>		<i>bắn</i>	‘tirer (fusil, arbalète) / to shoot’
<i>pəc</i>	<i>pìc</i>		<i>bắt, bắt</i>	‘cueillir / to pick, pluck’
<i>təŋʔ</i>	<i>latəŋʔ</i>		<i>đứng</i>	‘cloison / partition’
<i>bət</i>	<i>bət</i>		<i>mất</i>	‘perdre (objet) / to lose (object)’
<i>m.rəŋʔ</i>		<i>riŋʔ</i>	<i>rân</i>	‘pou / louse’
<i>məc</i>		<i>mít</i>	<i>mật</i>	‘fiel, bile / gall’
<i>t.kəŋ</i>		<i>kìŋ</i>	<i>gần</i>	‘près de / near of’
<i>k.vət</i>		<i>vít</i>	<i>quát</i>	‘fouetter / to whip’
<i>k.rəmʔ</i>		<i>karìmʔ</i>	<i>sấm</i>	‘tonnerre / thunder’
<i>jərʔ</i>		<i>jìlʔ</i>	<i>dậy</i>	‘se lever / to get up’

### 3.10. PVM u

PVM	Arem		Viet	
<i>u</i>	<i>u</i>		<i>o / ô</i>	
<i>-suk</i>	<i>ut<sup>h</sup>ok</i>		<i>tóc</i>	‘cheveux / hair’
<i>s.ʔuh</i>	<i>nuh</i>		<i>ô, tô</i>	‘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 *a<sup>2</sup>-â<sup>2</sup>* in Arem. If these correspondences were regular, Arem words should have the final **-h**.

PVM	Arem		Viet	
<i>-h</i>	tone <i>a<sup>2</sup></i>	tone <i>â<sup>2</sup></i>	<i>hỏi-ngã</i>	
<i>bəs</i>	<i>baj<sup>2</sup></i>		<i>bây</i>	‘piège, tonilière / a trap’
<i>ka:s</i>	<i>kə:l<sup>2</sup></i>		<i>cái</i>	‘moutarde / mustard’
...	<i>t<sup>h</sup>aŋ<sup>2</sup></i>		<i>thẳng</i>	‘droit / straight’
...	<i>ròŋ<sup>2</sup></i>		<i>rông</i>	‘vide / empty’
<i>kuh</i>	<i>ku<sup>2</sup></i>		<i>củ</i>	‘tubercule / tuber’
<i>foh</i>	<i>fo:<sup>2</sup></i>		<i>nhô</i>	‘arracher, désherber / to weed’
<i>k.cəh</i>	<i>kace:<sup>2</sup></i>		<i>giẻ</i>	‘torchon / duster’
<i>k.pa:s</i>	<i>kupə:l<sup>2</sup></i>		<i>vải</i>	‘étouffe / cloth’
<i>cis</i>	<i>cin<sup>2</sup></i>		<i>chí</i>	‘fil / thread’
<i>k.cih</i>		<i>cì:<sup>2</sup></i>	<i>giữ</i>	‘surveiller / to watch’
<i>k.cus</i>		<i>cù:<sup>2</sup></i>	<i>chổi</i>	‘balai / broom’

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 *sắc-nặng* according to two modalities (see tables below). In type A, the expected tone *hỏi* merged in the tone *sắc*, while the expected tone *ngã* merged in tone *nặng*. 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 *nặng*.

type A		type B	
<i>ngang</i>	<i>sắc (hỏi)</i>	<i>ngang</i>	<i>sắc (hỏi)</i>
<i>huyền</i>	<i>nặng (ngã)</i>	<i>huyền</i>	<i>nặng (ngã)</i>

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 Hạ (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 **\*/s/** 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 sesqui-syllable (see table 5.1.). Remember that in the 17th century, *v* was *ɸ / ɸě*, and *d* was *d / ɗ* (de Rhodes 1651, 1991).

- Pre-glottalized \*/**ʔ** **ɗ** **f**/
- Unproblematic units \*/**ʔ** **h** **m** **n** **ɲ** **v** **j** **r** **l**/
- Clusters with **-r-**
- Clusters with **-l-**

### 5.1. Table of the evolutions of PVM \*/**p-b t-d c-ʃ k-g**/ and \*/**s**/

PVM	Arem	Vietnamese		
		*monosyll.	*sesqui-syllable	
			modern	1651
<i>p - b</i>	<i>p</i>	<i>b</i> [b]	<i>v</i> [v]	< <i>ɸ/ɸě</i> [β]
<i>t - d</i>	<i>t</i>	<i>đ</i> [d]	<i>d</i> [z/j]	< <i>d/dě</i> [ð]
<i>c - ʃ</i>	<i>c</i>	<i>ch</i> [c]	<i>gi</i> [z/j]	< <i>gi</i> [j]
<i>k - g</i>	<i>k</i>	<i>c/k</i> [k]	<i>g/gh</i> [ɣ]	< <i>g/gh</i> [ɣ]
<i>s</i>	<i>t<sup>h</sup></i>	<i>t</i> [t]	<i>r</i> [r/z/j]	< <i>r</i> [r]

#### Labials **p-b**

PVM	Arem	Vietnamese		
<i>p-b</i>	<i>p</i>	<i>b</i>	<i>v</i>	
<i>pa:</i>	<i>pæ:</i>	<i>ba</i>		‘trois / three’
<i>pəp<sup>?</sup></i>	<i>pjɲ<sup>?</sup></i>	<i>bắn</i>		‘tirer (fusil, arbalète) / to shoot’
<i>po:n<sup>?</sup></i>	<i>puən<sup>?</sup></i>	<i>bốn</i>		‘quatre / four’
<i>k.pa:s</i>	<i>kupɛ:l<sup>?</sup></i>		<i>vải</i>	‘éttoffe / fabric’
<i>k.pu:r</i>	<i>apɯ:l</i>		<i>vôi</i>	‘chaux / lime’
<i>k.pa:<sup>?</sup></i>	<i>pæ:<sup>?</sup></i>		<i>vá</i>	‘rapiécer, réparer / to mend, repair’

#### Alveolars **t-d**

<i>t-d</i>	<i>t</i>	<i>đ</i>	<i>d</i>	
<i>tɯ:c</i>	<i>tɯ:c</i>	<i>đốt</i>		‘allumer / to light’
<i>ta:ɲ</i>	<i>ta:ɲ</i>	<i>đan</i>		‘tisser / to weave’
<i>l.ta:<sup>?</sup></i>	<i>atæ:<sup>?</sup></i>	<i>đá</i>		‘pierre / stone’
<i>k.ta:l<sup>?</sup></i>	<i>atæ:l<sup>?</sup></i>		<i>dái</i>	‘scrotum / scrotum’
<i>təɲ<sup>?</sup></i>	<i>latəɲ<sup>?</sup></i>		<i>đứng</i>	‘cloison / partition’
<i>k.ta:w</i>	<i>tɛ:w</i>		<i>dao</i>	‘sabre, couteau / sword, knife’

#### Palatals **c-ʃ**

<i>c-ʃ</i>	<i>c</i>	<i>ch</i>	<i>gi</i>	
<i>-ci:m</i>	<i>ici:m</i>	<i>chim</i>		‘oiseau / bird’
<i>k.car<sup>?</sup></i>	<i>cal<sup>?</sup></i>	<i>cháy</i>		‘incendie / fire’
<i>ʃa:l</i>	<i>ciəl</i>	<i>chài</i>		‘filet-épervier / cast net’
<i>k.ʃə:ɲ</i>	<i>kacɛ:ɲ</i>		<i>giường</i>	‘lit / bed’
<i>C.ca:m</i>	<i>cæ:m</i>		<i>giàm</i>	‘emprisonner / to imprison’
<i>k.caj<sup>?</sup></i>	<i>kacaj<sup>?</sup></i>		<i>giấy</i>	‘papier / paper’

Velars **k-g**

<i>k-g</i>	<i>k</i>	<i>c / k</i>	<i>g / gh</i>	
<i>t.kɑ:c</i>	<i>takɑ:c</i>	<i>cát</i>		‘sable / sand’
<i>kɑ:n</i>	<i>kɑ:n</i>	<i>con</i>		‘fils, fille / son, daughter’
<i>ke:ʔ</i>	<i>ke:ʔ</i>	<i>cái</i>		‘féminin / féminin’
<i>r.kɑ:</i>	<i>lakæ:</i>		<i>gà</i>	‘poulet / chicken’
<i>c.kɑ:ŋ</i>	<i>cikæ:ŋ</i>		<i>gang</i>	‘empan / handspan’
<i>c.gi:ʔ</i>	<i>cikèjʔ</i>		<i>gây</i>	‘bâton, canne / walking stick’

Fricative alveolar **s**

<i>s</i>	<i>t<sup>h</sup></i>	<i>t</i>	<i>r</i>	
<i>sɑ:mʔ</i>	<i>t<sup>h</sup>æ:mʔ</i>	<i>tám</i>		‘huit / eight’
<i>sa:j</i>	<i>t<sup>h</sup>a:j</i>	<i>tai</i>		‘oreille / hear’
<i>si:</i>	<i>t<sup>h</sup>i:</i>	<i>tay</i>		‘main / hand’
<i>p.səŋʔ</i>	<i>ut<sup>h</sup>ŋʔ</i>		<i>răn</i>	‘serpent / snake’
<i>p.se:ʔ</i>	<i>ut<sup>h</sup>e:ʔ</i>		<i>rái</i>	‘loutre / otter’
<i>k.saŋ</i>	<i>at<sup>h</sup>aŋ</i>		<i>răng</i>	‘dent / teeth’

5.2. Pre-glottalized **ʈ ɖ ʃ**

PVM	Arem	Viet	
<i>ʈ</i>	<i>b</i>	<i>m</i>	
<i>c.ɓu:c</i>	<i>cibuc</i>	<i>mút</i>	‘sucrer (sucre) / to suck (sugar)’
<i>ɓah</i>	<i>abah</i>	<i>mũa</i>	‘vomir / to vomit’
<i>ɓɑ:jʔ</i>	<i>ba:jʔ</i>	<i>muối</i>	‘sel / salt’
<i>ɓu:k</i>	<i>buk</i>	<i>múc</i>	‘puiser / to draw (water)’
<i>t.ɓaŋ</i>	<i>abaŋ</i>	<i>mãng</i>	‘pousses (bambou) / bamboo shoot’

<i>ɖ</i>	<i>d</i>	<i>n</i>	
<i>ɖɑ:k</i>	<i>dæ:k</i>	<i>nước</i>	‘eau / water’
<i>ɖam</i>	<i>dam</i>	<i>năm</i>	‘cinq / five’
<i>ɖi:p</i>	<i>di:p</i>	<i>nếp</i>	‘(riz) glutineux / sticky (rice)’
<i>k.ɖe:</i>	<i>kade:</i>	<i>nai</i>	‘cerf sambar / stag’
<i>ɖɑ:</i>	<i>da:</i>	<i>no</i>	‘rassasié / satiated’

<i>f</i>	<i>t</i>	<i>nh</i>	
<i>fuk</i>	<i>tuk</i>	<i>nhúc</i>	‘arbre à amadou / touchwood’
<i>faw</i>	<i>taw</i>	<i>nhau</i>	‘placenta / placenta, afterbirth’
<i>foh</i>	<i>to:ʔ</i>	<i>nhỏ</i>	‘arracher / to snatch’

### 5.3. Evolutions of others initials

PVM	Arem	Viet
ʔ	ʔ	# [ʔ]
h	h	h [h]
m	m	m [m]
n	n	n [n]
ɲ	ɲ	nh [ɲ]
ŋ	ŋ	ng(h) [ŋ]
v	v	v [v]
ʃ	ʃ	d [z/j]
r	r	r [r/z/j]
l	l	l [l]

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

### 5.4. Clusters with -r-

PVM	Arem	Viet	
<i>C.r-</i>	<i>Cr-/r-</i>	<i>s-</i>	
<i>g.rəjʔ</i>	<i>cirɛjʔ</i>	<i>sợi</i>	‘fil / thread’
<i>p.rək</i>	<i>para:k</i>	<i>sóc</i>	‘écureuil / squirrel’
<i>k.rəjʔ</i>	<i>kara:lʔ</i>	<i>sói</i>	‘loup / wolf’
<i>s.ranʔ</i>	<i>ranʔ</i>	<i>sắn</i>	‘manioc / cassava’
<i>(C.rəŋ)</i>	<i>kəra:ŋ</i>	<i>song</i>	‘rotin / rattan’
<i>c.ru:</i>	<i>cirù:</i>	<i>sâu</i>	‘profond / deep’
<i>p.ru:ʔ</i>	<i>parawʔ</i>	<i>sáu</i>	‘six / six’
<i>p.ri:</i>	<i>pari:</i>	<i>say</i>	‘ivre / drunk’
<i>c.ra:ŋ</i>	<i>ʳra:ŋ</i>	<i>sườn</i>	‘côtes / ribs’

### 5.5. Clusters with -l-

PVM	Arem	Viet	
<i>C.l-</i>	<i>l-/tl-/Cl-</i>	<i>tr-</i>	
...	<i>lù:mʔ</i>	<i>trộm</i>	‘voler, dérober / to steal’
<i>k.lən</i>	<i>ulan</i>	<i>trăn</i>	‘python / python’
<i>p.le:ʔ</i>	<i>ule:ʔ</i>	<i>trái</i>	‘fruit / fruit’
<i>k.la:</i>	<i>cilə</i>	<i>trưa</i>	‘midi / midday’
<i>p.ləŋ</i>	<i>ʳləŋ</i>	<i>tranh</i>	‘herbe à pailote / thatch-grass’
<i>b.ləj</i>	<i>tləj</i>	<i>trời</i>	‘ciel / sky’
<i>k.ləjʔ</i>	<i>tləjʔ</i>	<i>trói</i>	‘ligoter / to tie up’
<i>k.lam</i>	<i>tlam</i>	<i>trăm</i>	‘cent / hundred’
<i>k.lo:nʔ</i>	<i>tluənʔ</i>	<i>trốn</i>	‘s’ enfuir / to run away’

## 6. Finals consonants

### 6.1. Table of the evolution of final consonants

PVM	Arem	Viet
<i>p</i>	<i>p</i>	<i>p</i>
<i>t</i>	<i>t</i>	<i>t</i>
<i>c</i>	<i>c</i>	<i>t</i>
<i>k</i>	<i>k</i>	<i>c</i>
<i>ʔ</i>	<i>-ʔ</i>	<i>(sắc năng)</i>
<i>h</i>	<i>h</i>	<i>(hỏi ngã)</i>
<i>s</i>	<i>h</i>	<i>i / y (hỏi ngã)</i>
<i>m</i>	<i>m</i>	<i>m</i>
<i>n</i>	<i>n</i>	<i>n</i>
<i>ɲ</i>	<i>ɲ</i>	<i>n</i>
<i>ŋ</i>	<i>ŋ</i>	<i>ng(h)</i>
<i>r</i>	<i>l</i>	<i>i / y</i>
<i>l</i>	<i>l</i>	<i>i / y</i>

### 6.2. Final palatals PVM *c ɲ*

PVM	Arem	Viet	
<i>t.kə:c</i>	<i>takə:c</i>	<i>cát</i>	‘sable / sand’
<i>rə:c</i>	<i>rùə:c</i>	<i>ruôt</i>	‘intestins / intestines’
<i>c.bu:c</i>	<i>cibu:c</i>	<i>mút</i>	‘sucrer (sucre) / to suck (sugar)’
<i>p.səɲ</i>	<i>ut<sup>h</sup>ɲ<sup>ʔ</sup></i>	<i>rắn</i>	‘serpent / snake’
<i>ta:ɲ</i>	<i>ta:ɲ</i>	<i>đan</i>	‘tisser / to plait’
<i>m.rəɲ<sup>ʔ</sup></i>	<i><sup>n</sup>riɲ<sup>ʔ</sup></i>	<i>rận</i>	‘pou / louse’

### 6.3. Final fricatives PVM *h s*

PVM	Arem	Viet	
<i>t.pah</i>	<i><sup>m</sup>pah</i>	<i>vả</i>	‘giffler / to slap’
<i>cəh</i>	<i>acəh</i>	<i>chẻ</i>	‘fendre (du bois) / to split (wood)’
<i>-bah</i>	<i>abah</i>	<i>mửa</i>	‘vomir / to vomit’
<i>la:s</i>	<i>liəh</i>	<i>lưỡi</i>	‘langue / tongue’
<i>mu:s</i>	<i>mùh</i>	<i>mũi</i>	‘nez / nose’
<i>C.kas</i>	<i>akæ:h</i>	<i>gãi</i>	‘gratter (de l’ongle) / to scrape (of nail)’

### 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.

PVM	Arem	Viet	
<i>pər</i>	<i>pal</i>	<i>bay</i>	‘voler / to fly’
<i>har</i>	<i>hæ:l</i>	<i>hai</i>	‘deux / two’
<i>t.kar<sup>ʔ</sup></i>	<i>ikal<sup>ʔ</sup></i>	<i>gáy</i>	‘chanter (coq) / to crow (rooster)’
<i>ʃa:l</i>	<i>ciəl</i>	<i>chài</i>	‘filet-épervier / cast net’
<i>k.ta:l<sup>ʔ</sup></i>	<i>atæ:l<sup>ʔ</sup></i>	<i>đái</i>	‘scrotum / scrotum’
<i>k.mo:l<sup>ʔ</sup></i>	<i>umùəl<sup>ʔ</sup></i>	<i>mối</i>	‘termite / white ant’

## 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.

Notwithstanding 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

- Alves, Mark. 2006. Linguistic Research on the Origin of the Vietnamese Language: An Overview. *Journal of Vietnamese Studies* 1(1-2): 104-130.
- Barker, Muriel A. & Milton E. Barker. 1970. Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua* 24: 268-85.
- Đặng Nghiêm Vạn, Chu Thái Sơn & Lưu Hùng. 1986. *Les ethnies minoritaires du Vietnam*. Editions en langues étrangères, Hanoi. [Arem are included in the ethnic group Chứt, pp. 38-42].
- Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies* 15: 139-154.
- Enfield, N.J. & Gérard Diffloth. 2009. Phonology and sketch grammar of Kri, a Vietic language of Laos. *Cahiers de Linguistique – Asie Orientale* 38(1): 3-69.
- Ferlus, Michel. 1975. Vietnamien et proto viet-muong", *ASEMI* 6(4): 21-55.
- Ferlus, Michel. 1979. Lexique thavung-français, *Cahiers de Linguistique, Asie Orientale* 5: 71-94.
- Ferlus, Michel. 1982. Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien. *Cahiers de Linguistique – Asie Orientale* 11(1): 83-106. [traduit en vietnamien: Sự biến hóa của các âm tắc giữa (obstruantes médiales) trong tiếng việt, *Ngôn ngữ* 1981(2): 1-22.].
- Ferlus, Michel. 1992. Histoire abrégée de l'évolution des consonnes initiales du Vietnamien et du Sino-Vietnamien. *MKS* XX: 111-125.
- Ferlus, Michel. 1996. Langues et peuples viet-muong", *Mon-Khmer Studies* 26 (Special Volume Dedicated to Dr. David D. Thomas on The Occasion of his Sixty-Fifth Birthday): 7-28.
- Ferlus, Michel. 1997. Problèmes de la formation du système vocalique du vietnamien. *Cahiers de linguistique – Asie Orientale* 26(1): 37-51.
- Ferlus, Michel. 1997. Le maleng brô et le vietnamien", *Mon-Khmer Studies* 27 (Hommages à Paul K. Benedict): 55-66.
- Ferlus, Michel. 1998. Les systèmes de tons dans les langues Viet-Muong. *Diachronica* XV: 11-27.
- Ferlus, Michel. 1999. Les disharmonies tonales en Viet-Muong et leurs implications historiques. *Cahiers de linguistique – Asie Orientale* 28(1): 83-100.
- Ferlus, Michel. 2004. The Origin of Tones in Viet-Muong. *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistic Society 2001*, Edited by Somsong Burusphat, Arizona State University: 297-313.
- Gage, William. 1985. Vietnamese in Mon-Khmer perspective. In: *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*. Edited by Surya Ratanakul, David D. Thomas & Suwilai Premsirat: 493-524. Bangkok: Mahidol University.
- Gregerson, Kenneth J. 1969. A Study of Middle Vietnamese Phonology. *Bulletin de la Société des Etudes Indochinoises* 44(2): 131-193.
- Haudricourt, André-Georges. 1953. La place du vietnamien dans les langues austroasiatiques. *Bulletin de la Société de Linguistique de Paris* 49(1): 122-8.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242: 69-82.
- Kasuga, Atsushi. 1994. *A Study on Arem Phonology*. Master's thesis submitted to Tokyo University.

- Kasuga, Atsushi. 2008. *Arem Vocabulary*. Manuscript, organized in order Vietnamese-Arem-English.
- Maspero, Henri. 1912. Etudes sur la phonétique historique de la langue annamite: Les initiales. *Bulletin de l'Ecole Française d'Extrême-Orient* 12(1): 1-127.
- Nguyễn Tài Cẩn. 1995. *Giáo trình lịch sử ngữ âm tiếng Việt* [Course in historical phonetics of the Vietnamese language]. Hà Nội: Nhà xuất giáo.
- Nguyễn Văn Khang, Bùi Chí & Hoàng Văn Hành. 2002. *Từ Điển Mường-Việt* [Mường-Việt Dictionary]. Hà Nội, Nhà xuất bản văn hóa dân tộc.
- Nguyễn Văn Lợi. 1993. *Tiếng Rục* [The Rục language]. Hà Nội, Nhà xuất bản khoa học xã hội.
- Rhodes, (Père) Alexandre de. 1651. *Dictionarium annamiticum, lusitanum, et latinum*. Rome. Reprinted with a translation into Modern Vietnamese: Viện Khoa Học Xã Hội tại T.P. Hồ Chí Minh, *Từ Điển Annam-Lusitan-Latinh*, 1991, Nhà xuất bản khoa học xã hội.
- Suwilai Premrirat. 2000. *So (Thavung) preliminary Dictionary*. Institute of Language and Culture for Rural Development, Mahidol University, Thailand, Department of Linguistics and Applied Linguistics, University of Melbourne, Australia.
- Taylor, Kate W. 2001. On being Muonged. *Asian Ethnicity* 2(1): 24-34.
- Thompson, Laurence C. 1973. Proto-Viet-Muong Phonology. *Austroasiatic Studies*, Oceanic Linguistics Special Publication n° 13, Edited by Philip N. Jenner, Laurence C. Thompson and Stanly Starosta, part II: 1113-1203.
- Trần Trí Dõi. 1995. *Thực trạng kinh tế và văn hóa của ba nhóm tộc 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]
- Trần Trí Dõi. 1999. *Nghiên cứu ngôn ngữ các dân tộc thiểu số Việt Nam* [Study of languages of ethnic minorities in Vietnam]. Nhà xuất bản đại học quốc gia Hà-Nội. [Arem: demography p. 110; vocabulary p. 16 and 37]
- Vương Hoàng Tuyên. 1963. *Các dân tộc nguồn gốc Nam-Á ở miền bắc Việt-Nam* [Ethnic groups of Austro-Asiatic origin in North Vietnam]. Nhà xuất bản Giáo Dục, Hà-Nội. [Arem: vocabulary p. 71; Compare list of a hundred words in Viet-Muong languages, including Arem; see fold-out page V-VIII, end of the book]

# 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 /tʃ/, *j* represents /dʒ/, *ñ* the palatal nasal, and *ng* 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 /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ respectively), while word-finally it represents the glottal stop /ʔ/ (*soh* /sɔʔ/ ‘fruit’), and in all other contexts represents the glottal fricative /h/. Orthographic *y* also represents three different sounds depending on where it occurs: as the single onset of a syllable it represents the palatal approximant /j/ (*yap* /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 /ŋ/ ‘REF’ and *pyn* represents /pŋ/ ‘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

	<u>TB nominalization, V-final</u>
Derivational	[V-NMZ]NOUN [V-NMZ]ADJ
Clausal	[(NP)... V-NMZ]NP
	<u>Pnar nominalization, V-initial</u>
Derivational:	[NMZ-V]NOUN [NMZ V]ADJ
Clausal:	[NMZ V... (NP)]NP

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.<sup>1</sup>

#### 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.<sup>2</sup> Example (1) is of a verbless clause where *jing-* derives the event nominal ‘regarding’ by prefixing to the verb *ya-toh* ‘have.relationships’ (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-təʔ*      *u=ni*      *u=wəʔ*      *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*    *ʃim*    *kti*    *nəʔ*    *u*      *ja*      *ka=dʒiŋ-sɲʃar*    *sa*      *ʃi-sen*  
 IRR    take    hand    IMM    3S.M    BEN    F=NMZ-rule      once    one-each  
*ha*      *ka=kti*    *ka*      *yong*    *oo*  
*ha*      *ka=kti*    *ka*      *jəŋ*      *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=aiñ*    *wa*      *da*      *chna*    *ki=nong-pyllai*  
*ka=aŋ*      *wa*      *da*      *ʃna*      *ki=nəŋ-pʰlaj*  
 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.

<sup>1</sup> 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.

<sup>2</sup> 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.

- (4) *ka=yu-spong toh u=æm ko*  
*ka=ju-spəŋ tɔʔ u=ɛm kɔ*  
 F=NMZ-wrap right NF=have 3S.F  
 ‘the turban is necessary’ (lit. ‘the turban, is right to have it’) [TACJ\_133]

### 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=ræp u=khih*  
*bian i=pntʰɔr u=rɛp 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=pŋ-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?*  
*tɛ pʰi kʰiʔ pʰi nɔŋ*  
 NVIS 2PL work 2PL CONF  
 ‘so you, you work, right?’ [AIJ\_012]

- (8) *he-i=jooħ i=por man ko ka=khih*  
*he-i=dʒoʔ i=pɔr man kɔ 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 hæħ i=khih yong i*  
*i=ni hæʔ 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*  
 mntu da dɛp u=jap tɛ ɔŋ 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 khræh i mynstæp*  
 daw k<sup>h</sup>rɛʔ i mntɛp  
 IRR prepare 1PL tomorrow  
 ‘we will prepare tomorrow’ (intention, uncertain)

- (11b) *u=khræh i mynstæp*  
 u=k<sup>h</sup>rɛʔ i mntɛp  
 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=paɗ  
 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.<sup>3</sup> 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 worshippers’ [AIJ\_159]

<sup>3</sup> 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*.

