

# AUSTROASIATIC STUDIES

Papers from ICAAL 4

Mon-Khmer Studies Journal  
Special Issue No. 2

Volume editors:  
Sophana Srichampa  
Paul Sidwell



## Pacific Linguistics E-8

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia. Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

---

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm Ross and I Wayan Arka (Managing Editors),  
Mark Donohue, Nicholas Evans, David Nash, Andrew Pawley,  
Paul Sidwell, Jane Simpson, and Darrell Tryon

---

Copyright in this edition is vested with the authors

First published 2011

National Library of Australia Cataloguing-in-Publication entry:

Author: International Conference on Austroasiatic Linguistics (4th., 2009 : Salaya, Thailand).  
Title: Austroasiatic studies [electronic resource]: papers from ICAA1 4, Mon-Khmer Studies Journal Special issue no. 2 / edited by Sophana Srichampa and Paul Sidwell.  
ISBN: 9780858836419 (Electronic document)  
Subjects: Austroasiatic languages--Congresses.  
Other Authors/ Srichampa, Sophana.  
Contributors: Sidwell, Paul J.  
SIL International.  
Australian National University. Research School of Pacific and Asian Studies.  
Pacific Linguistics  
Mahāwitthayālai Mahidon.

Typeset by the editors and assistants

---

Published concurrently by:



Pacific Linguistics  
School of Culture, History and Language  
College of Asia and the Pacific  
The Australian National University



Research Institute for Languages and Cultures of Asia  
Mahidol University  
Nakhon Pathom, Thailand



SIL International  
7500 West Camp Wisdom Road  
Dallas, Texas, USA

## Contents

Preface.....	iv
The Syntax of Khasi Adverbial Clauses.....	1
George BEDELL	
Deponent Verbs and Middle-Voice Nouns in Temiar.....	11
Geoffrey BENJAMIN	
Toward Proto Pearic: Problems and Historical Implications.....	38
Michel FERLUS	
Goal-marking in Munda with special reference to the Lower Munda languages.....	52
Arun GHOSH	
A Synopsis of Mal Phonetics.....	69
THERAPHAN L-Thongkum, CHOMMANAD Intajamornrak, KANITHA Putthasatien, YUPAPORN Huadsiri, and SUPAPORN Phalipat	
Mon Converging Towards Thai Models:.....	81
Patrick McCORMICK	
A Study of Language Use and Literacy Practices to Inform Local Language Literature Development among Khmu in Thailand.....	98
Michelle M. MILLER & Timothy M. MILLER	
Multiple Uses of the Pronoun <i>de</i> : in Kmhmu'.....	112
R. Anne OSBORNE	
About <i>ʔaɔj</i> in Contemporary Khmer.....	124
Denis PAILLARD	
Pearic, a Dying Branch of Austroasiatic Languages and Its Struggle for Survival.....	138
SUWILAI Premsrirat	
Aspects of Ho Phonetics and Phonology.....	154
Anna PUCILOWSKI	
Notes on Glottal Constriction in Gorum.....	174
Felix RAU	
Comitative PP as Head in Externally-Headed Relative Clauses in Khasi.....	184
Karumuri v. SUBBARAO & Gracious M. TEMSEN	

## Preface

### History of International Conference on Austroasiatic Linguistics (ICAAL)

The field of Austroasiatic linguistics has had a remarkable history, not the least for having no regular international conference for decades, even though there was sufficient scholarship to support an international journal, *Mon-Khmer Studies*. The field had developed dramatically in the 1960s - with special being paid attention to Indo-China, there was a tremendous flow of new field data. And it was in this circumstance that the ICAAL was first mooted in 1972.

The suggestion was first raised by Norman Zide, long time Munda languages scholar based at the University of Chicago, who had been present at earlier related meetings. These were two gatherings held at the School of Oriental and African Studies (London) in 1961 and 1965, the papers published as: *Linguistic Comparison in South East Asia* (1963, H. L. Shorto, ed.), and *Indo-Pacific Linguistic Studies* (1965, G. B. Milner and E. J. A. Henderson eds. *Lingua* vols. 14 and 15). While these meetings had broad agendas, there was a strong representation by Austroasiatic, due in large part of scholars such as Harry Shorto and Norman Zide.

Lawrence Reid, then at Hawai'i, took on the main effort of organizing a conference dedicated to Austroasiatic, and was successful in attracting a substantial grant from the US National Science Foundation to help bring scholars from around the world to Hawai'i. Subsequently, the first ICAAL was held at the University of Hawai'i in Honolulu during the first week of January in 1973. It was an historic gathering, not just a tremendous success as a meeting, but a programmatic turning point for the field. The two conference volumes that came out in 1976, as *Austroasiatic Studies* (P. N. Jenner, L. C. Thompson, S. Starosta eds. Oceanic Linguistics special publications No. 13) presented some 50 papers in 1343 pages. The contributions established the state of the art with treatments of classification, grammatical description, phonological and morphological history, and language contact, among others. It was the single most important publication for Austroasiatic linguistics for the next 30 years, until the appearance of Shorto's posthumous *A Comparative Mon-Khmer Dictionary* in 2006.

That first meeting also decided to designate *Mon-Khmer Studies* (MKS) the principle journal of the field. Previously MKS had been produced intermittently out of Saigon, where it had mainly served SIL associated scholars. Subsequently the journal moved to Hawai'i and a broader editorial board established. Today MKS is based at Mahidol University in Thailand, and is still published with cooperation of (now) SIL International. Curiously, we note that at the first ICAAL it had been suggested to change the title of MKS to *Austroasiatic Studies*, but this did not follow through, and instead that title was given to the ICAAL proceedings.

Despite the great success of ICAAL 1, circumstances had begun to turn. The dramatic political realignment of Indo-China in 1975 saw attention turn away from the region generally, and within the US (which had funded the first meeting) linguistics was increasingly in the grip of the generative movement which was scarcely interested in empirical language studies. The existing momentum was sufficient that a second meeting was held at Central Institute of Indian Languages in Mysore in 1978, and attended by a few of those who had been present 5 years earlier in Hawaii, plus numerous South Asian scholars. This second ICAAL was held as a post-plenary session of the *International Congress of Anthropological and Ethnological Sciences*. David Stampe and Gerard Diffloth agreed to take on the task of editing the proceedings, but unfortunately that project was abandoned after various delays. As luck would have it, a bound set of mimeographed papers from the meeting is kept at the Central Institute of Indian Languages in Mysore, and a mostly readable photocopy made its way into the Cambridge University library. Versions of several of the papers did make it into press as journal articles or book chapters, but a gap remains in the history of ICAAL.

Subsequently there were no independent international Austroasiatic meetings. There were various special sessions at conferences devoted to other Asian language families (such as the annual International Conferences on Sino-Tibetan Languages and Linguistics) and this ad hoc arrangement served the scholarly community during the lean years of the 1980s and 1990s. But with the new millennium, and the resurgence of language documentation and emerging interdisciplinary interest from archaeology, genetics etc. the need for a regular Austroasiatic conference became imperative. This was recognized by Gérard Diffloth, Michel Ferlus and George van Driem in a discussion they held in 2001, and later Diffloth and van Driem were able to organize an informal gathering, which they called the ICAAL 3 Pilot Picnic, at Siem Reap in June 2006 (with support from Dutch national research council (NOW) and l'École française d'extrême-orient).

The Pilot Picnic was such a success that it paved the way for the rebirth of ICAAL with a full blown conference (ICAAL 3) which was held at Deccan College Post-Graduate & Research Institute in Pune, November 26-28, 2007. The principal organizer was Professor Keralapura Shreevinasaiah Nagaraja, and the meeting was also supported by the Central Institute of Indian Languages in Mysore. The proceedings have been edited for publication by Prof. Nagaraja, and will be published as *Austro-Asiatic Linguistics: In memory of R. Elangaiyan* (CIIL Publications).

#### ICAAL 4

In 2009 ICAAL 4 was held at the Research Institute for Language and Culture of Asia, Salaya campus of Mahidol University (Thailand) 29-30 October 2009. There were some 92 registered participants, delivering 70 papers over two rather long days (the full program can be seen online at <http://icaal.org/>). Reflecting the meeting theme of "An Austroasiatic Family Reunion" participants came from a wide range of Asian countries including Thailand, Malaysia, Vietnam, Laos, Myanmar, India, Bangladesh, Nepal, Singapore and China, as well as western nations.

Participants from the host institution made an especially significant contribution discussing projects which are documenting and revitalizing endangered languages of Thailand, under the leadership of Prof. Suwilai Premsrirat (ret.), who has worked tirelessly for Thailand minorities during her distinguished career. Comparative and historical papers were also prominent, since there is increasing attention to the role that linguistics can play in discovering the prehistory of Southeast Asia. Papers hoping to discover the Austroasiatic homeland, and other historical questions, were delivered by George van Driem (Netherlands), Paul Sidwell (Australia), Michel Ferlus (France), Roger Blench (Britain), Gerard Diffloth (Cambodia). In connection with these there was also discussion of the application of new statistical methods in language classification (methods originally developed for genetics), with papers read by Russel Gray (New Zealand), Jerry Edmondson (USA) and Kenneth Gregerson (USA). Generally it was the most important meeting of historical linguists on these questions since the first ICAAL meeting in 1973.

Very important was the participation of scholars from South Asia, made possible by financial support provided the Centre for Research in Computational Linguistics (Bangkok), which was also a partner in organising and running the meeting. All up there were nine papers concerning the Munda, Khasi and Nicobar languages spoken in India and elsewhere in South Asia. This marked a high point in realising the conference theme, intended to bring scholars together who might otherwise never meet.

The Keynote address was delivered by retired Professor of Southeast Asian Linguistics at Cornell, Franklin Huffman. Prof. Huffman worked in Southeast Asia during the 1960s and 1970s, and among his achievements were the development of extensive teaching materials, bibliographic resources, and a widely used dictionary of Cambodian (for example, it used by the United Nations Transitional Authority in Cambodia 1992-93). So important was this work, that some have credited him with "saving the Cambodian language" after the chaotic Khmer-Rouge years destroyed so much educational and cultural infrastructure.

A special highlight of the meeting, marking the real continuity of the movement over four decades, was a special presentation of gifts honoring senior figures in the field, including several who had been present at the 1973 Hawaii meeting. Honored were: Geoffrey Benjamin, Theraphan L-Thongkum, Gerard Diffloth, Suwilai Premsrirat, Franklin Huffman, and Michel Ferlus.

At the business meeting on the final day participants resolved to launch a new project, a major handbook of Mon-Khmer languages with chapters on history and typology, and sketch grammars of representative languages. Mathias Jenny was nominated as chief editor, and an initial group of contributors nominated. It is expected that the handbook will take several years to produce, and ultimately become one of the canonical texts of the field.

Finally, the need for an international committee of scholars to ensure the continuity and standards of ICAAL was discussed, and the following names were nominated and accepted by consensus: Arun Gosh (India) Brian Migliazza (USA) David Stampe (USA) Doug Cooper (Thailand) Gerard Diffloth (Cambodia) Jerry Edmondson (USA) Mathias Jenny (Switzerland) Patricia Donegan (USA) Paul Sidwell (Australia) Roger Blench (UK) Sophana Srichampa (Thailand). Members of this ICAAL International Committee assisted with review of papers in this volume (in addition to various external reviews) and assist with assessing and ranking abstract submitted to the conference.

### **This volume**

The present volume begins with the MKS editorial board accepting the invitation of the ICAAL 4 business meeting to produce the ICAAL proceedings as a series of MKS special volumes, with papers subject to the usual MKS review processes. Principle responsibility for editing and manage the review process fell to Sophana Srichampa and Paul Sidwell, and Kirk Person also provided special assistance with copyediting and reviewing. Staff of Mahidol University provided administrative and typesetting assistance, especially Suttalak Soonghangwa, who deserves special mention. All papers went through a review process, and we thank those anonymous readers who contributed so much to helping contributors with their paper.

In order to emphasise the continuity of the ICAAL movement we decided to retain the *Austroasiatic Studies* name as part of the title. Our intention is to produce these as an ongoing series in a timely fashion, keeping pace the conference meetings, which are now being held every 2 years. And a note of clarification concerning the numbering in the sub-title; MKS launched a monograph series in 2000 with Ken Smith's *Sedang Dictionary* as the first of these "special issues". Others were planned, but did not come to fruition, so this seems appropriate to breath new life into the series with the ICAAL proceedings. MKS also published another (self-contained) series, being Aj. Suwilai's five volume Khmu dialect dictionaries.

It is our dear hope that this new momentum, reflected in the success of the resurgent ICAAL meetings, will be maintained, and that the new vigour surrounding the study of Austroasiatic languages will continue to grow.

Dr. Paul Sidwell  
Assoc. Prof. Sophana Srichampa  
Bangkok, June 2011

# The Syntax of Khasi Adverbial Clauses

George Bedell  
Payap University

## Introduction.

Khasi is a Mon-Khmer language spoken primarily in eastern Meghalaya State, India. According to *Ethnologue* (Lewis 2009) the number of speakers in India is 865,000; there is considerable internal diversity however and it is likely that ‘Khasi’ includes more than one language. The Khasi understood in this paper is Standard Khasi as used for literary purposes, and recorded in Roberts (1891), Nissor Singh (1904) and the standard Khasi translation of the Christian Bible. The examples cited are taken from *Ka Khubor jong ka Jingieit* (2000), and given in the orthography used there, unless otherwise noted. The numbers indicate chapter and verse in *Ka Gospel U Mathaios* (The Gospel According to Matthew). This translation was made in the nineteenth century and is sometimes difficult for modern speakers to understand; it remains in common use.

## Subordinate Clauses.

Subordinate clauses in Khasi are most often marked with the particle *ba*, either alone or with another preceding particle prefixed. Clauses headed with *ba* may have a variety of interpretations: Roberts (1891; §244, pp. 205-6) mentions three. Sentence (1) illustrates a complement clause, and sentences (2) and (3), adverbial clauses.

- (1) *U Kpa jong phi uba ha bneng u tip*  
3SM father of youPL 3SM=Cin sky 3SM know  
*[ba phi dokam ia kine kiei kiei baroh].*  
C 2PL need OBJ 3PL=this 3PL=Q 3PL=Q all  
‘your Father in heaven knows [(that) you need all these things]’ (6:32)

- (2) *ngan leh ia kaba bha aiu, [ba*  
1S=FUT do OBJ 3SF=C good what C  
*ngan ioh ia ka jingim bymjukut]?*  
1S=FUT get OBJ 3SF ACT=alive endless  
‘what good thing should I do [(in order) to get eternal life]?’ (19:16)

- (3) *Phi bakla, [ba phim tip ia ki jingthoh,*  
2PL wrong C 2PL=NEG know OBJ 3PL ACT=write  
*lymne ia ka bor U Blei].*  
or OBJ 3SF power 3SM God  
‘you are wrong, [because you do not know the scriptures or God’s power]’ (22:29)

There are also relative clauses, exemplified by (4).

---

Bedell, George. 2011. “The Syntax of Khasi Adverbial Clauses.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAALA. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.1-10.  
Copyright vested in the author.



- (4) *ki khynnah [ba ki iapyrta ha ka templ],*  
 3PL child C 3PL COLL=shout in 3SF temple  
 ‘the children [(who are) shouting in the temple]’ (21:15)

A complement clause as in (1) is one which serves as an argument of a verb (here *tip* ‘know’). Adverbial clauses as in (2) and (3) modify a verb (*leh* ‘do’ in (2)) or the main clause (*phi bakla* ‘you are wrong’ in (3)). A relative clause as in (4) is one which modifies a noun (here *khynnah* ‘children’).

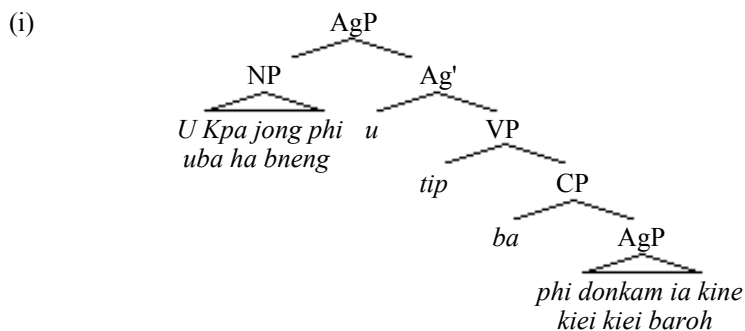
It is generally the case that the complement of *ba* can serve as an independent main clause; corresponding to (1) to (4) are (5) to (8).

- (5) *phi donkam ia kine kiei kiei baroh*  
 2PL need OBJ 3PL =this 3PL=q 3PL=q all  
 ‘you need all these things’
- (6) *ngan ioh ia ka jingim bymjukut*  
 1S=FUT get OBJ 3SF ACT=alive endless  
 ‘I will get eternal life’
- (7) *phim tip ia ki jingthoh, lymne ia*  
 2PL=NEG know OBJ 3PL ACT=write or OBJ  
*ka bor U Blei*  
 3SF power 3SM God  
 ‘you do not know the scriptures or God’s power’
- (8) *ki iapyrta ha ka templ*  
 3PL COLL=shout in 3SF temple  
 ‘they are shouting in the temple’

The initial *phi* in (5) and (7), *nga* in (6) and *ki* in (8) are markers of subject-verb agreement and characteristic of independent clauses. The future tense *-n* is suffixed to the agreement marker in (6), and the negative *-m* is suffixed to it in (7). See Bedell (2008) for discussion of this agreement.

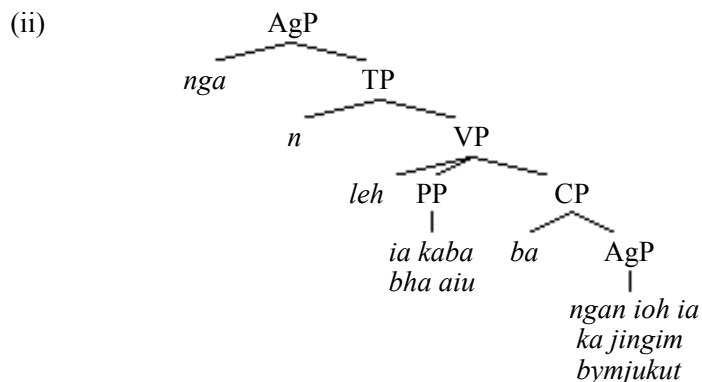
### Structures 1.

As shown in (i), (ii) and (iv), we take *ba* to belong to the category C (conjunction or complementizer) which combines with a complement clause (AgP).

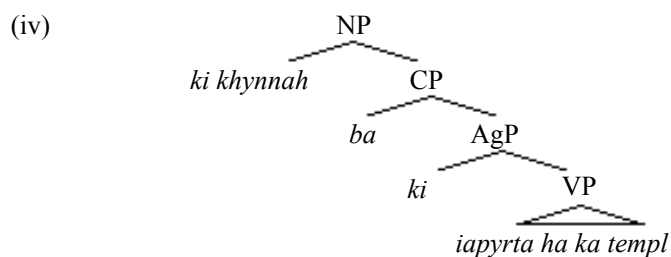


The resulting CP may occupy an argument position as in (i).





Its syntactic position in (ii) is similar but there it is not an argument.



Its position in (iv) is quite distinct, attached to NP rather than VP.

#### Adverbial Clauses.

(9) to (11) further illustrate Khasi adverbial clauses; each of them consists of a well formed clause as in (12) to (14) preceded by a conjunction *haba* ‘when’, *naba* ‘for’ or *kumba* ‘as’.