- (13b) [ki wa mane] yong ka  
 ki wa mane jɔŋ ka  
 3PL NMZ worship GEN 3S.F  
 ‘the worshipers of her’ / ‘the worshippers belonging to her’

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=paid* ‘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).<sup>4</sup> 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*  
 m tɔʔ ka wa bha  
 NEG be 3S.F NMZ be.good  
 ‘it isn’t good’ [BMPJ\_037]
- (15) *ki=tæ ki wa lai skur*  
 ki=tɛ ki wa laj skur  
 PL=NVIS 3PL NMZ go school  
 ‘the school-goers’ [BPVM\_007]
- (16a) *tæ kam-tæ ki=lɔk wa maya*  
 tɛ kam-tɛ ki=lɔk wa maja  
 NVIS like-NVIS PL=friend NMZ love  
 ‘so in that case dear/beloved friends...’ [BPDJ\_044]
- (16b) *ha-dein wa da æm ka=kur soo kpoh...*  
 ha-den wa da ɛm ka=kur so kpɔʔ  
 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

<sup>4</sup> 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 tɔʔ 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 pɲ-man je-i=tu tɛ tɔʔ wa  
 happen 3PL NMZ CAUS-happen DAT-N=MDIST NVIS be NMZ  
*[ym tip ki u=keiñ]*  
 m tip ki u=keɲ  
 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.*  
 ɛm ar ɲut ʃi lɔk ki u=nik wa u=sɪŋ  
 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=tɛ tɔʔ u=nik tɔʔ u wa malik u wa jɔʔ  
 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 ɛm tɛ u=sɪŋ tɔʔ u wa doʔ m tɔʔ  
 3S.M NMZ have NVIS M=Singh be 3S.M NMZ be.poor NEG be  
*[u [wa yoh]]*  
 u wa jɔʔ  
 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 referents 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

- Benjamin, Geoffrey. 1976. An outline of Temiar grammar. *Oceanic Linguistics Special Publications* 129–187.
- Bickel, Balthasar. 1999. Nominalization and focus constructions in some Kiranti languages. In *Topics in Nepalese Linguistics*, ed. Yogendra P. Yadava and Warren G. Glover. Royal Nepal Academy.
- Bradley, David. 1980. Phonological convergence between languages in contact: Mon-Khmer structural borrowing in Burmese. In *Proceedings of the Sixth Annual Meeting of the Berkeley Linguistic Society*, 259–267.
- Burenhult, Niclas. 2005. *A grammar of Jahai*, volume 566. Pacific Linguistics.
- Comrie, Bernard, and Sandra A. Thompson. 1985. Lexical nominalization. In *Language Typology and Syntactic Description*, ed. Timothy Shopen, volume III: Grammatical Categories and the Lexicon, 349–398. Cambridge: Cambridge University Press.
- Costello, Nancy A., and Sulavan Khamluan. 1998. Affixes in katu of the lao pdr. *Mon-Khmer Studies* 31.
- Genetti, C. 2011. Nominalization in Tibeto-Burman languages of the Himalayan area. In *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. Foong Ha Yap, Karen Grunow-Hårsta, and Janick Wrona, volume 96, 163–193. John Benjamins.
- Gerner, Matthias. 2012. The typology of nominalization. *Language and Linguistics* 13:803–844.
- Hopple, Paulette M. 2003. The structure of nominalization in burmese. Doctoral Dissertation, University of Texas at Arlington.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge University Press.
- Matisoff, J. A. 1972. Lahu nominalization, relativization, and genitivation. In *Syntax and Semantics*, ed. John P. Kimball, 237–257. Academic Press.
- Morev, Lev N. 2006. Cognitive and structural aspects of abstract nominalization in Tai and some other isolating languages of Southeast Asia. *Mon-Khmer Studies* 36:139–146.
- Noonan, Michael. 1997. Versatile nominalizations. In *Essays on Language Function and Language Type*, ed. Joan Bybee, John Haiman, and Sandra A. Thompson, 373–394. John Benjamins.
- Parkin, Robert. 1991. A guide to austroasiatic speakers and their languages. *Oceanic Linguistics Special Publications* i–213.
- Watters, David. 2008. Nominalization in the Kiranti and Central Himalayish languages of Nepal. *Linguistics of the Tibeto-Burman Area* 31:1–4.
- Yap, Foong Ha, Karen Grunow-Hårsta, and Janick Wrona. 2011. Introduction: Nominalization strategies in Asian languages. In *Nominalization in Asian languages: Diachronic and typological perspectives*, ed. Foong Ha Yap, Karen Grunow-Hårsta, and Janick Wrona, volume 96, 1–57. John Benjamins.

# 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, p<sup>h</sup>, t, k, k<sup>h</sup>, 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, p<sup>h</sup>r, k<sup>h</sup>r, pw, kw, p<sup>h</sup>w, k<sup>h</sup>w/. Final consonants are limited to /p, t, c, k, m, n, ɲ, 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

<i>p<sup>h</sup> p b</i>	<i>t<sup>h</sup> t d</i>	<i>c<sup>h</sup> c</i>	<i>k<sup>h</sup> k</i>
<i>m</i>	<i>n</i>	<i>ɲ</i>	<i>ŋ</i>
<i>f</i>	<i>s</i>		<i>h</i>
<i>w</i>	<i>r l</i>	<i>j</i>	

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

	Front	Back unrounded	Back rounded
Close	<i>i i:</i>	<i>u u:</i>	<i>u u:</i>
Close-mid	<i>e e:</i>	<i>ɤ ɤ:</i>	<i>o o:</i>
Open	<i>ɛ</i>	<i>a a:</i>	<i>ɔ</i>
Diphthongs	<i>ia</i>	<i>ua</i>	

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)

<i>p p<sup>h</sup> b</i>	<i>t t<sup>h</sup> d</i>	<i>c c<sup>h</sup></i>	<i>k</i>	<i>ʔ</i>
<i>f</i>	<i>s</i>		<i>x</i>	<i>h</i>
<i>m</i>	<i>n</i>	<i>ɲ</i>	<i>ŋ</i>	
<i>w</i>	<i>l</i>	<i>r</i>	<i>j</i>	

The velar fricative /x/ in Tai Lue is the equivalent of the aspirated velar stop /k<sup>h</sup>/ 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, t<sup>h</sup>w, xw, tr, t<sup>h</sup>r/. 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/.

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

	Front	Back unrounded	Back rounded
High	<i>i i:</i>	<i>u u:</i>	<i>u u:</i>
Mid	<i>e e:</i>	<i>ɤ ɤ:</i>	<i>o o:</i>
Low	<i>ɛ ɛ:</i>	<i>a a:</i>	<i>ɔ ɔ:</i>

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 /ɔ:/ and /ɛ:/ 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 /ɛ:/ and /ɔ:/ 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 /ɛ:, ɔ:/ as /ia, ua/ suggests that these diphthongs developed from an older /ɛ:/ and /ɔ:/. 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
<i>pɛ:t<sup>2</sup></i>	<i>piat<sup>1</sup></i>	‘eight’
<i>sip<sup>1</sup> sɔŋ<sup>1</sup></i>	<i>sip<sup>2</sup> suan<sup>3</sup></i>	‘twelve’
<i>mɛŋ<sup>4</sup> sa:p<sup>2</sup></i>	<i>mian<sup>3</sup> sa:p<sup>1</sup></i>	‘cockroach’
<i>lɛm<sup>1</sup></i>	<i>liam<sup>1</sup></i>	‘sharp-pointed’
<i>sɔɔk<sup>2</sup></i>	<i>suak<sup>1</sup> tɔ<sup>2</sup></i>	‘elbow’
<i>hɔn<sup>6</sup></i>	<i>ruan<sup>2</sup></i>	‘hot’

The number ‘two’ forms an exception as it is not pronounced consistently. As a monosyllable the vowel /ɔ:/ is retained, while in numeral compounds it is adjusted to Muak Sa-aak /ua/, as in *sip<sup>2</sup> suan<sup>3</sup>* ‘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

Muak Sa-aak; for example, Lue *hok*<sup>2</sup> ‘six’ > Muak Sa-aak *rɔk*<sup>2</sup>. 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<sup>h</sup>r-/ or /pr-/, clusters which do not occur in Tai Lue: for example, Lue *p<sup>h</sup>ung*<sup>3</sup> ‘bee’ > Muak Sa-aak *p<sup>h</sup>rxŋ*<sup>2</sup>. In some places where /r/ is found in Muak Sa-aak, /r/ is also found in standard Thai: Muak Sa-aak *ruan*<sup>2</sup> ‘hot’ > Thai *rɔ:n*, Muak Sa-aak *rx:n*<sup>1</sup> ‘[CLSF] house’ > Thai *ruan* ‘house’, Muak Sa-aak *rx*<sup>3</sup> ‘boat’ > Thai *rua*. In other apparent loanwords containing /r/ in Muak Sa-aak, the /r/ 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 /r/ in places that the other Tai languages do not, including consonant clusters with /r/. 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 /r/ 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 /r/ or clusters with /r/ 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 /r/ in any of these languages, showing that /h/ is not systematically being replaced with /r/. 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 /r/ in comparison to Tai Lue, Tai Ahom, and Tai Aiton (Tabassum and Morey 2009).

Muak Sa-aak	Thai	Tai Lue	Tai Ahom	Tai Aiton	gloss
<i>ro</i> <sup>3</sup> <i>caj</i> <sup>3</sup>	<i>hǔa caj</i>	<i>hoo</i> <sup>1</sup> <i>caj</i> <sup>1</sup>	<i>ru u chau</i>		heart
<i>rɔk</i> <sup>2</sup>	<i>hòk</i>	<i>hok</i> <sup>2</sup>	<i>ruk</i>		six
<i>ha</i> <sup>2</sup>	<i>hâ</i>		<i>ha</i>		five
<i>p<sup>h</sup>rxŋ</i> <sup>2</sup>	<i>p<sup>h</sup>ûŋ</i>	<i>p<sup>h</sup>ung</i> <sup>3</sup>	<i>phrung, phreng</i>	<i>phruŋ</i> <sup>3</sup>	bee
<i>prut</i> <sup>2</sup>				<i>phrû</i> <sup>9</sup>	sorcerer

The existence of the rhotic in other Tai languages suggests that Muak Sa-aak has not replaced /h/ with /r/ 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 /r/ 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)

Muak Sa-aak	Hu	U	Lamet	Northern Kammu	Southern Kammu	
<i>*voiceless</i>						
<i>p<sup>h</sup>xj<sup>3</sup></i>	<i>phij</i>	<i>phèt</i>	<i>pij</i>	<i>pij</i>	<i>pij</i>	‘to shoot’
<i>tam<sup>1</sup></i>	<i>thaj</i>	<i>thán</i>	<i>tá:j</i>	<i>tá:j</i>	<i>ta:j</i>	‘to weave’
<i>k<sup>h</sup>a:p<sup>1</sup></i>	<i>khàp</i>	<i>khap</i>	<i>ká:p</i>	<i>ká:p</i>	<i>ka:p</i>	‘jaw’
<i>*voiced</i>						
<i>pri<sup>2</sup></i>	<i>pvi?</i>	<i>qi</i>	<i>priü?</i>	<i>pri?</i>	<i>bri?</i>	‘forest’
<i>puce<sup>2</sup> ta:k<sup>1</sup></i>	<i>phltàk</i>	<i>?atǎŋ</i>	<i>pltà:k</i>	<i>ktá:k</i>	<i>kda:k</i>	‘palm (of hand)’
<i>ka:ŋ<sup>3</sup></i>	<i>kàŋ</i>	<i>káã</i>	---	<i>kà:ŋ</i>	<i>ga:ŋ</i>	‘house’
<i>kak<sup>2</sup></i>	<i>kák</i>	<i>kàk</i>	<i>kàk</i>	--	--	‘to bite’

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 *bii<sup>1</sup>* ‘gallbladder’ > Muak Sa-aak *bi:<sup>3</sup>*, Lue *bup<sup>1</sup>* ‘hit’ > Muak Sa-aak *bup<sup>2</sup>*, Lue *da?<sup>1</sup>* ‘kick’ > Muak Sa-aak *da:<sup>2</sup>*, Lue *dɔn<sup>1</sup>* ‘month’ > Muak Sa-aak *dɔ:n<sup>3</sup>*, Lue *dɛn<sup>4</sup>* ‘border’ > Muak Sa-aak *dian<sup>3</sup>*.

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 *bil<sup>3</sup>* ‘forget’ may be found in U *pin* and has been reconstructed as proto-Mon-Khmer *\*[b]iir*; this may be seen in Palaung *bir* and Praok *pi[ɔm]*. Muak Sa-aak *dɔŋ<sup>3</sup>* ‘quiet’ has the proto-Mon-Khmer reconstruction *\*dǎk*, seen also in Palaung *dɔ?* and Riang-Lang *dǎk*. Muak Sa-aak *bɔt<sup>2</sup>* ‘cloud’ may be linked to the proto-Mon-Khmer reconstruction *\*?ut*, *\*?uut*, which does not contain the voiced initial, but the realizations of *?bɔt* in Bo Luang Lawa and of *pú:t* 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: *pal<sup>1</sup> k.tit<sup>2</sup>* ‘(be) equal’, *mu:l<sup>1</sup>* ‘boundary’, *jul<sup>1</sup> jil<sup>3</sup>* ‘crush to powder’, *pɔl<sup>1</sup>* ‘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

Tai Lue	Muak Sa-aak	gloss
<i>na:</i> <sup>3</sup>	<i>na:</i> <sup>2</sup> <i>ɲa:</i> <sup>j3</sup>	‘face’
<i>k<sup>h</sup>a:</i> <sup>3</sup>	<i>k<sup>h</sup>a:</i> <sup>2</sup>	‘slave’
<i>ɲa:</i> <sup>4</sup>	<i>t.ɲa:</i> <sup>2</sup>	‘sesame seed’
<i>ju:</i> <sup>6</sup>	<i>ju:</i> <sup>2</sup>	‘push’

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.ɲa:*<sup>2</sup> ‘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: *na:*<sup>2</sup> ‘face’ coming from Tai Lue, and *ɲa:*<sup>j3</sup>, ‘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 allotone 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 /ɛ:/, /ɔ:/ 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 /ɛ:/ and /ɔ:/. 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

- Hall, Elizabeth. 2013. A phonological description of Muak Sa-aak. *Mon Khmer Studies* 42: 26-39.
- Hall, Elizabeth. 2010. A phonology of Muak Sa-aak. MA Thesis, Payap University: Thailand.
- Hopple, Paulette. 2007. Notes re: the visit of Plang from Mong Yawng. ms. (Unpublished)
- Hudak, Thomas John, ed. 1996. *William J. Gedney's The Lue Language: Glossary, Texts and Translations*. Ann Arbor, Michigan: Center for South and Southeast Asian Studies, University of Michigan.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2013. *Ethnologue: languages of the world*, seventeenth edition. Dallas: SIL International. Online version: <http://www.ethnologue.com/>.
- Narumol Charoenma. 1982. The phonologies of a Lampang Lamet and Wiang Papao Lua. *Mon-Khmer Studies* 11: 35-45.
- Shorto, Harry Leonard, Paul Sidwell, and Doug Cooper. 2006. *A Mon-Khmer comparative dictionary*. Ed. Christian Bauer. Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Svantesson, Jan-Olof. 1988. "U." *Linguistics of the Tibeto-Burman Area*, vol. 11, no. 1, pp. 64-133.
- Svantesson, Jan-Olof. 1991. "Hu - a language with unorthodox tonogenesis", in *Austroasiatic Languages, Essays in honour of H. L. Shorto*, ed. J.H.C.S. Davidson, pp. 67-80. School of Oriental and African Studies, University of London.
- Tabassum, Zeenat and S.D. Morey. 2009. 'Linguistic features of the Ahom *Bar Amra*,' in S. Morey and M. Post (eds) *North East Indian Linguistics II*. Delhi: Cambridge University Press, India, 70-89.

# Conceptual metaphors of Vietnamese taste terms

Rujiwan 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*<sup>1</sup>

**Keywords:** methaphor, 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	<i>chua</i>	lắm.
woman	3 <sup>rd</sup>	voice	<i>sour</i>	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

1. In this study, metaphorical concepts are written as devised by George Lakoff. The metaphorical concepts are written from target domain to source domain (TARGET DOMAIN IS/AS SOURCE DOMAIN).

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.

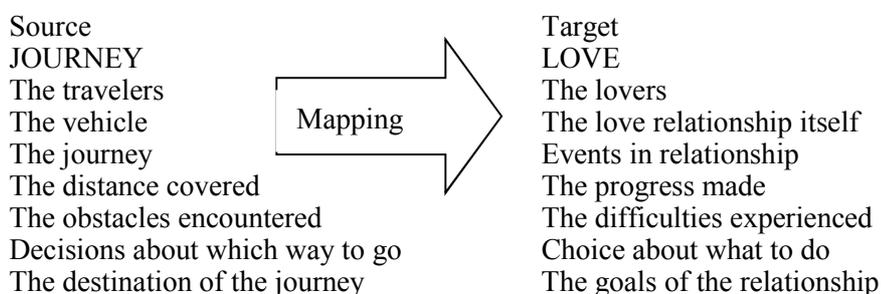
Lakoff and Johnson (1980) gave an example of metaphorical language use about time of English speakers. It is found that there are several expressions about time in English that express temporal metaphor.

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:



(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

<i>Source</i>	<i>Target</i>
- Container	- Body
- Heat of liquid	- Anger
- Level of heat	- Level of anger

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

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

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
- Ly Toan Tanga and Kieu Van Le Thi (2013) A cross-cultural study of conceptualizing internal body organs in SEA languages.

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 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 term *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       3<sup>rd</sup>            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*  
 on     face                woman    3<sup>rd</sup>        brimful    some    drop    tear        bitter  
*cay*.  
*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    nhiều    *cay*        *đắng*,    bấy        nhiều    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) *giọng* ‘voice’+ *chua* ‘sour’

Cô            ấy            *giọng*    *chua*            lắm.  
woman    3<sup>rd</sup>            voice    sour            very

‘That woman has a very high pitched voice.’

In (6) *chua* follows a noun *giọng* ‘voice’ forming a compound.

7) *nói* ‘speak’+ *ngọt* ‘sweet’

Nói            ngon            *nói*            *ngọt*.  
speak        delicious    speak        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.<sup>1</sup>

<sup>1</sup> [ ] indicates a semantic attribute of a word.

<i>Chua</i> 'sour'	<i>Đắng</i> 'bitter'
[A taste of unripe fruit]	[Mouth-sticking taste]
[Able to be cooked]	[Oral perception]
[Acidic taste]	[Perceived taste]
[A unique taste]	[Taste in the mouth]
[Irritating throat]	[Taste of gallbladder]
[Oral perception]	[Unfavorable]
[Perceived taste]	[Unsavory]
[Sharp taste]	
[Taste of lemon]	
[Taste of star apple]	
[Taste of tamarind]	
[Tingling in the teeth]	
<i>Mặn</i> 'salty'	<i>Ngọt</i> 'sweet'
[Be full-flavored]	[A taste of ripe fruit]
[Healing and preventing oral diseases]	[Clear as sugary syrup]
[Intense taste]	[Good taste]
[Make thirsty]	[Liking]
[Oral perception]	[Oral perception]
[Perceived taste]	[Perceived taste]
[Preserving food for a long time]	[Savory]
[Taste of salt]	[Sugary]
[Taste that difficult to change]	[Taste of sugar]
[Valuable]	

### 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<sup>2</sup>. *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.

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

<i>Source Domain</i>	<i>Target Domain</i>
Taste term	feelings
<i>chua</i> ‘sour’+ <i>chát</i> ‘tart’	bitterness
[Perceived taste] [Perceived taste]	[Perceived feelings]
[Oral perception] [Oral perception]	[Mental perception]
[Acidic]	[Corrosive feeling]

**Diagram 1:** metaphorical mapping process of *chua* in *chua chát*

From the above semantic mapping, attributes of *chua* ‘sour’ and its compounding element, *chát* ‘tart’, 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. *Chua* transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings], from [oral perception] to [mental perception] and from [acidic] to [corrosive feeling]. *Chát* also transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings] and from [oral perception] to [mental perception].

- 12) Cô ấy cảm thấy *chua xót* lắm khi nhìn thấy những đứa  
 She that feel *sour sting* very when see some CL  
 trẻ mồ côi này.  
 child orphan this  
 ‘The woman felt sad when she saw these orphans.’

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

<i>Source Domain</i>	<i>Target Domain</i>
Taste term	feelings
<i>chua</i> ‘sour’ + <i>xót</i> ‘to sting’	painfulness
[Perceived taste] [Cause sharp pain]	[Perceived feelings]
[Oral perception]	[Mental perception]
[Acidic]	[Corrosive feeling]
	[Painful]

**Diagram 2:** metaphorical mapping process of *chua* in *chua xót*

From the above semantic mapping, attributes of *chua* ‘sour’ and its compounding element, *xót* ‘to sting’, 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. *Chua* transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived feelings], from [oral perception] to [mental perception] and from [acidic] to [corrosive feeling]. *Xót* transfers its semantic attributes from the source domain [cause sharp pain] to the target domain [painful].

- 13) Đó là một bài học *đắng cay* dành cho người ích kỷ như  
 that is one lesson *bitter spicy* for give person selfish like  
 vậy.  
 this  
 ‘That is a bitter lesson for a selfish person.’
- 14) Những nỗi ngọt bùi và *cay đắng*.  
 some feeling sweet crunchy and *spicy bitter*  
 ‘The sweets and the bitter of life’

*Đắng cay* or *cay đắng* means suffering. *Đắng* is bitter and *cay* is spicy. When the two words co-occur as a single unit, it metaphorically expresses mentally suffering feelings. The transfer of its semantic attributes can be illustrated as follows.

<i>Source Domain</i>		<i>Target Domain</i>
Taste term		feelings
<i>đắng</i> ‘bitter’ + <i>cay</i> ‘spicy’		suffering
[Perceived taste]	[Perceived taste]	[Perceived feelings]
[Oral perception]	[Oral perception]	[Mental perception]
[Unfavorable]	[Corrosively hot]	[Unpleasant]
		[Corrosive feeling]
		[Suffering]

**Diagram 3:** metaphorical mapping process of *đắng* in *đắng cay*

<i>Source Domain</i>		<i>Target Domain</i>
Taste term		feelings
<i>cay</i> ‘spicy’ + <i>đắng</i> ‘bitter’		suffering
[Perceived taste]	[Perceived taste]	[Perceived feelings]
[Oral perception]	[Oral perception]	[Mental perception]
[Corrosively hot]	[Unfavorable]	[Unpleasant]
		[Corrosive feeling]
		[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    3<sup>rd</sup>    love    together    *salty*    *warm*    very  
 ‘That couple loves each other warmly.’

*Mặn nồng* means warm and timely love. *Mặn* is salty and *nồng* 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.

<i>Source Domain</i>		<i>Target Domain</i>
Taste term		feelings
<i>mặn</i> ‘salty’ + <i>nồng</i> ‘warm’		warm and timely love
[Perceived taste]	[Perceived feelings]	[Perceived feelings]
[Oral perception]	[Body perception]	[Mental perception]
[Be full-flavored]	[A fairly high degree of hot]	[Good Feeling]
		[Closely]
		[Affectionate]

**Diagram 5:** metaphorical mapping process of *mặn* in *mặn nồng*

From the above semantic mapping, attributes of *mặn* ‘salty’ and its compounding element, *nồng* ‘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]. *Nồng* 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.

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

**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*  
 through know many difficulty woman 3<sup>rd</sup> already feel get  
*được những ngọt bùi.*  
 get some 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.

<i>Source Domain</i>	<i>Target Domain</i>
Taste term	feelings
<i>ngọt</i> ‘sweet’ + <i>bùi</i> ‘crunchy’	fun or happiness
[Perceived taste] [Perceived taste]	[Perceived feelings]
[Savory] [Pleasingly]	[Favorable]
	[Enjoy]

**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 [savory] 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 *ngot*, 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                3<sup>rd</sup>            *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.

<i>Source Domain</i>	<i>Target Domain</i>
Taste term	habits or manners
<i>chua</i> ‘sour’	an overconfident manner of a woman or a woman dressed in outlandish color.
[Perceived taste]	[Perceived personality]
[Sharp taste]	[Strike the eyes]
[A unique taste]	[A unique personality]
	[Self-confidence]

**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.

<i>Source Domain</i>	<i>Target Domain</i>
Taste term	habits or manners
<i>chanh</i> ‘lemon’+ <i>chua</i> ‘sour’	overconfident manner
[Fruit]                    [Perceived taste]	[Perceived personality]
[Acidic taste]            [Acidic taste]	[Strike the eyes]
[A unique taste]        [A unique personality]	[Self-confidence]

**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    3<sup>rd</sup>            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-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.

<i>Source Domain</i>		<i>Target Domain</i>	
Taste term		habits or manners	
<i>chua</i> ‘sour’ + <i>ngoạ</i> ‘rude’		an overconfident manner of a woman and exploit.	
[Perceived taste]	[High pitch]	[Perceived personality]	
[Irritating to throat]	[Irritating to ears]	[Strike the eyes]	
[A unique taste]		[A unique personality]	
		[Self-confidence]	

**Diagram 10:** metaphorical mapping process of *chua* in *chua ngoạ*

From the above semantic mapping, attributes of *chua* ‘sour’ and its compounding element, *ngoạ* ‘rude’, 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. *Chua* transfers its semantic attributes from the source domain [perceived taste] to the target domain [perceived personality], from [irritating to throat] to [strike the eyes] and from [a unique taste] to [a unique personality]. *Ngoạ* transfers its semantic attributes from the source domain [irritating to ears] to [strike the eyes]

- 21) Ngoài mặt thì ngọt ngào, trong bụng thì thâm độc.  
 outside face CON suave inside stomach CON wicked  
 ‘To be suave on the outside and to be wicked within.’

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

<i>Source Domain</i>		<i>Target Domain</i>	
Taste term		persons	
<i>ngọt</i> ‘sweet’+ <i>ngào</i> ‘to mix in syrup’		gentle and sweet	
[Perceived taste]	[Sweet]	[Perceived personality]	
[Sugary]	[Outside]	[Outside aspects]	
		[Favorable]	

**Diagram 11:** 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 habits/manners 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 personality] and from [sugary] to [favorable]. *Ngào* transfers its semantic attributes from the source domain [outside] to the target domain [outside aspects].

### *Qualities metaphor*

From the investigation, *chua* and *ngọt* metaphorically express qualities and they occur with other taste terms and other words. In proverbs, only *chua* expresses qualities. This is shown in the following example.

- 22) Cô ấy là người dẫn chương trình thì không thể có  
 woman 3<sup>rd</sup> is person guide program CON cannot have  
*giọng chua* được.  
 voice sour get  
 ‘That woman is an MC, so she cannot use a high-pitched voice.’

*Giọng chua* means a high-pitched voice. *Giọng* means voice and *chua* means sour. When the two words co-occur as a single unit, it metaphorically represents a high-pitched voice. The transfer of its semantic attributes can be illustrated as follows.

<i>Source Domain</i>		<i>Target Domain</i>	
Taste term		qualities	
<i>giọng</i> ‘voice’ + <i>chua</i> ‘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.  
 CL teacher 3<sup>rd</sup> always speak sweet with student  
 ‘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.

<i>Source Domain</i>		<i>Target Domain</i>	
Taste term		state of being	
<i>nói</i> ‘speak’+ <i>ngọt</i> ‘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.  
 voice speak sweet mix in syrup  
 ‘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.

<i>Source Domain</i>		<i>Target Domain</i>	
Taste term		qualities	
<i>ngọt</i> ‘sweet’+ <i>ngào</i> ‘to mix in syrup’		sweet	
[Oral perception]	[Conformity]	[Auditory perception]	
[Savory]	[Outside]	[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.