- (9) *haba u mynsiem bymkhuid u la mih noh*  
 when 3SM spirit unclean 3SM PAST go out

*na u briew,*  
 from 3SM person

‘when an unclean spirit has gone out of a man’ (12:43)

- (10) *naba u hap jyndei ha ka ding, jyndei ha*  
 For he fall much into 3SF fire much into

*ka um ruh.*  
 3SF water also

‘for he falls often into the fire and often into the water’ (17:15)

- (11) *kumba U Jonah u la shong lai sngi bad lai*  
 as 3SM Jonah 3SM PAST sit three day and three

*miet ha ka kpoh u dohkha khraw,*  
 night in 3SF belly 3SM fish big

‘as Jonah spent three days and nights in the belly of the whale’ (12:40)

- (12) *u mynsiem bymkhuid u la mih noh na u riew*  
 3SM spirit unclean 3SM PAST go out from 3SM person

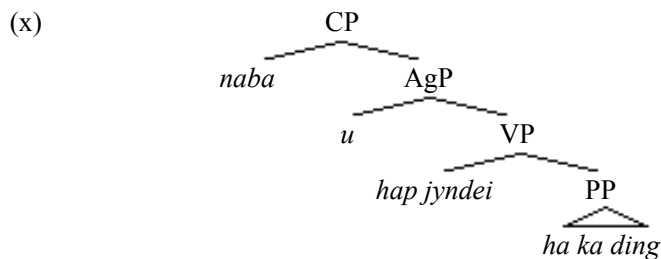
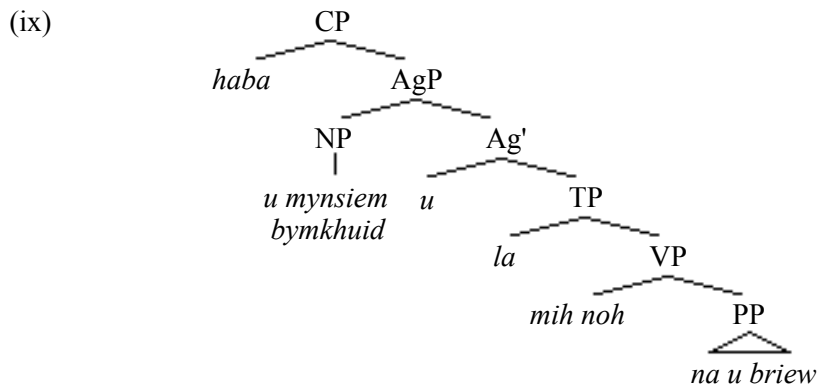
‘an unclean spirit went out of a man’

- (13) *u hap jyndei ha ka ding, jyndei ha ka um ruh*  
 he fall much into 3SF fire much into 3SF water also  
 ‘he falls often into the fire and often into the water’
- (14) *U Jonah u la shong lai sngi bad lai miet*  
 3SM Jonah 3SM PAST sit three day and three night  
*ha ka kpoh u dohkha khraw,*  
 in 3SF belly 3SM fish big  
 ‘Jonah spent three days and nights in the belly of the whale’

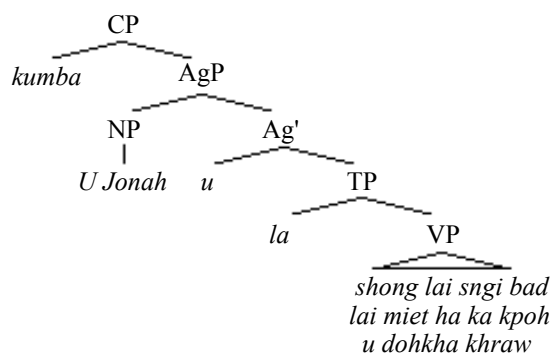
Roberts (1891) lists *naba* and *kumba* among ‘copulative conjunctions’ (p. 122). Though he briefly discusses the syntax of conjunctions (pp. 203-209), he says nothing about the internal structure of clauses containing them. Rabel (1961) also lists them as conjunctions, noting that they are compounds of a preposition and *ba* (p. 72). She does not discuss their syntax. Nagaraja (1985) discusses them under *ba* clauses (8.2.8, p. 97). They are composed of a preposition or prepositional adverbial followed by *ba*, but it is unclear what syntactic category *ba* or its compounds belong to. They are apparently not conjunctions.

### Structures 2.

As shown in (ix), (x) and (xi), we take *haba*, *naba* and *kumba* to belong to C, parallel to *ba* in (ii). The complement clauses in (9) and (11), unlike (2) and (12), contain overt subjects.



(xi)

**Prepositions.**

*Haba* in (9) appears to consist of the preposition *ha* ‘in, into’ combined with *ba*; *naba* in (10) appears to consist of the preposition *na* ‘from’ with *ba*, and *kumba* in (11) appears to consist of the preposition *kum* ‘like’ with *ba*. These prepositions are illustrated in (15) through (20).

(15) *na ka um*  
 from 3SF water  
 ‘out of the water’ (3:16)

(16) *na u briew*  
 from 3SM person  
 ‘out of a man’

(17) *ha ka ding*  
 into 3SF fire  
 ‘into the fire’

(18) *ha uta u briew*  
 to this 3SM person  
 ‘to this man’ (12:13)

(19) *kum u thiat,*  
 like 3SM yeast  
 ‘like yeast’ (13:33)

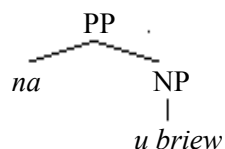
(20) *kum u briew uba u trai iing,*  
 like 3SM person 3SM-C 3SM owner house  
 ‘like the owner of a house’ (13:52)

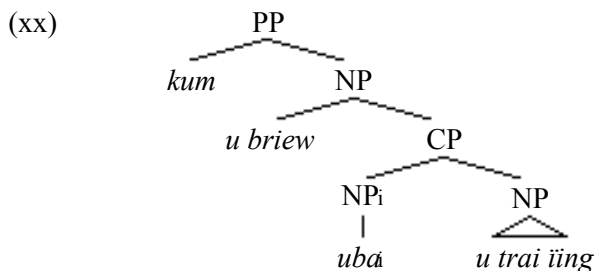
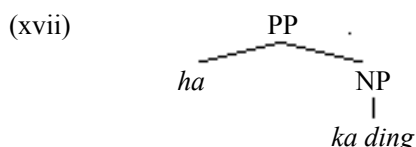
Note that (16) and (17) appear as constituents of (10) and (9).

**Structures 3.**

We take prepositions to be heads of prepositional phrases which take a complement NP as in (xvi), (xvii) and (xx).

(xvi)





Note that the conjunctions *haba*, *naba* and *kumba* belong to a distinct syntactic category from the related prepositions *ha*, *na* and *kum*. Though clearly similar, the meanings do not correspond exactly. As conjunctions, *haba* and *naba* lack any directional component.

### Deictics.

In addition to the conjunctions *haba*, *naba* and *kumba*, the prepositions *ha*, *na* and *kum* also appear attached to deictic stems *-ta* ‘that’, *-ne* ‘this’ and interrogative *-no*, as in (21) through (29). For discussion of Khasi deictics, see Bedell 2009c, and of Khasi interrogatives, Bedell 2009b.

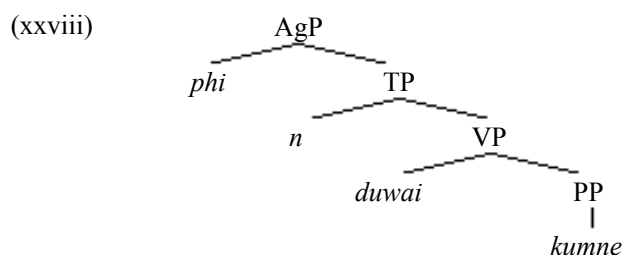
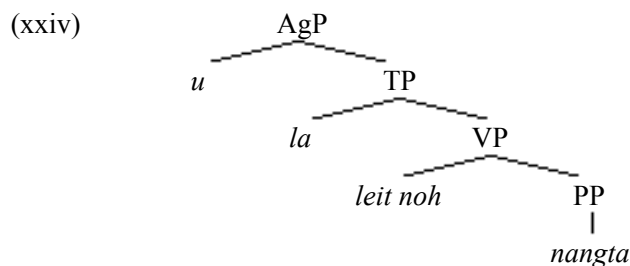
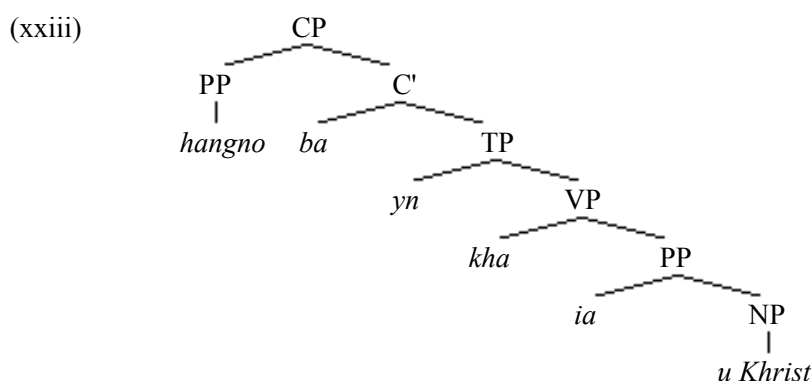
- (21) *u la don hangta weibriew.*  
 3SM PAST be there one-person  
 ‘he stayed there alone’ (14:23)
- (22) *um don hangne;*  
 3SM=NEG be here  
 ‘he is not here’ (28: 6)
- (23) *hangno ba yn kha ia U Khrist.*  
 Where C FUT give-birth OBJ 3SM Christ  
 ‘where the Christ will be born’ (2: 4)
- (24) *u la leit noh nangta.*  
 3SM PAST go away from-there  
 ‘he went away from there’ (19:15)
- (25) *To leit noh nangne hangtai,*  
 IMP go away from-here there  
 ‘go away from here to over there’ (17:20)
- (26) *nangno te ka la ioh ki ñiut?*  
 from-Q and 3SF PAST get 3PL weed  
 ‘and where did the weeds come from?’ (13:27)
- (27) *uban hikai kumta ia ki briew,*  
 3SM=C=FUT teach like-that OBJ 3PL person  
 ‘who teaches people like that’ (5:19)
- (28) *phin duwai kumne:*  
 2PL=FUT pray like-this  
 ‘you will pray like this’ (6: 9)

- (29) *kumno ba u leit hapoh ka templ,*  
 like-Q C 3SM go into 3SF temple  
 'how he went into the temple' (12: 4)

Roberts (1891) lists *hangne* and *nangta* as 'adverbs of place' (p. 117) and *kumno* as an 'adverb of manner' (p. 114). They are not mentioned in his discussion of adverbial syntax. Rabel (1961) analyzes them as pronouns with a locative base (*hang-*, *nang-*) or preposition (*kum*) followed by a demonstrative (*-ne*, *-ta*) or interrogative (*-no*) base (pp. 67-69). Nagaraja (1985) analyzes them as adverbials, but otherwise follows Rabel (4.4 and 4.5, pp. 44-47).

#### Structures 4.

In structures like (xxiii), (xxiv) and (xxviii), we take the combination of a preposition and a deictic or interrogative stem to be a compound preposition which takes no complement. The prepositional properties are retained in these compounds unlike the compound conjunctions.



#### Agreement.

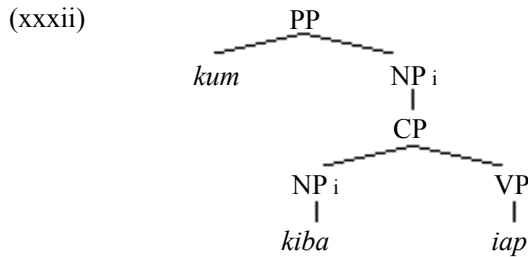
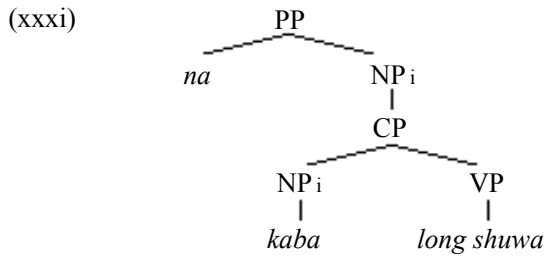
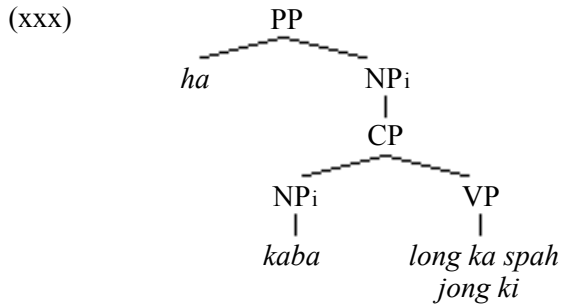
*Haba*, *kaba* and *kumba* contrast with *ha kaba* in (30), *na kiba* in (31) and *kum kiba* in (32). These are not adverbial clauses at all, but rather relative clauses without an overt head noun. The *ka-* or *ki-* which intervenes between the preposition and *ba* is the agreement characteristic of Khasi relative clauses. In most such clauses it is optional (as in (4) above), but appears to be required after a preposition. For discussion of Khasi relative clauses see Bedell 2009a.

- (30) *ha kaba long ka spah jong phi*, (6:21)  
 at 3SF-C be 3SF riches of youPL  
 ‘where your riches are’
- (31) *shibun na kiba long shuwa* (19:30)  
 many from 3PL-C be first  
 ‘many who are first’
- (32) *kum kiba iap*.  
 Like 3PL-C die  
 ‘like dead people’ (28: 4)

Neither Roberts (1891) nor Rabel (1961) appear to recognize *ha kaba* or *na kiba* as a distinct construction. Nagaraja (1985) regards them as like *haba* or *naba*, but with an additional ‘pronominal marker’ (8.2.8, p. 98).

### Structures 5.

Note the contrast between the relative clause structure in (xxx) and the embedded question structure in (xxiii), both expressed in English by ‘where’ which may be an interrogative or relative pronoun.



The Khasi adverbial clause type in (9), (10) and (11), though common, does not exhaust the possibilities. There is, for example, a temporal conjunction *mynba* ‘when’ as in (33).

- (33) *mynba u la thngan*,  
 when 3SM PAST be-hungry  
 ‘when he was hungry’ (12: 3)

There is an associated deictic as in (34) and interrogative as in (35), but *myn* is not used as a preposition.

(34) *uba mynta u long,*  
 3sm=c now 3sm be  
 ‘which now exists’ (6:30)

(35) *mynno ba ngi la iohi ba me la thngan, ...?*  
 when C 1PL PAST see C 2SM PAST be-hungry  
 ‘when did we see that you were hungry?’ (25:37)

There are also many prepositions which do not form conjunctions as well as conjunctions which are not related to a preposition. An example of the latter is *lada* ‘if’.

(36) *lada un sngew ia me,*  
 if 3SM=FUT listen OBJ youSM  
 ‘if he listens to you’ (18:15)

### Abbreviations

1S	first person singular
1PL	first person plural
2SM	second person masculine singular
2SF	second person feminine singular
2PL	second person plural
3D	third person diminutive
3SM	third person masculine singular
3SF	third person feminine singular
3PL	third person plural
ACT	action nominalizer
AGT	agent nominalizer
C	complementizer
CAUS	causativizer
CLASS	classifier
COLL	collective
DIM	diminutive
EMPH	emphatic
FUT	future tense
IMP	imperative
LOC	locative
NEG	negative
OBJ	object marker
Q	interrogative
PAST	past tense
PROX	proximate
SUBJ	subjunctive
VOC	vocative
youSM	you (singular masculine)
youSF	you (singular feminine)
youPL	you (plural)



**References**

- H. Roberts, *A Grammar of the Khasi Language*, London: Kegan Paul, 1891 (reprint New Delhi: Mittal 1995).
- U Nissor Singh, *Khasi-English Dictionary*, Shillong, 1904 (reprint New Delhi: Gyan 2003).
- L. Rabel, *Khasi, a Language of Assam*, Baton Rouge: LSU Press, 1961.
- K. S. Nagaraja, *Khasi, a Descriptive Analysis*, Pune: Deccan College Post-Graduate & Research Institute, 1985.
- Ka Khubor jong ka Jingieit* (The Message of Love: the New Testament in Khasi), Bangalore: Ka Bible Society jong ka Ri India, 2000.
- G. Bedell, 'The Syntax of Agreement in Khasi', presented to NEILS 3, 2008.
- G. Bedell, 'Agreement in Khasi Relative Clauses', presented to NEILS 4, 2009a.
- G. Bedell, 'The Syntax of Khasi Questions', presented to NEHU International Conference, 2009b.
- G. Bedell, 'The Syntax of Deixis in Khasi', presented to SEALS 19, 2009c.
- Lewis, M. Paul (ed.), *Ethnologue: Languages of the World*, Sixteenth edition. Dallas, Tex.: SIL International, 2009. Online version: <http://www.ethnologue.com/>

# Deponent Verbs and Middle-Voice Nouns in Temiar

Geoffrey Benjamin  
Nanyang University

## Abstract

The infix *-a-* in Temiar (Central Aslian, Peninsular Malaysia) covers a wide range of functions, both productive and non-productive. In Temiar morphology generally it serves iconically to indicate an embedded object. As a productive infix, it forms the middle voice of verbs. The main non-productive function of *-a-* is as a frozen infix in (1) a set of non-inflecting ‘deponent’ verbs and (2) a set of disyllabic (‘middle-voice’) nouns. The same semantic dimension underlies both, namely that the verbs and nouns in question indicate that the subject or entity is thought of as being simultaneously its own agent and patient. Culturally, the special attention paid to ‘middle-voice’ processes relates to a central Temiar interest in the specifically dialectical character of Self–Other relations.

## 1. Introduction

Temiar, currently spoken as a first language by approximately 27,000 people in northern Peninsular Malaysia, belongs to the Central Aslian (Senoic) division of Mon-Khmer (Diffloth 1975; Benjamin 1976a, 2004; Matisoff 2003). It is also spoken by a few thousand other people as a lingua franca among the northern Peninsular Orang Asli (Aborigines) and some Malays. Based on my own summary account of Temiar grammar (Benjamin 1976b), a moderately extensive secondary literature has emerged, concerned mostly with the rich reduplicative morphology exhibited by the Temiar verb (Tables 1 and 2).<sup>1</sup> However, the majority of those analyses are based on an approach to the Temiar verb that I later revised. In my 1976 account I treated the verbal infix *-a-* as indicating what I called the simulfactive mode. I now regard this infix as indicating, not mode (or *Aktionsart*), but *voice* (or diathesis), i.e. participant orientation. Specifically, the infix *-a-* indicates that the verb is in the middle voice, as part of a series that also includes an uninflected base form as well as an inflected causative in *-r-*.<sup>2</sup>

### 1.1 Inflecting verbs

Tables 1 and 2 illustrate the forms taken by the two different morphological categories of inflecting verbs: those with a monosyllabic base and those with a sesquisyllabic base. Each form is given twice, first in

---

<sup>1</sup> The secondary studies include McCarthy (1982: 208–221), Broselow & McCarthy (1983: 38–43), Ter Mors (1983: 280–288), Sloan (1988: 319–324), Itô (1989: 252–254), Spencer (1991:151, 155), Carstairs-McCarthy (1992:84–85), Shaw (1994: 122–127), Spaelti (1997: 112, 134–135, 162–165), Gafos (1998a: 233–249, 1998b: 517–522, 1999: 75–107), Walther (2000), Raimy (2000: 145–151), Hendricks (2001: 290), Idsardi & Raimy (n.d), Frampton (2004:160–163) and Yap (2009).

<sup>2</sup> I first presented this revised paradigm in Benjamin 2001: 114. So far, the only secondary accounts to have incorporated it are Matisoff 2003 and Yap 2009. A fuller discussion of Temiar morphology is presented in Benjamin 2011; the present paper derives from an earlier draft of that study. Commentators have employed several different labels for the reduplicative pattern exhibited by the imperfective forms in Tables 1 and 2. My preference is for Matisoff’s self-explanatory iconic term ‘incopyfixation’ (Matisoff 2003: 28–32).

---

Benjamin, Geoffrey. 2011. “Deponent Verbs and Middle-Voice Nouns in Temiar.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.11-37.  
Copyright vested in the author.

a strictly phonemic orthography and second in an orthography identical to my own practical orthography for Temiar. The phonemic orthography indicates the morphology more clearly, but is harder to read than the practical orthography. (Except where otherwise noted, I employ the practical orthography in the main body of this paper.) The glosses indicate some of the semantic distinctions signaled by the various forms.