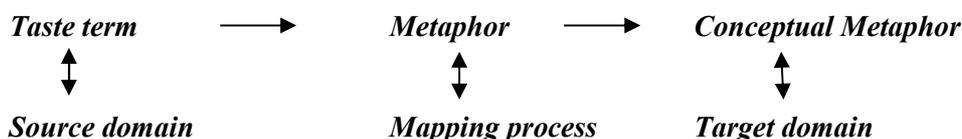
### 2.3 Metaphorical concepts of the Vietnamese taste terms

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)

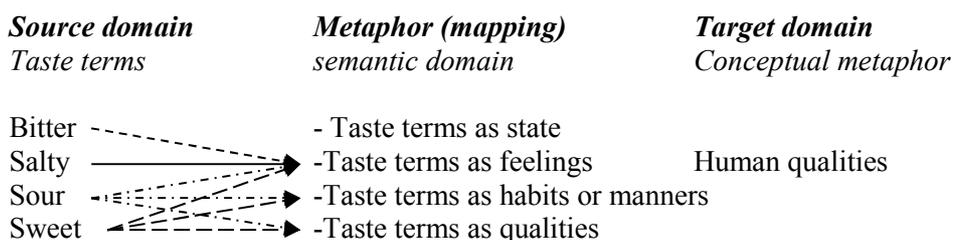
Metaphor that reflects the concepts of language users is called a conceptual metaphor. From the above-mentioned examples, the relationship between the taste terms, metaphor and concept can be illustrated in the following figure.



**Figure 2:** Process of conceptual metaphor of taste terms

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

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 domain: 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

- Beckhouse, A.E. (1994). *The Lexical Field of Taste: A Semantic Study of Japanese Taste Terms*. London: Cambridge University Press.
- Croft, W and Cruse D.A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Đặng Chân Liêu và bạn. (1999). *Từ điển Việt-Anh*. Nhà xuất bản thành phố Hồ Chí Minh.
- Hòang Khê. (2009). *Từ điển chính tả*. Nhà xuất bản Đà Nẵng.
- Hoàng Văn Hành và Nguyễn Vũ. (2003). *Từ điển tiếng Việt*. Nhà xuất bản từ điển bách khoa.
- Johnson, Mark. (1987). *The Body in the Mind*. Chicago: University of Chicago Press.
- Lakoff, George and Johnson, Mark. (1980). *Metaphor We Live By*. Chicago and London : University of Chicago Press.
- Lakoff, George and Johnson, Mark. (1980). "The Metaphorical Structure of The Human Conceptual System" *Cognitive science* 4. (p.195-208).
- Lakoff, George and Johnson, Mark. (1993). *Language and Thought 2<sup>nd</sup> edition*. Cambridge University Press.

- Ly Toan Thang. (2011). "The Vietnamese Expression of BODY and SOUL: A Cognitive and Cultural Linguistic Study" *Mon-Kmer Studies Journal Special Issue No.3*.(P.127-133)
- Nguyễn Đình Hoa. (1993). *Từ điển Việt-Anh*. Nhà xuất bản Giáo dục.
- Nguyễn Như Ý. (2003). *Từ điển chính tả học sinh*. Nhà xuất bản Giáo dục.
- Nguyễn Quốc Khánh và bạn. (2009). *Từ điển chính tả tiếng Việt dành cho học sinh*. Nhà xuất bản từ điển bách khoa.
- Nguyễn Sanh Phúc và nhóm công tác. (2000). *Từ điển Việt-Anh*. Nhà xuất bản Văn hóa thông tin.
- Nguyễn Văn Xô. (2009). *Từ điển tiếng Việt*. Nhà xuất bản Thanh niên.
- Phuong Vi Vo. (2003). Conceptual Metaphor of 'Love' in Vietnamese : An Analysis of *Selected Poems and Songs, Past and Present*. M.A. Thesis (Linguistics), Payap University, Chiang Mai, Thailand.
- Salalamba, Chatchawadee. (2000). "Cognitive Linguistics Theory." *A Course book for Ling 603 Semantic System of Thai Language*. Bangkok: Department of Linguistics, Thammasart University. (Mimeograph).
- Salalamba, Chatchawadee. (2005). "Metaphor According to Cognitive Theory." *Journal of Liberal Arts*, Vol. 5 No.1 January-June. (p.1-19)
- Sekuler, Robert and Blake, Randolph. (1990). *Perception*. New York. McGraw-Hill.
- Srichampa, Sophana. (2002). *Vietnamese Grammar*. Bangkok : Ten May Production.
- Srichampa, Sophana. (2006). "Vietnamese Language and Culture: Vietnamese Worldviews from Proverbs." *Mon-Kmer Studies Journal*, vol. 36, (p.103-120).
- Thepkanjana, Kingkarn. (1998). "The Role of Metaphor in Language and Thought". *Thought*. January, (p.12-23).
- Thomson, Lawrence C. (1987). *Vietnamese Reference Grammar*. In Stephen O' Harrow(ed.), *Mon-Kmer Studies XIII-XIV*.USA : University of Hawaii Press.
- Ungerer, F and Schmid, H-J. (1996). *An Introduction to Cognitive Linguistics*. London: Longman.
- Viện ngôn ngữ học. (2006). *Từ điển tiếng Việt*. Nhà xuất bản Đà Nẵng.

### Websites/online resources

- [http://en.wikipedia.org/wiki/Sugar\\_addition](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. Somsong 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 *\*a:*, broadly schematized as *\*a:* > *\*e:* > *\*iə* > *\*i:* 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, prt, pry, mra, kjm, pnz, 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, Kmhmú', 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, Premrirat 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'inic, Pogibenko & Bui Khánh-Thê (1990) on Ksingmul, Rischel (1989a,b, 1995, 2007) and Rischel & Egerod (1987) and Egerod & Rischel (1987) on Mlabri, Bui Khánh Thê (2000) on Phong/Kaniang, Ferlus (1970) and Đặng Nghiễm Vạn (1983) on Thai Hat/Õdu, 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
- tonogenesis and registrogenesis within Khmu dialects (Lindell et al. 1979, Premssirat 2001, 2004).

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:

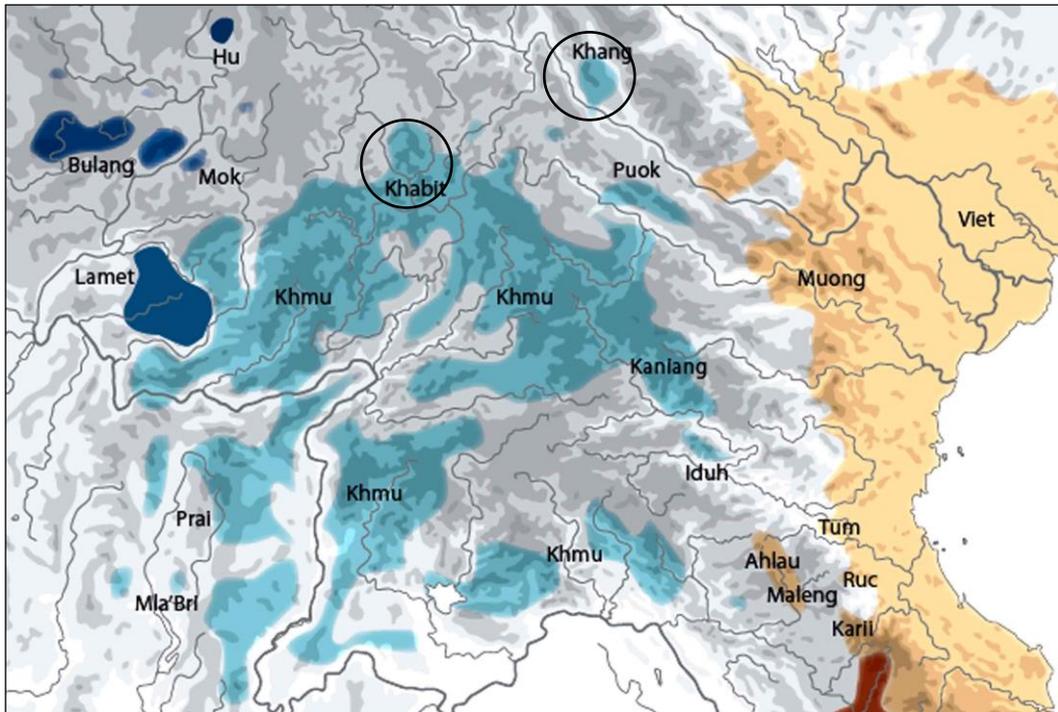
**Table 1:** Khmuic languages listed by Proschan (1996)

Language	Alternate names	Language	Alternate names
Kmhmu	(many <sup>1</sup> )	Mlabri	Phi Tong Luang
Phong	Tay Phong	Theen	Kha Sam Liam
Thin	Mal, Pray	Iduh	Tay Hat
Ksing Mul	Puok, Pou Hok	Khang	Mang U
Bit	Khabit		

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:

<sup>1</sup> 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/uə*<sup>2</sup>) that strongly hints at subgrouping

<sup>2</sup> Interpretation of the sources is somewhat problematic, but it would appear that *u* and *uə* are effectively indicating the same phoneme, a high unrounded non-front vowel.

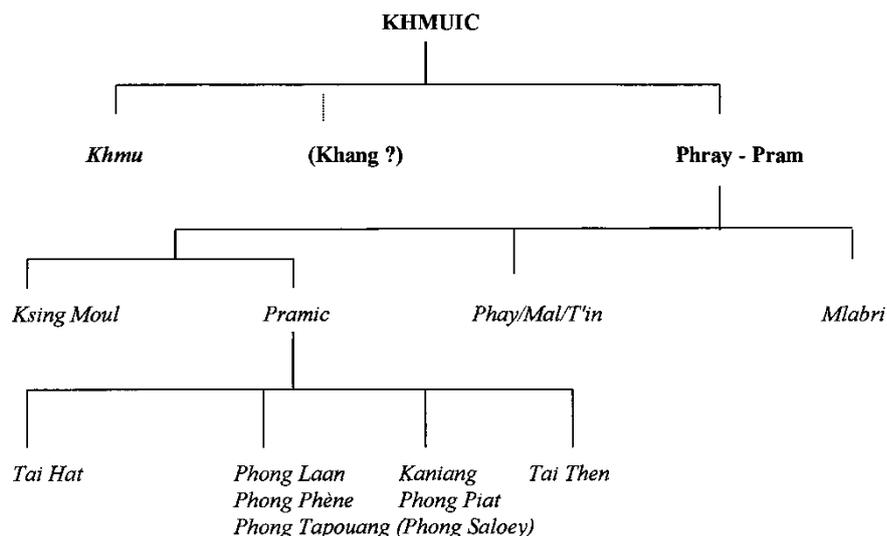
**Table 2:** Lexical comparisons supporting Palaungic classification of Khabit, Khang (plus other forms bracketed)

	Khmuic					Palaungic				Khasian
	Khmu Cuang	Mlabri	Mal	Phong	Khsing-Mul	Khabit	Khang	Palaung	Lamet	Khasi
‘eye’	<i>mat</i>	<i>mat</i>	<i>mat</i>	<i>mat</i>	<i>mat</i>	<i>pʰa:j</i>	<i>ɲaj<sup>2</sup></i>	<i>ɲaj</i>	<i>ɲa:j</i>	<i>mat</i>
‘fire’	--	<i>ʔu:lh</i>	<i>ʔo:h</i>	<i>ʔo:s</i>	<i>(həlɲoŋ)</i>	<i>tɛʰɲal</i>	<i>ɲal<sup>2</sup></i>	<i>ɲər</i>	<i>ɲal</i>	<i>(diŋ)</i>
‘sated’	<i>biʔ</i>	<i>biʔ</i>	<i>piʔ</i>	<i>kbe:j</i>	<i>(ʔkiŋ)</i>	<i>su:k</i>	<i>si<sup>2</sup></i>	<i>huʔ</i>	<i>sa:k</i>	<i>(hun)</i>
‘blood’	<i>ma:m</i>	<i>mɛ:m</i>	<i>miam</i>	<i>mi:m</i>	<i>miəm</i>	<i>s<sup>2</sup>nuəm</i>	<i>num<sup>2</sup></i>	<i>snam</i>	<i>na:m</i>	<i>sna:m</i>
‘laugh’	<i>k<sup>h</sup>ras</i>	--	<i>khieh</i>	<i>krih</i>	<i>khliə</i>	<i>k<sup>h</sup>na<sup>h</sup></i>	<i>ɲaj<sup>1</sup></i>	--	<i>kəna:s</i>	<i>rk<sup>h</sup>ie</i>
‘moon’	<i>(moŋ)</i>	<i>kiʔ</i>	<i>(thuaʔ)</i>	<i>ki:</i>	<i>(bluən)</i>	<i>(tɛ<sup>2</sup>riəŋ)</i>	<i>(khiaŋ<sup>1</sup>)</i>	<i>(kiər)</i>	<i>kheʔ</i>	<i>(bnaj)</i>
‘water’	<i>(ʔom, ʔo:k)</i>	<i>(ɲa:k)</i>	<i>(ʔo:k)</i>	<i>(paʔaŋ)</i>	<i>(hɔ:t)</i>	<i>ʔo:m</i>	<i>ʔom<sup>2</sup></i>	<i>ʔom</i>	<i>ʔo:m</i>	<i>ʔum</i>

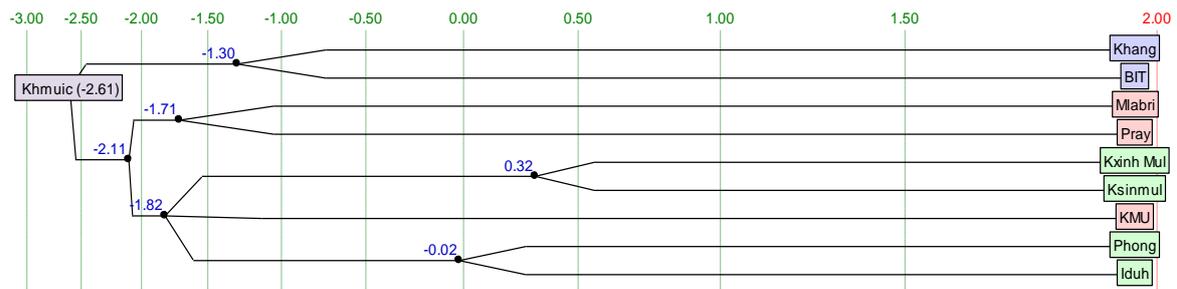
The Khmuic branch is readily defined by a very specific phonological innovation, the loss of pAA medial *\*h*, which is evident in the reflexes of ‘blood’ in the table 2 (and other etyma including ‘sated’ and ‘moon’ in Table 2). Unfortunately, Khabit and Khang reflexes of the specific etyma expected 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 *\*ɲal* ‘fire’, *\*ɲa:j* ‘eye’ and *\*sa:k* ‘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 from 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)



**Figure 3:** Khmuic languages tree from Peiros (2004)

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 correspondences<sup>3</sup> underlying the reconstruction are set out at Table 3.

**Table 3:** Khmuic consonant correspondences underling the reconstruction of Sidwell (2013)

Initials	K-Cuang	Mlabri	Mal	Khsing-Mul	Phong
*p	p	p	ph	p	p
*t	t	t	th	t	t
*c	s	ch	s	c	s
*c <sub>1</sub>	c	ch	s	c	s
*k	k	k	kh	k	k
*ʔ	ʔ	ʔ	ʔ	ʔ	ʔ
*b	b	b~b̥	p~mp	b	b
*d	d	d	t~nt	d	d
*j	ʃ	ʃ	c~ɲc	c	j, s (/C)
*g	g	g	k~ŋk	g, k (/R)	g, k (/R)
*m	m	m	m	m	m
*n	n	n	n	n	n
*ɲ	ɲ	ɲ	ɲ	ɲ	j
*ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
*w	w	w	w	w	w
*r	r	r	j	g, l (/p <sub>-</sub> ), r (/C <sub>-</sub> )	r
*l	l	l	l	l	l
*j	j	j	j	z	j
*s	h	th~ch	s	s	s
*h	h	h	h, Ø (/C)	h	h, Ø (/C)

<sup>3</sup> These are the correspondences established with multiple etymologies and confirmed with external comparisons.

Finals						
*p	p	p	p	p	p	p
*t	t	t	t	t	t	t
*c	c	c	c	c	c	c
*k	k	k	k	k	k	k
*ʔ	ʔ	ʔ	∅	∅	∅	∅
*m	m	m	m	m	m	m
*n	n	n	n	n	n	n
*ɲ	ɲ	ɲ	ɲ	ɲ	ɲ	ɲ
*ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
*w	w	w	w	w	w	w
*r	r	r	ʉ~w	l	r	r
*l	l	l	l	l	l	l
*j	j	j	j	j	j	j
*s	s, h (/i, e_)	lh	h, jh	h	s	s
*h	h	h	h	h	h	h
*∅	∅	∅	∅	∅	∅	∅

The proposal offered here is that the developments of pKhmuic *\*a:*, *\*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 *\*a:*, *\*a:*<sub>1</sub>, *\*a*, *\*a*<sub>1</sub> developments

	‘blood’	‘stone’	‘bone’	‘weave’	‘tongue’	‘tiger’
pKhmuic	*ma: <sub>1</sub> m	*gla: <sub>1</sub> ŋ	*jʔa: <sub>1</sub> ŋ	*ta: <sub>1</sub> n	*hnta:k	*rwa:j
Khmu Chuang	ma:m	gla:ŋ	cʔa:ŋ	ta: <sub>1</sub> n	hnta:k	rwa:j
Mlabri	mɛ:m	-	ʃiʔɛ:ŋ	-	-	rwa:j
Mal, Pray	miə̃m	lhiaŋ	ʔiə̃ŋ	tha: <sub>1</sub> n	nta:k	wa:j
Khsing-Mul	miə̃m	ʔə̃liə̃ŋ	-	ta: <sub>1</sub> n	hə̃lta:k	hə̃wa:j
Ōdu	mim	gliŋ	jon ʔiŋ	ta: <sub>1</sub> n	hta:	roj
Pong	mi:m	kli:ŋ	sʔi:ŋ	ta: <sub>1</sub> n	ta:ʔ	rawa:j
	‘wing’	‘thick’	‘year’	‘medicine’	‘arrow’	‘bitter’
pKhmuic	*pna: <sub>1</sub> r	*-ba: <sub>1</sub> l	*-na: <sub>1</sub> m	*crna: <sub>1</sub> m	*kam	*caŋ
Khmu Chuang	pni:r	hmbil	nim	srnim	kam	caŋ
Mlabri	hnɔ:r	-	hnɔ:m	-	-	-
Mal, Pray	panəu	mpal	-	nam	kham	saŋ
Khsing-Mul	phə̃lnal	-	-	-	kam	cuŋ
Ōdu	knɔ:r	bil	-	cnɔ:m	-	caŋ
Pong	ʔanna:r	-	-	-	kam	saŋ

The situation is that there appears to be two distinct correspondences each for *\*a:* and *\*a*, for which the notation *\*a:*, *\*a:*<sub>1</sub>, *\*a*, *\*a*<sub>1</sub> is adopted. pKhmuic *\*a:* and *\*a* have reflexes that are essentially unchanged across the branch, while *\*a:*<sub>1</sub>, *\*a*<sub>1</sub> have phonologically marked reflexes: beyond Khmu reflexes of *\*a:*<sub>1</sub> show fronted and raised vowels, while for *\*a*<sub>1</sub> 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

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  $*a$ ,  $*a_1$ ,  $*a$ ,  $*a_1$  offered here is as follows:

- $*a_1$  is a regular reflex of pAA  $*a$ ; 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.
- $*a_1$  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. Ódu *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  $*a_1$  correspondence seems to be relevant for Khmuic internal branching. If, on general phonetic grounds, we assume that there was a raising of pAA  $*a$  in the sequence  $*a > *ɛ > *iə > *i$ , the branching/subgrouping indicated in the following figures (both the Venn diagram and family tree modified from Chazée (1999) are indicated:

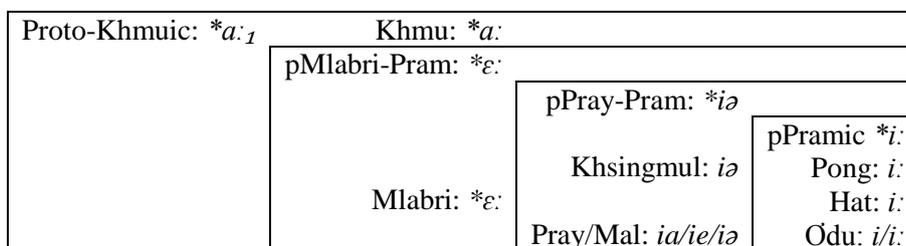


Figure 4: Venn diagram representation of  $*a_1$  developments

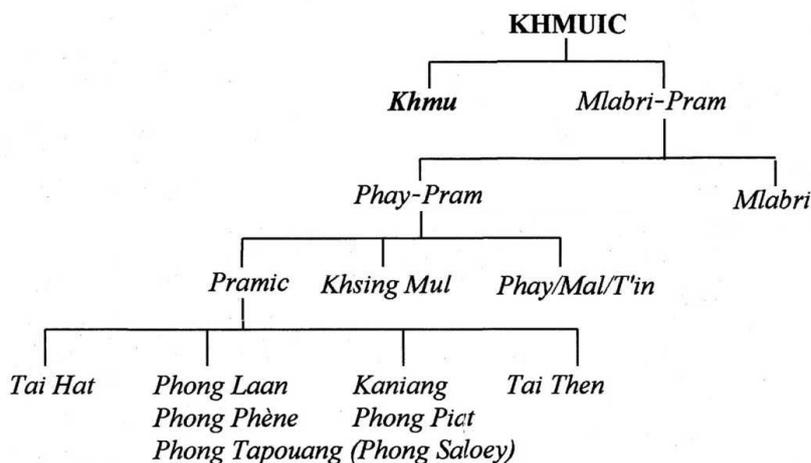


Figure 5: Khmuic tree based on  $*a_1$  developments

### 3. Homeland and migration

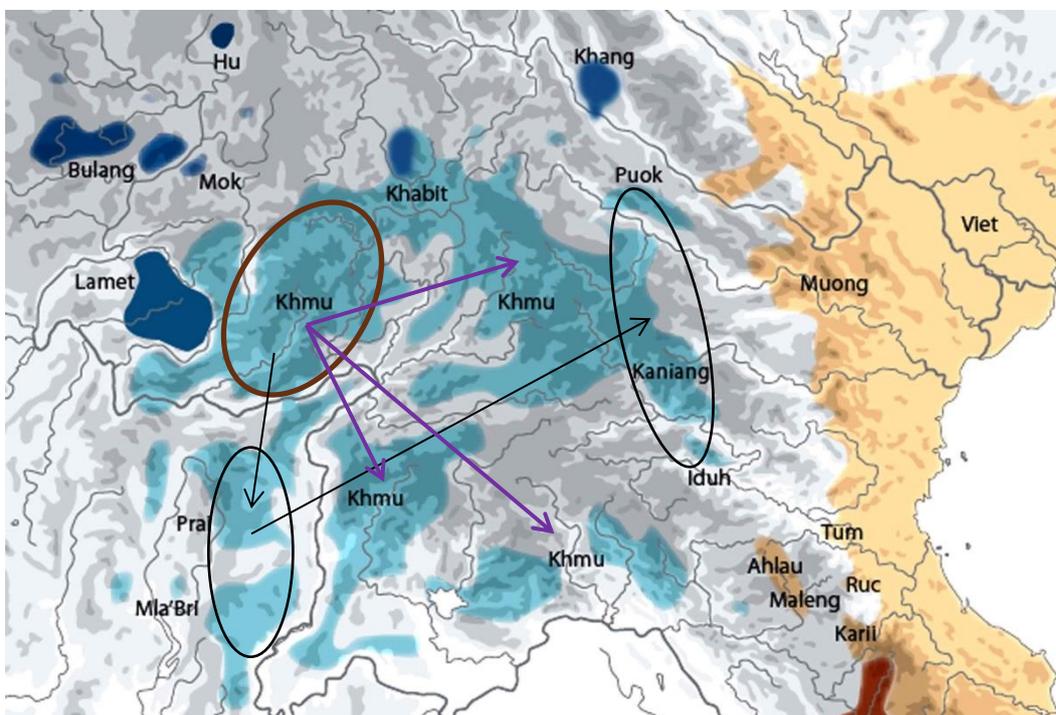
The proposed classification we have thusly arrived at is strongly nested, with a primary split between Khmu and a Mlabri-Pram sub-branch that diversified into the rest of the branch. However, the geographical interpretation is problematic, as we do not have an obvious centre of diversity, which would 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, Premssirat 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 *\*a:*, *\*a:ɿ*.

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 *\*a:*) 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

- Anderson, Gregory. 2006. Mon-Khmer Languages. In Keith Brown (ed.) *Encyclopedia Of Languages and Linguistics* (second edition), Amsterdam, Elsevier. pp.234-239.
- Bùi Khánh Thế. 2000. The Phong language of the Ethnic Phong which live near the Melhir Muong Pon Megalith in Laos. In *Pan-Asiatic Linguistics: the fifth International Symposium on Languages and Linguistics*. Ho Chi Minh City. pp. 199-253.
- Chazée, Laurent. 1999. *The Peoples of Laos: Rural and Ethnic Diversities*. Bangkok, White Lotus.
- Đặng Nghiêm Vạn. 1983. Mảng / Óđu vocabulary. Unpublished manuscript prepared in Hanoi for Michel Ferlus.
- Delcros, Henri and Jean Subra. 1966. *Petit dictionnaire du langage des Khmu' de la région de Xieng-Khouang*. Vientiane, Mission Catholique (Mimeograph).
- Diffloth, Gérard. 1982. Subclassification of Palaungic and notes on 'P'uman'. Paper presented at the 15th Sino-Tibetan Conference, August 1982. Beijing.
- Diffloth, Gérard. 2001. Map: Watersheds and Austroasiatic Languages (unpublished).
- Diffloth, Gérard. 2003. Austroasiatic Languages, *The New Encyclopædia Britannica* (15th edition) Macropedia. Vol 22. pp.701-704.
- Diffloth, Gérard. ms. The Mon-Khmer family of Languages: an Introduction (unpublished).
- Egerod, Søren and Jørgen Rischel. 1987. A Mlabri-English vocabulary. *Acta Orientalia* 48.35-88.
- Ferlus, Michel. 1970. manuscript wordlist of Mang U and Óđu (unpublished).
- Filbeck, David. 1971. *T'in: a historical study*. PhD thesis, Indiana University.
- Filbeck, David. 1978. *T'in: a historical study*. Canberra, Pacific Linguistics (B-49) Australian National University.
- 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 Lagrée, et publié...sous la direction de...Francis Garnier*. 2 volumes. Paris, Hachette.
- Haudricourt, André-Georges. 1953. La place du vietnamien dans les langues austroasiatiques. *Bulletin de la Société de Linguistique de Paris* 49.1:122-128.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242:69-82.
- Haudricourt, André-Georges. 1965. Mutation consonantique en Mon-Khmer. *Bulletin de la Société Linguistique de Paris*, 60:160-72.
- Kingsada, Thongpheth & Ryuichi Kosaka 1999. Khabit. In Kingsada, Thongpheth & Tadahiko (eds.) *Basic vocabularies of the Languages spoken in Phongxaly*, Lao P. D. R., Tokyo, ILCAA. pp.297-317
- Lindell, Kristina, Jan-Olaf Svantesson and Damrong Tayanin. 1981. Phonology of the Kammu dialects. *Cahiers de Linguistique, Asie Orientale* 9:45-71.
- Maspéro, Henri. 1955. Matériaux pour l'étude de la langue t'èng. *Bulletin de l'Ecole Française d'Extrême Orient*, 47:457-507.