VOICE	ASPECT			VERBAL NOUN
	Perfective	Imperfective	Progressive	
<b>Base</b>	<i>gəl</i> [gəl] 'sit' (completed act)	<i>glgəl</i> [gɛlgəl] 'sit' (incomplete act)	<i>bar-glgəl</i> [bar-gɛlgəl ~ bə-gɛlgəl] 'currently sitting'	<i>glnəl</i> [gɛlnəl] ~ <i>nlgəl</i> [nɛlgəl] 'a sitting' <sup>3</sup>
<b>Middle</b>		<i>gagəl</i> [gagəl] 'sit' (uncontrolledly, or all together at once)	<i>bar-gagəl</i> [bar-gagəl ~ bə-gagəl] 'currently seated' (uncontrolledly)	<i>gnagəl</i> [gɛnagəl] 'a sitting' (uncontrolledly)
<b>Causative</b>	<i>trgəl</i> [tɛrgəl] 'make (someone) sit' (completed act)	<i>trlgəl</i> [tɛrlgəl] 'make (someone) sit' (incomplete act)	<i>bar-trlgəl</i> [bar-tɛrlgəl ~ bə-tɛrlgəl] 'currently making (someone) sit'	<i>trngəl</i> [tɛrngəl] 'a making (someone) sit'

Table 1: Monosyllabic: *gəl* 'to sit'

VOICE	ASPECT			VERBAL NOUN
	Perfective	Imperfective	Progressive	
<b>Base</b>	<i>slɔg</i> [səlɔg] 'sleep' (completed act)	<i>sɔlgɔg</i> [sɛglɔg] 'sleep' (incomplete act)	<i>bar-sɔlgɔg</i> [bar-sɛglɔg ~ bə-sɛglɔg] 'currently sleeping'	<i>snɔlg</i> [sɛnɔg]; <i>sngɔlg</i> [səngɔlg] 'a sleeping'
<b>Middle</b>		<i>salɔg</i> [salɔg] 'fall asleep' (uncontrolledly)	<i>bar-salɔg</i> [bar-salɔg ~ bə-salɔg] 'currently falling asleep; sleep (with someone)'	<i>snalɔg</i> [sənalɔg] 'a falling asleep'
<b>Causative</b>	<i>srlɔg</i> [sɛrlɔg] 'put (someone) to sleep' (completed act)	<i>srgɔlg</i> [sɛrgɔlg] 'put (someone) to sleep' (incomplete act)	<i>bar-srgɔlg</i> [bar-sɛrgɔlg ~ bə-sɛrgɔlg] 'currently putting (someone) to sleep'	<i>sɛrlɔg</i> [sɛrɛnɔg] 'a putting (someone) to sleep'

Table 2: Sesquisyllabic: *səlɔg* 'to sleep' (also: 'lie down', 'marry')

The imperfective aspect is formed by copying the verb's final consonant as an infix, leftwards, before the final syllable. The middle voice is formed by infixing *-a-* into the same position. With monosyllabic verbs (Table 1) the initial consonant is copied as well, in both the imperfective and the middle forms. (For the morphophonemic changes and morphological processes involved, see Benjamin 1976b: 143–145, 168–170.)

<sup>3</sup> During my initial period of fieldwork in the 1960s and 1970s verbal nouns with prefixed word-initial *n-* were heard only in the state of Perak. More recently, this form is also being heard across the watershed in Kelantan, alongside the older infixal form. Could this indicate an increasing intolerance of infixes by people living in more modern circumstances?

## 1.2 Non-inflecting verbs

Temiar also possesses many disyllabic verbs that exhibit almost none of the inflectional possibilities displayed by the monosyllabic and sesquisyllabic verbs.<sup>4</sup> Their sole inflection is nominalization with *-n-*; but they also form a progressive-cum-imperfective with the proclitic *bar-* ~ *bə-*.<sup>5</sup> As Tables 3 and 4 illustrate, these fall into two subtypes: those with *-a-* in the first syllable (Table 3) and those with some other non-predictable vowel, such as *-o-* or *-e-* (Table 4).<sup>6</sup> For reasons explained later, I shall refer to most of the verbs with *-a-* in the first syllable as ‘deponent’ verbs.<sup>7</sup>

Perfective	ASPECT		VERBAL NOUN
	Imperfective, Progressive		
<i>halab</i> [halab] ‘travel downriver’	<i>bar-halab</i> [bar-halab ~ bə-halab] ‘currently travelling downriver’		<i>hnalab</i> [hənalab] ‘a journey downriver’

Table 3: Deponent: *halab* ‘to travel downriver’

Perfective	ASPECT		VERBAL NOUN
	Imperfective, Progressive		
<i>golap</i> [golap] ‘carry on shoulder’	<i>bar-golap</i> [bar-golap ~ bə-golap] ‘currently carrying on shoulder’		<i>gnolap</i> [gənoɫap] ‘a carrying on shoulder’

Table 4: Disyllabic: *golap* ‘to carry on one’s shoulder’

## 2. The Temiar middle voice

Before proceeding to the primary topic of this paper – deponent verbs and middle-voice nouns – let me first outline the normal productive usage of the middle voice in Temiar. Temiar is the only Aslian language that makes use of *-a-* as a fully productive affix in the inflection of the verb; its close relative Lanoh also employs this affix, but less productively. For the reasons outlined below, it seems clear that ‘middle

<sup>4</sup> These include a moderate number of verbs with more than three consonants, such as *sindul* ‘to float’, or verbs with a further presyllable, such as *səmaŋar* ‘to snore’.

<sup>5</sup> As indicated in Tables 1 and 2, the ‘progressive’ clitic *bar-* is almost universally attached to the reduplicated imperfective form of the verb. However, I recorded one or two cases in which *bar-* was attached instead to the unreduplicated perfective form of the verb, as in *wɛ-bə-kooy kɛʔ* [2DU-PROG-roast.PFV fish] ‘they were roasting the fish’ (instead of the expected *wɛ-bə-keɣkooy kɛʔ*). Since these instances were rare, it is impossible to determine whether they were intentional or simply speakers’ mistakes.

The following glossing conventions and abbreviations are employed in this paper: \* ‘non-occurring or hypothetical form’; - (hyphen) ‘proclitic’; <> ‘infix’; 1 ‘1st person’; 2 ‘2nd person’; 3 ‘3rd person’; ACC ‘accusative’; C<sup>i</sup> ‘initial consonant’; DU ‘dual’; EXPR ‘expressive’; INCL ‘inclusive’; INT ‘intensive’; IPFV ‘imperfective’; MID ‘middle voice’; NOM ‘nominative’; PFV ‘perfective’; PL ‘plural’; PROG ‘progressive’; SG ‘singular’; VET ‘vetative’ (negative imperative).

<sup>6</sup> In my earlier study (Benjamin 1976b: 138), I described the minor-syllable vowel as containing a single ‘phoneme’, /ə/, with a variety of allophonic pronunciations. It would have been better simply to describe the various pronunciations of what is, in fact, a non-phonemic epenthetic transition. These pronunciations vary between [ə], [i] and zero, depending on the following consonant and the vowel in the word-final major syllable. In a strictly phonemic transcription, therefore, these ‘minor vowels’ should be omitted, but it is still useful to incorporate them into a practical orthography. In the past, I have tended to write them all as ə, but in future I shall write *i* instead, where appropriate: *sikeʔ* ‘pandanus’ instead of *səkeʔ* (phonemically /skeʔ/). In the majority of Temiar dialects, the predictable non-phonemic epenthetic vowel is regularly pronounced [ɛ] in closed syllables (i.e., before -CCV-), as written in my practical orthography.

<sup>7</sup> As noted in Section 3.2 and the Appendix below, some verbs in *-a-* do not qualify as deponent, and should preferably be placed in the ‘disyllabic’ category.

voice’ best describes the character of these *-a*-inflected forms. ‘Simulfactive’, the term I employed previously (Benjamin 1976b: 172–174), refers to just one of the uses of the middle voice. I discuss it below under the label of ‘all-together’ middle.

The infix *-a-*, which also occurs in a range of non-verbal contexts, signals the notional meanings OBJECT or SALIENT OTHER. I hold that, in doing so, it reflects the iconic effects of contrasting deictic gestures made by the vocal organs:

- Opening the mouth wide, to address oneself to the rest of the world
- Closing the mouth in self-contemplation, as if in temporary retreat from the world.

The relatively open mouth position, which signifies the directing of one’s attention to the objective realm of OTHER (a generalization, perhaps, of ‘you’ [*hããʔ*]-deixis), is expressed phonetically in Temiar by the low vowel *a*, the back consonants *ʔ* and *h*, and vowel nasality (i.e., velic opening). The relatively closed mouth position, which conversely signifies the more subjective SELF-focused, ‘I’-deixis realm, is expressed phonetically by the high vowel *i* and the front consonants *m*, *y*, *c* and *r*.<sup>8</sup>

Elsewhere (Benjamin 2011), I have argued that this iconic pattern is manifested in a significant degree of phonetic and morphological syncretism in Temiar between (1) pronouns, (2) demonstratives (deictic particles), (3) definitizing clitics, (4) the inflection of human nouns for number, (5) the marking of the sentence for mood and orientation, (6) the marking of discourse continuity and (7) the inflections of the verb for voice and aspect. However, there is no space here to discuss these other features. The present paper is concerned only with the morphology of certain verb forms as well as related features displayed by some nouns.

As displayed in Tables 1 and 2, the middle voice of the Temiar verb is formed by inserting the OTHER-referring infix *-a-* into the presyllable. This carries with it several interconnected semantic implications. First, it indicates that the middle voice is regarded as object-incorporating, and hence heavily intransitive.<sup>9</sup> Second, these features imply in turn that the middle voice refers primarily to dynamic situations, and to events rather than processes. Third, intransitive actions or processes not proceeding beyond the self are more likely to be thought of as punctiliar than actions that bring about a change in an external object, which take more time and hence are more likely to be durative or iterative. Thus the employment of a verb in the middle voice carries with it the implication that the situation it refers to is a DETERMINATE EVENT that has become an OBJECT of attention in its own right.

This prototypically punctiliar quality of the middle voice explains the asymmetry in Tables 1 and 2, where only the base-form and causative of the verb possess distinctive inflections for the perfective and imperfective aspects. Given its peculiar semantic implications, therefore, the Temiar middle voice simply does not require a special imperfective-aspect inflection.<sup>10</sup> I say ‘*protoypically* punctiliar’ because the semantic features under discussion here are connected by strong tendencies – ‘elective affinities’, so to speak (following Max Weber) – rather than strict rules of co-occurrence. The features are therefore graded in character rather than discrete. For example, non-punctiliar usages of the middle voice do sometimes occur, as in the ‘progressive’ forms with *bar-* mentioned below.

<sup>8</sup> For a detailed description of the Temiar sound-system see Benjamin 1976b: 130–153. For further discussion of iconicity in Temiar and elsewhere, see Benjamin 2011. Note that the iconicity under discussion here relates to oral-articulatory gesture rather than the acoustic properties of the sounds produced by those gestures (cf. Jakobson & Waugh 1979: 182).

<sup>9</sup> Although I am keeping to the traditional term ‘middle voice’ (having studied a little Classical Greek in secondary school), the more recent term ‘unaccusative’ (Levin & Hovav 1995) often serves as a better characterization of these usages. Other authors, such as Comrie (1981) use ‘anti-causative’ for the same set of features.

<sup>10</sup> An appropriate morphological pattern could nevertheless be constructed for an imperfective-middle inflection: *gəl* ⇒ *glgəl* [gɛlgɛl] ⇒ *\*glagəl* [\*gɔlagəl] ‘sit’; *səlɔg* ⇒ *sglɔg* [sɛglɔg] ⇒ *\*sgalɔg* [\*sɔgalɔg] ‘sleep’. Although these patterns are not found as regular inflections of the Temiar verb, they are often employed in the highly iconic lexical class known as ‘expressives’, widely employed by speakers of Temiar and other Aslian languages (cf. Diffloth 1972, 1976b, Benjamin 1976b: 177–178).

Given these semantic underpinnings, there are at least three circumstances in which Temiar speakers might choose to employ the middle voice:<sup>11</sup>

- The action referred to by the verb is performed with reference to some *salient other*.
- The *object* of the action or process is *incorporated within* the verb as part of its meaning. It is this pattern that is sometimes referred to as ‘unaccusativity’ (Levin & Hovav 1995).
- The action or event referred to by the verb is a particular *determinate event* treated as a noteworthy object of comment.

### 2.1 ‘Salient-other’ middles

The middle voice is frequently employed when the verb is regarded as involving two or more actors directing their actions simultaneously to each other or to some common focus. There are at least two such usages: ‘all-together’ and reciprocal constructions. In these constructions, the actors treat each other as SALIENT OTHERS, as marked by the *-a-* infix of the verb.

In the following examples, the ‘all-together’ middle-voice constructions (1a) and (1c) are contrasted with their non-middle equivalents, (1b) and (1d):

- (1) a. *ʔun-wawɛʔ*.  
3PL-depart.MID.  
‘They all departed together.’
- b. *ʔun-wɛʔ*.  
3PL-depart.PFV.  
‘They departed.’
- c. *ʔun-bə-salɔg*.  
3PL-PROG-sleep.MID.  
‘They are all sleeping together.’  
[Said of a group of cats and kittens.]
- d. *ʔun-bə-seglɔg*.  
3PL-PROG-sleep.IPFV.  
‘They are sleeping.’

In the following example (2), the first appearance of the verb *tap* ‘to plant’ is in the ‘all-together’ middle-voice form (*tatap*), but in its second appearance it is in the non-middle perfective form (*tap*). (This example, like many of the others in this paper, is taken from my collection of stories recorded in the 1960s.)

- (2) *ʔun-bə-tatap*      *səlaay*      *wɛ-naʔ*.      *ʔun-tap*      *naar*      *ʔis*.  
3PL-PROG-plant.MID      swidden      2DU-that      3PL-plant.PFV      two      day.  
‘They were all planting the two-of-them’s swidden together. They planted for two days.’

‘All-together’ verbs correspond to the ‘simulfactive’ meaning that I ascribed to this form of the verb in my original account (Benjamin 1976b: 172–173). This was supported to some extent by those Temiar-speakers who explained the *-a-* inflection by saying that it marked a plurality of actors.<sup>12</sup> Clearly, actions

<sup>11</sup> On the middle voice as *primarily* semantic and only secondarily syntactic, see Watkins 1976: 309. Some authors, such as Arce-Arenales *et al.* 1994, distinguish between ‘voice’ as a syntactic category and ‘diathesis’ as the equivalent semantic category.

<sup>12</sup> Means (1998: 12) seems to have been told something similar by her committee of Temiar speakers. She refers to the infixed *-a-* form of the verb as the ‘plural’, and provides a purportedly ‘plural’ form for many of the verbs listed in her dictionary. But these *-a-* forms occur so frequently with singular subjects in ordinary Temiar speech (as in

carried out in or by social groups are very likely to be both determinate and simulfactive, as well as indicative of the attention that the individual actors must be paying to each other: hence the bringing together of these various meanings of *-a-* in the ‘all-together’ middle-voice inflection. They are not necessarily punctiliar, however, as ‘all-together’ middles often occur with dual or plural subjects attached to the verb via the ‘progressive’ clitic *bar-*, as in (1c) and (3). I suspect therefore that these ‘all-together’ middles may be underlain by a more fundamentally ‘reciprocal’ semantic.

The reciprocal actions of two individuals are expressed by linking a dual-number pronoun to a middle-inflected verb via the proclitic *bar-* ~ *bə-* ‘progressive’, as in (3a).<sup>13</sup> If the event had been non-reciprocal, with the two protagonists handing food out to third parties rather than to each other, the verb would have been in the non-middle imperfective, as in (3b).

- (3) a. *Ləpas naʔ wɛ-bə-ʔaʔog ʔə-lah.*<sup>14</sup>  
 After that 2DU-PROG-give.MID 3SG-lah.  
 ‘After that, they were giving each other [some]’.
- b. *Ləpas naʔ wɛ-bə-ʔɛgʔog ʔə-lah.*  
 After that 2DU-PROG-give.IPFV 3SG-lah.  
 ‘After that, they were giving [some, to others]’.

## 2.2 Object-incorporating (unaccusative) middles

As illustrated in the specially constructed form-sentences (4a)–(4e), the semantic interpretation of a verb inflected for middle voice depends largely on the properties of its grammatical subject. When the subject is animate and singular the interpretation will usually be that the verb refers to a body-move (4a, 4b) or to some emotional response (4c, as contrasted with 4d), either (agentively) reflexive or (non-agentively) medio-passive. In the absence of a positively agentive marking with the ‘intensive’ infix *-m-*, it is not always possible to tell which meaning is intended. The reason for this ambiguity is not hard to find. To say that the object of an action is incorporated within the verb (as indicated by the infix *-a-*) is to imply that the action (medio-passively, ‘unaccusatively’) does not proceed beyond the subject’s self. This is much the same as saying that the action is thought of as having no external source. Middle voice and intensive mode are therefore unlikely to co-occur, and I have not yet found any examples of such a combination. The simultaneous combination of agentive planning and non-agentive medio-passive ‘emotion’ depicted in (4e) is therefore unlikely ever to be more than a hypothetical example.

- (4) a. *Na-gagəl.*  
 3SG-sit.MID.  
 ‘He sat.’  
 (What did he do?:) ‘He sat down.’  
 (What happened to him?:) ‘He became seated.’

---

examples (7)–(9), below) that the ‘plural’ analysis cannot be correct – regardless of the attempts by my and Means’s Temiar respondents to explicate their language. Consequently, some of the *-a-* forms of the verb given in Means’s dictionary do not actually occur in the language, since they are semantically inappropriate or impossible, while those that do occur are actually middle-voice, not ‘plural’, forms. In Temiar, only adjectives and human nouns inflect for plurality – though not through infixation of *-a-*, which paradoxically indicates the *singular* number in human nouns (Benjamin 1976b: 185)!

<sup>13</sup> Burenhult (2005: 104) reports that Jahai, a Northern Aslian language, employs the *C<sup>a</sup>-a-* morphology as a ‘reciprocal’ inflection in a few verbs. Although he doesn’t say so, this usage was probably borrowed from Temiar, a language that most Jahais can speak.

<sup>14</sup> The phrase *ʔə-lah* is an idiomatic expression marking the declaratory character of the sentence that it terminates, especially in story-telling. There is no satisfactory way of translating it into English.



- b. *Na-salɔg.*  
3SG-lie.MID.  
'He lay down.'  
(Agentive:) 'He lay down and slept.'  
(Non-agentive:) 'He fell asleep.'
- c. *Na-yayaap.*  
3SG-weep.MID.  
'She burst uncontrolledly into tears.'
- d. *Na-yaap.*  
3SG-weep.PFV.  
'She wept.'
- e. *\*Nam-yayaap.*  
3SG<INT>-weep.MID.  
\*'She intentionally burst uncontrolledly into tears.'

An authentic text-example is provided in (5), which tells of the final success of two hungry men in hunting a bird to eat. The details of the cooking are narrated in a string of straightforward perfective verbs. But the final drinking of the soup is reported in the 'emotional' middle voice (*hahuj*, not *huj*), indicating that they were so hungry that they couldn't help but drink the soup up, and were therefore acting (medio-passively) under compulsion.

- (5) *Gəd*      *ʔi-kəwɛ̃ɛs*      *ha-cɔ̃s*      *naʔ*,      *gəd*      *yəəl*,  
Cut.PFV    NOM-child    ACC-bird    that,    cut.PFV    finish,  
*na-pəʔɔ̃l*      *ma-ʔoos*,      *na-didih*      *ʔɔ̃k*      *naʔ*,  
3SG-cook.PFV    to-fire,      3SG-boil.PFV    water    that,  
*wɛ-koh*,      *wɛ-koh*,      *wɛ-bə-hahuj*.  
3DU-pour.PFV,    3DU-pour.PFV,    3DU-PROG-drink.MID.  
'The son cut up the bird, finished cutting he cooked it on the fire, the water boiled, they poured it out and they drank it up.'

With an inanimate subject or with a verb that carries an inherently passive or adversative meaning there is little chance of ambiguity, for then the non-agentive interpretation of the middle voice *must* apply. It is this that gives rise to the passive-like absolute constructions exemplified in (6a), as contrasted with the agentive, non-middle, employment of the same verb in (6b):

- (6) a. *Kəbəəʔ*      *doh*      *na-wawɔ̃g*.  
fruit      this      3SG-open.MID.  
'This fruit split open.'
- b. *ʔi-wɔ̃g*      *kəbəəʔ*      *doh*.  
1SG-open.PFV    fruit      this.  
'I opened this fruit.'

From a semantic point of view (6a) constitutes a passive, as the grammatical subject governing the verb (*na*- '3SG', anaphoric to 'fruit', the topic) is actually the patient. This spontaneous, 'non-controlled' use of the middle voice is the nearest thing to a passive construction in Temiar, which possesses no overt passive inflection.