- Mikami, Naomitsu. 2003. A Khang Phonology and Wordlist. In Hiromi Ueda (ed.), *Reports on Minority Languages in Mainland Southeast Asia*, 1-42. *Endangered Languages of the Pacific Rim*, Faculty of Informatics, Osaka Gakuin University.
- Parkin, Robert. 1991. *A Guide to Austroasiatic Speakers and Their Languages*. Oceanic Linguistics Special Publications No.23. Honolulu, University of Hawaii Press.
- Peiros, Ilija J. 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (doktorskaja dissertacija).
- Pogibenko, T. G. and Bui Khánh-Thê. 1990. *Iazyk Ksingmul. Materialy sovetsko-v'etnamskoj lingvističeskoj ekspeditsii 1979 goda*. Moscow, Nauka.
- Preisig, Elisabeth, Somseng Sayavong & Suksavang Simana'. 1994. *Khmhu'-Lao-French-English Dictionary*. Vientiane, Ministry of Information and Culture.
- Premisrat, S. 1993. *Thai-Khmu-English Dictionary*. Institute of Language and Culture for Rural Development, Mahidol University, Salaya.
- Premisrat, Suwilai. 2001. Tonogenesis in Khmu Dialect of SEA. *Mon-Khmer Studies* 31:47-56.
- Premisrat, Suwilai. 2002. *Thesaurus of Khmu Dialects in Southeast Asia*. Institute of Language and Culture for Rural Development Mahidol University, Thailand.
- Premisrat, Suwilai. 2004. Register complex and tonogenesis in Khmu dialects, *Mon-Khmer Studies* 34:1-17
- Proschan, Frank. 1996. A Survey of Khmuic and Palaungic Languages in Laos and Vietnam. *Pan-Asiatic Linguistics*, 3: 895-919.
- Rischel, Jørgen. 1989a. Fifty years of research on the Mlabri language: A reappraisal of old and recent fieldwork data, *Acta Orientalia* 50: 49-78.
- Rischel, Jørgen. 1989b. Can the Khmuic component in Mlabri ('Phi Tong Luang') be identified as Old T'in?, *Acta Orientalia* 50: 79-115.
- Rischel, Jørgen. 1995. *Minor Mlabri: A Hunter-Gatherer Language of Northern Indochina*. Copenhagen: Museum Tusulanum Publishers.
- Rischel, Jørgen. 2007. *Mlabri and Mon-Khmer*. Historisk-filosofiske Meddelelser 99, Copenhagen.
- Rischel, Jørgen. and Søren Egerod. 1987. 'Yumbri' (Phi Tong Luang) and Mlabri, *Acta Orientalia* 48: 19-33.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics.
- Sidwell, Paul. 2013. Proto-Khmuic. <http://sealang.net/Monkhmer>.
- Smalley, William. 1963. Notes on Kraisi's and Bernatzik's word lists, *Journal of the Siam Society* 51.2: 189-201.
- Smalley, William. 1973. Bibliography of Khmu?. *Mon-Khmer Studies* 4:23-32
- Suwilai Premisrat. 2004. Register complex and tonogenesis in Khmu dialects. *Mon-Khmer Studies* 34:1-17
- Svantesson, Jan-Olof. 1983. *Kammu Phonology and Morphology*. Travaux de l'institut de Linguistique de Lund 18. Liber Förlag
- Svantesson, Jan-Olof. 1989. Tonogenic Mechanisms in Northern Mon-Khmer. *Phonetics*, 46:60-79.

# 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 *kɔ* ‘give’, which occurs in the pattern A V G T. With other ‘ditransitive’ predicates, such as *həbɑh* ‘show’ and *pəciəʔ* ‘feed’, the recipient G is obligatorily marked by the oblique marker *kɔ*, which is homonymous with the verb *kɔ* ‘give’, and the structure is A V *kɔ* 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 *ciəʔ* ‘eat’, *chan* ‘love’, and causatives like *həcɔt* ‘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).<sup>1</sup>

(1) Restricted transitives

<i>cao hvəʔ</i>	‘return house’	‘return home’
<i>ʔa phja</i>	‘go market’	‘go to the market’
<i>ceh dāc</i>	‘descend water’	‘go down into the water’
<i>məŋ phèə</i>	‘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 it is semantically related, without an overt marker such as the oblique *kv* or the locative *dəə*. Besides undirected motion verbs, the class of intransitive verbs also includes expressions like *mip* ‘be happy, enjoy’, *toc* ‘sleep’, and *khjət* ‘die’.

(2) Intransitives

<i>*kwac phja</i>	‘walk market’	intended: ‘walk to the market’
<i>*nùm hvəʔ</i>	‘exist house’	intended: ‘be at home’
<i>*mip puə</i>	‘happy fair’	intended: ‘enjoy the fair’
<i>*khjət kəhaŋ</i>	‘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:

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

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

1 All Mon examples are from the author’s own field notes or published texts, such as journals. Where no source is indicated, the examples are elicited with native speakers. Mon examples are phonemised in IPA, other cited examples are given as in the sources.

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. An 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**: [ $\pm$ VOL,  $\pm$ INST,  $\pm$ 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).



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 *ʔa* ‘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 *ʔa* ‘go’ functions as an orientation verb, together with the directional *cao* ‘return’. The transitivity value of the two directionals *cao* and *ʔa* 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) *ʔuə ʔa rən ciəʔ kwəŋ*.  
1SG go buy eat snack  
‘I went to buy a snack to eat.’
- (14) *dəh kwac cao ʔa hvəʔ*.  
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) *dəh pən hənt klv*.  
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) *dəh pəc kəʔ pnt ʔa*.  
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) 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 *kɔʔ* ‘get; can’ and *tɛh* ‘hit, come into contact; know how to do, do correctly’ are found in core serialization, that is, they occur after the patient argument.<sup>2</sup> 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) *dɛh pàc hɔpnt na kɔʔ*.  
 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 transitivization 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 *ʔa* ‘movement away from origo; go’ and *klɔŋ* ‘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 *cvp* ‘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’, *kriɔp* ‘run’, *pɔn* ‘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 *pɛk* ‘follow, move behind someone or something’. As with other verbs of motion, direction of movement is not part of the semantics of *pɛk*, 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’, *tvn* ‘move up’, *lùp* ‘move in’, *tɛt* ‘move out’, *cao* ‘move back’ and *cvp* ‘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) *lùp cvp lɔŋ.sì*.  
 enter arrive PN  
 ‘They arrived inside Lounzi.’ (WW2Monland)
- (19) *poj tɔʔ khreʔ ceh ʔa*.  
 1PL PL step.proudly descend go  
 ‘We walked down proudly.’ (KM\_SR)

2 Both *kɔʔ* ‘get’ and *tɛh* ‘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 *tət* ‘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, *tət* ‘exit’ is combined with *ʔa* ‘go’ (less commonly *klɿŋ* ‘come’) to introduce a goal.

The orientation verb *ʔa* ‘go’, less commonly *klɿŋ* ‘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 *ʔa* ‘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, *ʔa* ‘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

Directional			Orientation		
Basic	Causative	Movement	Basic	Causative	Movement
<i>ceh</i>	<i>phjeh</i>	‘down’	<i>ʔa</i>	<i>na</i>	‘away from Col’
<i>tøn</i>	<i>pətøn~hətøn</i>	‘up’	<i>klɿŋ</i>	<i>nèŋ</i>	‘toward Col’
<i>lùp</i>	<i>plop~həplup</i>	‘into’			
<i>tət</i>	<i>pətət~hətət</i>	‘out’			
<i>cao</i>	<i>phjao</i>	‘back’			
<i>cvp</i>	<i>cvp~həcvp</i>	‘arrive’			

In most cases, the causative directionals 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 *hə-* in some colloquial varieties, as in *hətət* ‘take out’. The shape of *həcvp* ‘bring to’ (for the expected, but unattested *\*phjpp*) 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 *\*<piñcup>*, or is built after analogous forms in other verbs. In modern Mon too, *cvp* ‘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 <rañ> ‘bring’, which already in Old and Middle Mon functions as venitive V2. The morphological causative of Old Mon <tlün> ‘come’ (modern Mon *klɿŋ*) survives in Modern Mon as *kəlɿŋ* with the specialized meaning ‘welcome, receive’. In *na* ‘take away’, Old Mon <rinʔār> ‘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 *\*nèə*. The form <rinʔār> is seen as a contraction of <rañ ʔār>, which came to serve as causative form of <ʔār> ‘go’ (Shorto 1971:318). In Middle Mon the shortened form <nʔā> 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.ɳàc*    *kwac*    *cao*    *ʔa*    *phèə*.  
 child    walk    return    go    school.  
 ‘The child walked back to school.’

- (21) *rə̀ə*      *kok*    *phjao*                      *na*                      *hə̀ə?*                      (\**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 *klɿŋ*) 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 *hù?* *ʔon*, example (23) would normally be interpreted that Lamaing burned down completely.

- (22) *krip*    *ceh*                      *ʔa*,    *krip*    *ton*                      *plə̀n*.  
 run    descend      go    run    ascend      again  
 ‘We ran down, then we ran up again.’ (KM\_SR)
- (23) *lə̀mə̀ŋ*      *tao*    *ʔa*    *hù?*    *ʔon*.  
 PN              burn    go    NEG    few  
 ‘Lamaing burned down a lot.’ (WW2Monland)
- (24) *ʔu.phɿ.ʔə̀ŋ*    *tə̀?*    *lùə*    *ʔa*    *lə̀-kə̀h*                      *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) *də̀h*    *həm*                      *ʔa*    *ʔə̀rè*                      *bət*    *ba*    *nə̀dì*.  
 3    speak                      go    language      about two    hour  
 ‘He continued to talk for about two hours.’
- (26) *pə̀rao*      *chan*    *poj*    *jə̀h*                      *ba*    *də̀h*    *wət*                      *ʔa*    *ja?*.  
 matter      love    1PL    person      two    3    forget                      go    FOC  
 ‘She has forgotten our love.’ (song lyrics)
- (27) *lùp*    *klɿŋ*                      *dɿŋ*    *sem*    *nə̀?*                      *kə̀h*                      *tə̀h*    *ʔa*    *pə̀ŋ*.  
 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) *ʔə̀ŋkəlòc*    *kə̀h*                      *mù?*                      *dak*    *klɿŋ*                      *mù?*,    *dak*    *klɿŋ*                      *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 Khae and the king return to the palace, so either can be seen as subject of *cao ʔa* ‘go back’. There is no syntactic clue that one or the other would be favored as subject.

- (29) *tək.khe pək ʔa ʔekərət cao ʔa nòn.*  
 PN follow go king return go palace  
 ‘Tok Khae 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) *ʔəmè kɔ kon kwac ʔa phèə.*  
 mother give offspring walk go school  
 ‘The mother made her child walk to school.’

- (31) *déh hədiəŋ klɔ tət ʔa nù kɔ klɔʔ.*  
 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 *ton* ‘go up’ and *cvp* ‘arrive’ are combined with the regular transitive forms of the orientation verb.

- (32) *bèʔ ʔadi kòh həlèh ton na.*  
 REF PN MEDL CAUS.free ascend CAUS.go  
 ‘Adi let them go up there.’ (WW2Monland)

- (33) *hwèʔ kəmməthan nɔʔ kɔʔ pəlɔŋ na cvp, kɔʔ tɔp*  
 corpse corpse PROX get convey CAUS.go arrive get bury  
*dəə səoʔsan tɔʔ kòh kləj mənih hùʔ kɔʔ.*  
 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 directionals

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) *déh tək phjeh 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 <rañ> ‘bring’ can be combined with the basic directionals, as in <rañ fīt> ‘bring out’, <rañ lop> ‘bring in’, <rañ tlūñ> ‘bring here’. If functioning as main predicate with causative meaning, directionals appear in the causative form, as <niman plop rañ> ‘invite to enter’ (Shorto 1971:312). In this example, the verb <rañ> ‘bring (here)’ appears as directional. By Middle Mon, directionals show at least partial transitivity harmony in multi-verb predicates, as in <nʔā bacuip> ‘bring to’ and <bak

plon phyau nʔār> ‘send back home’ (lit. ‘follow-return-CAUS.return-CAUS.go’) (Shorto 1971:318).<sup>3</sup> In modern Mon, as seen above, the system of directionals is fully developed into intransitive and transitive forms. The transitive forms of the directionals 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) *ʔəmè kəwac phjao na kon.*  
 mother CAUS.walk CAUS.return CAUS.go offspring  
 ‘The mother made her child walk back.’
- (36) *dəh hələh nəŋ kon.cao ʔəcùʔ mùə.*  
 3 CAUS.free CAUS.come grandchild old.man one.  
 ‘They released the grandchild of an old man.’
- (37) *cəre kəh plon nəŋ dɔŋ sem nɔʔ.*  
 secretary MEDL CAUS.exceed CAUS.come land Thai PROX  
 ‘The secretary brought him across (the border) here to Thailand.’ (WW2Monland)
- (38) *lə-kəh dəh pələŋ na həcɔp cɔp həəʔ*  
 TEMP-MEDL 3 convey CAUS.go CAUS.arrive arrive house  
*ʔjkkəthaʔ ʔəŋ.thon.*  
 chairman PN  
 ‘Then they brought me to the house of Chairman Aung Hton.’ (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 *kəpəh* ‘gather’ is volitional, the second *kɔʔ* ‘get’ non-volitional, but both combine with the transitive orientation *na*.

- (39) *kəpəh na ʔəpət toə dəh kɔʔ na həməj.kao.*  
 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) *dəh tək pətət nəŋ 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 *pək* ‘move behind someone or something’ (consistently glossed as ‘follow’ for convenience) gets the interpretation ‘chase’ by the use of the transitive directional verb *phjeh* ‘bring down’. If the intransitive *ceh* ‘go down’ were used, the reading would be ‘follow down’ (see section 4.3).

- (41) *səthi kəh kəh dəh pək phjeh nù həəʔ.*  
 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) *ʔaca pənəh tok həbāh hvm pjah 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) *mənih pjùʔ~pjùʔ kəh dēh rùj hətəŋ nəŋ.*  
 man old~RED MEDL 3 choose CAUS.ascend CAUS.come  
 ‘The old people chose [him].’ (WW2Monland)

- (44) *mìʔ.kon.pləm jəm rəh.rəŋ na svm nɔm 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 klŋy kok, khju nəŋ ləc, khju ləc lè.*  
 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) *dēh pəlvəm na hvəʔ.*  
 3 CAUS.destroyed CAUS.go house  
 ‘They destroyed the house.’
- (47) *kja kəh kəpak na taj dɔə wəə hə-ʔvt ʔa raʔ.*  
 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.

- (48) *dɛh pɛk ʔa ka.*  
 3 follow go car  
 ‘He drove a car (away from here).’
- (49) *dɛh pɛk ʔa rəə tɔʔ.*  
 3 follow go friend PL  
 ‘He followed his friends.’ ‘He went with his friends.’

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).

- (50) *dɛh pɛk na kləə.*  
 3 follow CAUS.go cow  
 ‘He is driving the cattle.’

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) *ʔɛŋkəlɔc kəh mɨʔ dək klɨŋ mɨʔ, dək klɨŋ la.*  
 English MEDL what ride come what ride come donkey  
 ‘What did the English ride coming here? They rode donkeys.’ (WW2Monland)

Movement of Patient foregrounded → transitive DIR

- (52) *kɨʔ nɛŋ Ø tɔə.teh wì nù pəŋaʔ dək nɛŋ kɔ klɨŋ.*  
 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) *kjàn kɨʔ ʔa sɔt pì pɔn mɛʔ.*  
 PN get go fruit bael four CL  
 ‘Kyan got four bael quinces (and went away).’ (KM\_SR)

Movement of Patient foregrounded → transitive DIR

- (54) *dɛh kɨʔ na həmaj.kao hɔəʔ pɔj*  
 3 get CAUS.go flowerpot house 1PL  
 ‘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

- (55) *dəh ciəʔ ʔa pɿŋ ba pəŋan.*  
 3 eat go cooked.rice two dish  
 ‘He ate two dishes of rice.’ (he is full now)

Affectedness of Patient foregrounded → transitive DIR

- (56) *dəh ciəʔ na pɿŋ ba pəŋan.*  
 3 eat CAUS.go cooked.rice two dish  
 ‘He ate two dishes of rice.’ (the rice is gone now)

If the affected patient is visible, as in (58), the venitive orientation verb is used, while the andative form is used if the affected patient is removed from the center of interest. Sentence (59) could be uttered in a situation where there was an annoying dog in the neighborhood, and someone removed it by killing it. Sentence (57) is appropriate if the dog-killer is back from his activity without the dead dog.

Affectedness of Agent foregrounded → intransitive DIR

- (57) *dəh həcət klɿŋ klə.*  
 3 CAUS.die come dog  
 ‘He killed a dog.’ (he is here now)

Affectedness of Patient foregrounded → transitive DIR

- (58) *dəh həcət nəŋ klə.*  
 3 CAUS.die CAUS.come dog  
 ‘He killed a dog.’ (the dead dog is here/visible)

- (59) *dəh həcət na klə.*  
 3 CAUS.die CAUS.go dog  
 ‘He killed the dog.’ (the dog is gone/invisible)

The use of the different forms of directionals (and, less systematically, other non-initial verbs in complex predicates) allows for an assessment of semantic transitivity in Mon in terms of affectedness. We will briefly sum up the findings and see their implication for the notion of transitivity in the following section.

## 5. Transitivity in Mon - conclusions

In general terms, Mon can be described as a fundamentally intransitive language in Nichols’ (1982:458; see also Margetts 1999:61ff) terminology. The transitivity value (syntactic and semantic) of a verb can be changed in Mon by adding a causativizing morpheme, both to intransitive and transitive stems, a morphological process that is only marginally productive in Mon. If no morphological causative is available for a verb, a periphrastic causative construction, made up of a main verb and the preverbal causative (permissive, jussive) (*paʔ*) *kv* ‘(do) give’, is used. There is no grammaticalized means to detransitivize a transitive verb.

Mon has been shown to be a language that makes clear distinction between syntactic and semantic transitivity. Syntactically transitive verbs can be either intransitive or transitive. Transitive verbs are further divided into monotransitive and ditransitive, according to the number of arguments they license. Monotransitive verbs further have a subclass of ‘restricted transitives’, which grammatically combine only with a restricted set of objects, while unrestricted transitive verbs felicitously combine with any NP as object (though possibly with semantically nonsensical outcomes).

Syntactic transitivity in Mon is clearly a discreet notion, any one verb being either intransitive or transitive. Semantic transitivity, on the other hand, is scalar notion, with an expression being more or less transitive, depending on a number of language-dependent factors. The findings of this study strongly suggest that the main factor defining high transitivity. We have seen in several examples above that agency, intention, or volitionality of the agent are not crucial factors in determining the semantic transitivity value. In Kittilä’s analysis of transitive events, it is

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 monoclausal 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 hədiəŋ klɔ̃ tət ʔa nù kɔ̃ klɔ̃? ra,*  
 3 chase dog exit go ABL OBL garden FOC  
 ‘He chased the dogs out of the garden.’

*chaʔ.kə̀h klɔ̃ tɔʔ kə̀h hùʔ tət pùh.*  
 but dog PL MEDL NEG exit NEG  
 ‘...but the dogs didn’t go out.’

(61) *dəh hədiəŋ pətət na klɔ̃ nù kɔ̃ klɔ̃? ra.*  
 3 chase CAUS.exit CAUS.go dog ABL OBL garden FOC  
 ‘He chased the dogs out of the garden.’

\**chaʔ.kə̀h klɔ̃ tɔʔ kə̀h hùʔ tət 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 directionals 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 *tʰwəʔ* ‘exit’ would be ungrammatical.

(62) *di=lu-gá kʰwè-dwe-go ʔein-é-teʰàn-dè-gá.ne*  
 PROX=person-SBJ dog-PL-OBJ house-front-garden-inside-ABL  
*təʰauʔ-tʰouʔ-laiʔ-te, m̀aun-tʰouʔ-laiʔ-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 found 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:**

- Aikhenvald, Alexandra Y. Multi-verb constructions: setting the scene. In Alexandra Y. Aikhenvald & Pieter C. Muysken (eds.) *Multi-verb constructions*. Leiden: Brill, 1-26.
- Bisang, Walter. 1992. *das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer*. Tübingen: Gunter Narr Verlag.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56, 251-99.
- Jenny, Mathias. 2005. *The verb system of Mon*. Zurich: ASAS.
- Kittilä, Seppo. 2002. *Transitivity: towards a comprehensive typology*. Turku: University of Turku.
- LaPolla, Randy. 2010. Transitivity harmony in the Rawang language of northern Myanmar. In de Beuzeville, L. and P. Peters (eds), *From the southern hemisphere: parameters of language variation - proceedings of the 2008 Conference of the Australian Linguistics Society*. <http://ses.library.usyd.edu.au/handle/2123/7107>
- Margetts, Anna. 1999. *Valence and transitivity in Saliba, an Oceanic language of Papua New Guinea*. Nijmegen: MPI Series in Psycholinguistics.
- Næss, Åshild. 2007. *Prototypical transitivity*. Amsterdam: John Benjamins.
- Nichols, Johanna. 1982. Ingush transitivization and detransitivization. *Berkeley Linguistics Society* 8, 445-462.
- Shorto, Harry L. 1971. *a dictionary of the Mon inscriptions from the sixth to the sixteenth centuries*. London: Oxford University Press.
- Valenzuela, Pilar M. 2010. Multi-verb predicates and transitivity harmony in Shipibo-Konibo. In Alexandra Y. Aikhenvald & Pieter C. Muysken (eds.) *Multi-verb constructions*. Leiden: Brill, 185-212.

# Some Syntactic Aspects of Lexical Anaphors in Select Munda Languages<sup>1</sup>

**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 anaphors, 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

---

<sup>1</sup> We'd like to thank our language consultants Subodh Hansda and Sanat Hansda for providing data on Santali, Thakur Munda Prasad for providing data on Mundari and Pradip Bodro providing data for on Ho. These data were collected during our fieldtrips to Shantiniketan and Ranchi in 2011 by J. Mayuri and Karumuri .V. Subbarao. We are also thankful to Felix Rau and Paul Sidwell for their helpful comments.

(Tibeto-Burman) has a complex form of nominal reflexive as in (1a), Telugu (Dravidian) has a simplex form of nominal reflexive as in (1b) and Ho has a verbal reflexive as in (1c):<sup>2</sup>

Manipuri

1a. *ai-na*        ***i-sa-na***        ***i-sa-bu***        *t<sup>h</sup>agat-i*  
*I-nom*        **I-self-nom**        **I-self-acc**        praise-fut  
 ‘I will praise myself.’

Telugu

1b. *prati.va:qi-ki:*        ***tanu***        *goppava:qiga:* *anipista:q*  
 every.fellow-dat        **self.nom**        great.person        appears  
 ‘Every person thinks that he is great.’

Ho

1c. *am*        *aṛsi-re-m*        *nel-ke-n-a*  
*you*        mirror-in-2s.SAM        see-pst-**vr**-[+fin]  
 ‘You saw yourself in the mirror.’

There are languages in the subcontinent that have both a complex (polymorphemic) form, as in the example (1a) above, and a simplex (monomorphemic) form of the nominal anaphor as in (1b). Both forms are permitted in a ‘subcategorized position’ (Chomsky 1965) (Subbarao 2012 for details). A verbal reflexive implies a morphosyntactic encoding of the verbal predicate itself, as illustrated by (1c). In Ho the detransitivizing marker *-n-* performs the function of a verbal reflexive.

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 *-ɔk-* 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

<sup>2</sup> The reflexive marking elements are in bold; co-reference is encoded by italics.

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. *pəɭ<sup>h</sup>uə-kɔ*      ***akote-kɔ***      *sarhao-ε-n-a*  
 student-pl      **themselves-SAM**      praise-pst-**vr**-[+fin]  
 ‘The students praised themselves.’

Ho

3. *ɖobro*      ***ajek<sup>ʔ</sup>te***      *arsɪ-re*      *nel-ke-n-a*  
 Dobro      **himself**      mirror-in      see-pst-**vr**-[+fin]  
 ‘Dobro saw himself in the mirror.’

Mundari

4. *hɔn-kɔ*      ***akote***      *arsɪ-re-kɔ*      *nel-ke-n-a*  
 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<sup>3</sup>:

5. *pəɭ<sup>h</sup>uə-kɔ-kɔ*      *sarhao-ε-n-a*  
 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.

<sup>3</sup> 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

Pronoun	gloss	reflexive	postposition (near, about, with etc.)	gloss
<i>uni</i>	he/she	<i>ac<sup>?</sup></i>	-X	himself/herself
<i>akɔ</i>	they	<i>akɔ</i>	-X	themselves
-	-	<i>a:pre</i>	-X	self's

**Table 2:** Forms of nominal anaphors in Ho

pronoun	gloss	reflexive	postposition (near, about, with etc.)	gloss
<i>ae</i>	he/she	<i>ajek<sup>?</sup></i>	-X	himself/herself
<i>aŋ</i>	I	<i>aŋ</i>	-X	myself
<i>aben</i>	you	<i>aben</i>	-X	yourself
<i>akɔ</i>	they	<i>akɔ</i>	-X	themselves
<i>ape</i>	you (pl)	<i>ape -ape</i>	-X	yourselves
<i>abu</i>	we (incl)	<i>abu -abu</i>	-X	ourselves

**Table 3:** Forms of anaphors in Mundari

pronoun	gloss	reflexive	postposition (near, about, with etc.)	gloss
<i>akɔ</i>	they	<i>akɔ</i>	-X	themselves
<i>ae</i>	he/she	<i>aja<sup>?</sup></i>	-X	himself/herself
-	-	<i>apna</i>	-X	self's

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 ŋel-ε-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-kɔ-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<sub>i</sub> siŋboŋa-ta<sup>?</sup>a-re-m<sub>i</sub> 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.<sup>4</sup> 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)<sup>5</sup>:

Ho

9. *qobro aj-lɔ 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 *-q-*.

Ho

- 10a. *un-kin-kin isin-akaw-q-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 *-q-*.

- 10b. *un-kin joma:g<sup>2</sup>-kin 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.<sup>6</sup>

### 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 detransitized 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<sub>i</sub> duwər-ε<sub>i</sub> j<sup>h</sup>i:c<sup>2</sup>-ke-d-a*  
 Baha door-SAM open-pst-[+tr] - [+fin]  
 ‘Baha opened the door.’

- 11b. *duwər j<sup>h</sup>i:c<sup>2</sup>-ε-n-(j)a*  
 door open-pst- [-tr] -[+fin]  
 ‘The door opened.’

Ho

- 12a. *sumi silpin kesek<sup>2</sup>-ke-q-a*  
 Sumi door close-pst-[+tr] -[+fin]  
 ‘Sumi closed the door.’

<sup>4</sup> For evidence in support of this claim, see Subbarao 2012.

<sup>5</sup> For further details see Mayuri *et al* 2014.

<sup>6</sup> 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. *silpin*    *ajak<sup>2</sup>-te*    *kesek<sup>2</sup>-je-n-a*  
 door    self-by    close-pst-[-tr]-[+fin]  
 ‘The door closed by itself.’

#### Mundari

- 13a. (*əŋ*)    *duwər*    *nik<sup>2</sup>-ke-d-a-ŋ*  
 (I)    door    open- pst-[+tr]-[+fin]-SAM  
 ‘I opened the door.’

- 13b. *duwər*    *nijə-ka-n-a*  
 door    open-pst-[-tr]-[+fin]  
 ‘The door opened.’

### 6 Special verb forms in reflexive constructions

As mentioned earlier, Santali, Mundari and Ho have a special verb form,<sup>7</sup> when the verb carries a verbal reflexive. Santali verb roots are marked by the verbal suffix *-ɔk<sup>2</sup>-<sup>8</sup>*, Mundari and Ho by *-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 *-n-* that manifests reflexivity<sup>9</sup>. Moreover, the unique verbal reflexive suffix *-ɔk<sup>2</sup>-/-en-* may be dropped<sup>10</sup> without effecting the grammaticality.

#### Santali

14. *ɨŋ*    *arsi-rɛ-ŋ*    *ŋɛl(ɔk) -ka-n-a*  
 I    mirror-in-SAM    look(REFL)-pres-vr-[+fin]  
 ‘I am looking at myself in the mirror.’

#### Mundari

15. *hɔn-kɔ*    *akɔ-te*    *arsi-re*    *nel(en)-ke-n-a*  
 child-pl    they-acc    mirror-in    see(REFL)-pst-vr-[+fin]  
 ‘The children saw themselves in the mirror.’

#### Ho

16. *ɖɔbrɔ*    *aj ajte*    *gojen/gok<sup>2</sup>(en) -ja-n-a*  
 Dobro    himself    kill(REFL)-pst-vr-[+fin]  
 ‘Dobro killed himself.’

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

**Table 4:** List of verb stems in Santali

verb	gloss	reflexive	gloss
<i>ŋɛl</i>	see	<i>ŋɛl-ɔk<sup>2</sup></i>	to see oneself
<i>ɔl</i>	write	<i>ɔl-ɔk<sup>2</sup></i>	to write to oneself
<i>ɛm</i>	give	<i>ɛm-ɔk<sup>2</sup></i>	to give oneself
<i>goc</i>	kill	<i>guy-uk</i>	to kill oneself

(Ghosh 2008:70)

<sup>7</sup> 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.

<sup>8</sup> According to Ghosh (2010: 88), *-ɔk* in Santali functions both as a passive and a reflexive. *ɔk* is marked as middle voice by Neukom (2001: 64) and he also says, “the suffix *-ɔk* loses its vowel after stems ending in a vowel, eg., *idi* ‘take’ + *-ɔk*’-> *idi-k*’ ‘be taken’. Santali also has a verbal suffix *-joŋ* indicating medio-passive (Ghosh 2008: 70. The nature of the morpheme *-joŋ* requires to be investigated further.

<sup>9</sup> *-n-* and *-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).

<sup>10</sup> Optionality is indicated by parentheses.

**Table 5:** List of verb stems in Mundari

verb	gloss	reflexive	gloss
tusiŋ	to put on	<i>tusiŋ-en</i>	to put on oneself
lel	to see	<i>lel-en</i>	to see oneself

(Anderson: 2007: 46, Ghosh 2008: 129)

**Table 6:** List of verb stems in Ho

verb	gloss	reflexive	gloss
em	to give	<i>em-en</i>	to give oneself
tam	to strike	<i>tam-en</i>	to strike oneself
atom	to move	<i>atom-en</i>	to move oneself
bai	to make	<i>bai-n</i>	to make for oneself
abung	to wash	<i>abung-en</i>	to wash oneself
dul	to pour	<i>dul-en</i>	to pour over oneself

(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-subcategorized 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,<sup>11</sup> when the verbal reflexive is present as in (17b), (18b) and (19b).

#### Santali

17a. *hɔn-kɔ<sub>i</sub>*      *acʰ tɛ<sub>i</sub>*      *arsi-re*      *ɲel-ke-d-a-kɔ*  
 child- pl      **themselves**      mirror-in      see-pst-[+tr] -[+fin] -SAM  
 ‘The children saw themselves in the mirror.’

17b. *hɔn-kɔ<sub>i</sub>*      (*acʰ tɛ<sub>i</sub>*)<sub>i</sub>      *arsi-re*      *ɲel-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<sub>i</sub>*      *akɔte<sub>i</sub>*      *arsi-re-kɔ*      *lel-ke-ɖ-a*  
 child-pl      **themselves**      mirror-in-SAM      see-pst-[+tr]-[+fin]  
 ‘The children saw themselves in the mirror.’

18b. *hɔn-kɔ<sub>i</sub>*      (*akɔte*)<sub>i</sub>      *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. *ɖobro<sub>i</sub>*      *ajɛkʰtɛ<sub>i</sub>*      *arsi-re*      *nel-ke-ɖ-a-j*  
 Dobro      **himself**      mirror-in      see-pst-[+tr]-[+fin]-SAM  
 ‘Dobro saw himself in the mirror.’

<sup>11</sup> The optionality of the nominal entity is depicted parenthetically.

- 19b. *ḍobro<sub>i</sub>* (*ajək<sup>ʔ</sup>te*)<sub>i</sub> *arsi-re* *nel-ke-n-a-j*  
 Dobro **himself** mirror-in see-pst-**vr**-[+fin]-SAM  
 ‘Dobro saw himself in the mirror.’

### 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/-ḍ-* 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. *hɔpna<sub>i</sub>* *ac<sup>ʔ</sup>so:rrɛ* *bijŋ-kin<sub>j</sub>-ɛ<sub>i</sub>* *ɲɛl-kɛ-d-ɛ<sub>j</sub>(j)a*  
 Hopna **himself-near** snake-dual-SAM see-pst-[+tr]-OAM-[+fin]  
 ‘Hopna saw two snakes near himself.’

Mundari

21. *somri* *aj-ak<sup>ʔ</sup>* *japa-re* *mijan* *bij* *lel-ke-ḍ-a*  
 Somri **herself-gen** near-in one snake see-pst-[+tr]-[+fin]  
 ‘Somri saw a snake beside herself.’

Ho

22. *asai* *ajək<sup>ʔ</sup>-japak<sup>ʔ</sup>-re* *mijak<sup>ʔ</sup>/mɔjak<sup>ʔ</sup>* *bij* *nel-ke-ḍ-a*  
 Asai **himself-near-in** 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 (polymorphemic), 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. [*mɔntri<sub>i</sub>* *gidrə<sub>j</sub>* [*PRO<sub>j</sub>* *ac<sup>ʔ</sup>te<sup>\*</sup><sub>ij</sub>* *sarhawɛ*] *metak<sup>ʔ</sup>-dɛ-a*]  
 minister child **self.acc** praise tell-[+tr]-[+fin]  
 ‘The minister<sub>i</sub> asked the child<sub>j</sub> to praise self<sup>\*</sup><sub>ij</sub>.’

Anaphor in a *non-subcategorized position*

- 23b. [*baha<sub>i</sub>* *arel<sub>j</sub>* [*PRO<sub>ij</sub>* *ac<sup>ʔ</sup><sub>ij</sub>-lagid<sup>ʔ</sup>* *ca:* *benawɛ*] *metak<sup>ʔ</sup>-dɛ-a*]  
 Baha Arel **self-for** tea make tell-[+tr]-[+fin]  
 ‘Baha<sub>i</sub> asked Arel<sub>j</sub> to make some tea for self<sub>ij</sub>.’

## Mundari

Anaphor in a *subcategorized position*

- 24a. [*somra<sub>i</sub>* *sugaɽ<sub>j</sub>-ke* [*PRO<sub>j</sub>* *ajak<sup>2</sup><sub>\*i/j</sub>-ba:re-re* *baɽai-mente*] *kajej-tan-a*]  
 Somra Sugad-acc **self-about-in** praise-comp tell-pres-[+fin]  
 ‘Somra<sub>i</sub> asks Sugad<sub>j</sub> to praise self<sub>\*i/j</sub>.’

Anaphor in a *non-subcategorized position*

- 24b. [*somra<sub>i</sub>* *sugaɽ<sub>j</sub>-ke<sub>j</sub>* [*PRO<sub>i/j</sub>* *aj-lagid<sub>i/j</sub>* *ca:ha* *baj-mente*] *kajije-kan-a*]  
 Somra Sugad-acc **self-for** tea make-comp tell-pres-[+fin]  
 ‘Somra<sub>i</sub> asks Sugad<sub>j</sub> to make tea for self<sub>i/j</sub>.’

## Ho

Anaphor in a *subcategorized position*

- 25a. [*mɔntrii* *hɔn-kɔ-tej* [*PRO<sub>j</sub>* *ajak<sup>2</sup>te<sub>\*i/j</sub>* *sahra:o*]-*ici-ke-d-a*]  
 Minister child-pl- by **self** praise-caus-pst-[+tr]-[+fin]  
 ‘The minister<sub>i</sub> made the children<sub>j</sub> praise themselves<sub>\*i/j</sub>.’

Anaphor in a *non-subcategorized position*

- 25b. [*dɔbrɔ<sub>i</sub>* *sumi-ke<sub>j</sub>* *mita-di-(j)a* [*PRO<sub>i/j</sub>* *aj<sub>i/j</sub>-lagid<sup>2</sup>* *ca:j* *bajime*]]  
 Dobro Sumi-dat tell-[+tr]-[+fin] **self-for** tea make  
 ‘Dobro<sub>i</sub> asked Sumi<sub>j</sub> to make tea for self<sub>i/j</sub>.’

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<sub>i</sub>* [(*ac<sup>2</sup>-ac<sup>2</sup>-te<sub>i</sub>* *aɽi* *cɛhra:-e*] *monɛk<sup>2</sup>-a*]  
 Baha self-self-acc a lot pretty-SAM considers-[+fin]  
 ‘Baha considers herself pretty.’

## Mundari

- 26b. [*somri<sub>i</sub>* [(*aj<sup>2</sup>-aj<sup>2</sup>-te<sub>i</sub>* *ti:vi:-rɛ* *ɽuraŋ-tante*] *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<sub>i</sub>* [(*ajak<sup>2</sup>-ajak<sup>2</sup>-te<sub>i</sub>* *hosijər-ɛ*] *manati/baintana-tan-a*]  
 Asai self-self-acc clever-SAM consider-pres-[+fin]  
 ‘Asai considers herself clever.’

In (26a) in Santali, the form *ac<sup>2</sup>-ac<sup>2</sup>-te* occurs in a case-marked position where the matrix verb exceptionally case marks the derived object. If the postposition *-te* in *ac<sup>2</sup>-ac<sup>2</sup>-te* is treated as an emphatic, a [+transitive] verb such as *monɛk<sup>2</sup>* ‘consider’ would lack a direct object in the derived structure, which is a clear violation of the Theta Criterion (Chomsky 1981). Hence, *ac<sup>2</sup>-ac<sup>2</sup>* 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*<sup>2</sup> 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/-tε* 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/-tε* 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/-tε* 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. *upəl a:prε-ceta:nrε beja:r-aka-n-a-ε*  
 Upəl self-on angry-pst-[-tr] -[+fin] -SAM  
 ‘Upel was angry/ upset with herself.’

Mundari

28. *somra apna-te-gε naraj-men-a-e*  
 Somra herself-with-emph anger-is-[+fin] -SAM  
 ‘Somro is angry with herself.’

Ho

29. *sumi ajek-te kurkure/ra:gua-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 infixation. 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 ∅ 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

Ho			
verb	gloss	reciprocal	gloss
<i>capaḍa</i>	to slap	<i>ca-pa-paḍa</i>	‘slap1-vrec-slap2’
<i>til</i>	to pour	<i>ti-pi-l</i>	‘pour1-vrec-pour2’

Table 8

Santali			
verb	gloss	reciprocal	gloss
<i>dal</i>	‘beat’	<i>da-pa-l</i>	‘beat1-vrec-beat2’
<i>em</i>	‘give’	<i>ε-pe-m</i>	‘give1-vrec-give2’

Table 9

Mundari			
verb	gloss	reciprocal	gloss
<i>ɔl</i>	to write	<i>ɔ-pɔ-l</i>	‘write1-vrec-write2’
<i>nel</i>	to see	<i>ne-pe-l</i>	‘see1-vrec-see2’

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-ε*            *ɲel-aka-d-ε-a*  
 Arel            Baha-SAM        see-pst-[+tr]-OAM-[+fin]  
 ‘Arel saw Baha.’
- 31b. *arel*            *a:r*            *baha:*            *ɲε-pe-l-aka-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*            *cɔrɔp-ke-d-a*  
 Gangu            Asai-acc        kiss-pst-[+tr]-[+fin]  
 ‘Gangu kissed Asai.’
- 32b. *gangu*            *ɔndɔ*            *asai*            *cɔ-pɔ-rɔp-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]
- 33b. *somra*            *aɾ*            *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	OAM : Object Agreement Marker
[+tr] : transitive	pass : passive
[-tr] : intransitive	pl : plural
AA: Austro-Asiatic	pst: past
acc : accusative	SAM : Subject Agreement Marker
caus: causative	self ben: self benefactive
dat: dative	Sg : Singular
DR : Dravidian	TB: Tibeto-Burman
refl: reflexive	vr: verbal reflexive
IA: Indo-Aryan	vrec: verbal reciprocal
incl: inclusive	

### References

- Anderson, G. D. S. 2007. *The Munda Verb: Typological Perspectives*, Berlin/New York: Mouton de Gruyter.
- Ghosh, A. 2008. Santali. G. D. S. Anderson (Ed.), *The Munda Languages* (pp. 11-88). (London: Routledge).
- Bodding, P.O. 1929. *Materials for a Santali grammar, II (mostly morphological)*. Dumka: Santal Mission of the Northern Churches.
- Burrows, L. 1915/1980. *The Grammar of the Ho Language: An Eastern Himalayan Dialect*, New Delhi: Cosmos Publications, India.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Holland: Foris Publication.
- Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origins and Use*. New York: Praeger.
- Deeney, J. S. J. 1975. *Ho Grammar and Vocabulary*. Chaibasa, Bihar: Xavier Ho Publications.
- Ghosh, A. 2010. *Santali: A Look into Santali Morphology*. Delhi: Gyan Publishing House.
- Mayuri, J., K. V. Subbarao, G. Uma Maheshwar Rao, Sanat Hansda. 2014. 'Non-Nominative Subjects in Santali'. *South Asian Language Analysis-30*. University of Hyderabad.
- Minegishi, M, & G. Murmu (2001) *Santali basic lexicon with grammatical notes*. Institute for the Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. Neukom, L. 2001. *Santali*. Lincom Europa: Muenchen.
- Subbarao, K.V. 2012. *South Asian Languages: A Syntactic Typology*. New York: Cambridge University Press. Also, published in India by Cambridge India, Delhi.
- Subbarao, K.V. 2012. *South Asian Languages: A Syntactic Typology*. Web Materail. Free download available at: [www.cup.org/subbarao](http://www.cup.org/subbarao)

# Another look at serial verb constructions in Khmer

Stephen SELF<sup>1</sup>

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 per se (Wilawan 1993, 1995; Sak-Humphry 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 *kluən-æŋ* 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 jɔ:k kambət ka:t sac*  
Sokh take knife cut meat  
‘Sokh cut the meat with a knife.’

(1b) *kɲom tɨŋ siəwphəw ɲəoj Bill*  
1SG buy book give Bill  
‘I bought a book for Bill.’ (Spruiell 1988:252)

(1c) *baək bəŋɲuəc ɲəoj kɲom pha:ŋ*  
open window give 1SG also  
‘Open the window for me, will you?’ (Huffman, Promchan & Lambert 1970:139)

<sup>1</sup> 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.

- (1d) *knom caol ba:l təw la:n*  
1SG throw ball go car  
'I throw the ball at the car.' (Spruiell 1988:252)
- (1e) *knom wiej ckae slap*  
1SG beat dog die  
'I struck the dog dead' (Mikami 1981:110)
- (1f) *koət təw Waikiki roəm ciəmuəj jə:ŋ*  
3SG go Waikiki dance with 1PL  
'He goes to Waikiki to dance with us.' (Sak-Humphry 1995:181)
- (1g) *viə ba:n mək lau:p sdap niw kra:om 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.<sup>2</sup>

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

SVC Type	Verb Types		Word Order
	V1	V2	
SHARED AGENT	TR	TR	A V1 T <sub>1</sub> V2 T <sub>2</sub>
SHARED THEME	TR	INTR	A V1 T=T V2
SHARED AGENT & SHARED THEME	TR	TR	A V1 T V2 A V1 V2 T
PIVOT	TR	INTR	A V1 T=A V2

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 jək kandiəw ka:t srow*  
3SG take sickle cut rice.plant  
'He/she takes the sickle to cut the rice plant.' (Sak-Humphry 1995:189)
- (2b) *knom wiej kəw bajk*  
1SG hit glass break  
'I hit the glass and it broke.'
- (2c) *ko:n baoh sɔmʔa:t pteəh*  
child sweep clean<sub>CAUS</sub> house  
'The child sweeps the house clean.' (Schiller 1989:408–409)
- (2d) *ba:ŋ jək kasaet ʔa:n*  
older.brother take newspaper read  
'The older brother took the newspaper to read it.' (Wilawan 1995:61)
- (2e) *wiə noəm kme:ŋ-kme:ŋ ruət lə: 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

<sup>2</sup> 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 Leipzig 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 *ʔɔ:k* ‘take’ or *praə* ‘use’, may also be approaching such grammaticalized status. Grammaticalization, or the lack thereof, however, is an epiphenomenon 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).<sup>3</sup> 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.

Monoclausality 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.<sup>4</sup> 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) display 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

<sup>3</sup> 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

<sup>4</sup> 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 & Reuland 1993:658).

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-bounded behavior of the bimorphemic reflexive *kluən-æŋ* 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; Spruiell 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.<sup>5</sup>

**Table 2:** SVCs in MSEA

Austroasiatic	Tai	Miao-Yao	Tibeto-Burman
<i>Chrau</i> (Thomas 1971)	<i>Standard Thai</i> (Iwasaki & Ingkaphirom 2005; Rangkupan 2007)	<i>White Hmong</i> (Jarkey 1991)	<i>Pwo Karen</i> (Kato 2003)
<i>Jeh</i> (Gradin 1976)	<i>Black Tai</i> (Fippinger 1975)		<i>Kayah Li</i> (Solnit 1997)
<i>Kammu</i> (Holmer 2005) <i>Semelai</i> (Kruspe 2004)	<i>Nung</i> (Saul & Wilson 1980) <i>Lao</i> (Enfield 2007)		
<i>Vietnamese</i> (Mikami 1981; Thompson 1984; Sophana 1997)			

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

<sup>5</sup> 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.<sup>6</sup> These faults render the article problematic as an assessment of SVCs in Khmer.<sup>7</sup>

More perplexing still, two in extenso treatments of Khmer SVC behaviour per se, Wilawan (1992; 1995) and Sak-Humphry (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.<sup>8</sup> 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 Sak-Humphry (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 Sak-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:

<sup>6</sup> E.g. Mikami (1981:106) example (12c): #/\**knom ba:n koət təw psa: ʔəoj* (intended meaning: 'I got him to go to the market for me').

<sup>7</sup> 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.

<sup>8</sup> Sak-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.

Sak-Humphry (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.<sup>9</sup> 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.<sup>10</sup>

To make matters worse, in his recent reference grammar of Khmer, Haiman (2011) contradicts the view of both Wilawan and Sak-Humphry 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<sup>11</sup> that behave quite differently from the types of sentences under consideration here (Haiman 2011:271).<sup>12</sup>

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.

<sup>9</sup> 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 tɨw bɔːŋ haəj tɨw psaː phaːŋ*  
1SG go bank and go market also

‘I go to the bank and then [I] also go to the market.’ (Ehrman & Sos 1972:73)

<sup>10</sup> 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 Sak-Humphry’s to distinguish between infinitival and sentential complements. We shall have more to say about her examples below.

<sup>11</sup> 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”.

<sup>12</sup> 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-constructional 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.<sup>13</sup>

### 3.1 Monoclausality in Khmer SVCs

Monoclausality 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-controller within its same minimal clause (such as PRO). Thus, in establishing the monoclausality 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.

<sup>13</sup> 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 *kluən-æŋ* 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 *kluən-æŋ* is clause-bounded (i.e. it must find its antecedent within its own minimal clause). Unlike the English reflexive, however, Khmer *kluən-æŋ* 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* *nīw* *pteah* *kluən-æŋ*  
 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* *nīw* *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) *#knom* *baŋ* *Mora* *səmlap* *koət*  
 1SG shoot Mora kill 3SG  
 #‘I shoot Mora [and] kill him [i.e. another person]’

Example (3a) shows that *kluən-æŋ* 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 *kluən-æŋ* 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* *tha:* *nīəŋ* *sra-ləŋ* *kluən-æŋ*  
 Mora tell Sokh say girl love self  
 ‘Mora tells Sokh that the girl loves herself / \*Mora / \*Sokh.’
- (4b) *Mora* *prap* *Sokh* *tha:* *nīəŋ* *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) *knom* *twə:* *ʔəoj* *koət* *səmlap* *kluən-æŋ*  
 1SG make give 3SG kill self  
 ‘I make him kill himself / \*me.’
- (4d) *knom* *twə:* *ʔəoj* *koət* *səmlap* *koət*  
 1SG make give 3SG kill 3SG  
 ‘I make him kill him [i.e. another person].’

In (4a), *kluən-æŋ* 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 *kluən-æŋ* 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 ban sɔmlap kluən-æŋ*  
Sokh shoot CAUS.die self  
'Sokh shoots and kills himself.'

(5b) *Sokh ban sɔmlap 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 *kluən-æŋ* 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 *mɔ:k* 'come' and *tɨw* '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 jɔ:k kambət (tɨw) sɔmlap kluən-æŋ*  
Sokh take knife (go) CAUS.die self  
'Sokh takes the knife to (go) kill himself.'

(6b) *Sokh jɔ:k kəmpɰlɔ:ŋ (tɨw) ban 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 tha: kru:-pɛ:t tɨw tɨn ce:k ɲam niw pteah kluən-æŋ*  
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) *ɲnom noəm kru:-pɛ:t tɨw tɨn ce:k ɲam niw pteah kluən-æŋ*  
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 tha: ɲnom noəm kru:-pɛ:t tɨw tɨn ce:k ɲam*  
Mora tell Sokh say 1sg take doctor go buy banana eat  
*niw pteah kluən-æŋ*  
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 *kluən-æŋ* and the subject of the embedded clause, *kru:-pɛ:t* 'doctor'. Two of these verbs have their own individual complements (*tɨn ce:k* 'buy bananas' and *niw 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 kru:-pɛ:t* '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 *kluən-æŋ* 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).

### 3.1.2 Ruling out coordination

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 (Sak-Humphry 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 preceded by the focus-marking word *ki:* and followed by the relativizer *dael*.

- (8a) *siəwphəw nih kme:ŋ mə:k psa: tɨn*  
 book DEM child come market buy  
 ‘These books the child comes to the market to buy.’ (Sak-Humphry 2005:186)
- (8b) *sro:w nih Sokh jɔ:k kandiəw ka:t*  
 rice.plant DEM Sokh take sickle cut  
 ‘This rice plant Sokh takes the sickle to cut.’ (Sak-Humphry 2005:190)
- (8c) *kme:ŋkme:ŋ nih wiə noəm ruət lə: phnum*  
 children.children DEM 3SG take run up mountain  
 ‘These children, he took [them] walking on the mountain.’ (Sak-Humphry 2005:193)
- (9a) *ki: siəwphəw dael kme:ŋ mə:k psa: tɨn*  
 FOC book REL child come market buy  
 ‘It is the books that the child comes to the market to buy.’ (Sak-Humphry 2005:187)
- (9b) *ki: kandiəw dael Sokh jɔ:k ka:t sro:w*  
 FOC sickle REL Sokh take cut rice.plant  
 ‘It is the sickle which he takes to cut the rice plant.’ (Sak-Humphry 2005:191)
- (9c) *ki: kme:ŋkme:ŋ nih dael wiə noəm da:ə lə: phnum*  
 FOC children.children DEM REL 3SG take walk up mountain  
 ‘It is these children whom he took walking on the mountain.’ (Sak-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

Sak-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) \**tɨn siəwphəw nih kme:ŋ mə:k psa:*  
 buy book DEM child come market  
 Intended meaning: ‘To buy these books, the child comes to the market.’

- (11b) \**ki:*      *tɨŋ*              *siəwphəw* *daɛl*      *kme:ŋ*      *mɔ:k*      *psa:*  
 FOC    buy              book      REL      child      come      market  
 Intended meaning: ‘It is to buy books that the child comes to the market.’

Sak-Humphry 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. Sak-Humphry’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.

- (15a) *seckdej* *ka:*      *ciə*      *cambɔ:ŋ*      *rəbɔh*      *neak*      *damnaeu*      *ki:*      *hawh*  
 matter    NMLZ    COP      principal    GEN      person      travel    FOC      fly  
*haəu*      *ptɔal*      *ko:l*      *daw*      *kantae*      *lawn*      *kantae*      *ləʔɔ:*  
 fly      direct      destination      goal      get      fast      get      good  
 ‘The main concern of the traveler is (this): to get to his destination, the quicker, the better.’ (Haiman 2011:247)

- (15b) *daɛl*      *kɨnom*      *dəŋ*      *nuh*      *ki:*      *tha:*      *niəŋ*      *nuy*      *mɔ:k*  
 REL      ISG      know      DEM      FOC      COMP      girl      FUT      come  
 ‘What I know is this: she will come.’ (Haiman 2011:248)

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 *cəŋ* ‘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 *cəŋ* 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 *tha:* ‘say’ permit subsequent verbs to appear with or without such overt coreferential subjects.

- (16) *Sina<sub>i</sub>* *prap* *tha:*              (*wiə<sub>i</sub>*) *cəŋ*      (\**wiə<sub>i</sub>*)      *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 *cəŋ* take a necessarily infinitival complement with null subject PRO.