The employment of the middle voice to indicate absolute meanings is still productive, as exemplified by (7). This was said by a Temiar as he stopped his car for a few minutes to allow it to cool

down. He could have employed the simple perfective form *geej* ‘to burn’, but this would have indicated neither the suddenness nor the apparent absence of external cause expressed by the middle-voice form that he chose.

- (7)        *ʔi-ʔiid*                    *karetɔɔh*    *na-gageej*.  
 1SG-fear.that.PFV    car                    3SG-burn.MID.  
 ‘I was afraid the car would burn up.’

Some verbs almost always occur in the middle-voice form, even though their ordinary perfective base forms are also occasionally used. Of these, *dado* ‘flee’ is an example; the non-middle forms *do* and *dɛʔdo* ‘run’ are also employed, but infrequently. Presumably, ‘fleeing’ is something that Temiar do under compulsion, either in terror or mock terror (as in children’s play), rather than as a fully ‘agitive’ act.<sup>15</sup>

### 2.3 ‘Determinate-happening’ middles

As already noted, the object-incorporating property of the *-a-* infix can also be used to indicate an event’s noteworthiness or unusual character. Examples of this usage are relatively rare, especially when unassociated with ‘all-together’, reciprocal, absolutive or ‘emotional’ meanings. One such example is (8). The middle-voice usage here (*lalɔɔs*, instead of the expected *lɔɔs*) expresses my story-teller’s affected surprise that the protagonist would dare to do what he had just warned his wife not to do.

- (8)        *Na-tipuuʔ*    *ma-leh*,    *na-lalɔɔs*,                    *na-pɛdpood*    *ma-kəbəəʔ*    *tɛɛʔ*.  
 3SG-deceive    to-wife    3SG-return.MID,    3SG-eat.IPFV    to-fruit                    earlier.  
 ‘He deceived his wife, and – believe it or not! – went back and ate the fruit.’  
 [He had earlier (*tɛɛʔ*) forbidden her to eat the fruit, pretending it was poisonous.]

A simpler example, extracted from a conversation that I overheard, is given in (9a); contrast this with the normal non-middle expression in (9b).

- (9)    a.    *Na-tatəd*.  
           3SG-stand.MID.  
           ‘He stood to attention’  
       b.    *Na-təd*.  
           3SG-stand.PFV.  
           ‘He stood.’

### 3. Deponent (inherently middle) verbs

So far, I have been discussing examples of the productive inflectional use of the verbal infix *-a-*. However, as already mentioned, Temiar also possesses a large number of *non*-inflecting, mostly disyllabic, verbs. The sole inflection they exhibit is nominalization with *-n-*.<sup>16</sup>

<sup>15</sup> Cf. Blagden (1906: 620), entry G44, where various forms of this verb are given in different Temiar dialects: *dädó* [*dɛʔdoʔ*] and *dadu* (*daduk*); *dado* (*dadok*), in the phrases ‘run off!’ *em da dok* [*ʔem-dado* ‘PL.INCL<INT>-run.MID!’ or perhaps *ham-dado* ‘2SG<INT>-run.MID!’], and ‘don’t run away’, *ed da dok* [probably, *ʔa-dado* ‘VET-run.MID’].

<sup>16</sup> This lack of inflection is puzzling. In Semai, the closely related Central Aslian language, such verbs do exhibit reduplicative inflectional processes. Gérard Diffloth has suggested to me that this is because Temiar has had less time than Semai to absorb these items, which presumably are largely of Austronesian provenance, and from Chamic in particular. (The verb *golap* in Table 4 is such an import from Chamic.) My own lexicostatistical data (Benjamin 1976a: 73) confirm that Semai has had more contact with Malay than Temiar has had. Thus, (1) disyllabic verbs, especially those (like *golap*) with something other than *-a-* in the first syllable should be examined for Austronesian etymology, and (2) forms with *-a-* in the first syllable should be examined for evidence of having been altered from some other vowel. While proposal (1) must await another occasion, proposal (2) is mentioned briefly in Section 5.1 below.

Ancient Greek and Latin possessed a class of verbs that carried an apparently active meaning despite an overtly middle (in Latin grammars, usually labeled ‘passive’) morphology. Traditionally, these verbs are referred to as ‘deponents’, because they appear to ‘lay aside, shed’ (Latin, *deponere*) the middle or passive sense suggested by their morphology. Temiar too has such a class of deponents, consisting of a number of non-inflecting disyllabic verb-stems possessing the full vowel *a* in the first syllable but with meanings that appear on initial examination to be wholly active. These verbs lack almost all inflection, so that *halab* ‘to travel downriver’ (Table 3), for example, possesses no perfective *\*həlab*, no imperfective *\*heblab*, and no causative *\*herlab* or *\*həreblab*. They do, however, take the progressive clitic *bar-* ~ *bə-*, which generally serves with these verbs to express durative or imperfective meanings, as in *bə-halab* ‘travelling downriver’. They also form verbal-nouns by *-n-* infixation, as in *hənalab* ‘a journey downriver’. Other examples are: *satəh* ‘to collapse’, *carəəh* ‘to walk down-hill’ and *gabag* ‘to sing’.

But do these deponent verbs really possess active meanings? Just as with some Latin deponents (Baldi 1974: 19), closer examination suggests that these verbs are interpretable as having inherently middle-voice meanings.<sup>17</sup> For some of the verbs this is not hard to accept. When travelling downriver (*halab*), one passively lets one’s raft be carried by the current, and the only action the rafter takes is to steer the craft away from obstructions; the meaning in this case falls into the medio-passive category. When a house collapses (*satəh*), that is something it undergoes, not something it does: the meaning is essentially the same as that of the absolutive unaccusative middles discussed earlier. Similarly, going downhill (*carəəh*) in Temiar country is hard for walkers to control, as they descend precipitously through the undergrowth. *Carəəh*, then, is something one *suffers* to some extent, and the middle-voice morphology is again quite appropriate.<sup>18</sup> In contrast, the verb *təŋuh* ‘go uphill’, which undeniably denotes an action, lacks the *a* vowel. Indeed, I doubt whether the middle-voice form *\*təŋuh* would be acceptable to a Temiar speaker, as it would imply an unacceptable combination of active and passive meanings.<sup>19</sup>

But what is one to make of the middle-voice morphology of *gabag* ‘to sing’? Surely, singing is unambiguously an *activity*? It is here that culturally derived ideas intervene, for traditional Temiar songs are not primarily thought of as instances of simple self-expression or public entertainment. Temiars sing either to quietly entertain a visiting spirit-guide positioned somewhere near their own dream-soul (*rəwaaq*), or while performing ceremonially as mediums through whom visiting spirit-guides sing out their message to the whole community. In the former case the singer will usually be transiting gradually from dreaming to waking (typically at daybreak), and the singing will in effect be self-directed. This kind of singing is semantically like a reflexive body-move (cf. Diffloth 1974), for which the middle-voice inflection – if the language possesses one – is the appropriate form of the verb. On the other hand, where the singer performs publicly as a spirit-medium, he (occasionally, she) will typically be in a light trance and his singing will be conceived of as a passive experience issuing from a supposedly external agentive source, the spirit-guide. Here again, the inherently middle-voice morphology of the verb *gabag* is quite appropriate to the meaning that singing possesses in Temiar culture.<sup>20</sup>

This analysis of the verb *gabag* both explains and gains support from a puzzling ethnographic curiosity. During communal song-sessions, Temiars often shout encouragement to the lead-singer with the phrase *ma-rii’ hah* ‘to your self!’. One would expect this literally to mean ‘keep it to yourself!’ (i.e. ‘quieten

<sup>17</sup> Diffloth (1976a: 242–243) presents a list of disyllabic Semai verbs with *-a-* in the first syllable. Although the *-a-* infix is not productive in Semai (unlike Temiar), many of the meanings there can nevertheless be analyzed as ‘deponent’ in character.

<sup>18</sup> The likely etymology of *carəəh* is instructive: Mon *carəh* ‘to fall’, Mon *garəh* ‘to push, shove, fall down in pieces’, *carəh da’* ‘to drop water on the ground’, Khmer *jrəh* (*chrüh*) ‘tomber, se détacher’ etc. (Pinnow 1959: 91, no. 106). Temiar therefore has *added* the *-a-*, reinterpreting the experience of precipitate downhill movement as something like ‘to fall and be made to fall’. (However, there is also a rarely used Temiar verb *cəreəh* ‘to chop bamboo down at its base’ that could be related to *carəəh*.)

<sup>19</sup> Conceivably, *təŋuh* could be employed for ‘all climbing a hill together’, but I have yet to hear this usage.

<sup>20</sup> For more on the musical context, see Roseman 1984, who incorporates my analysis of *gabag* ‘sing’ on p. 422. For a directly parallel instance from Malay, see my discussion of the choice between the older middle-voice form *bernyanyi* and the now more usual active-voice *menyanyi*, both meaning ‘sing’ (Benjamin 1993b: 378). Present-day Temiars refer to their own and others’ pop-music ‘singing’ as *nəŋii’*, borrowing the Malay word.

down!’), but the attendant behavior makes it clear that even greater volume and enthusiasm are being called for. The point, of course, is that the Temiars themselves conceive of public singing-out as a self-directed, ‘middle’ activity. Here the dialectical entanglement of SELF and OTHER that is so characteristic of Temiar culture (see below) infuses both language and behavior simultaneously.

Further examples of middle-voice deponent verbs and adjectives (which in Temiar are stative verbs) are given in the Appendix below. The listing also gives suggested reasons for the middle-voice semantics exhibited by these verbs. It may seem strange that many of these deponent verbs are stative, for I earlier characterized the middle voice as being prototypically dynamic. The answer, it seems to me, is that these stative deponent verbs represent states that are assumed to be in *dynamic* equilibrium, held in place by the simultaneous pull of two opposed forces. Moreover, these verbs are non-inflected lexical items complete in themselves, rather than inflected middle-voice forms selected from within an inflectional paradigm that offers such other choices as causative or imperfective.

### 3.1 Partly-inflecting (semi-deponent) verbs

In the previous sections I discussed only two basic morphological classes of verbs in Temiar: inflecting and non-inflecting. However, there also exist verbs that fall between these two poles by exhibiting a restricted range of inflections. These ‘semi-deponent’ verbs fall into two subtypes: some follow a middle–imperfective distinction (Table 5), while others exhibit a middle–causative distinction (Table 6). In both cases, the least-inflected form of the verb is a deponent in *-a-*; there is no simple base form in *-ə-*.

VOICE	ASPECT			VERBAL NOUN
	Perfective	Imperfective	Progressive	
Base	—	<i>hɣrɛk</i> [hɛŋrɛk] ‘clearing up’ (incomplete)	<i>bar-hɣrɛk</i> [bar-hɛŋrɛk~ bə-hɛŋrɛk] ‘currently clearing up’	<i>hɛŋrɛk</i> [həŋrɛk] ‘a clearing-up’
Middle	<i>harɛk</i> [harɛk] ‘cleared up’ (complete)	—	—	—
Causative	—	—	—	—

Table 5: Partly-inflecting (type 1, base–middle): *harɛk* ‘to clear up (sky, river)’

VOICE	ASPECT			VERBAL NOUN
	Perfective	Imperfective	Progressive	
Base	—	—	—	—
Middle	<i>kaleek</i> [kaleek] ‘to prop, be propped’	—	<i>bar-kaleek</i> [bar-kaleek~ bə-kaleek] ‘currently propping’	<i>knaleek</i> [kənaleek] ‘a prop’
Causative	<i>krleek</i> [kɛrleek] ‘set a prop in place’ (completed act)	<i>krɛleek</i> [kɛrɛleek] ‘setting a prop in place’ (incomplete act)	—	—

Table 6: Partly-inflecting (type 2, middle–causative): *kaleek* ‘to prop’

Typical Temiar props (*kənaleek*, Table 6) are the struts set athwart a dug-out canoe, or the scaffolding (also *kənaleek*) jammed between the ground and a house or platform to stop it collapsing, or the bridges on a stringed instrument. What these have in common is that they simultaneously hold two objects apart while being held in place by those objects. Such props thus affect while being affected – which is exactly what Benveniste (1971: 149) identifies as the chief criterion of the middle voice, as opposed to the

active or passive voice. The verb *kaleek* therefore deserves its inherently middle-voice morphology; an active-voice base-form morphology (*\*kaleek*) would not be appropriate. The causative *kerleek*, meaning ‘to set a prop or scaffold in place’, is in regular use, but the semantic implications of choosing between this and the semi-deponent middle voice are not always obvious, as illustrated in (10).

- (10) *Na-kaleek*            *ya-ha-jəhũʔ*    *ʔə-naʔ*,    *na-kaleek*  
 3SG-scaffold.MID    yes-ACC-tree    3SG-that,    3SG-scaffold.MID  
*gəgəgəg*        *bəlas*    *ma-jalək*    *ʔəh*.  
 until [EXPR]    all.way    to-top        3SG.  
 ‘He scaffolded the tree, he scaffolded all the way up to its top.’

Here, my story-teller chose not to employ the causative *kerleek*, even though that might seem to have been the appropriate form. Instead, he employed the semi-deponent middle-voice form *kaleek*, regardless of the accusatively marked object (*ha-jəhũʔ* ‘the tree’) that it appears to govern. His aim, presumably, was to emphasize how the climber was supported *himself* as he ascended and to downplay what was being done to the tree. If, on the other hand, the story had required the protagonist to prop up a rickety house to prevent it collapsing, the narrator would probably have chosen the causative forms *kerleek* or *kəreŋleek*.

Further examples of such ‘type 2’ partly-inflecting verbs are:

- *catək* ‘to close’ (said of a door closing by itself: absolutive, non-agentive, intransitive)’, *certək* ‘to close’ (said of someone closing a door: agentive, transitive)
- *lakəʔ* ‘to break off, come off’ (said of something falling off spontaneously: absolutive, non-agentive, intransitive), as opposed to *lerkəʔ* ‘to break something off, remove’ (agentive, transitive)ʔ.

An overheard example of the latter is given in (11). This reported on the success of a surgeon in repairing the damage done to a Temiar road-accident victim.

- (11) *Na-lerkəʔ*            *bəsiiʔ*    *num-gəntək*.  
 3SG-come.off.CAUS    iron        from-ear.  
 ‘He removed metal from the ear.’

### 3.2 Non-deponent verbs in -a-

As noted earlier, there are several non-inflecting Temiar verbs with *-a-* in the initial syllable, but which do not exhibit any clearly middle-voice semantic. (See the ‘Unclassified’ section in the Appendix.) Examples are *sapood* ‘to wrap, make a parcel (an Austronesian loan)’ and *dalag* ‘to call (someone)’. Further investigation may yet reveal an underlying middle-voice semantic in many of these verbs. But, given that [a] is the most common vowel in the languages of the world, it is not surprising that this too sometimes occurs in the presyllable of non-deponent verbs, along with such vowels as [o] or [e].

## 4. Verbs in -a- and Temiar culture

As we have seen, there are at least four classes of verbs with *-a-* in their presyllable:

- Productive middle-voice forms, as part of a fully inflecting verb paradigm that also contains active and causative forms: *gagəl-gəl-tergəl* ‘sit’ (Table 1) and *saləg-sələg-serləg* ‘sleep’ (Table 2).
- True ‘deponents’, non-inflecting or partly-inflecting verbs with an underlying middle-voice semantic: *halab* ‘travel downriver’ (Table 3), *hareek* ‘clear up’ (Table 5).
- Partly-inflecting verbs that are agentless (and intransitive) in their primary middle-voice form, but which also possess an agentive causative-transitive inflection: *lakəʔ* ‘to break off (intrans)’, *lerkəʔ* ‘to break off (trans)’, but no *\*ləkəʔ* (Table 6).

- Non-inflecting *-a-* verbs with (as yet) no discernible middle-voice semantic: *dalag* ‘to call (someone)’.

This pattern discloses a more general theoretical issue. According to Comrie (1981: 161),

while the genuine derived causative [in Temiar, verbs in *-r-* ~ *ter-*] may be a productive process, the derived anti-causative [in Temiar, verbs in *-a-*] will not be, since one cannot iteratively reduce the degree of transitivity of a predicate: once it is intransitive, that is necessarily the end of the process.

But the Temiar middle voice in *-a-* does precisely that, as in the fully productive series *gagəl* (middle) – *gəl* (active) – *tergəl* (causative), where *gəl* ‘sit’ (Table 1) is indeed intransitive. Thus, the Temiar middle-causative axis is not primarily syntactic, but *semantic*, referring to the contrast between ‘inside’ actions and external ones, with the neutral unmarked action, *gəl*, represented by the root-form of the verb. This suggests that in Temiar the syntactic valency schema is calqued upon the semantic relations, rather than the other way round. So, while it may be syntactically irrational to further ‘reduce’ an intransitive verb like *gəl*, it is still a *meaningful* thing to do, given the framework of understanding that Temiar-speakers operate within.<sup>21</sup>

The predominant orientational mode of Temiar culture is formed of the dialectical interplay of SELF and OTHER. This dialectical pattern infuses their social interactional style (Benjamin 1994: 44–47), their religious life (Benjamin 1993a: 271–273), and their musical structures (Roseman 1984). The closest that Temiars ever come to talking explicitly about this aspect of personhood is in discussing the animating subjectivities (‘souls’) that are said to inhabit a wide range of beings, including people. It is then that their ideas about the mutual entangling of SELF and OTHER become patent.

Such ‘animate’ beings are said each to possess two souls, one associated with its upper part (the head-hair roots of humans and animals, the leaves of trees, and the summits of mountains) and the other with its lower part (the heart, breath and blood, the roots, and the subterranean mass). Dreamers and spirit-mediums report that upper-body souls when encountered as spirit-guides (*gonig*) are like young men or women in appearance, but that spirit-guides derived from lower-body souls appear as tigers. In other words, upper-body souls are seen as familiar, domestic and SELF-like, while lower-body souls are seen as strange, wild and OTHER-like. Yet it is a person’s heart-*cum*-tiger soul, the *hup*, that is claimed to be the source of his or her will and agency: it is one’s *hup* that makes one do things or, alternatively, lacks the desire to do something. Tigers, of course, are clearly OTHER. But it is also possible to perceive as ‘other’ the usually autonomous beating of one’s heart (also *hup*)<sup>22</sup> or one’s breathing (*hemnum*, the *-n-*infix reduplicated form of *hup*), since these can be directly monitored by the individual without their needing to be controlled. The head-soul, *rəwaay*, on the other hand, is clearly SELF-like in its association with the incessant but unobserved growth of the hair – the marker of bodily integrity. But the *rəwaay* is also cast in the role of a patient-like, non-controlling experiencer of whatever befalls the individual in dreams, trance and sickness – one form of which carries the (imperfective) verbal label *reywaay*, ‘to suffer uncontrolled soul-loss’.

Thus, for the Temiars, the controlling *hup* (the ‘I’) is an autonomous OTHER inside the person, while the experiencing *rəwaay* (the ‘me’) is an equally autonomous, but non-controlling, SELF. In other words, *hup* beliefs imply that one’s actions are at the same time something that one undergoes, while *rəwaay* beliefs imply that the things one undergoes (growth, dreaming etc) are at the same time one’s actions (as in one’s dreams, which Temiars sometimes talk of as if they were activities that they can monitor). The individual’s empirical self or felt subjectivity is thus portrayed as a dialectical SELF-and-OTHER composite.

This cultural connection goes some way towards explaining why it is the ‘objective’ marker *-a-* that indicates the middle voice in Temiar, whereas many other languages express the middle-voice meanings through a transparently ‘subjective’ marking. The *-r-* in the Malay middle-voice prefix *ber-* also follows the

<sup>21</sup> The increasingly common (British) English expression ‘I was sat there’ is an interesting parallel to the Temiar *gagəl* ‘sit.MID’, in that it too is a (covertly) middle-voice construction involving an already intransitive verb.

<sup>22</sup> In stricter anatomical usage, as when butchering an animal, some Temiars associate *hup* with the liver rather than with the heart.

same SUBJECTIVE pattern (Benjamin 2009: 306–310). But since the subject of a middle-voice verb is as much affected as affecting, the option exists for it to be encoded semantically as objective (*-a-*) rather than subjective, even though that objectivity refers back to the grammatical subject.