The verb *cəŋ* ‘want’ may also occur with sentential complements and a complementizer (the verb *ʔəoj* ‘give’), but only when the subject of the complement is distinct from that of *cəŋ* (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 *cəŋ* appears at first much like the shared agent-shared theme type of potential SVCs given in **Table 1** and (2c) above.

- (17) *kɲom* *cəŋ* *niʔjiej* *piəsa:* *khmae*  
 1SG want speak language Khmer  
 ‘I want to speak Khmer.’

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

- (18) *kɲom* *cəŋ* *səmlap* *kluəŋ-əŋ*  
 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) *ʔki:* *səmlap* *kluəŋ-əŋ* *dael* *kɲom* *cəŋ*  
 COP CAUS.die self REL 1SG want  
 Intended meaning: ‘It is to kill myself that I want.’
- (19b) *ka:* *dael* *kɲom* *cəŋ* *ki:* *səmlap* *kluəŋ-əŋ*  
 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 preceeded by the focus-marker *ki:*, 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 V1, *kɲom* ‘1SG’. Hence, Fisher’s (1988) analysis is most likely correct: same-subject uses of *cəŋ* take open or infinitival complements whose null subject (PRO) is controlled by the subject of *cəŋ* 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.<sup>14</sup> 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 *ba:n* ‘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

<sup>14</sup> 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 *bontæ* ‘but’ between V1 and the negator-V2 pair (20e-g). This move makes the conative semantics of V1 even more explicit.

- (20a) *knom niʔjiej piəsa: khmae min ba:n te*  
 1SG speak language Khmer NEG be.ble NEG  
 ‘I cannot speak Khmer.’
- (20b) \**knom min niʔjiej piəsa: khmae ba:n te*  
 1SG NEG speak language Khmer be.able NEG
- (20c) *knom rək ka:bup min khəj*  
 think look.for wallet NEG see  
 ‘I can’t find (lit. look for, don’t find) [my] wallet.’ (adapted from Bisang 2009:800)
- (20d) *knom sdap min lu: te*  
 1SG listen NEG hear NEG  
 ‘I can’t hear [you] (lit. listen, not hear).’ (Smyth 2008:107)
- (20e) *knom niʔjiej piəsa: khmae bontæ min ba:n lə:*  
 1SG speak language Khmer but NEG be.ble good  
 ‘I [try to] speak Khmer, but cannot [speak it] well.’
- (20f) *knom rək ka:bup bontæ min khəj*  
 think look.for wallet but NEG see  
 ‘I looked for [my] wallet, but didn’t find [it].’
- (20g) *knom sdap bontæ min lu:*  
 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.<sup>15</sup>

- (21a) *mə:l min khəj*  
 look NEG see  
 ‘did not see’ (Maspero 1915:408)
- (21b) *sdap min lu:*  
 listen NEG hear  
 ‘did not hear’ (Haiman 2011:291)
- (21c) *kit min khəj*  
 think NEG see  
 ‘did not figure out’ (Huffman, Promchan & Lambert 1970:187)
- (21d) *ho:p 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

<sup>15</sup> But see Durie (1997:322–323) for a discussion of lexicalization of SVC combinations.

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 mɔ:k sa:la: riən*  
 student must come school study  
 ‘The student must come to school and study.’
- (22b) *nisət (min) trəw (\*min) mɔ:k sa:la: riən te*  
 student NEG must NEG come school study NEG  
 ‘The student must not come to school and study.’
- (23a) *kambət trəw jɔ:k ka:t sac \*(daoj bɔ:ŋ)*  
 knife undergo take cut meat (by older.brother)  
 ‘The knife was taken to cut the meat by older brother.’
- (23b) *sac trəw ke: jɔ:k kambət ka:t*  
 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*, *ʔat*, *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 kra:om da:əmchə: te:*  
 3SG sit NEG cry under tree NEG  
 \*‘She sits not crying under the tree.’ (Sak-Humphry 2005:182)
- (24b) *wiə min ɔŋkuj jum kra:om da:əmchə: te:*  
 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 *cəŋ*, 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 cəŋ (\*min) sɔmlap kluən-æŋ te:*  
 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” *nij*. 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 *nij* shows up, however, the tense value for both verbs is necessarily the same: future. By contrast, the perfective aspect marker *haaj*, which is itself a grammaticalized verb meaning ‘to finish’,<sup>16</sup> follows the verb it inflects and can only occur at the end of both verbs in a multi-verb construction. In fact, *haaj* 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 *traw* ‘must’ to demonstrate that the various verbs in a potential Khmer SVC must also agree in modality.

- (26a) *kɲom nij jɔ:k kambət (nij) ka:t sac*  
 1SG FUT take knife (FUT) cut meat  
 ‘I will take the knife and cut meat.’
- (26b) *kɲom jɔ:k kambət ka:t sac haaj*  
 1SG take knife cut meat PFV  
 ‘I took the knife and cut meat.’
- (26c) *kɲom traw jɔ:k kambət ka:t sac*  
 1SG must take knife cut meat  
 ‘I must take the knife and cut meat.’
- (26d) *\*kɲom jɔ:k kambət haaj ka:t sac*  
 1SG take knife PFV cut meat  
 \*‘I took the knife [and] cut meat [now].’
- (26e) *\*kɲom jɔ:k kambət nij ka:t sac*  
 1SG take knife FUT cut meat  
 \*‘I took the knife, wil cut meat.’
- (26f) *\*kɲom jɔ:k kambət traw ka:t 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 monclausality 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)

<sup>16</sup> As pointed out by the reviewer, independent uses of *haaj* as a lexical verb are rare in Khmer, though Haiman (2011:186) reports it does occur in the idiom *kɲom taw haaj* (literally ‘I go finish’), meaning something like *I can do everything* or *You know me*.

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.<sup>17</sup> 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 bəŋkoəp kru:-bəŋriən*  
student listen follow command teacher-CAUS.learn  
‘The students obey [their] teacher’s commands.’ [SVC]

(28b) *nisət sdap haəj-niŋ tam bəŋkoəp kru:-bəŋriən*  
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

<sup>17</sup> In this way, the conjunction *haəj-niŋ*, which intriguingly contains the word *niŋ* that can also function as a future-tense auxiliary verb, appears to 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 SVCs 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 *kluən-æŋ* 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 SVCs 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 SVCs encode single complex events. It is thus demonstrated that the structures in question display all of the “characteristic and diagnostic features of SVCs” 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.<sup>18</sup> The question is whether the language makes a distinction between “compact” SVCs of the sort detailed here and much longer (and possibly less coherent) “narrative” SVCs similar to those which Pawley (2008) has described for the Papuan language Kalam.<sup>19</sup>

## References

- Aikhenvald, Alexandra Y. & Robert M. W. Dixon. 2006. *Serial verb constructions a cross-linguistic typology*. Oxford, NY: Oxford University Press.
- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In: Alexandra Y. Aikhenvald and R. M. W. Dixon (eds.), *Serial verb constructions. A cross-linguistic typology*, 1–68. Oxford: Oxford University Press.
- Anderson, Gregory D. S. 2010. *Auxiliary Verb Constructions*. Oxford: Oxford University Press.
- Bisang, Walter. 1992. *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer: vergleichende Grammatik im Rahmen der Verbserialisierung, der Grammatikalisierung und der Attraktorpositionen*. Gunter Narr Verlag.
- Clark, Marybeth. 1989. Hmong and areal South-East Asia. *Papers in Southeast Asian Linguistics No.11, Southeast Asian Syntax* 11. 175–230.
- Bisang, Walter. 2009. Serial Verb Constructions. *Language and Linguistics Compass* 3. 792-814.
- Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke and Felix K. Ameka. 2007. Principles of event segmentation in language: The case of motion events. *Language* 83. 495–532.
- Chao, Yuan Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley & Los Angeles: University of California Press.
- Clark, Marybeth. 1992. Serialization in mainland Southeast Asia. *The Third International Symposium on Language and Linguistics*. 145–159.
- Durie, Mark. 1997. Grammatical structures in verb serialization. In A. Alsina & J. Bresnan (eds.), *Complex predicates*, 289–354. Stanford, CA: CSLI Publications.
- Ehrman, Madeline Elizabeth & Kem Sos. 1972. *Contemporary Cambodian: A grammatical sketch*. Washington, D.C.: Foreign Service Institute.
- Enfield, N. J. 2007. *A grammar of Lao*. Berlin; New York: Mouton de Gruyter.
- Fippinger, Jay W. 1975. Black Tai sentence types: a generative semantic approach. In Jimmy G. Harris & James R. Chamberlain (eds.), *Studies in Tai linguistics in honor of William J. Gedney*, 130–169. Bangkok, Thailand: Central Institute of English Language. <http://www.sil.org/resources/archives/3399> (30 June, 2014).

<sup>18</sup> See, e.g., Ehrman and Sos (1972:25).

<sup>19</sup> A similar distinction is found in Van Staden and Reesink (2002; Foley 2004:132) between “component” and “narrative” serialization.

- Fisher, Karen. 1985. The syntax and semantics of anaphora in Khmer. Ithaca, NY: Cornell University M.A. thesis.
- Fisher, Karen. 1988. Agreement and The Distribution of Anaphora. In Michael Hammond, Edith Moravcsik & Jessica Wirth (eds.), *Studies in Syntactic Typology*, 25–36. John Benjamins.
- Foley, William. 2004. The notion of “event” and serial verb constructions: Arguments from New Guinea. *SEALS XIV Papers from the 14th annual meeting of the Southeast Asian Linguistics Society*. 129–156. (14 July, 2014).
- Gradin, Dwight. 1976. The verb in Jeh. *The Mon-Khmer Studies Journal* 5. 43–75.
- Haiman, John. 2011. *Cambodian: Khmer*. (London Oriental and African Language Library 16). Amsterdam/Philadelphia: John Benjamins Publishing.
- Holmer, Arthur. 2005. Verb serialization in Kammu. *Department of Linguistics, Lund University Working Papers* 51. 65–84. (30 June, 2014).
- Huffman, Franklin E. 1967. An outline of Cambodian grammar. Ithaca, NY: Cornell University Ph.D. Dissertation.
- Huffman, Franklin Eugene, Charan Promchan & Chhom-Rak Thong Lambert. 1970. *Modern spoken Cambodian*. Ithaca, NY: Cornell University, SEAP Publications.
- Iwasaki, Shōichi & Preeya Ingkaphirom. 2005. *A Reference Grammar Of Thai*. Cambridge University Press.
- Jarkey, Nerida. 1991. *Serial Verbs in White Hmong: A Functional Approach*. Sydney, NSW, Australia: University of Sydney. □
- Kato, Atsuhiko. 2003. Pwo Karen. In Graham Thurgood, & Randy J. LaPolla (eds.), *The Sino-Tibetan Languages*, 632–648. (Routledge Language Family Series). London and New York: Routledge.
- Kroeger, Paul. 2004. *Analyzing Syntax: A Lexical-Functional Approach*. Cambridge: Cambridge University Press.
- Kruspe, Nicole. 2004. *A Grammar of Semelai*. Cambridge: New York: Cambridge University Press.
- Kuteva, Tania. 2001. *Auxiliation: An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University Press.
- Lord, Carol. 1993. *Historical Change in Serial Verb Constructions*. Amsterdam, Philadelphia, PA: John Benjamins Publishing.
- Maspero, Georges. 1915. *Grammaire de la langue khmère (cambodgien)*. Paris, Impr. nationale. <http://archive.org/details/grammairedelala00maspgoog> (1 July, 2014).
- Meyer, Cynthia Ford. 1992. The semantics of periphrastic causatives in Khmer. *Southwest Journal of Linguistics* 11(1.2). 29–60.
- Migliazza, Brian. 1996. Mainland Southeast Asia: A unique linguistic area. *NOL* 75. 17–25.
- Mikami, Naomitsu. 1981. Serial verbs constructions in Vietnamese and Cambodian. *Gengo Kenkyu (Journal of the Linguistic Society of Japan)* 79. 95–117.
- Pawley, Andrew. 2008. Compact versus narrative serial verb constructions in Kalam. In Gunter Senft (ed.), *Serial verb constructions in Austronesian and Papuan languages*. (Pacific Linguistics). Canberra, ACT, Australia: Australian National University (ANU). <https://researchers.anu.edu.au/publications/24053> (14 July, 2014).
- Rangkupan, Suda. 2007. The Syntax and Semantics of GIVE-Complex Constructions in Thai. *Language and Linguistics* 8(1). 193–237.
- Reinhart, Tanya & Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24(4). 657–720. (2 July, 2014).
- Roberts, John R. 2012. Serial verbs in English: An RRG analysis of catenative verb constructions. *Functions of Language* 19(2). 201–234.
- Ross, John Robert. 1967. Constraints on variables in syntax. Cambridge, MA: Massachusetts Institute of Technology Ph.D. Dissertation. <http://dspace.mit.edu/handle/1721.1/15166> (2 July, 2014).

- Sak-Humphry, Chany. 1995. Serial and completive verbs in Khmer. In M. Alves (ed.), *Southeast Asian Linguistics Society III*, 171–204. Tempe, Arizona: Arizona State University.
- Saul, Janice E. & Nancy Freiberger Wilson. 1980. *Nung grammar*. 1ST edition. Vol. 62. (Summer Institute of Linguistics Publications in Linguistics). Dallas, TX : Arlington: University of Texas at Arlington.
- Schiller, Eric. 1989. On the phrase structure of serial verb constructions. *Proceedings of the Annual Meeting of the Chicago Linguistics Society (CLS)* 25. 405–419.
- Sells, Peter. 1987. Aspects of Logophoricity. *Linguistic Inquiry* 18(3). 445–479. (3 July, 2014).
- Senft, Gunter. 2004. What do we really know about serial verb constructions in Austronesian and Papuan languages? In Isabelle Bril & Ozanne-Rivierre Françoise (eds.), *Complex predicates in Oceanic languages*, 49–64. Berlin, Germany: Mouton de Gruyter.
- Seuren, Pieter AM. 1990. Serial verb constructions. In Brian D. Joseph & Arnold M. Zwicky (eds.), *When Verbs Collide: Papers from the 1990 Ohio State Mini-Conference on Serial Verbs*, vol. 39, 14–33. (Ohio State Working Papers in Linguistics). Columbus, OH: Ohio State University Press. <http://files.eric.ed.gov/fulltext/ED330217.pdf#page=25> (11 July, 2014).
- Smyth, David. 2008. *Colloquial Cambodian*. London and New York: Routledge.
- Solnit, David B. 1997. *Eastern Kayah Li: Grammar, Texts, Glossary*. Honolulu, HI: University of Hawaii Press.
- Sophana, Srichampa. 1997. Serial verb constructions in Vietnamese. *Mon-Khmer Studies* 27. 137–144. □
- Spruiell, William. 1988. Giving and causation: The functions of ʔaoy in Khmer. *LACUS Forum* 15. 53–68. □
- Staden, Miriam van & Gerard P. Reesink. 2002. A functional approach to verb serialization. *Paper presented at the Ninth International Conference on Austronesian Linguistics*. Canberra, ACT, Australia. □
- Starosta, Stanley. 1988. *The Case for Lexicase: An Outline of Lexicase Grammatical Theory*. London and New York: Pinter Publishers.
- Thomas, David D. 1971. *Chrau grammar*. Vol. 7. (Oceanic Linguistics Special Publication). University of Hawaii Press.
- Thompson, Laurence C. 1984. A Vietnamese grammar. *The Mon-Khmer Studies Journal* 13-14. 1–367.
- Valin, Robert D. Van & Randy J. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge University Press.
- Wilawan, Supriya. 1992. The so-called serial verb constructions. *The Third International Symposium on Language and Linguistics*, 1237–1254. Bangkok, Thailand: Chulalongkorn University.
- Wilawan, Supriya. 1995. Reanalysis of serial verb constructions proper. *Mon-Khmer Studies* 24. 53–68.

# 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.<sup>1</sup> The Nicobar Archipelago is a chain of twenty two islands, with thirteen of them inhabited, lying North to South in the Bay of Bengal.<sup>2</sup> The language is spoken by the Nicobarese presently inhabiting the three Islands, namely, Nancowry, Katchal and Kamorta of the archipelago.<sup>3</sup> 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.

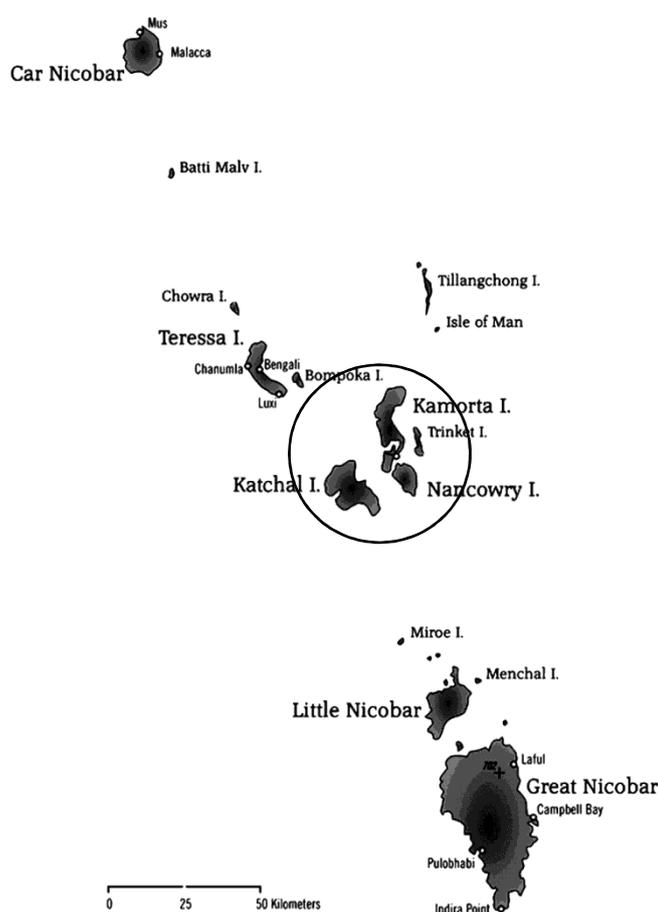
The data made use of for the paper have been drawn from the data collected as part of the Andaman Commissioned Project, a collaborative program entered into by the Union Territory Administration of the Andaman and Nicobar islands with the Central Institute of Indian Languages, Mysore. The objective of 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

<sup>1</sup> It is also known otherwise as Nancowry or Central Nicobarese.

<sup>2</sup> The thirteen inhabited islands are Car Nicobar, Chowra, Teressa, Bompoka, Nancowry, Katchal, Kamorta, Trinket, Tillong Chong, Kondul, Pulomilo, Little Nicobar and Great Nicobar. Among them Tillong Chong is devoid of ethnic inhabitants.

<sup>3</sup> Till 2004, just prior to the tsunami, the speakers of the language were spread across four islands, the fourth one being the Trinket. After the tsunamic 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.<sup>4</sup> 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 19<sup>th</sup> and December 26<sup>th</sup> 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.<sup>5</sup> Hence, attempts have already been initiated for collecting data from three more languages also with the author visiting the area of Takahanilāhngö (Great Nicobarese) and his colleague Winston Cruz, the areas of Sanënyö (Chowra) and Lamòngsë (Kondul). All the data thus collected are the property of Central Institute of Indian Languages, Mysore and they are marked as Andaman Commissioned Project data base in order to differentiate them from the others.



**Map:** Nicobar Archipelago with Muöt area circled.

<sup>4</sup> He, and only he, was authorized by the Tribal Council of the Island to work as informant.

<sup>5</sup> To determine the number of Nicobarese languages, a survey was conducted as part of this collaborative program between August 11<sup>th</sup> and October 24<sup>th</sup> 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), Sanënyö (Chowra), Lurö (Teresa), Lamòngsë (Kondul) and Takahanilāhngö (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.<sup>6</sup> 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<sup>7</sup>

	Bilabial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
<b>Plosive</b>	p (p)			t (t)	c (ch)	k (k)	ʔ (k̚)
<b>Nasal</b>	m (m)			n (n)	ɲ (ny)	ŋ (ng)	
<b>Lateral</b>				l (l)			
<b>Fricative</b>		f (f)	s (s)	ɹ (r)		x (h)	
<b>Approximant</b>		v (v)			j (y)		

### 2.2. Simple vowel

The nine simple vowels and their phonetic description can be inferred from table-2.

Table 2<sup>8</sup>

High	Front	Central	Back	
			Rounded	Unrounded
	i (i, ī)		u (u, ū)	ɯ (eu, eū,)
<b>High-mid</b>	e (ě, ē)		o (o, ō)	
<b>Mid</b>		ə (ō, öö)		
<b>Low-mid</b>	ɛ (e, ě)		ɔ (ò, ô)	
<b>Low</b>				ɑ (a, ā)

All these function as nucleus of root and as well as affixal syllables.

### 2.3. Nasalized simple vowel

Except ɔ, all the other eight simple vowels are found to attest their nasalized counterparts. They are ĩ (iñ, ñi), ũ (uñ, ñu), ũ̄ (euñ, eñu), ě (ěñ, ñě), ã (òñ, ñò), ẽ (eñ, ñe), õ (òñ, ñò) and ã (añ, ña).<sup>9</sup> They all are found to occur as nucleus of root syllables only.

<sup>6</sup> Stress is perceived, but not found to be phonemic.

<sup>7</sup> What are given in round brackets against consonants are their equivalents in Muöt orthography.

<sup>8</sup> 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, iō), ua (ua, uā), uö (uö, uō), euə (euö, eüö), ea (ea, eā), eə (eö, eō), oə (oö, oō).<sup>10</sup> Among them, ua, uö, euə and oə are found beginning with back vowels, while iə, eə and ea 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 (iön, iōñ), uã (uañ, uāñ), uö̃ (uön, uōñ), euə̃ (euön, eüön), eã (eañ, eāñ).<sup>11</sup> 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 t̚ ‘dental plosive’ at the syllable initial position (compare, for example, the syllable initial t̚ of the monosyllabic word t̚o:p ‘drink’ in the sentence 39 with the syllable final t̚ of the monosyllabic word nə:t ‘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.<sup>12</sup> The tenseness extends to full length in the case of open syllables (see, for example, the tensed vowels u: and e: of the respective open syllables tu: and fe: of the disyllabic words katu: ‘stay’ and ʔufe: ‘plural’ in sentences 25 and 36 respectively) and to half length in the case of closed syllables (see, for example, the half tensed vowels ə: and i: of the closed syllables of the monosyllabic words kə:n ‘get up’ and ci:n ‘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 i: and e: of the respective root syllables of the monosyllabic words ci:n ‘what’ and ne:n ‘past’ in sentences 24 and 39 respectively with the unstressed lax vowels i and e of the respective affixal syllables -si- and -se- of the word juə:nsise ‘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 tɑ- in the word tɑnã:ŋe ‘there’ in sentence 19a with ə of -ŋə in the word kaji:ŋe ‘go’ in sentence 30) or of suffixal ones in word medial position (compare, for example, ɑ of -ŋɑ- in the word xali:ŋase ‘prepare’ in sentence 34 with ə of -tə in the word tə:ŋtə ‘reach’ in sentence 32).

When nasalized, ɑ ‘low back unrounded vowel’ is found to become ã ‘low front unrounded vowel’ (compare, for example, ɑ of ma:t ‘sociative’ in sentence 25 with ã: of ʔã:cə? ‘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 ua: of the open syllable of the monosyllabic word cuə: ‘what’ of section 4.3.1) or to half length in closed syllable (see, for example, the diphthong ua: of the closed syllable juə:ŋ of the trisyllabic word juə:nsise ‘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 o:ə of the monosyllabic word ko:ən ‘child’ in the sentence 36) or with the final ones (see, for example, the diphthong oə: of the monosyllabic word tə:ək ‘toddy’ in sentence 39).

<sup>9</sup> What are given in round brackets against nasalized simple vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

<sup>10</sup> What are given in round brackets against complex vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

<sup>11</sup> What are given in round brackets against nasalized complex vowels are their equivalents in terms of laxness and tenseness in Muöt orthography.

<sup>12</sup> 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 Röpstorff (1875), **A Dictionary of the Nancowry Dialect of the Nicobarese Language** by De Röpstorff (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 Röpstorff (1875)

De Röpstorff (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.

Röpstorff (1875) <sup>13</sup>	Modern Muöt	IPA	Translation	Reference
katōm	Katōm	kaʈoːm	‘how many’	(p64)
kin-kee-en?	Chīn	ciːn	‘what’	(p109)
kahæ	Kahēñ	kaxēː	‘when’	(ibid.)
joa	Chū	cuː	‘where’	(ibid.)
kyouja?	Tai chua	ʈaːj cuaː	‘why’	(ibid.)

<sup>13</sup> Stands for the orthography made use of by De Röpstorff (1875).

#### 4.2. Treatment of interrogation by De Röpstorff (1884)

De Röpstorff (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.

Röpstorff (1884) <sup>14</sup>	Modern Muöt	IPA	Translation
Tiī	Chī	ci:	‘who’
Tiīn	Chīn	ciˈn	‘what’
Tiang tiī	Chöng chī	cəŋ ci:	‘whose’
Katōm	Katōm	kaɬəˈm	‘how many, how much’

(ppxvii-xviii)

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

<b>Röpstorff (1884)</b>	Tiīn	paiū
<b>Modern Muöt</b>	Chīn	payuh
<b>IPA</b>	ciˈn	paɟuːx
<b>Gloss</b>	what	man <sup>15</sup>
<b>Translation</b>	‘Who (what man) is there?’	

##### 6. Tiīn io me

<b>Röpstorff (1884)</b>	Tiīn	io	me
<b>Modern Muöt</b>	Chīn	yòk	měh
<b>IPA</b>	ciˈn	jɔːʔ	mɛː
<b>Gloss</b>	what	do want	you
<b>Translation</b>	‘What do you want?’		

##### 7. Tiīn wētié

<b>Röpstorff (1884)</b>	Tiīn	wē	tié
<b>Modern Muöt</b>	Chīn	vīk	cheūñ
<b>IPA</b>	ciˈn	viːʔ	cw:ə̃
<b>Gloss</b>	what	shall do	I
<b>Translation</b>	‘What shall I do?’		

<sup>14</sup> Stands for the orthography made use of by De Röpstorff (1884).

<sup>15</sup> Copula verb is found to have elided.

## 8. Tiang tī

<b>Röepstorff (1884)</b>	Tiang	tī
<b>Modern Muöt</b>	Chöng	chī
<b>IPA</b>	cəŋ	ci:
<b>Gloss</b>	poss	who <sup>16</sup>
<b>Translation</b>	'Whose is it?'	

## 9. Katōm kamehæwe kākāt

<b>Röepstorff (1884)</b>	Katōm	kamehæwe	Kă	kăt
<b>Modern Muöt</b>	Katōm	kamahēnvō	kōk	kōt
<b>IPA</b>	kaʈoːm	kamaxɛːvə	kəːʔ	kəːt
<b>Gloss</b>	how many	month	q	here <sup>17</sup>
<b>Translation</b>	'In how many months will you be here?'			

(ibid.)

As is seen, in Tiān paiū 'who (what man) is there?' the form Tiān is made use of to elicit information concerning human noun and as seen in Tiān io me 'what do you want?' as well as in Tiān wētié '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 Tī '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.