This interpretation is supported by the manner in which the Temiar middle and causative voices<sup>23</sup> relate to each other as the two poles of the valency schema. (As already noted, there is no inflectional passive voice in Temiar.) Whereas the middle voice carries a ‘subjective’ meaning through an iconically expressed incorporation of the syntactic *object* into the verb (as *-a-*), the causative voice expresses the ‘objective’ meaning of getting some *other* source or agency to do something through an iconically expressed incorporation into the verb of a SELF-referring marker, the high consonant *-r-*. An example would be: *reŋkaaʔ na-cətək* (door 3SG-close.MID) ‘the door closed’ as opposed to *ha-crətək [cɛrətək] reŋkaaʔ* (2SG-close.CAUS.PFV door) ‘you closed the door’ (i.e. ‘you caused the door to close’). Thus, the dialectical SELF–OTHER deixis pervades the semantic and grammatical organization of the Temiar verb, and is given phonic expression through iconicity.<sup>24</sup>

I end this paper by showing that this pattern extends to Temiar nouns as well as verbs.

### 5. Middle-voice (unaccusative) nouns and nominalizations

The idea that middle-voice nouns can exist would seem to confound linguistic common sense. However, if it is accepted that nouniness and verbiness are scalar variables rather than absolutes (Sasse 2001), then the idea should be less surprising.

In Temiar there exist many disyllabic nouns with *-a-* in the first syllable, as well as deverbalized forms combining *-a-* with the *-n-* nominalizer. On semantic grounds, a high proportion of these qualify as middle-voice nouns, in that they refer to entities that can be thought of as simultaneously both *acting* and *acted upon*, or as being both their own *source* and *undergoer*.<sup>25</sup> A categorized list of these nouns is given in the Appendix. The largest category – at least 135 items – consists of mammals, reptiles, birds and arthropods. As legged or winged animals, these *make themselves* move, and can therefore be thought of as their own source and undergoer. The middle-voice morphology is therefore entirely appropriate. Fish names, on the other hand (not listed in the Appendix), exhibit very few forms in *-a-*, perhaps because they are thought of simply as undergoers of the flow of water in which they find themselves.

Fewer plant-linked words contain the vowel *-a-* in the first syllable. This somewhat reduces the likelihood that they are interpreted as middle-voice nouns. But to the extent that these few words might be so interpreted, I suggest that it would be on the grounds that plants, in shaping themselves, are the undergoers of their own actions. ‘Actions’ may seem inappropriate when talking of plants, but (as pointed out earlier) Temiar animism does indeed accord agentive communicable-with subjectivity to several plant species.

Other categories of nouns with *-a-* in the first syllable include body parts, which can be thought of as both moving and being moved.<sup>26</sup> A small but significant category consists of words for human relations, which Temiars see as dialectical in character (cf. Benjamin 1994). The list of words for physical objects

---

<sup>23</sup> Linguists do not usually refer to the causative as a ‘voice’ of the verb, preferring to regard it as a derivational formation. On intuitive grounds, however, it seems appropriate to include the causative in whatever grammatical category the active, passive or middle are assigned to, and these have traditionally been labeled the ‘voices’ of the verb. The purely syntactic term ‘valency’ would also be applicable, but for its failure to capture some of the semantic properties that are at issue here. There is, however, a good partial precedent for the usage I am following here in Lehmann’s reference (1974:184), while discussing the middle and passive, to ‘other meanings comparable to voices, such as the causative.’ Jakobson (1957:4) too would seem to support this view: ‘voice characterizes the relation between the narrated event and its participants without reference to the speech event or to the speaker.’

<sup>24</sup> For further discussion of the dialectical mode of coherence that underlies Temiar culture, see Benjamin 1994: 46–47.

<sup>25</sup> Just as with verbs, there are many nouns with this shape that have no apparent middle-voice meaning – at least not that I can discern as of this writing. I list some of these at the end of the Appendix.

<sup>26</sup> Diffloth (1974: 128–131) discusses the same semantic issue in Semai, a close relative of Temiar, where syntactic analysis shows that there too body-moves are not accorded a purely agentive interpretation.



in *-a-* contains many in which the middle-voice semantic is highly appropriate, in that they refer to entities that hold themselves in place or which produce their own reaction (as explained in the Appendix below).

Examples of middle-voice nouns include:

- *layeg* ‘night’: because, as English puts it, night falls, happening spontaneously into existence of its own accord. This word is both a noun (as in *layeg teeʔ* ‘last night’) and a verb (as in *həj na-layeg* ‘already 3SG-night’, i.e. ‘night has fallen’).
- *sagub* ‘cloud’: because clouds appear to bring themselves into being.

### 5.1 Etymological considerations

Etymological evidence suggests that *-a-* has been (deliberately?) inserted into certain Temiar words, both nouns and verbs, at some stage in their history, to accord better with the middle-voice semantic framework I have been discussing. Consider the following Mon-Khmer cognate-sets and derivational series.<sup>27</sup>

- *gatiʔ* ‘snail’, cf. Shorto’s Proto Mon-Khmer reconstruction *\*gtooʔ*, but on very restricted data (Shorto 2006: 85): because a snail both supports and is supported by its own shell. Also, snails are animals, and hence self-moving.
- *jalaaʔ* ‘thorn’, cf. Mon *jərlaʔ*, *jəlaʔ* (Shorto 1971: 124), Jahai *jəlaʔ* (Niclas Burenhult, p.c.) and Semnam (Lanoh) *jilaaʔ* (Burenhult & Wegener 2009: 291): because a thorn both snags and gets snagged on people’s clothing.
- *jalək* ‘tree-top’, cf. Old Mon *clon /clɔŋ/* ‘highest peak, apex, spire’ (Shorto 1971: 114): because a tree grows its own ‘summit’.
- *kalɔʔ* ‘stupid, silly, ignorant, dumb’, cf. Middle Mon *kamlau*, *kəmlaw* ‘dumb’ (Shorto 1971:31): because dumbness is something one undergoes despite oneself, not something one does.
- *kaŋiiʔ* ‘to court, flirt’, cf. Old Mon *(ka)ŋi ~ guŋi* ‘to embellish’ (Shorto 1971:78): among Temiars this is always reciprocal, the original meaning probably being ‘to adorn each other’.<sup>28</sup>
- *kawaaʔ* ‘kinsperson’, cf. Spoken Mon *kəwa* ‘companion(s)’ (Shorto 1962: 227): in Temiar, kinspersons are necessarily *reciprocally* so.
- *januʔ* ‘chief’, cf. Old Mon *jnok* ‘to be great, high-ranking’ (Shorto 1971: 128): Temiar chiefs become and remain such only through continuing dialectical relations with their followers.
- *sakɔʔ* ‘white-haired, grey-haired’, cf. Old Mon *skɔʔ*, Modern Mon *həkɔʔ* ‘grey haired’ (Shorto 2006: 74): because hair becomes grey of its own accord.

### 5.2 Middle-voice deverbal nominalizations

Temiar verbs are regularly nominalized to form verbal nouns by the infixation or prefixation of *(-)n-*, as displayed in Tables 1–4. Less productively, the infix *-n-* also occurs in a set of full nouns denoting physical or abstract entities. An example is *cəner* ‘knife’, derived from the base-form of the verb *cər* ‘to pare’. However, several such nouns relate more to the middle-voice of the verb, retaining or inserting the characteristic *-a-* infix along with the middle-voice semantic that it marks, as illustrated in the following examples:

- *bənatək*, *mənatək* ‘eyelet-loop (on a basket)’, from *batək* ‘to make an eyelet-loop’: because it both supports and hangs from the strap that goes through it.

<sup>27</sup> See also footnote 18, on the Mon-Khmer etymology of *carəəh* ‘to go downhill’.

<sup>28</sup> A photograph of Temiar ‘allo-adornment’ appears as plate 3 of Wavell *et al.* 1966, opposite p.19.

- *canaa* ‘food’, from *caa* ‘to eat’: ‘that which undergoes eating’. Contrast this with the verbal noun *ce’naa* ‘an eating’.
- *canuuŋ* ‘beater’, from *cuuk* ‘to beat, hammer’: because a beater rebounds in the user’s hand. Contrast this with *cəner* ‘knife’, above, which is not \**caner* because a knife acts more ‘transitively’ and consequently suffers no ‘feedback’.<sup>29</sup> (See also Matisoff 2003: 27.)
- *ganas* ‘tool for reaching fruit down from a tree’, from *gas* ‘to reach fruit down from a tree’: because a *gas*-ed fruit falls of its own accord, and the actor and tool ‘undergo’ the fruit-fall. This is probably an Austronesian loan: cf. Dempwolff 1938: 53, \**gat* ‘entzweisein’, Toba Batak, Javanese *gas*.
- *laney* ‘knowledge’, from *lek* ‘to know’: because in a non-literate society knowledge is attained by undergoing experience, not through active study.
- *mənanuu* ‘size, bigness’, from *mənuu* ‘to be big’: ‘that which has become or been made big’.
- *panəh* ‘shamanic dancing space’, from *pəh* ‘to hold a séance’: the portion of the house where both the shaman and the house itself undergo possession by a spirit-guide, which is also sometimes referred to as one’s *pəh*.
- *sənaləg* ‘marriage, married state’, ‘undergoing marriage’, from *sələg* ‘to marry’. Contrast this with the verbal noun *sənləg* ‘marrying’.

### References

- Arce-Arenales, Manuel, Melissa Axelrod & Barbara A. Fox. 1994. ‘Active voice and middle diathesis.’ In: Barbara Fox & Paul J. Hopper (eds), *Voice: Form and Function*, Amsterdam: Benjamins, pp. 1–21.
- Baldi, Philip. 1974. ‘Reciprocal verbs and symmetric predicates.’ In: Michael W. La Galy *et al.* (eds), *Papers from the Tenth Regional Meeting*, Chicago: Chicago Linguistic Society, pp. 17–26.
- Benjamin, Geoffrey. 1976a. ‘Austroasiatic subgroupings and prehistory in the Malay Peninsula.’ In: Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds), *Austroasiatic Studies, Part I*, Honolulu: University Press of Hawaii, pp. 37–128.
- . 1976b. ‘An outline of Temiar grammar.’ In: Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds), *Austroasiatic Studies, Part I*, Honolulu: University Press of Hawaii, pp. 129–187.
- . 1993a. ‘Temiar.’ In: Paul Hockings (ed.), *Encyclopedia of World Cultures, Volume 5: East and Southeast Asia*, Boston: G. K. Hall/MacMillan, pp. 265–273. (Also: [www.everyculture.com/East-Southeast-Asia/Temiar.html](http://www.everyculture.com/East-Southeast-Asia/Temiar.html).)
- . 1993b. ‘Grammar and polity: the cultural and political background to Standard Malay.’ In: W. A. Foley (ed.), *The Role of Theory in Language Description*, Berlin: Mouton de Gruyter, pp. 341–392.
- . 1994. ‘Danger and dialectic in Temiar childhood.’ In: Josiane Massard-Vincent & Jeannine Koubi (eds), *Enfants et Sociétés d’Asie du Sud-est*, Paris: L’Harmattan, pp. 37–62.
- . 2001. ‘Orang Asli languages: from heritage to death?’ In: Razha Rashid & Wazir Jahan Karim (eds), *Minority Cultures of Peninsular Malaysia: Survivals of Indigenous Heritage*. Penang: Malaysian Academy of Social Sciences (AKASS), pp. 101–122.
- . 2004. ‘Aslian languages’, ‘Aslian: characteristics and usage.’ In: Asmah Haji Omar (volume editor), *The Encyclopedia of Malaysia, Volume 12: Languages and Literatures*, Kuala Lumpur: Archipelago Press, pp. 46–49.
- . 2009. ‘Affixes, Austronesian and iconicity in Malay.’ *Bijdragen tot de taal-, land- en volkenkunde* 165: 291–323 (also [www.kitlv-journals.nl/index.php/btlv/article/view/3653/4420](http://www.kitlv-journals.nl/index.php/btlv/article/view/3653/4420)).
- . 2011. ‘Temiar morphology (and related features): a view from the field.’ 49pp. Unpublished typescript.
- Benveniste, Emile. 1971. *Problems in General Linguistics*. Translated by Mary Elizabeth Meek. Coral Gables: University of Miami Press.

<sup>29</sup> The morphophonological changes involved in the derivation of *canuuŋ* from *cuuk* and of *laney* from *lek* are explained in Benjamin 1976b: 143.

- Blagden, C. O. 1906. 'Language.' In: W. W. Skeat & C. O. Blagden, *Pagan Races of the Malay Peninsula*, volume 2, London: MacMillan, pp. 379–775.
- Broselow, Ellen & John J. McCarthy. 1983. 'A theory of internal reduplication'. *The linguistic review* 3: 25–88.
- Burenhult, Niclas & Claudia Wegener 2009. 'Preliminary notes on the phonology, orthography and vocabulary of Semnam.' *JSEALS: Journal of the Southeast Asian Linguistics Society* 1: 283–312.
- Carstairs-McCarthy, Andrew. 1992. *Current Morphology*. London/New York: Routledge.
- Comrie, Bernard. 1981. *Language Universals and Language Typology*. Oxford: Blackwell.
- Diffloth, Gérard. 1972. 'Notes on expressive meaning.' In: Paul Peranteau *et al.* (eds), *Papers from the 8<sup>th</sup> Regional meeting, Chicago Linguistic Society*: 440–447.
- . 1974. 'Body moves in Semai and in French.' In: Michael W. La Galy *et al.* (eds), *Papers from the 10<sup>th</sup> Regional Meeting, Chicago Linguistic Society*, pp. 128–138.
- . 1975. 'Les langues mon-khmer de Malaisie: classification historique et innovations.' *Asie du sud-est et monde insulinde* 6 (4): 1–19.
- . 1976a. 'Minor-syllable vocalism in Senoic languages.' In: Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds), *Austroasiatic Studies, Part I*, Honolulu: University Press of Hawaii, pp. 229–247.
- . 1976b. 'Expressives in Semai.' In: Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds), *Austroasiatic Studies, Part I*, Honolulu: University Press of Hawaii, pp. 249–264.
- Frampton, John. 2004. *Distributed Reduplication*. Unpublished monograph, February 2004. Northeastern University. Downloaded from: [www.math.neu.edu/ling/DR/DR\\_March1\\_04.pdf](http://www.math.neu.edu/ling/DR/DR_March1_04.pdf).
- Gafos, Adamantios I. 1998a. 'Eliminating long-distance consonantal spreading.' *Natural language and linguistic theory* 16: 223–278.
- . 1998b. 'A-templatic reduplication.' *Linguistic enquiry* 29: 515–527.
- . 1999. *The Articulatory Basis of Locality in Phonology*. New York: Garland.
- Hendricks, Sean. 2001. 'Reduplication without prosodic templates: a case from Semai.' *Journal of East Asian linguistics* 10: 287–306.
- Idsardi, William J. & Eric Raimy. N.d (2006?). 'Representational economy.' No pagination [73 pages]. Downloaded from: <http://mendota.english.wisc.edu/~raimy/papers/idsardiraimyrepecon.pdf>.
- Itô, Jungko. 1989. 'A prosodic theory of epenthesis.' *Natural language and linguistic theory* 7: 217–259.
- Jakobson, Roman. 1957 [1971]. 'Shifters, verbal categories and the Russian verb.' Pages 130–147 in his *Selected Writings, Volume 2*. The Hague: Mouton.
- & Linda R. Waugh. 1979. *The sound shape of language*. Brighton: Harvester Press.
- Lehmann, Winfred. 1974. *Proto-Indo-European Syntax*. Austin: Texas University Press.
- Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity*. Cambridge MA: MIT Press.
- Matisoff, James A. 2003. 'Aslian: Mon-Khmer of the Malay Peninsula', *Mon-Khmer studies* 33:1–58.
- McCarthy, John. 1982. 'Prosodic templates, morphemic templates, and morphemic tiers.' In: Harry van der Hulst & Norval Smith (eds), *The Structure of Phonological Representations (Part 1)*, Dordrecht: Foris, pp. 191–223.
- Means, Nathalie. 1998. *Temiar-English, English-Temiar Dictionary*. (Edited by Gordon P. Means.) St Paul: Hamline University Press.
- Pinnow, Heinz-Jürgen. 1959. *Versuch Einer Historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Harrassowitz.
- Raimy, Eric. 2000. *The Morphology and Morphophonology of Reduplication*. Berlin: Mouton de Gruyter.
- Roseman, Marina. 1984. 'The social structuring of sound: the Temiar of Peninsular Malaysia.' *Ethnomusicology* 28: 411–445.
- Sasse, Hans-Jürgen. 2001. 'Scales between nouniness and verbiness.' In: Martin Haspelmath *et al.* (eds), *Language Typology and Language Universals, Part 1*. Berlin: Walter de Gruyter, pp. 495–509.
- Shaw, Patricia A. 1994. 'The prosodic constituency of minor syllables.' In: *Proceedings of the Twelfth West Coast Conference on Formal Linguistics*: 117–132. Stanford: Stanford Linguistics Association.
- Shorto, H. L. 1962. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press.
- . 1971. *A Dictionary of the Mon Inscriptions from the Sixth to the Sixteenth Centuries*. London: Oxford University Press.
- . 2006. *A Mon-Khmer Comparative Dictionary*. Canberra: Pacific Linguistics.

- Sloan, Kelly. 1988. 'Bare-consonant reduplication: implications for a prosodic theory of reduplication.' *Proceedings of the Eleventh West Coast Conference on Formal Linguistics*, 319–330. Stanford: CSLI Publications.
- Spaelti, Philip. 1997. 'Dimensions of variation in multi-pattern reduplication.' Unpublished PhD thesis in Linguistics, University of California, Santa Cruz.
- Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Blackwell.
- Ter Mors, Christine. 1983. 'Affix to X. Broselow & McCarthy versus Marantz: a reinterpretation.' *The linguistic review* 3: 275–297.
- Walther, Markus. 2000. 'Temiar reduplication in one-level prosodic morphology.' *Proceedings of the SIGPHON 2000 Workshop on Finite-State Phonology*: 13–21, Luxembourg, 6 August 2000. Downloaded from: [www.arxiv.org/pdf/cs.CL/0008015](http://www.arxiv.org/pdf/cs.CL/0008015).
- Watkins, C. 1976. 'Towards Proto-Indo-European syntax: problems and pseudoproblems.' In: Sanford B. Steever *et al.* (eds), *Papers from the Parasession on Diachronic Syntax*, Chicago Linguistic Society, pp. 305–326.
- Wavell, Stuart *et al.* 1966. *Trances*. London: Allen & Unwin.
- Yap Ngee Thai. 2009. 'Non-exhaustive syllabification in Temiar.' *JSEALS: Journal of the Southeast Asian Linguistics Society* 1: 267–282.

## Appendix

### Lists of deponent verbs and middle-voice nouns

This list was re-checked in the field in October 2010, but it is still not exhaustive. Malay loanwords have been excluded. Many of the words listed here were characterized by my respondents as ‘old’ (*manah*). Other words that I had collected previously – sometimes just two years earlier! – were now judged not to exist at all, and are accordingly not listed here. These changing evaluations may indicate that under modern conditions such middle-voice meanings are thought to be less noteworthy than previously.

### Verbs

- ʔadees* ‘to exercise (one’s body), keep moving around’: one does this to oneself  
*ʔaluur* ‘to shoot/be shot backwards, of felled branch’: it undergoes its own action  
*ʔaroʔ* ‘to talk nonsense, to joke; to flirt, entice’: a socially reciprocal activity  
*ʔayir* (1) ‘to plait the interwoven bottom on a bag’: the strands hold themselves in place  
*babɔh* ‘to act undecidedly’: one undergoes one’s own indecision  
*bacuuh* ‘to imitate (an animal’s sound); to echo’: because the animal sounds through the imitator; because the sound comes back upon the sounder  
*bagəd* ‘to try something new for the first time’: both an experience and an activity  
*bakɔʔ* ‘set a *bakɔʔ* trap’: such traps are self-springing  
*baleel* ‘to roll something up in a wrapper’: because the wrapping then holds itself in place  
*balaaw* ‘to fail to catch game when hunting’: the hunter undergoes the failure  
*balec* ‘to abuse, speak angrily to someone’: such behavior rebounds on the abuser  
*baləc* ‘to not believe’ (religion); ‘to refuse (to do something)’: such behavior rebounds on the refuser  
*batək* ‘to make a loop (*bənatək*) (for the straps of a basket)’: the loop holds itself in place (see the main text)  
*bawek* ‘to bend a springy object to shape’: the bender both pulls and is pulled by the object  
*cabəl* ‘to join things together by insertion’: the parts both join and are joined  
*caciiʔ* ‘to dance (standing still while moving arms and hands)’: the dancer moves her body while experiencing her centre of gravity unmoved  
*cadog* ‘to decorate one’s hair or earlobe with flowers or ornaments’: the inserted decoration holds itself in place  
*cadɔʔ* ‘to sit cross-legged’: the legs hold each other in place  
*cakəəd* ‘to be sticky, to stick’: sticky objects both ‘stick’ and ‘are stuck’  
*calēb* ‘to have specks in one’s eye’: an undergone experience  
*calɔʔ* ‘to pour away’: the pourer senses the vessel getting lighter, thereby experiencing the pouring  
*calɔh* ‘to fruit, set seed (rice, millet)’: the plant undergoes the fruiting  
*cananɥ* ‘to beat a gong’: the beating rebounds through the arm and ear; once started, the gong keeps itself sounding  
*carəəh* ‘to descend a slope’: one undergoes one’s own descending (see the main text)  
*carɔəd* ‘to be interlocked’: the parts both lock and are locked  
*catūd* ‘to be physically obstructed’ (by a crowd or object): one both obstructs and is obstructed  
*cayaanɥ* ‘to open one’s legs wide (as in childbirth or intercourse)’: the legs are held in place by their own friction against the flooring or ground  
*gabag* ‘sing’: the singer both sings and is sung through (see the main text)  
*gacuuh* ‘to collide (a body part) into something’: the body both knocks and is knocked  
*gadis* ‘to hurry around requesting things’: one undergoes the momentum of one’s own movement  
*galag* ‘to eat greedily; to act enthusiastically’: one is overcome by one’s own greed or enthusiasm  
*galek* ‘to scratch, tickle’: the nails are resisted by the flesh  
*galook* ‘to beat bamboo stampers (in song sessions)’: the bamboos bounce back into the hands  
*gapid* ‘to be trapped or tightly squeezed (as in a tree-fork)’: one undergoes the experience (semi-deponent):  
*gəpid* ‘to squeeze’, but no \**gəpid*)

- gatah* ‘to chip away (with a small knife)’: the action rebounds on the hand  
*hahəəj* ‘to burst out laughing’: overcome by one’s own uncontrolled laughter  
*halab* ‘to transport/be transported downriver’: the raft is transported by the river’s flow (see the main text)  
*haləəg* ‘to laugh (in company)’: because such laughing is catching  
*haruj* ‘to cease raining’: the rain appears to stop itself  
*hayaj* ‘to nearly do something, narrowly avoid doing something’: experienced rather than carried out deliberately  
*jajeeh* ‘to be abundant, in excess; to leave a surplus’: one finds oneself unable to finish (or reproducing too fast) through no positive action of one’s own  
*jaləb* ‘to set (sun)’: the sun appears to do so of its own accord  
*jaweeəg* ‘to stride’: one gets caught up in the rhythm of the movement  
*kabəəd* ‘to hug tightly (climbing a tree or riding piggy-back)’: climbers hold themselves in place  
*kabəəd* ‘to send off tightly hugging side-shoots (plant, as on a strangling fig)’: the plant forms itself  
*kaceed* ‘to not know’: ignorance is a state one finds oneself in, not through any actions of one’s own  
*kacēəd* ‘to boil over, flow over (water)’: the vessel undergoes the overflow; ‘to give fire to (from flame pr spark)’; the object undergoes the fire  
*kacəj* ‘to crush (a louse etc) with one’s fingernail’: one experiences the ‘crunch’ of the insect  
*kadeeb* ‘to climb something, clasping with hands and feet’: the climber both holds on and is held  
*kakəəh* ‘to beat with hand or stick, without cutting’: the action rebounds on the beater’s hand  
*kalag* ‘to eat greedily’: the eater is driven by his own greed  
*kaleek* ‘to prop’: it both supports and is supported (see the main text)  
*kaləən* ‘to feel pity’: the pitier both extends pity and experiences it  
*kalut* ‘to feel sorrow, mournful’: the sorrower both extends sorrow and experiences it  
*kamum* ‘chew’: the food resists the chewing.  
*kaŋəb* ~ *kaŋəb* ‘to blink’: both an action and something experienced  
*kaŋii* ‘to court, flirt’: a reciprocal action (see the main text)  
*kapej* ~ *kapeej* ‘to claw (with both paws, like a tiger or bear)’: the claws undergo resistance  
*kapək* ‘to kill by beating’: the action rebounds on the killer’s hand  
*kapus* ‘to overflow’: the vessel undergoes the overflow  
*karaaw* ‘to cry out’: an emotional response that happens to the crier  
*karəh* ‘to dry off, dry out’: both an action and a process  
*lalee* ‘to assimilate oneself; ethnically mixed (person, population)’; *Gəb lalee* ‘person with a Temiar mother and a Malay father’; *Kuy lalee* ‘type of Malay language used in some song-ritual lyrics’: assimilation is something one both does and undergoes  
*lapəej* ‘to be encrusted (eye after sleep)’: it happens to the waker  
*lapood* ‘to be soaking wet’: this is undergone  
*malar* ‘to follow around listlessly (like a child)’: the child cannot help itself  
*mameej* ‘to suffer a kind of headache (believed to be caused by dew)’; an undergone experience; also a middle-voice noun  
*naŋaah* ‘to feel sympathy or sorry for someone’: one cannot help doing so  
*pacəh* ‘to hurry on ahead, hurry up’: one gets caught up in the rhythm of the movement  
*pasuu* ‘to not like someone’: regarded as rebounding onto the initiator  
*payaa* ‘to not want to court someone, reject someone’: regarded as rebounding onto the initiator  
*ragəh* ‘to travel in a hurry’: one gets caught up in the rhythm of the movement  
*sabəə* ‘to catch (of a trap)’: the trap releases itself  
*sagoob* ‘to hold food in a bulging gullet (snake, lizard)’: the gullet and food hold each other in place  
*saloh* ‘to lodge, stay with someone’: regarded as a reciprocal activity  
*sahuur* ‘to faint; to die’: an undergone process  
*sapood* ‘to wrap’: because the package keeps itself wrapped  
*sarəəg* ‘to give way (soft ground)’: the ground and the walker undergo the subsidence  
*satəh* ‘to collapse, break off’, ‘release (of fall trap)’: the object undergoes its own collapse (see the main text)  
*tabag* ‘to hang something up for storage’: the item hangs itself  
*tadə* ‘to hold one’s hand out’: an action in expectation of receiving something; or an action in muscular dynamic equilibrium

*tajeeh* ‘to suspect (whether)’: suspicions come uninvited  
*tajəl* ‘to hang’ (trans): the item hangs itself  
*takil* ‘to throw something strongly; to bowl (as in baseball, cricket)’: the thrower experiences the follow-through momentum  
*takɔʔ* ‘to cut bamboo (not wood) into short sections’: the cutter’s experiences the rebound  
*takuuʔ* ‘to hurl (a spear); to travel quickly (airplane)’: it covers both active and middle meanings – to hurl and be hurled  
*talay* ‘to postpone (an activity)’: one is then under obligation to do it later  
*taləʔ* ‘(foot) to slip or slide on muddy ground’: the foot undergoes the slipping  
*taloot* ‘to be untrustworthy, deceitful’: this will rebound on the liar  
*talɔʔ* ‘to try food for taste, try out a blowpipe, test something’: one experiences the sensation of newness  
*taraʔ* ‘to stop’: one brings oneself to a stop  
*tarah* ‘to plane flat’: the hand undergoes the resistance of the wood  
*tareel* ‘to fall off (dried up umbilical cord)’: it undergoes its own action  
*tayes* ‘to point’: because the pointer experiences the arm’s weight  
*tayɔʔ* ‘to lift (a heavy object)’: because the lifter both lifts and undergoes the pull of the weight  
*yayah* ‘to be crazy, wild’: one undergoes the emotion

### Stative verbs, Adjectives

These are conditions or circumstances that the source appears to undergo.

*ʔalooy* ‘unripe, fresh’  
*babuh* (1) ‘rotten (wood)’  
*bahul* (1) ‘to be a big eater’: he is driven by his own greed  
*gaʔɔʔ~gaʔəʔ* ‘to be shrunken with hunger (belly)’  
*gadal~kadal* ‘hard pan (soil), firm (ground), caked (powder)’ (semi-deponent, ~ *kəldal*, but no \**kədal*)  
*gahēel* ‘out of breath, weary’  
*galah* ‘straight (hair)’  
*gatɔʔw* ‘thin (body, cane)’  
*gawɔʔk* ‘skinny’  
*gayuh* ‘rough or soft (stony or overgrown ground)’  
*haraaw* (1) ‘hairy, frizzled (hair, fur, spines, clothing)’  
*hayɔʔ* ‘light (weight)’  
*jaleeg* ‘restless, promiscuous (staying in a different place each night)’  
*jami* ‘swollen-cheeked’  
*kalɔʔ* ‘stupid, silly, ignorant, dumb’ (see the main text)  
*karək* ‘to feel very cold; rigor mortis’  
*karuk* ‘withered, dead (plant), weak, near death (person)’  
*katüüd* ‘to swell (a boil, a tree-canker)’  
*lapood* ‘soaking wet’  
*latah* ‘bald in front’  
*latəb* ‘bald on top (with hair at the sides)’  
*manah* ‘to be old, former’ (of things, not animals or humans)  
*papah* ‘to be the wrong way round, upside down’  
*papɔʔ* ‘crazy, mad, insane’  
*rahem* ‘round-faced’  
*ratih* ‘diligent (in work)’: carried away by one’s enthusiasm.  
*rawəj* ‘finished, used up’  
*sagɔʔl* ‘thirsty’  
*sakɔʔl* ‘white-haired, grey-haired’ (see the main text)  
*saraaʔ ma-* ‘to depend on someone, to be up to someone to decide’: dependency is mutual  
*sayeep* ‘dead, dried out (bamboo)’  
*takel* ‘giggly, laughing a lot’: giggling is uncontrollable

*talur* ‘to be slippery (ground)’  
*wawar* ‘to be out of true alignment’

## Nouns

### Mammals and reptiles

These are all legged creatures that move themselves through their own actions.

It has proved difficult to identify many of these animals more closely. Many of the animal names are avoidance names or taboo names, and are therefore not necessarily the common ones.

*ʔabir* ‘a squirrel’  
*ʔabɔɔŋ* ‘a land rat’  
*ʔacaam* ‘a large squirrel’  
*ʔaceel* ‘a large squirrel’  
*ʔacɛɛl* ‘a large squirrel’  
*ʔacoh* ‘common grey-bellied squirrel’  
*ʔadɛɛŋ* ‘a small squirrel’  
*ʔagəəj* ~ *ʔagəəc* ‘common Malay squirrel, plantain squirrel’  
*ʔahɔl* ‘flat-headed cat’  
*ʔajɔh* ‘rusa deer (*Cervus timorensis*)’  
*ʔajɔ̃r*, *ʔajoor* ‘common tree-shrew’  
*ʔakuub* ‘forest squirrel’  
*ʔalaam* ‘elephant’  
*ʔalaay* ~ *ʔalaaj* ‘elephant’  
*ʔaləəp* ‘elephant’  
*ʔalul* ‘elephant’  
*ʔamaah* ‘a small frog’  
*ʔamɔɔʔ* ‘goat-antelope’  
*ʔamug* ‘a white squirrel’  
*ʔaŋaŋŋ* ‘a small red-tailed squirrel’  
*ʔapəəp* ‘a large civet cat’  
*ʔapɔŋ* ‘pig-tailed macaque (*Macaca nemestrina*)’  
*ʔarɔŋ* ‘barking deer (*Muntiacus muntjak*)’  
*ʔataah* ‘horse-tailed squirrel’  
*ʔataŋ* ~ *ʔataŋŋ* ‘tiger’  
*ʔatũd* ‘elephant’  
*badɔɔt* ‘bearded pig’  
*bageɛt* ‘monitor lizard (*Varanus rudicollis*)’  
*bakaŋŋ* ‘small-toothed palm-civet’, ‘bear civet’  
*bapaak* ‘a white-bellied dark-backed squirrel’  
*bareɛw* ‘tapir’  
*basɔh* ‘white-eyed gibbon’  
*bateɛw* ‘a gibbon’  
*bawaaj* ‘pig-tailed macaque’  
*caʔɛ̃ɛʔ* ‘a small house-mouse’  
*cadɛɛʔ* ‘a small squirrel’  
*capɛɛg* ‘solitary large male (of monkeys)’  
*dalɔk* ‘a lizard’  
*dariʔ* ‘soft-shelled river turtle’  
*hadaaʔ* ‘short-tailed mongoose’  
*hagaab* ‘two-horned rhinoceros’  
*hareŋ* ‘monitor lizard (*Varanus salvator*)’



*hayum* ‘bamboo rat’  
*jajɔɔʔ* ‘common palm civet; tiger-civet’  
*jaŋŋŋ* ‘Muller’s rat’  
*kabɔŋ* ‘a frog sp.’  
*kabuc* ‘monitor lizard’ (*Varanus salvator*)  
*kabuk* ‘a green snake’  
*kajɛɛʔ* ‘flat-backed land tortoise’  
*kahuul* ‘tadpole’  
*karããj* ‘a river-turtle, small box-tortoise’  
*kasinj* ‘rusa deer (*Cervus timorensis*)’  
*kaweeb* ‘Malayan sun bear’  
*kayiiʔ* ‘flying lemur’  
*mamuug* ‘tiger’  
*pariʔ* ‘a monitor lizard’  
*payããd* ‘a monitor lizard’  
*sakɔl ~ sagɔɔl* ‘cream-colored giant squirrel’  
*taʔŋŋ* ‘wild pig’  
*tabɛɛg* ‘bullfrog’  
*tabəək ~ tabɔɔk* ‘dusky leaf-monkey’  
*tagaat* ‘a snake’  
*tahɔɔr* (1) ‘striped ground squirrel’  
*tajuulʔ* ‘snake’  
*talam* ‘elephant’  
*tapɛɛl* ‘pygmy flying squirrel’  
*tarook* ‘a lizard’  
*tawɔɔh* ‘white-handed gibbon’

### Birds and bats

These are all winged creatures that move themselves through their own actions.

*ʔahah* ‘great slatey woodpecker’  
*ʔakɛ̃b* ‘stonechat’  
*ʔakɛɛg* ‘red-headed trogon’  
*ʔakuul* ‘trogon (generic)’  
*ʔatɛ̃b* ‘grey-chinned minivet’  
*bajɔw* ‘cotton teal’  
*baryɛt* (a kind of bird?)  
*bayɔɔj* ‘a small night-calling bird’  
*caban* ‘black-crested yellow bulbul’  
*cabew* ‘chestnut-collared kingfisher’  
*cacɛr* ‘red-headed tailor bird’  
*danɔw* ‘lesser short-wing’  
*gagoob* ‘large scimitar babbler’  
*kakaʔ* kingfisher’  
*akeh* ‘hornbill (various kinds)’  
*kakuuh* ‘black hornbill’, ‘white-billed hornbill’  
*kasaaʔ* ‘grey-headed tree-babbler’  
*kasar* ‘black-headed bulbul’  
*kawɛ̃ɛ̃d* ‘fruit bat’, ‘flying fox’  
*lasar* ‘large bat, flying fox’  
*pahɔɔŋ* ‘black wood-partridge’  
*payɛh* ‘ferruginous babbler’  
*rabɔɔl* ‘weaver finch’

*sabaat* ‘hawk owl’  
*sagur* ‘a small bird’  
*sayah* (1) ‘red-headed tree-babbler’  
*sayɔl* ‘mountain nun-thrush’  
*sayɔɔl* ‘false vampire bat’  
*taʔəj* ‘hornbill sp.’, ‘wrinkled hornbill’  
*tadoor* ‘red jungle fowl’  
*tagək* ‘gold-whiskered barbet’  
*tahãār* ‘southern pied hornbill’  
*tahɛɛr* ‘grey and buff woodpecker’  
*talɔɔd* ‘Scop’s owl’  
*tapər* ‘white-eared fruit bat’  
*tatoh* ‘maroon woodpecker’, ‘greater yellow-naped woodpecker’  
*tayɛt* ‘grey wagtail’ (or ‘blue-throated bee-eater’?)

### Arthropods and snails

These are all creatures that move themselves through their own actions.

*bahul* (2) ‘a small dung beetle’  
*bayɔɔj* ‘a bird’  
*capood* ‘a very large black, or blue or striped fly; a large biting fly’  
*capɔɔg* ‘polydesmoid millipede’  
*cawaas* ~ *cawãās* ‘earwig’  
*galul* ‘a large black mosquito’  
*garɛɛd* ‘a sunset-chirping cricket’  
*garɔɔk* ‘a night-chirping cricket’  
*garuuc* ‘termite’  
*gasəʔ* ‘red stinging ant’  
*gateɛʔ* ‘mite’  
*gatũʔ* ‘snail(shell)’ (see the main text)  
*jaleed* ‘firefly’  
*jareb* ‘a small noisy outdoor cricket’  
*jawiis* ‘a seasonal cicada’  
*kabed* ‘ant’  
*kajek* ‘a black biting ant’  
*kasɔɔd* ‘fire-ant’  
*lawəh* ‘a large red swarming tree-ant’  
*maɲaay* ‘scorpion’  
*padaaw* ‘hornet’  
*saley* ‘jewel beetle’  
*tabəl* ‘black honey bee’  
*talay* ‘a large beetle’  
*talãy* ‘black millipede’  
*talək* ‘bumble bee, carpenter bee’  
*tanɔɲ* ‘dragonfly’  
*taroʔ* ‘house lizard’  
*tawãäg* ‘butterfly’  
*tawiik* ‘spider’  
*wawah* ‘a moth’  
*yayeed* ‘an evening-sounding cicada’

### Body parts

Body parts can be thought of as both moving and *being* moved.

- ʔayir* (2) ‘itchy scalp, from louse feces’  
*bakoh* ‘male genitals’ (sometimes ‘penis’, ‘testicles’)  
*balək* ‘animal’s beak or horn’  
*cawək* ‘head, skull’  
*jakaʔ ~ cakaʔ* ‘lower jaw’ (also: ‘overhang’ of a roof)  
*kabooj* ‘large vesicular swelling’  
*kadəg* ‘hollow of the knee-joint’  
*kalar* ‘throat, glottis, esophagus, adam’s apple’  
*kalej* ‘testes’  
*kalōðr* ‘snout of pig, bear, cat, etc.’  
*kaŋəŋ* ‘elbow’  
*kapəŋ* ‘cheek’  
*karəb* ‘sternum’  
*karəl* ‘knee’  
*katoŋ* ‘knee-cap’  
*kaway* ‘wing’  
*kayɛʔ* ‘little finger’  
*kayood* ‘fetus’: an Other within one’s Self  
*lageʔ* ‘caul’ ‘placenta’, ‘afterbirth’; also ‘new-born baby’: an Other within one’s Self (Benjamin 1994: 52)  
*pacɛʔ* ‘still-born fetus’  
*panik* ‘navel’  
*sabook* ‘windpipe’  
*sakəb* ‘corpse’  
*sapal* ‘upper arm’  
*tabooʔ* ‘thumb, big toe’  
*taŋən* ‘neck’  
*tapaag* ‘palm, sole; “hand” of bananas’  
*tapaar* ‘back of the hand, instep of the foot’

### Plants and plant parts

Plants, in shaping themselves, can be regarded as the undergoers of their own ‘actions’.