<b>Röepstorff (1884)</b>	<b>Modern Muöt</b>	<b>IPA</b>	<b>Translation</b>	<b>Reference</b>
Tiū	Chū	cu:	'where'	(p14)
Kahæ	Kahēn	kaxɛ:	'when'	(ppxx-xxi)
Katōm	Katōm	kaʈoːm	'how many'	(ibid.)

Further, he speaks about another interrogative form kă 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 kă ina kalāh oknōk omtōm

<b>Röepstorff (1884)</b>	Léat	kă	ina	kalāh	oknōk	omtōm
<b>Modern Muöt</b>	Leāt	kōk	inān	kalāh	hungōnk	umtūm
<b>IPA</b>	leɑːt	kəːʔ	ʔinā:	kalaːx	xuŋɔ̃ːk	ʔumtuːm
<b>Gloss</b>	perf	q	2du	taste	food	all
<b>Translation</b>	'Have you (two) tasted all the food?'					

The insights obtained from the present paper enable to consider the form kă in the sentences 9 and 10 not as an independent interrogative particle, but as a demonstrative particle.

<sup>16</sup> The subject and copula verb of the sentence are found to have elided.

<sup>17</sup> 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,

Man (1889) <sup>18</sup>	Modern Muöt	IPA	Translation
Chī	Chī	ci:	‘who’
Ten chī	Tin chī	tiːn ci:	‘whom’
Tai chī	Tai chī	taːj ci:	‘by whom’
Lamōngto-chī, Lamōngto-ten-chī	Lamōngtōchī, Lamōngtō tinchī	laməŋta ciː, laməŋta tiːnci:	‘from whom’
Chamang-ta-chī	Chamōngtō chī	caməŋta ci:	‘whose’
Chū <sup>19</sup>	Chū	cu:	‘where’
Chūa, Chūan, Chin, Kâ, Ka, Kan	Chūa, Chūan Chīn,  Ka	cuaː, cuaːn, ciːn,  ka	‘what’
Chun	Chī	ci:	‘which’
Chūan-shī	Chuānsi	cuaːnsi	‘why’
Kâ-shīn	Kasī	kasi:	‘like what’
Ka-rīshe	Karīse	kaɪːse	‘how much’
Karām, Katōm	Karām, Katōm	kaɪaːm, kaɪoːm	‘how many’
Chin-lēang-dīo	Chin leāng rīöv	ciːn leaːŋ ɹiːöv	‘what else’

(ppxxvi-xxvii)

As evidence for their usage in utterances, he provides two sentences which are reproduced below as 11 and 12.

#### 11. Chamangta chī

<b>Man (1889)</b>	Chamangta chī	en eñh
<b>Modern Muöt</b>	Chamōngtō chī	nēk eñh
<b>IPA</b>	caməŋta ci:	nɛːʔɛːx
<b>Gloss</b>	whose	this <sup>20</sup>
<b>Translation</b>	‘Whose is this?’	

<sup>18</sup> Stands for the orthography made use of by Man (1889).

<sup>19</sup> The form is not found listed in the notes, but found listed in the dictionary part (P122).

<sup>20</sup> See footnote 15.

## 12. Lamòngto ten chī en eñh

<b>Man (1889)</b>	Lamòngto ten chī	en eñh
<b>Modern Muöt</b>	Lamòngtō tinchī	nēk eñh
<b>IPA</b>	laməŋta tɨŋci:	nɛːʔɛːx
<b>Gloss</b>	from whom	this <sup>21</sup>
<b>Translation</b>	'From whom (did you get) this?'	

(ibid.)

The present paper is of the view that all these twenty forms seem to be formed out of just four interrogative markers, namely, chī, chūa, chūan, chin and ka-.

## 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:

<b>Man (1889)</b>	<b>Modern Muöt</b>	<b>IPA</b>	<b>Translation</b>
Kâhē	Kahēñ	kaxɛː	'when (of future only)'
Kâhē-ta-shī	Kahēñ tō sī	kaxɛː t̚asi	'when (of a coming event)'
Kâhē-ta-tai	Kahēñ tōtai	kaxɛː t̚at̚aj	'when (future, of making, giving, &c.)'
Kâhē-ngalâh	Kahēñ ngō lah	kaxɛːŋalaːx	'when (future, of travelling)'
Kâhē-nga	Kahēñ ngō	kaxɛːŋə	'when (of past time, in reference to death, sickness)'
Kâhē-ngashī	Kahēñ ngō si	kaxɛːŋasi	'when (of some past event)'
Kâhē-ngatai	Kahēñ ngō tai	kaxɛːŋat̚aj	'when (of past time, in reference to making, giving, &c.)'
Kâhē-tashe, Kâhē-tarit, Kâhē-ngashe	Kahēñ tō se, Kahēñ tō ret, Kahēñ ngō se	kaxɛː t̚ase, kaxɛː t̚aɛt kaxɛːŋase	'when (of any past event)'
hē-chūa	Hēñ chūa	xɛːcua:	'when (at what time)'
Kâ-shanī-tashe	Ka sanīk tō se	kasaniːʔt̚ase	'how long (time)'
Kâ-shanī-latōh	Kasanīk lō tōh	kasaniːʔlat̚ox	'how old'
Kâ-rūa-hanga-heng	Kareūöt hōngō hěng	kaɽuəːtxaŋaxeŋ	'how long ago (today)'
Kâ-rūa-hanga-yan	Kareūöt hōngō yōn	kaɽuəːtxaŋajən	'how long ago (in the past)'
Karām-shuâ	Karām tō suâ	kaɽaːm̚tasua:	'how many times'
Kâ-inôa-atô	Ka inôaŋs ô tō	kaʔinôaːsəʔt̚o	'how soon'
Kâ-rūala-heng	Ka reūölō hěng	kaɽuəːlaxeːŋ	'what time (of today) is it?'
Kâ-rūala-hatôm	Ka reūölō hatôm	kaɽuəːlaxat̚om	'what time (of night) is it?'

(ppxxxii-xxxiii)

Of these, except hē-chūa '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:

<sup>21</sup> See footnote 15.

Man (1889)	Modern Muöt	IPA	Translation
Ka-rüa-hanga	Ka reüöt höngö	karuːətɕaŋə	‘how far’
Ka-ra-müa-hanga	Karameüöthöngö	kaɾamurːətɕaŋə	‘how near’

(pxxxiv)

The insights obtained from the present paper enable to view the form Ka-ra-müa-hanga ‘how near’ as the agentivized counterpart of Ka-rüa-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.

Man (1889)	Modern Muöt	IPA	Translation
Ka-rī	Ka rī	kariː	‘how big’
Ka-ra-mī	Karamī	kaɾamiː	‘how small’
Ka-rī-fāp	Karī fāp	kaiː faːp	‘how fat; how wide (of ship)’
Ka-ra-mī-fāp	Karamī fāp	kaɾamiː faːp	‘how narrow (of ship)’
Karī-tāk	Karī tak	kaiː ʈaːk	‘how wide (of plank)’
Karamī-tāk	Karamī tak	kaɾamiː ʈaːk	‘how narrow (of plank)’
Ka-chin-yāwa	Kachin yāvō	kacinjaːvə	‘how deep’
Ka-rüat	Kareüöt	karuːət	‘how long’
Kara-müat	Karameüöt	kaɾamurːət	‘how short (inanim.)’
Ka-rüala	Kareüölö	karuːələ	‘how high’
Ka-rüala-kōi	Kareüölö kui	karuːəlakuːj	‘how tall’
Kara-müala-kōi	Karameüölö kui	kaɾamurːəlakuːj	‘how short (anim.)’
Karüa-hanga-tai	Kareüö höngö tai	karuːətɕaŋəʈaːj	‘how far off (of object shot, speared, &c)’
Karī-tare-she, Karī-tat-she	Karī tö re se, Karī tet se	kaiːʈaːese, kaiːʈetse	‘how much more remains’
Karī-hata-she	Karī hö tö se	kaiːxatase	‘how much has been paid, delivered, &c’
Katōm-tare	Katōm tö re	kaʈoːmʈaːe	‘how many more’
Kâ-yan	Kayön	kaʈəːn	‘how, in what state of health’
Kâ-shin-mush-tai	Kasin müs tai	kasinmuːstəːj	‘in what style’
Ka-shī	Kasī	kasiː	‘what kind, what sort’

(ppxxxiv-xxxv)

To illustrate the usage of these forms in utterances, he has listed a sentence and the same is reproduced below as 13.

#### 13. Kâ yan ka an ta-linheñ

<b>Man (1889)</b>	Kâ	yan	ka an	ta-linheñ
<b>Modern Muöt</b>	Ka	yön	kök añn	tö linheñ
<b>IPA</b>	ka	jəːn	kəːʔ ʔãːn	ʈalinxɛː
<b>Gloss</b>	how	health	he	today <sup>22</sup>
<b>Translation</b>	‘How is he today?’			

(ibid.)

<sup>22</sup> 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)’, Kara-müat ‘how short (inanim.)’ and Kara-müala-kōi ‘how short (anim.)’ as agentivized counterparts of Ka-rī ‘how big’, Ka-rī-fāp ‘how fat; how wide (of ship)’, Karī-tāk ‘how wide (of plank)’, Ka-rüat ‘how long’ and Ka-rüala-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,

Man (1889)	Modern Muöt	IPA	Translation
Chūan-shī, Chūan-lāng-ngashī, Chūan-lāng-ngitô, Chūan-hän-ngashī, Chūan-hän-ngitô, Chūan-wī, Chūan-tai-chūa	Chūan si, Chūan löng ngö si, Chūan löng ngö tò, Chūan hän ngö si, Chūan hän ngö tò, Chūan vīk, Chūan tai chūa	cua'nsi, cua'nləŋŋasi, cua'nləŋaɬɔ, cua'nxã:ŋasi, cua'nxã:ŋaɬɔ, cua'nvi'ʔ, cua'nɬa:jcua:	‘why’

(pxxxv)

And to illustrate their usage in utterances he gives, what is reproduced below as 14, as the sample sentence.

#### 14. Chūan wī meñ ta watshī meñ met dalngatô

<b>Man (1889)</b>	Chūan	wī	meñ	ta watshī	meñ	met	dalngatô
<b>Modern Muöt</b>	Chūan	vīk	mēñ	tō vāt si	mēñ	m'it	rāl ngö tò
<b>IPA</b>	cua'n	vi'ʔ	mẽ:	ɬa va't si	mẽ:	m'it	.ɾaŋaɬɔ
<b>Gloss</b>	why	behave	svs	such way	you	aren't you	ashamed
<b>Translation</b>	‘Why do you behave in such a way? Aren't you ashamed of yourself?’						

(ibid)

#### 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).<sup>23</sup> 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. Tàu meñ kâ an

<b>Man (1889)</b>	Tàu	meñ	kâ an	
<b>Modern Muöt</b>	Tāv	mēñ	kōk ann	
<b>IPA</b>	ɬa:v	mẽ:	kə'ʔ ʔã'n	↗
<b>Gloss</b>	younger brother	your	he	sf <sup>24</sup>
<b>Translation</b>	‘Is he your younger brother?’			

<sup>23</sup> For a similar passing reference, see (p xi) of Introductory Remarks.

<sup>24</sup> See footnote 15.

## 16. Chī yô haiyüan

<b>Man (1889)</b>	Chī	yô	haiyüan
<b>Modern Muöt</b>	Chī	yòk	hayeüön
<b>IPA</b>	ci:	jɔːʔ	xajɯəːn
<b>Gloss</b>	who	future	hunt pigs
<b>Translation</b>	‘Who is going to hunt pigs?’		

## 17. Kâhē tashe meñ dāk

<b>Man (1889)</b>	Kâhē tashe	meñ	dāk
<b>Modern Muöt</b>	Kahēñ tō se	mēñ	reuk
<b>IPA</b>	kaxɛːtase	mɛː	ɹuːk
<b>Gloss</b>	when	you	come
<b>Translation</b>	‘When did you come?’		

(pplv-lvi)

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<sup>25</sup> and 19a-b<sup>26</sup> can be made use of for understanding such a function.<sup>27</sup>

## 18a. Öön in kōön mēñ inkānō in Merī

<b>Modern Muöt</b>	öön	in	kōön mēñ inkānō	in	Merī
<b>IPA</b>	ʔəː	ʔiːn	koːən mɛː ʔinkaːnə	ʔiːn	meiː
<b>Gloss</b>	cop	prox	your daughter	prox	Mary
<b>Translation</b>	Mary is your daughter				

## 18b. Öön in kōön mēñ inkānō in Merī?

<b>Modern Muöt</b>	öön	in	kōön mēñ inkānō	in	Merī	
<b>IPA</b>	ʔəː	ʔiːn	koːən mɛː ʔinkaːnə	ʔiːn	meiː	ʔ?
<b>Gloss</b>	cop	prox	your daughter	prox	Mary	sfī
<b>Translation</b>	Is Mary your daughter?					

## 19a. Yuāngsise uksök tōngānngē in Sipā nen

<b>Modern Muöt</b>	yuāngsise	uksök	tōngānngē	in	Sipā	nen
<b>IPA</b>	juɑːŋsise	ʔuksək	tɔŋɑːŋɛ	ʔiːn	sipaː	neːn
<b>Gloss</b>	prog	stand	there	prox	sheeba	pst
<b>Translation</b>	‘Sheeba was standing there’					

<sup>25</sup> The sentences are unmarked for present tense.

<sup>26</sup> It is because of the free word-order the marker for past tense occur sentence finally.

<sup>27</sup> The sentences 18a-b can also be found occurring with the elision of copula verb.

## 19b. Yuāngsise uksök tōngānge in Sipā nen?

<b>Modern Muöt</b>	yuāngsise	uksök	tōngānge	in	Sipā	nen	
<b>IPA</b>	juɑːŋsise	ʔuksəːk	tʰɑŋɑːŋe	ʔiːn	sipaː	neːn	ʔ?
<b>Gloss</b>	prog	stand	there	prox	sheeba	pst	sfi
<b>Translation</b>	‘Was Sheeba standing there?’						

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 ch-interrogative words<sup>28</sup> and ka- interrogative words.

## 5.2.1. Sentence with ch-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 chīn ‘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 öön kök önn inkōnyö?

<b>Modern Muöt</b>	chī	öön	kök	önn inkōnyö
<b>IPA</b>	ciː	ʔöː	kəːʔ	ʔöːn ʔinkəːŋə
<b>Gloss</b>	who	cop	dist <sub>3</sub>	he <sup>29</sup>
<b>Translation</b>	‘Who is he?’			

## 21. Chī öön tö ki ānn e in nyī önn inkōnyö?

<b>Modern Muöt</b>	chī	öön	tö ki ānn e	in	nyī önn inkōnyö
<b>IPA</b>	ciː	ʔöː	tə kiʔāːne	ʔiːn	niː ʔöːn ʔinkəːŋə
<b>Gloss</b>	which	cop	among these	prox	his house <sup>30</sup>
<b>Translation</b>	‘Which is his house?’				

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

<sup>28</sup> ch represents voiceless palatal consonant in Muöt orthography.

<sup>29</sup> The sentence can also be found occurring with the elision of copula verb.

<sup>30</sup> See footnote 29.

inflected for case relationship, here the possessive,<sup>31</sup> and the grammatical function of such case inflected form as object can be seen from the sentence 22 listed below.

22. Ān chōk chī öön ufē kök e?

<b>Modern Muöt</b>	ān chōk	chī	öön	ufē kök e
<b>IPA</b>	ʔā:cəʔ	ci:	ʔö:	ʔufe: kə:ʔε
<b>Gloss</b>	arrow	q	cop	those <sup>32</sup>
<b>Translation</b>	‘Whose arrows are those?’			

### 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 öön kök önn inkānō?

<b>Modern Muöt</b>	chīn	öön	kök	önn inkānō
<b>IPA</b>	ci:n	ʔö:	kə:ʔ	ʔö:n ʔinkɑ:nə
<b>Gloss</b>	who	cop	dist <sub>3</sub>	she <sup>33</sup>
<b>Translation</b>	‘Who is she?’			

24. Chīn öön in leāng mēn?

<b>Modern Muöt</b>	chīn	öön	in	leāng mēn
<b>IPA</b>	ci:n	ʔö:	ʔi:n	leɑ:ŋ mē:
<b>Gloss</b>	who	cop	prox	your name <sup>34</sup>
<b>Translation</b>	‘What is your name?’			

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 yök nö kateū in Mark tō öt Luang?

<b>Modern Muöt</b>	mat	chī	yök	nö	kateū	in	Mark	tō öt Luang
<b>IPA</b>	mɑ:t	ci:	jɔ:ʔ	nə	kaʔu:	ʔin	mɑ:ɪk	tə ʔə:t lua:ŋ
<b>Gloss</b>	soc	q	fut	svs	stay	prox	Mark	in kondul
<b>Translation</b>	‘With who will Mark stay in Kondul?’							

### 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.

<sup>31</sup> 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.

<sup>32</sup> The sentence is unmarked for present tense and can also be found occurring with the elision of copula verb.

<sup>33</sup> See footnote 29.

<sup>34</sup> See footnote 29.

## 26. Chuān yòk topre in mēn?

<b>Modern Muöt</b>	chuān	yòk	topre	in	mēn
<b>IPA</b>	cuɑːn	joːʔ	toːpɾe	ʔiːn	mɛːʔ
<b>Gloss</b>	what	will	drink	prox	you
<b>Translation</b>	‘What will you drink?’				

## 27. Chuān yuāngsise vīk ānn ufē?

<b>Modern Muöt</b>	Chuān	yuāngsise	vīk	ānn	ufē
<b>IPA</b>	cuɑːn	juɑːŋsise	viːʔ	ʔãːn	ʔufeː
<b>Gloss</b>	what	prog	do	dist <sub>1</sub>	they <sup>35</sup>
<b>Translation</b>	‘What are they doing?’				

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 kōk nyi mēn?

<b>Modern Muöt</b>	chū	ôt	kōk	nyi mēn
<b>IPA</b>	cuː	ʔoːt	kəːʔ	niː mɛː
<b>Gloss</b>	where	exis	dist <sub>3</sub>	your house <sup>36</sup>
<b>Translation</b>	‘Where is your house?’			

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 nō chīm kōk kinyônm?

<b>Modern Muöt</b>	chuānsi	nō	chīm	kōk	kinyônm
<b>IPA</b>	cuɑːnsi	nə	ciːm	kəːʔ	kiŋõːmʔ
<b>Gloss</b>	why	svs	cry	dist <sub>3</sub>	baby <sup>37</sup>
<b>Translation</b>	‘Why does the baby cry?’				

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 kahēn ‘when (in terms of period of time)?’, sentences with karām hōngō ‘when (in

<sup>35</sup> See footnote 25.

<sup>36</sup> The sentence is unmarked for present tense and can also be found occurring with the elision of existential verb.

<sup>37</sup> 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 karīhōt ‘how much (liquid)?’ and sentences with kareūōt ‘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 kahēn

They are sentences that begin with the interrogative word ka-hēn ‘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. Kahēn mēn yòk kayīngō nō hayeūōn in mēn?

<b>Modern Muöt</b>	kahēn	mēn	yòk	kayīngō	nō	hayeūōn	in	mēn
<b>IPA</b>	kaxɛː	mɛː	jɔːʔ	kajiŋə	nə	xajuəːn	ʔin	mɛː
<b>Gloss</b>	when	svs	will	go	purp	pig hunting	prox	you <sup>38</sup>
<b>Translation</b>	‘When will you go for pig hunting?’							

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 kahēn ‘when (in terms of period of time)?’ is found functioning as temporal adverb.

### 5.2.2.2. Sentence with karām hōngō

They are sentences that begin with the interrogative word karām hōngō ‘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. Karāmhōngō kōōngō in mēn tō òal hakī mōök?

<b>Modern Muöt</b>	Karāmhōngō	kōōng	in	mēn	tō òal hakī	mōök
<b>IPA</b>	kaɪaːmxaŋə	kəŋ	ʔin	mɛː	tə ʔaːl xakiː	məːk
<b>Gloss</b>	when	get up	prox	you	in the morning	will <sup>39</sup>
<b>Translation</b>	‘When will you get up in the morning?’					

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 karāmhōngō ‘when (in terms of hour)?’ is found to function 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 chōn ròh chōn tōng tō kōk matāi mēn in chōn?

<b>Modern Muöt</b>	kasī	chōn	ròh	chōn	tōng tō	kōk	matāi mēn	in	chōn
<b>IPA</b>	kasiː	cɔː	.ɔːx	cɔː	təŋtə	kəːʔ	matəːj mɛː	ʔin	cɔː
<b>Gloss</b>	how	svs	can	svs	reach	dist <sub>3</sub>	your island	prox	I <sup>40</sup>
<b>Translation</b>	‘How can I reach your island?’								

<sup>38</sup> The sentence can also be found occurring with the elision of serial verb subject.

<sup>39</sup> 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 tə ʔaːl xakiː ‘in the morning’ occur after the subject mɛː ‘you’ of the sentence.

<sup>40</sup> 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 òt in mēñ?

<b>Modern Muöt</b>	kasī	òt	in	mēñ
<b>IPA</b>	kasi:	ʔɔːt	ʔiːn	mɛː
<b>Gloss</b>	how	exis	prox	you <sup>41</sup>
<b>Translation</b>	‘How are you?’			

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 kōk insūön larōm in mēñ?

<b>Modern Muöt</b>	katāi	halīngöse	kōk	Insūön larōm	in	mēñ
<b>IPA</b>	kaʈaːj	xaliːŋase	kəːʔ	ʔinsuːən laɔːm	ʔiːn	mɛː
<b>Gloss</b>	how	prepare	dist <sub>3</sub>	pandanus bread	prox	you <sup>42</sup>
<b>Translation</b>	‘How do you prepare pandanus bread?’					

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 kaʈaːj ‘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 òt tāi in mēñ ufē kōk kōön?

<b>Modern Muöt</b>	karāmkui	òt	tāi	in	mēñ	ufē	kōk	kōön
<b>IPA</b>	kaɾaːmkuj	ʔɔːt	ʈaːj	ʔiːn	mɛː	ʔufeː	kəːʔ	kəːən
<b>Gloss</b>	how many	exis	dat	prox	2sg	pl	dist <sub>3</sub>	child <sup>43</sup>
<b>Translation</b>	‘How many children do you have?’							

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 karīsě

They are sentences that begin with the interrogative word karīsě ‘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 the quantity of human noun can be illustrated by the sentence 36 given below.

<sup>41</sup> See footnote 29.

<sup>42</sup> See footnote 25.

<sup>43</sup> See footnote 36.

36. *Karīsē* ô<sub>t</sub> t<sub>ai</sub> in m<sub>ēn</sub> uf<sub>ē</sub> k<sub>ōk</sub> k<sub>ōn</sub>?

<b>Modern Muöt</b>	karīsē	ô <sub>t</sub>	t <sub>ai</sub>	in	m <sub>ēn</sub>	uf <sub>ē</sub>	k <sub>ōk</sub>	k <sub>ōn</sub>
<b>IPA</b>	ka.i:se	ʔoːt	taːj	ʔiːn	mɛː	ʔufeː	kəːʔ	koːən
<b>Gloss</b>	how many	exis	dat	prox	2sg	pl	dist <sub>3</sub>	child <sup>44</sup>
<b>Translation</b>	‘How many children do you have?’							

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. *Karīsē* ô<sub>t</sub> ki ā<sub>nn</sub> n<sub>ôt</sub> t<sub>ō</sub> ā<sub>nne</sub>?

<b>Modern Muöt</b>	karīsē	ô <sub>t</sub>	ki	ā <sub>nn</sub>	n <sub>ôt</sub>	t <sub>ō</sub> ā <sub>nne</sub>
<b>IPA</b>	ka.i:se	ʔoːt	ki	ʔãːn	nɔːt	ta ʔãːne
<b>Gloss</b>	how many	exis	pl	dist <sub>1</sub>	pig	there <sup>45</sup>
<b>Translation</b>	‘How many pigs are there?’					

38. *Karīsē* m<sub>ēn</sub> umk<sub>ōm</sub>ō in hinyu<sub>āh</sub> in m<sub>ēn</sub>?

<b>Modern Muöt</b>	karīsē	m <sub>ēn</sub>	umk <sub>ōm</sub> ō	in	hinyu <sub>āh</sub>	in	m <sub>ēn</sub>
<b>IPA</b>	ka.i:se	mɛː	ʔumkoːmə	ʔin	xinjuaːx	ʔin	mɛː
<b>Gloss</b>	how much	svs	receive	prox	salary	prox	you <sup>46</sup>
<b>Translation</b>	How much salary you receive?’						

While seeking to elicit the quantity, the interrogative word *karīsē* ‘how many?, how much?’ in the sentences 37 and 38 is found to function as grammatical adverb.

5.2.2.7. Sentence with *karīhōt*

They are sentences that begin with the interrogative word *karīhōt* ‘how much (liquid)?’ and seek to elicit information concerning quantity of liquid nouns. The sentence 39 given below would illustrate such a function.

39. *Karīhōt* m<sub>ēn</sub> n<sub>ēn</sub> top in t<sub>ōk</sub> minye<sub>ūi</sub> in m<sub>ēn</sub>?

<b>Modern Muöt</b>	karīhōt	m <sub>ēn</sub>	n <sub>ēn</sub>	top	in	t <sub>ōk</sub>	minye <sub>ūi</sub>	in	m <sub>ēn</sub>
<b>IPA</b>	ka.i:xət	mɛː	neːn	toːp	ʔin	toːk	minjuːj	ʔin	mɛː
<b>Gloss</b>	how much	svs	pst	drink	prox	toddy	yesterday	prox	you <sup>47</sup>
<b>Translation</b>	‘How much toddy did you drink yesterday?’								

As is seen above, sentence 39 seeks to elicit the quantity of toddy, a liquid noun and while doing so, the interrogative word *karīhōt* ‘how much (liquid)?’ functions as grammatical adverb.

5.2.2.8. Sentence with *kareūōt*

They are sentences that begin with the interrogative word *kareūōt* ‘how long?’ and seek to elicit information concerning the length of nouns. The following sentence 40 can be taken for illustrating such a function.

<sup>44</sup> See footnote 36.

<sup>45</sup> 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 *ta ʔãːne* ‘there’ occur sentence finally.

<sup>46</sup> The sentence is unmarked for present tense and can also be found occurring with the elision of serial verb subject.

<sup>47</sup> See footnote 38.

40. Kareūöt neḵ òal matāi neḵ eñh?

<b>Modern Muöt</b>	kareūöt	neḵ	òal matāi	neḵ eñh
<b>IPA</b>	kɑɪuəˈt	nɛːʔ	ʔɑːl matɑːj	nɛːʔ ɛːxʔ
<b>Gloss</b>	how long	prox	village	this <sup>48</sup>
<b>Translation</b>	‘How long is this village?’			