- ʔadɛg* ‘a tasty wild tuber’  
*ʔajɛl* ‘a cane’  
*ʔapoos* ‘wild ginger’  
*ʔasaad* ‘a large squash’  
*ʔaseh* ‘millet (taboo name)’  
*ʔawaat* ‘bamboo’  
*ʔawən* ‘bamboo’  
*babuh* (2) ‘a small toadstool’  
*badək* ‘jelutong tree’  
*bajaaw* ‘a fruiting vine’  
*bayas* ‘a palm with edible pith’  
*bayoor* ‘a large secondary-forest tree’  
*cadag* ‘a long-leafed plant used for plaiting ritual crowns’  
*cakoob* ‘bark (of tree)’  
*caŋeh* ‘pulasan fruit’  
*galook* ‘large rotten limb of a tree, about to fall’  
*gareed* ‘a tree (provides the barkcloth used by menstruating women)’

*gayεek* ‘a wild fruit’  
*ha’og* ‘a fruit-bearing forest tree’  
*hakɔr* ‘a wild seasonal fruit’  
*hayul* ‘generic term for staple food crops (non-leafy)’  
*haraaw* (2) ‘a forest flower’  
*kabup* ‘calyx, bud’  
*kacuuu* ‘a bitter leaf eaten with betel’  
*kalok* ~ *kaluk* ‘the bark eaten with betel chew’  
*karuu* ‘tree-resin illuminant’: it exudes itself; it keeps itself alight.  
*katak* ‘node of bamboo; vertebra’: it separates the other parts, but is held in place by them  
*kawɔk* ‘a medicinal herb’  
*labɔk* ‘rolled-up leaf bud’, ‘young shoot (of wild ginger, bamboo)’  
*lalɔ* ‘a tree’  
*manaar* ‘a cane’  
*pasəg* ‘a species of cane’  
*palɔ* ‘firewood’  
*raniik* ‘a wild seasonal fruit’  
*rarɔh* ‘a wild seasonal fruit’  
*sakɔb* ‘the spiny sheath on bamboo’  
*sakool* ‘a long leaved vegetable’  
*sajɔŋ* ‘a bush, the stem of which is pounded for medicine’  
*sajɔg* ‘a mountain-growing bush’  
*sareεg* ‘a wild yam’  
*sayah* (2) ‘the chaff after winnowing rice grains’  
*taba* ‘springy branch’  
*tamud* ‘a sweet-smelling herb’  
*tanεg* ‘a leafy vegetable with a spicy taste’

### Human and spiritual relations

Temiar social relations are thought of as dialectically generating each other.

*ʔayad* ‘gang, group (of friends)’  
*balu* ‘widow, widower’  
*cacɔ* ‘grandchild, younger sibling’s child’  
*Karey* ‘thunder, Thunder deity’: both cause and result  
*kawaa* ‘kinsperson (primarily consanguineal)’ (see the main text)  
*lanɔy* ‘ghost’; also used of ‘terrorist’ during the Malayan Emergency  
*janu* ‘chief’ (see the main text)  
*lage* ‘newborn baby’ (also used to label various perinatal misfortunes)  
*pacɔg* ‘a spirit-invasion disease, causing a prickly sensation’: both cause and result.  
*paley* ‘spirit-medium’s hut’ (strictly, the palm-leaf used in its construction): it undergoes possession by the tiger-spirit  
*papəəd* ‘infant’ (plural *pɛdpəəd* ~ *pɛnpəəd*, but no singular \**pəpəəd*)  
*sabat* ‘a convulsive attack (associated with childhood and childbearing)’ (see Benjamin 1994: 54–55)  
*sape* ‘pollution caused by recent death’  
*sarɔɔ* ‘ghost of deceased’  
*sayēēd* ‘young child’ (plural *sɛdyēēd* ~ *sɛnyēēd*, but no singular \**səyēēd*)  
*tajɔr* ‘watery fingertip manifestation of one’s spirit-guide’s presence’  
*tanig* ‘a gout-like disease (caused by interaction with a river-spirit)’: both cause and result

### Physical objects and processes

These are thought of as being simultaneously their own source and undergoer.

- ʔabat* ‘sarong’: it holds itself in place  
*ʔalɛɛg* ‘plaited storage bag’: it forms its own shape as it fills  
*ʔapəð* ‘back-basket’: it is self-hanging  
*ʔapil* ‘mat’: it lays itself down  
*ʔapɔk* ‘tobacco pouch’: it takes its own shape  
*bakɔɔ* ‘large noose trap, set on tree branch’: it ‘springs’ itself  
*baliik* ‘sky, position above’: it appears to be self-supporting  
*baraa* ‘main beam (of house)’: because it holds itself in place  
*bawur* ‘fishing rod’: because it both pulls and is pulled  
*cacuh* ‘woven roof thatch’: it holds itself together  
*cagɔl* ‘pond’: it is self-forming  
*gadan* ‘winnowing tray’: it throws the grains and is thrown by them when they fall  
*gagid* ‘middle’: it falls between two limits while holding them apart  
*gahool* ‘a depression in the ground, valley’: it is self-forming  
*galeed* ‘open-weave rotan basket (for carrying water-tubes)’: it is self-hanging and weighed down by its contents  
*galɛl* ‘glowing ember’: it keeps glowing of its own accord  
*gasek* ‘brochette, sliver of wood for cooking on’, ‘splinter of bamboo (for tattooing)’: it both penetrates and is surrounded by what it penetrates  
*jalaɔ* ‘thorn’: it both snags and gets snagged on people’s clothing (see the main text)  
*jalɔk* ‘tree-top’: a tree grows its own top (see the main text)  
*lamuun* ‘bent-over sapling-spring in various kinds of trap’: it is held in dynamic equilibrium  
*layɛg* ‘night’ (both a noun and a verb): night falls, happening spontaneously into existence of its own accord (see the main text)  
*raboon* ‘a large backbasket’: it holds itself in place  
*ranɔ* ‘back-basket’: it holds itself in place.  
*sagub* ‘cloud’: clouds appear to bring themselves into being (see the main text)  
*salɔg* ‘a kind of backbasket (for use when trapping)’: it holds itself in place  
*samɔg* ‘blowpipe wadding’: it both blows and is blown  
*sape* ‘section of a house affected by death-pollution’: it both pollutes and is polluted  
*saroog* ‘plaited rice-bin’: it takes its own shape when filled  
*takoon* ‘pool, pond’: it is self-forming  
*takɔ* ‘small container (for bait, wadding)’: it is self-hanging

### Unclassified: not obviously deponent or middle-voice

A sampling of residual verbs and nouns in *-a-* that cannot as yet be explained in the terms discussed in this paper

- ʔabaag* ‘split-bamboo internode used as eating dish’  
*ʔacag* ‘to plan to do something’  
*ʔaləəh* ‘to guess’  
*ʔamɛs* ‘small’  
*ʔapɛt* ‘short’  
*ʔarap* ‘a possession’  
*dalag* ‘to call someone’  
*haləəh* ‘to guess’  
*kamaay* ~ *kaməəy* ‘to store food (against future hunger)’  
*kanɛɛ* ‘we excl.’  
*laʔəs* ‘dirty’

*lagoh* 'dark leafy vegetable, like a *bayas* leaf'  
*lakɔb* 'to fold (trouser cuff, corner of a page)'  
*lalah* 'open terrain'  
*lawɛɛŋ* 'flower-odor' (ritual language)  
*lawəg* 'to mix and scoop food up with the fingers'  
*lawud* 'to cook a stew'  
*marɛɛk* 'fish-weir platform'  
*palɔ* 'log of firewood'  
*paniŋ* 'eventually, in future'  
*rages* 'to serve oneself bit by bit from the different dishes at a communal meal'  
*sarag* 'to transplant (tobacco, wet rice)'  
*tabɛɛh* 'to prepare something (for use)'  
*tahɔɔr* (2) 'hole (animal's, in a nest or the ground)'  
*takaah* 'slope'  
*tarɔg* 'spear'  
*tayɛɛd* 'to pick something up between fingers and thumb'

# Toward Proto Pearic: Problems and Historical Implications

Michel Ferlus

Independant Researcher (ret. CNRS, France)

## Abstract

The reconstruction of Proto-Pearic phonology allows us to consider the following hypothesis: The contrast /tense vs lax/ in Vietic, Katuic and Pearic was formed under the influence of Ancient Chinese along the trade route leading from North-Vietnam to the gulf of Thailand.

## 1. Introduction

This text will examine of the essential facts of Pearic: the problem of the two-way shifts of the voiceless initial plosives, the origin of the final glottal stop, and the origin of the two layers of registers (creaky and breathy). I will present a chronology of phonetic changes and an attempt an explanation of their propagation along ancient trade roads through Southeast Asia.<sup>30</sup>

## 2. Ethnonyms in Pearic subgroup

‘Chong’, also ‘Song/Xong/Kasong’, originate from \***kjɔːʔ** as the genuine autonym of Pearic populations. It is attested as *Tchouang* within “*Tchouang thieves*” (*zhuàng zéi* 撞賊) in the Tcheou Ta-kouan description of Cambodia, from the end of 13rd century (Pelliot 1902: 156; 1951: 70). In modern Khmer: *jañ* ជាន់ **cɔːŋ** 1. “barbare”, 2. “nom de tribu à demi sauvage” (Guesdon 1930); and “Chong (name of a tribe), barbaric, wild” (Jacob 1974) with a strong derogatory connotation. The meaning “barbaric” perhaps allows a connection with *zhuàng* 壯 “Zhuang people”, also *zhuàng/chuáng* 撞.

‘Por’ and ‘Pear’ originate from Skt. *varṇa*- “color, caste” according to two treatments in Khmer.

‘Por’: from Skt. *varṇa*- through the treatment \***bər**> \***bər**> \***bɔːr**. It is attested in Khmer dictionaries as “color, appearance”; *bar* ៧៧ **pɔːr**> **pɔː** (Guesdon 1930: 1203).

‘Pear’: from Skt. *varṇa*- through the literate treatment \***bar**> khmer *bār* ប៉ា / *ba(r)ṇ* ប៉ាណ **pʰar** > **pəə** (Ferlus, 1981). Its use by Khmer and French administration justifies the present name of the so-called Pearic subgroup.

‘Pol’: from Skt. *bala* “army, guard”. The Pol were at the disposal of the King for the guard of monuments and other places (Bregues 1905), they were composed of war prisoners, convicts and mountain peoples. Contrary to appearances, Pol is not an alternative of Por.

‘Samrê’, also ‘Somre/Somræ/Somray’ **səmre:/səmraj**: formed of **sre**: “field, cultivate” infixed by **-m-**; meaning “cultivators”.

---

<sup>30</sup> Abbreviations used in this paper: MK: Mon-Khmer; PMK: Proto Mon-Khmer; PP: Proto Pearic; EPP: Early Proto Pearic; LPP: Late Proto Pearic; OC: Old Chinese; MC: Middle Chinese. T: tense; L: lax.

---

Ferlus, Michel. 2011. “Toward Proto Pearic: Problems and Historical Implications.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAALA. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.38-51.  
Copyright vested in the author.

‘Saoch’, or the Chung of Kompong Som. This exonym is used only by Khmer speakers. In Chung, it would be \*so:c. Saoch has been connected to the Khmer *s’uc* ស្លុក *s’o:c* “scarlet fever” (Diffloth, *see* in Isara 2007: 60). Guesdon (1930: 1841) gives “pimply”. Derogatory ethnic names are rare, so why would the Khmer, who used ‘Pear’ to name other groups, have marginalized Chung people with the derogative ‘Saoch’? Saoch was also thaicized into ‘Kha Saut’ and ‘Ut’ ឃ្នុត in King Rama V’s travelogue to name the Chung Yul of Thailand (Isara 2007: 60, 26-27).

‘Khamen Boran’ of Pursat (Bastian 1868: 264-6) spoke a Pearic language although this designation usually names the Khmer Daeum (Martin, personal communication).

The exonym ‘Suay’ of Kompong Speu Pear is shared to name some Katuic population.

**3. Pearic Languages and dialects:**

There are four generations of linguistic data:

- Before 1900: short vocabularies.
- From 1900 to 1970: consistent vocabularies collected by attentive investigators but not linguists, French in the majority, knowing the Khmer language or working with Khmer assistants. These vocabularies give a satisfactory idea of the consonantism and, a little less, vocalism.
- The data of Marie Martin, which recognizes the existence of register features.
- The 1980s marks a transition with the arrival of professional linguists in Pearic studies. We have from now good data and scientific analyses. The register system is now clearly described (e.g. Huffman 1985; Theraphan 1984, 1991).

Data of the pre-linguistics period make it possible to date the shifts of finals \*-r, \*-l and \*-s. French investigators normally distinguish clearly \*-r and \*-l which are noted -l and -rr by Baradat. The final spelled -s is not consistent, it represents \*-s as well as \*-h. It can be explained by Khmer writing in which final -s is generally pronounced -h these days. The rules of Khmer spelling were transliterated into Latin spelling, and there is also, in certain authors, an improper use of the empty -r, as in khmer ‘*ankar* អង្ករ រៀក *ʔəŋka*: “husked rice” in which -r is not significant.

The Proto Pearic of Headley (1985) was elaborated with pre-linguistic data. Despite this handicap, this work remains a milestone in the Pearic studies and a good basis for further researches.

Pearic languages	Authors	Date
Chong	Crawfurd	1828
Samreh d’Angkor	Bastian	1868
Khamen boran	Bastian	1868
Xong	Bastian	1868
Xong	Garnier	1873
Khamen boran	Garnier	1873
Kouys - Porrh	Harmand	1878-79
Saoch	Pavie	1881-82
Sâmre (Siem Réap), Pors (Pursat)	Moura	1883
Sâauch	[Leclère]	1900
Porr (Pursat)	Bregues	1905
Saoch	Pannetier	?



Peâr de Kranhung	Pannetier	?
Souy	Pannetier	?
Saoch	Ménétrier	1926
Chawng (Ban Dan Champon, Tratt)	Isarangura	1935
Pear des Cardamomes (Phnom Krevanh)	Morizon	1936
Chông de Siam (=Chong of Trat)	Baradat	1941
Péâr de l'Ouest	Baradat	1941
Péâr de l'Est (Peam Prus)	Baradat	1941
Kompong Suoi de Kompong Speu	Baradat	1941
Péâr de Thom	Baradat	1941
Saoch (Kompong Som)	Ellul	1968-
Saoc (Komong Som)	Purtle ( <i>in</i> Headley)	1970-
Chong ໂວ (Chanthaburi)	Martin	1974
Chong hœp (Chanthaburi)	Martin	1974
Sômray (Kranhung, Batdambang)	Martin	1974
Sômree (Peam Prus, Pursat)	Martin	1974
Chong (Klong Phlu, Chanthaburi)	Surekha	1982
Samree (Borai, Trat)	Theraphan	1984
Chong (Klong Phlu, Chanthaburi)	Huffman	1985
Chong (Klong Phlu, Chanthaburi)	Sirikarn	1987
Chong (Klong Phlu, Chanthaburi)	Saifon	1991
Chong (Klong Phlu, Chanthaburi)	Theraphan	1991
Chong (Klong Phlu, Chanthaburi)	Edmonson	1996
Kasong (Khlông Saeng, Trat)	Kunwadee	1996
Chong (Klong Phlu, Chanthaburi)	Siripen	2001
Samre (Trat)	Pornsawan	2001
Chong (Klong Phlu, Chanthaburi)	Isara	2002
Kasong (Khlông Saeng, Trat)	Noppawan	2003
Kasong (Khlông Saeng, Trat)	Sunee	2003
Chong (Klong Phlu, Chanthaburi)	Isara	2004
Chung Yul (Kampong Som)	Isara	2007
Chung Yuy (Thung Na, Kanchanaburi)	Isara	2007
Chung (Kanchanaburi & Kampong Som)	Isara	2009

Table 1: Chronology of Pearic data

#### 4. Register System of Pearic

Acoustic analyses of Chong by Theraphan (1991) once and for all demonstrated the existence of a four registers system in Pearic that preceding works foresaw to some extent (Martin 1974a; Surekha 1982; Huffman 1985).

The system combines the modal feature /clear/ with the marked features /breathy/ and /creaky/ to form four syllabic combinations.

R1	[CVC] <i>Clear modal</i>	R2	[CVʔC] <i>clear-creaky</i>
R3	[CVC̤] <i>breathy</i>	R4	[CVʔC̤] <i>breathy-creaky</i>

Table 2: Pearic registers

Examples (Siripen 2001):

- R1 **ka:k** “armpit”, **tʰa:k** “span”, **cʰɔ:k** “pound”, **kla:ŋ** “branch”
- R2 **kəta:ʔk** “tongue”, **kəna:ʔk** “crow”, **kəsɔ:ʔŋ** “ladder”, **kʰla:ʔŋ** “owl”
- R3 **tʰa:k** “water”, **kəla:ŋ** “ear”, **cʰɔ:k** “shine”, **kəpʰɔ:k** “hole”
- R4 **kəla:ʔk** “swallow”, **kəla:ʔŋ** “sand”, **cʰɔ:ʔŋ** “Chong”, **ŋa:ʔj** “far”

Syllables ending in -Ø, -ʔ and -h are only attested with R1 et R3.

**5. Reconstruction of initial plosives:**

As in many languages of the Southeast Asia, Pearic dialects underwent a phonetic restructuring of the initial plosives:

Low series (or series 2): devoicing of voiced initial plosives, generally **\*/b d j g/ > /p t c k/**; in Chong Klong Phlu **/pʰ tʰ cʰ kʰ/**. These shifts are associated with vowel raising and breathiness.

High series (or series 1): Comparison shows two types of treatment of voiceless initial plosives. In the first type, the most simple initials remain unchanged, **\*/p t c k/= /p t c k/** as in Khmer and in Thai. In the second type, the most characterized, it is **\*/p t c k/> /pʰ tʰ cʰ kʰ/**. A vowel lowering can be associated with this. An exception: Kompong Thom Pear, surrounded by Khmer speakers, only attests the simple shift.

According to the principle of the regularity of sound change, this two-ways treatment would be an anomaly. Headley (1985) had adopted a provisional solution by reconstructing two series of plosives, **\*/P T C K/** for the simple type, and **\*/p t c k/** for the characterized type. The state of knowledge at the time did not permit a solution to the problem.

Proto Pearic		Pearic dialects	Kg Thom	
Ferlus	Headley 1985			
<b>*p</b>	p <sup>1</sup>	*p	p <sup>h</sup>	p
	p <sup>2</sup>	*P	p	p
<b>*t</b>	t <sup>1</sup>	*t	t <sup>h</sup>	t
	t <sup>2</sup>	*T	t	t
<b>*c</b>	c <sup>1</sup>	*c	c <sup>h</sup>	c
	c <sup>2</sup>	*C	c	c
<b>*k</b>	k <sup>1</sup>	*k	k <sup>h</sup>	k
	k <sup>2</sup>	*K	k	k

Table 3: Proto-Pearic plosives according to Headley (1985)

In my notation, superscripts <sup>1</sup> and <sup>2</sup> are only used to differentiate the two types of correspondences.

Ferlus	Headley	examples
*p (p <sup>1</sup> )	*p	p <sup>h</sup> o:j <sup>R1</sup> [p <sup>h</sup> o:j] “after”, p <sup>h</sup> o:n <sup>R2</sup> [p <sup>h</sup> o:ʔn] “four”
*p (p <sup>2</sup> )	*P	pa:ŋ <sup>R1</sup> [pa:ŋ] “flower”, kəpa:t <sup>R2</sup> [kəpa:ʔt] “cotton”
*t (t <sup>1</sup> )	*t	t <sup>h</sup> a:n <sup>R1</sup> [t <sup>h</sup> a:n] “weave”, t <sup>h</sup> a:m <sup>R2</sup> [t <sup>h</sup> a:ʔm] “crab”
*t (t <sup>2</sup> )	*T	to:t <sup>R1</sup> [to:t] “head”, kəta:k <sup>R2</sup> [kəta:ʔk] “tongue”
*c (c <sup>1</sup> )	*c	ch <sup>o</sup> :j <sup>R1</sup> [ch <sup>o</sup> :j] “to plant”, ch <sup>i</sup> :n <sup>R2</sup> [ch <sup>i</sup> :ʔn] “ripe”
*c (c <sup>2</sup> )	*C	co:k <sup>R1</sup> [co:k] “run”, ca:t <sup>R2</sup> [ca:ʔt] “knife”
*k (k <sup>1</sup> )	*k	k <sup>h</sup> e:n <sup>R1</sup> [k <sup>h</sup> e:n] “son”, k <sup>h</sup> o:n <sup>R2</sup> [k <sup>h</sup> o:ʔn] “mouse”
*k (k <sup>2</sup> )	*K	ka:ŋ <sup>R1</sup> [ka:ŋ] “moon”, kə:t <sup>R2</sup> [kə:ʔt] “low”

The type \*/p t c k/ > /p<sup>h</sup> t<sup>h</sup> c<sup>h</sup> k<sup>h</sup>/, named “mutation germanique” by Haudricourt (1965), is sporadically attested in the Austroasiatic area. Apart from Pearic, it exists in Khasi and in Phay/Tin. Its scarcity compared to the simple type could allow one to consider it the standard shift in Pearic. As for the type \*/p t c k/ = /p t c k/, very largely spread, it is attested in particular in Khmer and Thai, languages in contact with Pearic. I think that this when this is the case in Pearic, it is due to the influence of Khmer. It is known that in Khmer and Thai the phenomena of the restructuring of initials occurred in second half of the seventeenth century. At this time the Pearic dialects would have formed a continuous territorial unit, except for the isolated Kompong Thom Pear in which the marked process did not occur.