As is seen above, sentence 40 seeks to elicit the length of a village and during the process the interrogative word *kareūöt* ‘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**

Markers of interrogation		
Sentence final rising intonation	Interrogative word	
	Ch-interrogative word	Ka-interrogative word
↗	Chī Chīn Chuān Chū Chuānsi	Kahēn Karāmhōngō Kasī Katāi Karāmkui Karīsē Karīhöt Kareūöt

## 7. Finding

1. All the three previous works reviewed here, Röpstorff (1875), Röpstorff (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, Röpstorff (1875), Röpstorff (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 Röpstorff (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 Röpstorff (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, *ka* and *kan* as variants of *kā* (cf. section 4.3.1). The present day language doesn’t seem to have such independent interrogative markers.

<sup>48</sup> 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 graphological, phonological, morphological, syntactic and semantic levels of linguistic investigation. The use of Ti and æ by Röpstorff (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 graphological discrepancy. Likewise, the lacuna being observed in Röpstorff (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 Röpstorff (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 meñ ta watshī meñ met dalngatō ‘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 meñ ‘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-mūa-hanga (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

- Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics* (sixth edition). Blackwell Publishers.
- De Röpstorff Frederick Adolph. 1875 (1987). *Vocabulary of Dialects spoken in Nicobar and Andaman Isles* (second edition). New Delhi, Madras: Asian Educational Services.
- \_\_\_\_\_. 1884. *A Dictionary of the Nancowry Dialect of the Nicobarese Language*. Calcutta: Home Department Press.
- Diffloth, Gerard. 1974. Austro-Asiatic Languages. *Encyclopedia of Britannica*, Vol.2.
- Hartman, PRK et al. 1973. *Dictionary of Language and Linguistics*. London: Applied Science publishers Ltd.
- Justin, Anstice. 1990. *The Nicobarese*. Calcutta: Seagull Books.
- Kloss, Boden C. 1902 (1971). *Andamans and Nicobars*. Delhi-7, India: 18-D Kamala Nagar, Vivek Publishing House.
- Lewis, M Paul. 2009. *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Tex: SIL International.
- Man, Edward Horace. 1889 (1975). *A Dictionary of the Central Nicobarese Language*. Delhi: Sanskaran Prakashak.
- Pearce, Michael. 2007. *The Routledge Dictionary of English Studies*. London and New York: Routledge.
- Radhakrishnan, R. 1981. *The Nancowry word, Phonology, Affixal morphology and Roots of A Nicobarese Language*. Linguistic Research Inc. P.O. Box 5677, Station ‘L’ Edmonton, Alberta, Canada T6C 4G1.
- \_\_\_\_\_. 2010. “Personal Pronouns in Muöt”. In Nagaraja K S and Kashyap Mankodi (eds.) *Austroasiatic Linguistics*. Mysore, India: Central Institute of Indian Languages.
- \_\_\_\_\_. 2011. “Demonstratives in Muöt”. In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp. 89-99.
- Rajasingh, v. R. 2013. “Negators in Muöt”. *Mon-Khmer Studies*, Vol.41 pp46-59.

### Abbreviations

2sg:	Second person singular
cf.:	Compare with
cop:	Copula verb
dat:	Dative case
dist <sub>1</sub> :	Distal demonstrative 1
dist <sub>3</sub> :	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 Srichampa 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:

Publication No. 1 of the Linguistic Circle of Saigon

Published jointly by The Linguistic Circle of Saigon and The Summer Institute of Linguistics 1964

Foreword by Nguyen Dinh Hoa, Director of Cultural Affairs, Ministry of National Education

Introduction by David D. Thomas

Transformational Paradigms of Bahnar Clauses, John E. Banker

Word Classes in Brou, John D. Miller

The Substantive Phrase in Brou, Carolyn P. Miller

Personal Pronouns in Pacoh, Sandra K. Watson

Bahnar Affixation, Elizabeth M. Banker

Bahnar Reduplication, Elizabeth M. Banker

Pacoh Phonemes, Richard Watson

A Survey of Austroasiatic and Mon-Khmer Comparative Studies, David D. Thomas

Nguyen Dinh Hoa wrote in his Foreword:

The Linguistic Circle of Saigon is happy to introduce its first publication, *Mon-Khmer Studies I*, jointly sponsored by the LCS and the Summer Institute of Linguistics.

The Linguistic Circle of Saigon was launched very informally in 1959 with the help of faculty members of the University of Saigon and their American colleagues of the S.I.L. Vietnam Branch. Thanks to Dr. Nguyen-Quang-Trinh, then Rector of the University, the group was able to hold monthly meetings at the Rectorate, and under the chairmanship of Dr. Le-van-Ly papers were read in Vietnamese, English and French at the informal gatherings of linguists and language teachers in the area. Attendance has been growing steadily, and it is hoped that as linguistics is here to stay the tradition of holding periodic meetings and hearing research papers will be maintained as part of Saigon's scholarly activities.

The articles contained in this special volume speak for themselves; their authors, hard-working, promising linguistic scientists, can be rightly proud of their fine contributions to linguistics in general and Vietnamese linguistics in particular.

—January 1964

As David Thomas notes in his Introduction, this first issue of *MKS* represented no more than the output of a linguistic workshop he had conducted in Hue (Vietnam) in 1963, but now in historical perspective, we observe that it signaled the beginning of a renewed effort in the documentation and study of Mon-Khmer (and other Southeast Asian) languages. Indeed, the origins of the Journal carry the seeds of many of the aspects of *MKS* fifty years later.

## 2. A Cooperative Effort

From the get-go *MKS* was to be the property of any and all who would lend a hand in carving out a space for the study of this little-known language family. Our good friend and linguist, Professor Nguyen Dinh Hoa, and colleagues at the University of Saigon and Ministry of Education were steadfast in their commitment early on to SIL's fledgling early steps to study the minority languages.

In later years I remember Dave's extensive efforts to establish a partnership for *MKS* with the University of Hawaii, which continued for a several years. The home that *MKS* has enjoyed at Mahidol University and its Institute of Language and Culture for Rural Development has been especially meaningful and stable. And the recent additional association with the Australian National University has further enhanced prospects for the future of the Journal.

### 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 MKS 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. I like the words of wisdom attributed to linguist, John R. “Haj” Ross, who having been through most of the generativist wars, said (approximately), “Let's try to imagine some way in which our opponent might just be right.”

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-Thư 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., *cái ý kiến, thật âu yếm*); *s1* and *s2* are the first and second syllables of a bisyllabic word/chunk (e.g., *ý kiến, âu yếm*). 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  $f_0$  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 tend to have a series of iambs 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 other 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  $F_0$  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 basic of auditory signal and spectrogram: cues for segmentation included final lengthening, pause,  $F_0$  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. *và* [and], *lại* [again], *ấy*[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*, *thật 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

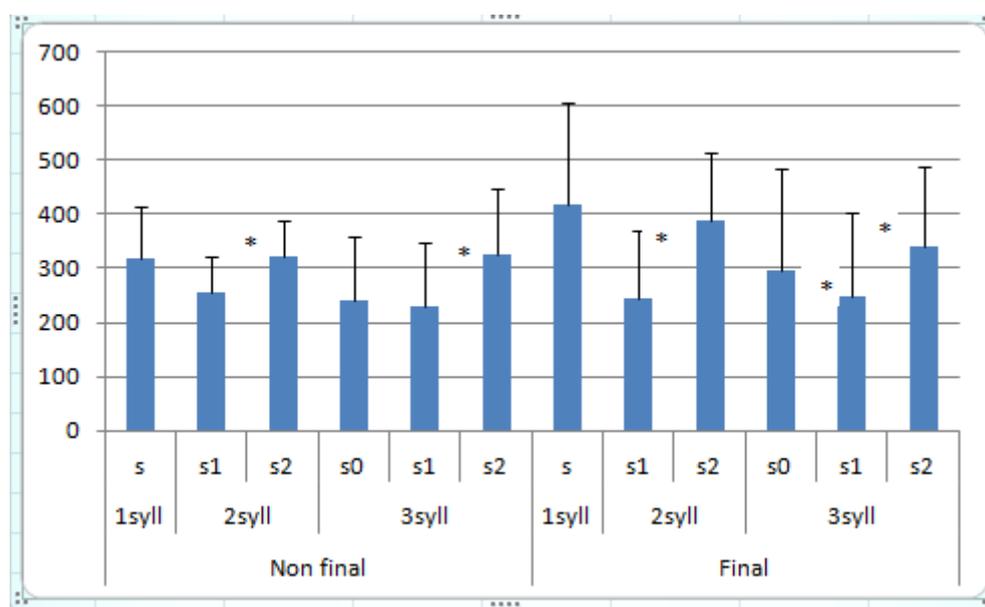
### 2.6.1. Prosodic units

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 monosyllable 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) 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.



**Figure 1:** Mean duration (ms). s, s0, s1 and s2 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  $p < 0.01$

### 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 (s0 and s1). Within bisyllabic words/chunks, the second syllables had stronger intensity than the first syllables. Within three-syllable phrases/chunks, the first syllable (s0) of a three-word phrase was not significantly different from the first syllable (s1) of the bisyllabic words and the second syllable (s2) also had stronger intensity than the first syllable (s1). 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.

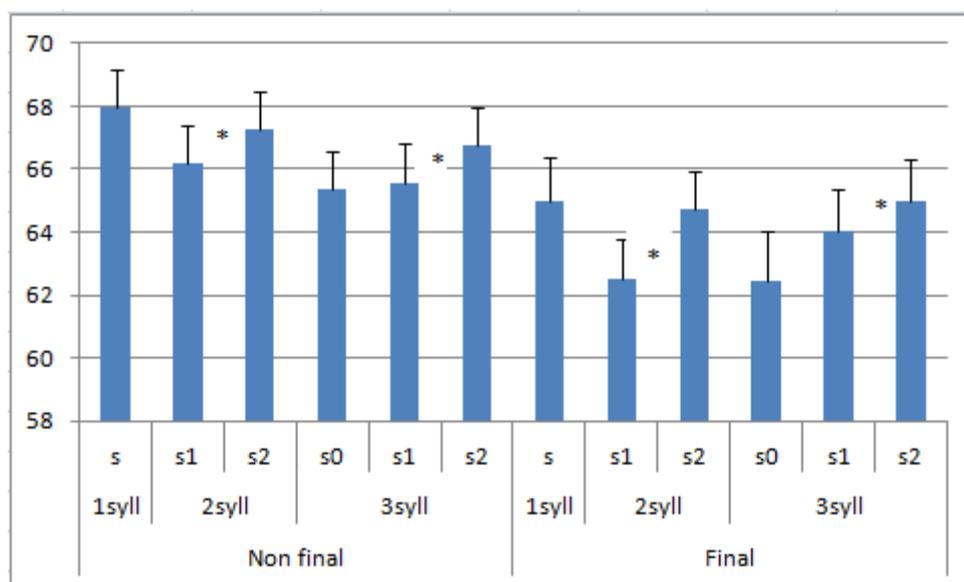


Figure 2: Mean intensity (dB). s, s0, s1 and s2 are syllable positions in segmented chunks. 1syll, 2syll, and 3syll: number of syllables in chunks. Non final and final are positions of chunks in utterances. The symbol \* means  $p < 0.1$ .

## 2.7. Discussion and conclusion

In summary the duration and intensity results mirror each other in indicating an iambic pattern ( $s_2 > s_1$ ) 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 ( $s_2 > s_1$ ) or tri-syllabic ( $s_2 > s_0 \sim s_1$ ) 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 Huynh (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

- Boersma P. & Weenink D. (2007). Praat: doing phonetics by computer (Version 4.5.18) [Computer program]. Retrieved March 30, 2007, from <http://www.praat.org/>
- Cao, X. H. (2003). *Vietnamese: issues of phonology, grammar and semantics*. Education press.
- Chaudhary, C. C. (1983). Word stress in Vietnamese: A preliminary investigation. *Indian Linguistics* 44:1-10.
- Do, T. D. (1986). *Elements pour une e'tude comparative de l'intonation en Francais et en vietnamien: L'accent de mots en vietnamien*. Memoire de DEA (Universite' de Paris 3 ILPGA).

- Durand, M. Maurice, and Nguyen, Tran Huan (1985). *An Introduction to Vietnamese Literature*, translated from the French by D.M. Hawke. New York: Columbia University Press.
- Gsell, R. (1980). Remarques sur la structure de l' espace tonal en Vietnamien du sud (Parler de Saigon). *Cahiers d'etudes Vietnamiennes*, 4 (Universite' Paris 7).
- Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds.), *Phonetics and Phonology, Vol 1: Rhythm and Meter*. San Diego: Academic Press. pp. 201-260.
- Hoàng, T. and Hoàng, M. (1975). Remarques sur la structure phonologique du Vietnamien. *Essais Linguistiques. Etudes Vietnamiennes* 40.
- Jones, R. B. and Huynh, S. T. (1960). *Introduction to Spoken Vietnamese*. Washington, D.C
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications
- Nguyễn, D. L. (1970). *A contrastive phonological analysis of English and Vietnamese* (Pacific Linguistics Series, No 8). Canberra: Australian National University.
- Nguyễn, T.A.T. and Ingram J.C. (2007a). Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America*, 112 (3), 1746-1757.
- Nguyễn, T.A.T. and Ingram J.C. (2007b). Stress and tone Sandhi in Vietnamese reduplications. *Mon-Khmer Studies*, 37, 15-40.
- Nguyễn, T.A. T. (2010). Rhythmic pattern and corrective focus in Vietnamese polysyllabic words. *Mon-Khmer Studies*, 39, 1-28.
- Nguyễn, T.A. T. (2013). Acoustic and perceptual correlates of Vietnamese folk poetry rhythmic structure. *Journal of the Southeast Asian Linguistic Society*, 6, 54-77.
- Schiering, R., B. Bickel, & K. Hildebrandt. (2010). "The prosodic word is not universal, but emergent," *Journal of Linguistics* 46, 657-709
- Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed), *Nordic Prosody II*, Trondheim: TAPIR.
- Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and M.-L. Kean (eds), *Juncture*, Anna Libri, PO Box 876, Saratoga, Calif. 107-129
- Thompson, L. (1965). *A Vietnamese Reference Grammar*. Honolulu: University of Hawaii Press.

### Appendix 1: the first excerpt of a short story

TÔI ĐI HỌC, a short story by Thanh Tịnh

Hàng năm cứ vào cuối thu, lá ngoài đường rụng nhiều và trên không có những đám mây bàng bạc, lòng tôi lại nao nức những kỷ niệm hoang mang của buổi tựu trường.

Tôi không thể nào quên được những cảm giác trong sáng ấy nảy nở trong lòng tôi như mấy cành hoa tươi mỉm cười giữa bầu trời quang đãng.

Những ý tưởng ấy tôi chưa lần nào ghi lên giấy, vì hồi ấy tôi không biết ghi và ngày nay tôi không nhớ hết. Nhưng mỗi lần thấy mấy em nhỏ rụt rè núp dưới nón mẹ lần đầu tiên đến trường, lòng tôi lại tung bùng rộn rã.

Buổi sáng mai hôm ấy, một buổi mai đầy sương thu và gió lạnh. Mẹ tôi âu yếm nắm tay tôi dẫn đi trên con đường làng dài và hẹp. Con đường này tôi đã quen đi lại lắm lần, nhưng lần này tự nhiên tôi thấy lạ. Cảnh vật chung quanh tôi đều thay đổi, vì chính lòng tôi đang có sự thay đổi lớn: Hôm nay tôi đi học.

Tôi không lội qua sông thả diều như thằng Quý và không ra đồng nô hò như thằng Sơn nữa.

Trong chiếc áo vải dù đen dài tôi cảm thấy mình trang trọng và đứng đắn.

Đọc đường tôi thấy mấy cậu nhỏ trạc bằng tôi, áo quần tươm tất, nhí nhảnh gọi tên nhau hay trao sách vở cho nhau xem mà tôi thèm. Hai quyển vở mới đang ở trên tay tôi đã bắt đầu thấy nặng. Tôi bặm tay ghi thật chặt, nhưng một quyển vở cũng chỉ ra và chênh đầu chúi xuống đất. Tôi xúc

lên và nắm lại cẩn thận. Mấy cậu đi trước o sách vở thiệt nhiều lại kèm cả bút thước nữa. Nhưng mấy cậu không để lộ vẻ khó khăn gì hết.

Tôi muốn thử sức mình nên nhìn mẹ tôi:

- Mẹ đưa bút thước cho con cầm.

Mẹ tôi cúi đầu nhìn tôi với cặp mắt thật âu yếm:

- Thôi để mẹ nắm cũng được.

Tôi có ngay cái ý kiến vừa non nớt vừa ngây thơ này: chắc chỉ người thạo mới cầm nổi bút thước.

## **Appendix 2**

An example of the segmented chunks

Tôi/ không thể nào/ quên được/ những cảm giác/ trong sáng/ ấy /nảy nở/ trong lòng tôi/ như/  
mấy cảnh/ hoa tươi /mím 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<sup>1</sup> 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<sup>1</sup> 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 covariation 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*

---

<sup>1</sup> 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. *aŋ<sub>i</sub>*    *uli*            *-ŋ<sub>i</sub>*    *jom*    *-ke*    *-d*    *-a*  
I        mango        1SG    eat     PST    TR    FM  
'I ate a mango'.

In the above sentence subject pronoun is *aŋ*. The object NP *uli* 'mango' agrees with the subject NP and the subject pronominal clitic *-ŋ* occurs to the right of the object NP *uli* as velar nasal *ŋ*.

2. *pro<sub>i</sub>*    *uli*    *-ŋ<sub>i</sub>*    *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)<sub>i</sub>*    *ma:di:*    *-m<sub>i</sub>*    *jom*    *-ke*    *-d*    *-a*  
you    meal        2SG    eat     PST    TR    FM  
'You have eaten meals'/'You ate meals'
4. *(ini)<sub>i</sub>*    *laq*    *-ai<sub>i</sub>*    *jom*    *-a*  
he    bread    3SG    eat     FM  
'He eats bread.'
5. *(inku)<sub>i</sub>*    *laq*    *-ko<sub>i</sub>*    *jom*    *-e*    *-a*  
they    bread    PL    eat     3SG    FM  
'They eat bread.'
6. *(ako)<sub>i</sub>*    *pa:iti*    *-ko<sub>i</sub>*    *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), *ako* 'they', an NP, is subject of the sentence and it can be dropped, because the agreement marker is present on object NP *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<sub>i</sub>*    *laq*    *-ko<sub>i</sub>*    *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. *ini(NH)* *seno* *-ja* *-n* *-a*  
 he go PST INTR FM  
 'He had gone/He went'.
9. *in-kin(H)* *seno* *ja* *-n* *-a*  
 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. *ale* *an* *-lagte* *vo:ta* *-ke* *-d* *-a*  
 we I for vote PST TR FM  
 (inclu) '\*We voted for me.' (Literal meaning)
- \*ale<sub>i</sub>* *an* *-le<sub>i</sub>* *sa:la:* *-ke* *-a*  
 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. *repo* *ini* *-e* *koʔari* *-tan* *-a*  
 Repo he POSS criticise PRS FM  
 'Repo criticises him'
13. *na:masi* *ini* *-e* *koʔari* *-tan* *-a*  
 Namsi he POSS criticise PRS FM  
 'Namsi criticises him'
14. *ini* *na:masi* *-ke* *koʔari* *-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:

15. *repo<sub>i</sub> ja:piɔ̄ -a -e<sub>i</sub>*  
 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:<sub>i</sub> -hon kita:b -e<sub>i</sub> paɣa:w -e -a*  
 boy man book 3SG read 3SG FM  
 'A boy reads a book'.
17. *ini<sub>i</sub> laɔ̄<sub>j</sub> -ai<sub>i</sub> jom -e<sub>j</sub> -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<sub>i</sub> na:masi -ke kita:b -e<sub>i</sub> ema:ɔ̄ -i -e*  
 Repo Namasi DAT book 3SG give PST FM  
 'Repo gave a book to Namsi.'
19. *(a<sub>n</sub>) (a<sub>m</sub> -ke) miad' uli -ŋ<sub>i</sub> em -le -ɔ̄ -me<sub>j</sub> -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. *a<sub>n</sub> hapnam -ko<sub>i</sub> -ŋ<sub>i</sub> nel -le -ɔ̄ -ko<sub>i</sub> -a*  
 I girl PL 1SG see PST TR 3SG FM  
 'I had seen girls'.

21. *ale<sub>i</sub> am<sub>j</sub> -le<sub>i</sub> dula:r-te -m<sub>j</sub> -a*  
 we you 2PL love PRS 2SG FM  
 ‘We love you.’
22. *ini<sub>i</sub> laq<sub>i</sub> -e<sub>i</sub> jom -a*  
 he bread 3SG eat FM  
 ‘He eats bread’.
23. *inku<sub>i</sub> laq<sub>i</sub> -ko<sub>i</sub> jom -a*  
 they bread 3PL eat FM  
 ‘They eat bread’.

In sentence (20), the pronoun *an* ‘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. *an -re aṭkarai -min -a [(an<sub>i</sub>) sugra:-n<sub>i</sub>]*  
 I PP believe be FM I good-1SG  
 ‘[I believe [myself to be good]]’
25. *ini [a:j-a:j-ge sugara:] aṭkarain ke -d -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 sepeq bugin -a*  
 that boy good FM  
 ‘That boy is good’.
27. *bugin sepeq*  
 good boy  
 ‘Good boy’.

In sentences (26) and (27), there is no agreement on the NP *sepeq* ‘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:

28. *so.ma: ama: ti:ji: a:buŋ -ke -d̥ -a -i*  
Soma your hand wash PST TR FM 3SG  
'Soma washed your hands.'
29. *so.ma: aŋa: ti:ji: abuŋ ke -d̥ -a -i*  
Soma my hand wash PST TR FM 3SG  
'Soma washed my hands'.

### Maithili (mai)

30. *si.ta: həm<sub>i</sub> -ər ha:th dho -l -ək<sub>i</sub>*  
Sita I GEN hand wash PST 1SG  
'Sita washed my hands'.

In (30), verb agrees with the possessor *həmər*, 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. *so.ma:<sub>i</sub> adana -a [S<sub>2</sub> cilke sa:ikal -ko hare -e<sub>i</sub> -a]*  
Soma know FM how bicycle ACC drive 3SG FM  
'Soma knows [how to ride a bicycle.]'
32. *so.ma: sanəŋ -tan -a [[S<sub>2</sub> am<sub>i</sub> gupui:-m<sub>i</sub> -e]*  
Soma wants PRS FM you fight 2SG 3SG  
'Soma wants [you to fight.]' (Literally: Soma wants that I and you fight)
33. *so.ma:<sub>i</sub> sanəŋ -tan -a [S<sub>2</sub> aŋ<sub>i</sub> nikul -eŋ<sub>i</sub> ol -e<sub>i</sub>] -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 3<sup>rd</sup> 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 [S<sub>2</sub> okona aŋa: juđi: jom -ke -d̥ -a]*  
that mango which my friend eat PST TR FM  
*ena soja: -ka -n -a*  
that useless PST INTR FM  
'[That mango [which my friend had eaten] was useless.]'
35. *en tebal marəŋ -gi -a [S<sub>2</sub> okona: -re am<sub>i</sub> jom -am<sub>i</sub>]*  
that table big EMP FM which PP you eat 2SG  
*rike -ta -d̥ -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 [ $\pm$ Animate]

Ho is very sensitive with regard to the feature [ $\pm$ Animate]. The inanimate NP has the object agreement marker *-e/-i* in the verb in [-past] tense. The 3<sup>rd</sup> 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. *a:je am -a seta: -ko -e ta:m -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 [ $\pm$ human]. The occurrence of subject agreement marker to the left of the subject depends on the feature [ $\pm$ 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<sub>i</sub> aŋ -ta hapnum -ko<sub>j</sub> -m<sub>i</sub> kul -le -d -ko<sub>j</sub> -a*  
 you I for girl PL 2SG send PST TR PL FM  
 ‘You sent girls to me’.
39. *aŋ<sub>i</sub> (am<sub>j</sub> -ke) mi:aq’ uli -ŋ<sub>i</sub> ema -le -d -me<sub>j</sub> -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<sub>i</sub> aŋ -ta: hapnum -ko<sub>j</sub> -m<sub>i</sub> kul -le -d -ko<sub>j</sub> -a*  
 you I for girl PL 2SG send PST TR 3PL FM  
 ‘You sent girls to me.’
41. *remo<sub>i</sub> a:j -ta:ʔre miqo banduk -e<sub>i</sub> jo:ga:w -le -d -a*  
 Remo he near one gun 3SG put PST TR FM  
 ‘Remo put the gun near him.’
42. *apuŋ<sub>i</sub> aŋ -ke am<sub>j</sub> -e<sub>i</sub> ema:d -mi<sub>j</sub> -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 infix 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

- Crystal, D. 1997. *A Dictionary of linguistics and Phonetics*, Oxford. Blackwell.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrech. Foris.
- \_\_\_\_\_. 1995. *The Minimalist Programme*. MA. MIT. Cambridge.
- Choudhary, P.K. 2005. *Anaphors in Nepali and Mundari*. Unpublished. M.phil. Thesis. University of Delhi.
- Choudhary, P.K. 2009. Negative Polarity Items in Mundari. *Indian Linguistics*: 70:79-89.
- Deeney, J. 2002. *Ho Grammar and Vocabulary*. Xavier Ho publications, Chaibasa.
- Haegeman, L. 1994. *Introduction to Government and Binding theory*. Blackwell. Oxford.
- Keenan, E.L. 1978. The syntax of verb final languages. *Syntactic Typology: Studies in the phenomenology of languages*, ed. W.P. Lehman, Pp.267-327. University of Texas Press. Austin.
- Koopman, H. 1987. "The genesis of Haitian: implications of a comparison of some features of the Syntax of Haitian, French and West-African languages: in P. Muysken et. al.(ed): *Universals Versus Substratum Influence in Creole Genesis*, Benjamins, Amsterdam. Pp. 231-259.
- Lehman, W.P. 1982. 1982. Literature and Linguistics: Text Linguistics. *The Literary Criterion* 17, 1:18-29.
- Lyons, J. 1968. *Introduction to Theoretical Linguistics*. CUP. London.
- Munda, R.D. 1969. Aspects of Mundari verb. *Indian Linguistics* 30:2.27-49.
- Osada, T. 1992. *A reference grammar of Mundari*. Tokyo Universities of Foreign Studies. Tokyo.
- Radford, A. 1998. 'Case, Agreement and under specification in Child English'. *Essex Research Reports in Linguistics* 20:68-80.

- Roy, S.C. 1912. *The Munda and their country*. City Book Society. Calcutta.
- Sinha, N.K. 1975. *Mundari Grammar*. (Grammar Series; 2). Central Institute of Indian Languages. Mysore.
- Steel, S. 1978. Word order variation: a typological study. In *Universals of Human Languages IV*. Edited by J.H. Greenberg, Ferguson and Moravcsik. 588-623. Calif.: CSLI. Stanford.
- Subbarao, K.V. 2001. Agreement in South-Asian languages. MS.University of Delhi.
- Zide, N.H. 1966. *Studies in Comparative Austro-Asiatic Linguistics*, Mouton & Co.London, The Hague, Paris.