It is thus necessary to re-examine the historical phonetics of Pearic by taking account the influence of the Khmer language.

## 6. Origin of final -ʔ in Pearic

Generally, final glottal stop -ʔ in Pearic does not correspond regularly to PMK -ʔ which is preserved in Khmu, Waic, Mon (except \*-iʔ > today *oa* -*ay*) and partially in Vietic.

	Chong	register	Khmu	*Waic	Mon
hand	ti:	R1	tiʔ	teʔ	toa <i>tay</i>
louse	ch <sup>i</sup> :	R1	seʔ	siʔ	coa <i>cay</i>
pestle	kəhi:	R1	cn <sup>d</sup> reʔ	ŋriʔ	
fruit	p <sup>h</sup> li:	R1	pleʔ	pliʔ	
eight	kəti:	R1		snteʔ	
rattan	se:	R1			riʔ <i>ri</i>
rice field	(wa:j3) se:	R1	<sup>h</sup> reʔ		
grand father	ta:	R1	taʔ	taʔ	
monkey	wa:	R1	<sup>h</sup> vaʔ	hwaʔ	
eat	ch <sup>a</sup> :	R1		saʔ	ceʔ <i>ca</i>
grass	kətu:	R1		tiʔ	
grandson	co:	R1	ɔʔ	siʔ	
dog	ch <sup>o</sup> :	R1	səʔ	səʔ	
sesame	kəŋo:	R3		ŋaʔ	
leg	p <sup>h</sup> lu:	R3	bluʔ		
forest, woods	p <sup>h</sup> ri:	R3	briʔ	breʔ	
banyan tree	(ne:m2) ch <sup>i</sup> :	R3	ɟriʔ		səa <i>jay</i>
day	k <sup>h</sup> əŋiʔ	R1 (<*R2)	sŋiʔ	sŋeʔ	təŋoa <i>tŋay</i>
ground, soil	t <sup>h</sup> eʔ	R1 (<*R2)	pteʔ	kteʔ	təeʔ <i>ti</i>
leaf	laʔ	R1 (<*R2)	<sup>h</sup> laʔ	hlaʔ	hlaʔ <i>sla</i>
stone, rock	k <sup>h</sup> əmoʔ	R1 (<*R2)		smoʔ	məʔ <i>tma</i>
dream	p <sup>h</sup> oʔ	R1 (<*R2)	mpoʔ	rmoʔ	kəpəʔ <i>lpa</i>
shell	k <sup>h</sup> ləʔ	R1 (<*R2)	kləʔ	nləʔ	kənaoʔ <i>knu</i>

<MM <i>kinlo'</i>				
paddy	(ha:j1) kɔʔ	R1 (<*R2)	ɲŋkoʔ	ɲŋkoʔ
rain (n.)	kəmaʔ	R3 (<*R4)	kmaʔ	mɔʔ <i>kma</i>
thorn	kəlaʔ	R3 (<*R4)	cɾlaʔ	həlɛʔ <i>khala</i>

Table 4: Distribution of final glottal stop

In the older transcriptions, the final -ʔ of Chong corresponds to a glottal constriction on the vowel while giving an impression of dissyllabism. Let us compare some transcriptions of Suôy Kompong Speu (Baradat 1941) with those of Chong Klong Phlu (Siripen 2001).

	Suôy (Baradat)	Chong (Siripen)
Ground, soil	<i>thé-é</i>	t <sup>h</sup> eʔ
Leaf	<i>sla-a</i>	laʔ
Stone, rock	<i>thmaû-aû</i>	k <sup>h</sup> əmoʔ
Acid, sour	<i>chô-ô</i>	c <sup>h</sup> oʔ
Skin	<i>trâlô-aû</i>	k <sup>h</sup> aloʔ

However, there are exceptions. It should not be forgotten that the former authors, in spite of their merits, were not professional linguists. Historically, the final -ʔ in Pearic comes from a glottal constriction, this is the reason why it does not correspond to PMK final glottal stop. Current -ʔ is only the result of a recent secondary development. By writing the glottal constriction as -ʔ-, one can propose the syllabic evolution CVʔV>CVʔ.

At this point, two important facts are highlighted: (i) Pearic -ʔ does not originate from PMK -ʔ, (ii) Pearic -ʔ (CVʔ) must be reinterpreted as a syllabic glottalization (creakiness) -ʔ- (CVʔV). As a consequence, final -ʔ must be removed from Early Proto Pearic. The system of final plosives thus had only four units \*/p t c k/.

In Chong (Siripen 2001), open syllables -Ø and syllables in -ʔ (only short vowels) are attested with only registers R1/R3. On the basis of preceding remarks, it is possible to reorganize the two sub systems (syllables -Ø and -ʔ) in a sole system while transferring syllables ending in -ʔ from R1/R3 into R2/R4 on the model of other final vowels (table 5). As a consequence, Pearic languages must be reconstructed without final -ʔ as in Khmer (Ferlus 1992), in Katuic and in Bahnaric (Sidwell 1998, 2005).

Syllables -Ø and -ʔ With two registers	R1 [CV]	R1 [CVʔ]
	R3 [Cʏ]	R3 [Cʏʔ]
↓		↓
Reorganization: syll. -Ø with four registers	R1 [CV]	R2 [CVʔ]<[CVʔV]
	R3 [Cʏ]	R4 [Cʏʔ]<[CʏʔV]

Table 5: Reorganization of Pearic registers by syllable types

Following this reorganization, the distribution of new open syllables in the four registers is of the same type as that of the closed syllables (table 6).

Syllables -Ø: four registers	R1 [CV]	R2 [CVʔV]
	R3 [Cʏ]	R4 [CʏʔV]

Closed syllables: four registres	R1	[CvC]	R2	[Cv <sup>2</sup> C]
	R3	[CṿC]	R4	[Cṿ <sup>2</sup> C]

Table 6: Distribution of Pearic registers by syllable types

The most recent Proto Pearic (Late Proto Pearic), the stage preceding the devoicing of voiced plosives initials, must be reconstructed with the contrast /creaky vs modal/. It will be explained later (see §.8), on the one hand, that this contrast does not come from the PMK, and on the other hand, that it can be explained by an ancient syllabic contrast /tense vs lax/ (henceforth /T vs L/), possibly due to the influence of Middle Chinese. At this stage, my hypothesis is based primarily on the geographical and temporal coincidence that three Mon-Khmer groups lying along a known Chinese trade route developed similar /T vs L/ contrasts.

### 7. Stages of the evolution from PMK to Pearic

It is now possible to present a chart showing the principal steps of the phonetic shifts of the Pearic branch from the PMK stage up to modern times.

Stages	States <i>Shifts</i>
(1) Proto Pearic Mon-Khmer	Syllables with final *-ʔ.
	<i>Loss of final *-ʔ &gt; formation of syllables -∅</i>
(2) Early Proto Pearic	No syllable with final -ʔ.
	<i>Formation of /tense vs lax/ (influence of Middle Chinese) which evolve to /creaky vs modal/. Note: Loss of final *-ʔ was also propagated in Bahnaric, Katuic and Khmer. In Monic, it affected only the rime *-iʔ (oa -ay in modern Mon).</i>
(3) Late Proto Pearic	No syllable with final -ʔ. Marked register /creaky/.
	<i>Devoicing of plosives initials: */b d ɟ g/ &gt; /p t c k/ (in general). Reinforcement of unvoiced plosives initials: */p t c k/ &gt; /p<sup>h</sup> t<sup>h</sup> c<sup>h</sup> k<sup>h</sup>/ (regular) or preservation as /p t c k/ (influence of Khmer). Formation of register contrast /clear vs breathy/.</i>
(4) Pearic at registral stage	No final -ʔ. Four registers system: /clear/, /creaky/, /breathy/ and /breathy-creaky/.
	<i>Recreation of final -ʔ in tense syllables -∅.</i>
(5) Current Pearic languages	New final -ʔ. Four registers system: /clear/, /creaky/, /breathy/ and /breathy-creaky/. Diversification by final consonants.

Table 7: from PMK to Pearic languages

### 8. Origin of creakiness

The creakiness is not very widespread in MK languages, it exists only in Vietic, in a part of Katuic and in Pearic (Diffloth 1989). I already showed how a syllabic contrast /T vs L/ of Old Chinese had been propagated in Vietic while modifying its phonation (Ferlus 2004). On this model, one can suppose that the

contrast /creaky vs modal/ in Pearic can be also explained by this old syllabic contrast /T vs L/ of OC. However, it is difficult to prove this categorically.

**8.1 Formation of contrast /T vs L/ in Ancient Chinese.**

The syllabic type in OC was (Cv)CV(C), a part of the vocabulary was made up of monosyllables CV(C) the other part of sesqui-syllables CvCV(C). The coalescence of initials in sesqui-syllables developed a tenseness /T/, while monosyllables became lax /L/. Thus syllabic contrast CvCV(C) vs CV(C) was coupled with contrast /T vs L/. The evolution was continued by the monosyllabization and the formation of a syllabic contrast CV(C)/T vs CV(C)/L in MC, associated with modifications of vocalic aperture, vowel lowering in T-syllables and vowel raising in L-syllables (Ferlus 2009).

Old Chinese	Middle Chinese	<i>transferred to Vietic, Katuic and Pearic</i>
CvCV(C) (tenseness)	CV(C)/T (v. lowering)	<b>T</b> (ense)
CV(C) (laxness)	CV(C)/L (v. raising)	<b>L</b> (ax)

Table 8: Chinese Register Development

It will consider here only the contrast /T vs L/ which I will argue was propagated into Vietic, Katuic and Pearic.

**8.2 Transfer and evolution of contrast /T vs L/ in Vietic (Ferlus 2004)**

	Early Proto Vietic syll.	Late Proto Vietic finals			
		*/p t c k/	*/s h/	*/ʔ/	*/m n p ŋ r l w j/
<b>T</b>	CvCVC	not affected	not affected	*ʔ > Ø (1)	glottalization mʔ nʔ ... (2)
<b>L</b>	CVC	not affected	not affected	ʔ unchanged (2)	not affected (1)
		sắc-nặng	hỏi-ngã	<sup>1</sup> ngang-huyền <sup>2</sup> sắc-nặng	<sup>2</sup> sắc-nặng <sup>1</sup> ngang-huyền

Table 9: Vietic Register Development

The most outstanding fact of Vietic is the creation of open syllables in Late PVM. This fact created conditions for the formation of the three fundamental tones represented by *ngang-huyền*, *sắc-nặng* and *hỏi-ngã* in Vietnamese. Of note: the feature /T/ is strong enough to cause the loss of final -ʔ, but not enough to affect the voiceless final plosives. The voiced finals were glottalized and are represented by *sắc-nặng* tones in Vietnamese.

### 8.3 Transfer and evolution of the contrast /T vs L/ in Katuic (Diffloth 1989)

	Early Proto Katuic syll.	Late Proto Katuic finals			
		*/p t c k/	*/m n ɲ ŋ/	*/r l s h w j/	*Ø
T	CvCVC	mʔ nʔ jʔ ʔ	mʔ nʔ ɲʔ ŋʔ	rʔ lʔ sʔ hʔ wʔ jʔ	vʔ
L	CVC	p t c k	m n ɲ ŋ	r l s h w j	Ø

Table 10: Katuic Register Development

Contrast /T vs L/ has affected only some dialects (Katang, Talan, Yir/Ong) in the East of Katuic. Other Katuic languages (Suoy, Kuy/Kuoy, Sô/Bru, ...) were not affected. To simplify, I did not take account of the vocalic length in the development of /T/. The effect of tenseness is more important in Katuic than in Vietic.

### 8.4 Transfer and evolution of the contrast /T vs L/ in Pearic

	Early Proto Pearic syll.	Late Proto Pearic finals				
		*/p t c k/	*/m n ɲ ŋ/	*/r l s w j/	*/h/	*Ø
T	CvCVC	pʔ tʔ cʔ kʔ	mʔ nʔ ɲʔ ŋʔ	rʔ lʔ sʔ wʔ jʔ	h	vʔ>vʔ
L	CVC	p t c k	m n ɲ ŋ	r l s w j	h	Ø

Table 11: Pearic Register Development

One can observe the re-creation of syllables -ʔ. Except for the syllables -h, all the others were glottalized under the effect of the tenseness. The Pearic languages are those where the effects of tenseness are generalized the most, but where the finals are the least corrupted. Exception, in Chung Yul, final plosives were nasalized at creaky register: /pʔ tʔ cʔ kʔ/> /mʔ nʔ ɲʔ ŋʔ/ while merging with nasals of the same register (Isara 2009). This phonetic change is recent because it also affects the final /sʔ>tʔ/ (>/nʔ/).

## 9. The Han trail and its Linguistic Implications

During the 3rd-8th centuries CE, Chinese texts reveal the existence of dependencies of the Chinese Empire, located between the Middle Mekong and the north of present Cambodia.

The Records of the Three Kingdoms (*sān guó zhì* 三國志) record that to the 3rd century, a state named T'ang-ming (*táng míng* 堂明), located north of present Cambodia, sent embassies to the emperor of China (Pelliot 1903: 251). This practice indicates a nominal authority of China over this area.

In his great treatise of geography (*shí dào zhì* 十道志), the author Kia Tan (*Jia dān* 賈耽), 8th century, details the land route from the Chinese possession of Kiao-tche (*jiāo zhǐ* 交趾; Sino-Vietnamese: *Giao chí*), the present north of Vietnam, and leading to the dependency of Wen-tan (*wén dān* 文單) (Pelliot 1904: 210). It is thought that Wen-Tan was just one of the names of Tchen-la (*zhēn là* 真臘), in other words Ancient Cambodia which extended farther north than present day Cambodia.

What would be the reason of the existence of these dependencies in an outlying region from China and linked to Kiao-Tche by roads cut through geographical obstacles? It is clear that the roads described in the texts were only those controlled by the Chinese, of the great transcontinental trade route connecting southernmost China to the gulf of Thailand, and becoming a sea route toward India by a portage through the Isthmus of Kra. This land route, a priori difficult, was essential to avoid the Cham who controlled the sea route from China to India by the strait of Malacca. Tatsuo Hoshino (2002) remarkably studied the trans-

Mekong route to the Wen-Tan, despite various difficulties of locating the places quoted by the Chinese sources. We will call the part of the transcontinental trade route located between Kiao-Tche and the gulf of Thailand the “Han Trail”.

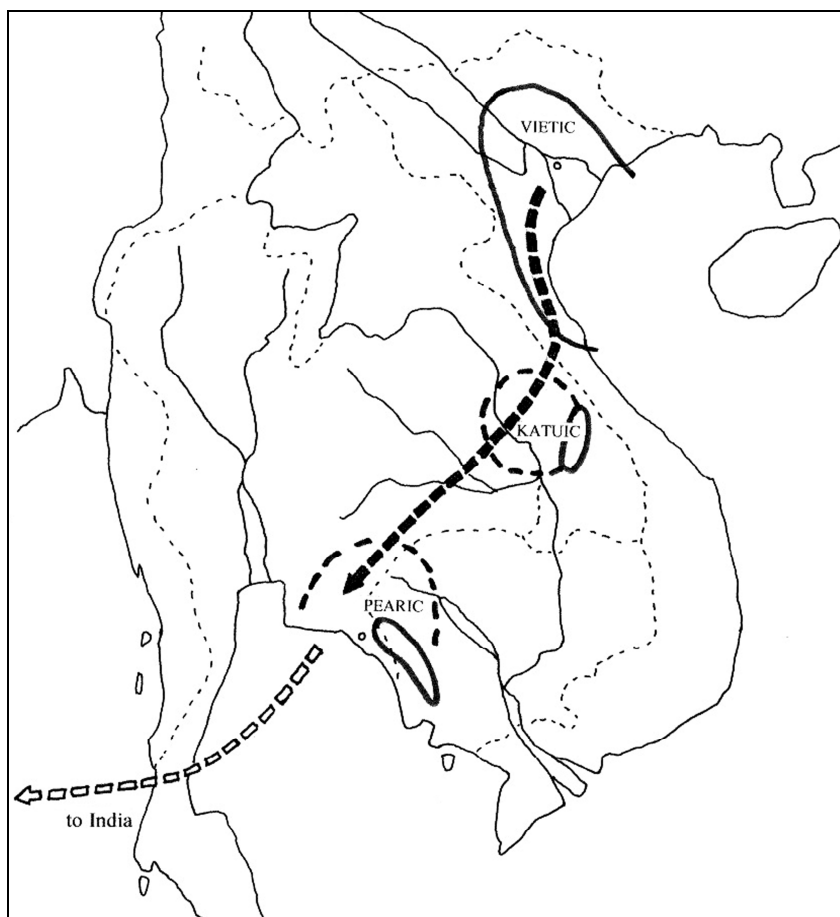


Figure 1: A tentative map of the trans-peninsular trade route or *Han trail*, leading from Kiao-Tche (ancient Vietnam) to the gulf of Thailand, and beyond to India.

What is the relation of the Pearic populations, now scattered in Cardamomes, with these trade route? According to the ethnologist Marie Martin, the oral traditions of Samre mention a Chong kingdom before the arrival of the Khmers. In addition, the Khmers of Chanthaburi had the memory of an old Chong capital located on present Phnom Sebap (Martin 1997: 70). These places, located between the Great Lake and the Gulf of Thailand, are the possible homeland of Pearic and a natural point of arrival of a trade route coming from central Indochina.

I have argued above that the contrast /T vs L/ of Ancient Chinese (OC>MC) had been transferred into Vietic, Katuic (partially) and Pearic. It can be objected that only a small part of Katuic attests this sound change, but I think that it is due to the expansion of the West Katuic, which was not influenced by Chinese. While at the same time the Tchen-la of Land<sup>31</sup> pushed back the East Katuic towards the margins (into the hills east of the Mekong). Let us recall that the basic population of Land Tchen-la before the unification of Cambodia was mainly the Bru (*pó lòu* 婆鑊) ethnic group as documented by Chinese authors (Ferlus 2005). However, these three linguistic groups are precisely located at the both ends and in the middle of this trade

<sup>31</sup> The Chinese sources distinguished Tchen-la of Land and Tchen-la of Water, apparently referring to Cambodia in-land from Cambodia around the Great Lake and lower Mekong.



route which during centuries was covered by Chinese travellers and traders. This coincidence between a linguistic fact, formation of /T vs L/ in Vietic, Katuic and Pearic, and a trade route where the Chinese carrying this contrast circulated, is sufficiently remarkable to deduce from it that this fact is not due randomly.

### References

- ASEMI: Asie du Sud-Est et Monde Insulindien  
 BEFEO: Bulletin de l'École Française d'Extrême-Orient  
 BSEI: Bulletin de la Société des Etudes Indochinoises  
 ILCRD: Institute of Languages and Culture for Rural Development  
 JSS: The Journal of the Siam Society  
 MKS: Mon-Khmer Studies
- Antelme, Michel Rethy & Hélène Suppya Bru-Nut. 2001. *Dictionnaire Français-Khmer*. Paris, L'Asiathèque.
- Baradat, R. 1941. *Les dialectes des tribus sâmrê*. Manuscrit de l'École Française d'Extrême-Orient, Paris. 267p.
- Baradat, R. 1941. Les sâmrê ou Peâr, population primitive de l'Ouest du Cambodge. *BEFEO* 41: 1-150.
- Bastian, Adolf. 1868. *Reise durch Kambojja nach Cochinchina*, tome 3. Herman Gostenoble, Jena. [samreh 89-90; khamen boran 264-265; xong 295-298]
- Blagden, Charles Otto. 1894. Early Indo-Chinese influence in the Malay Peninsula, as illustrated by some of the dialects of the aboriginal tribes. *Journal of the Straits Branch of the Royal Asiatic Society* 27: 21-56.
- Bregues, Docteur Jean. 1905. Notes sur les populations de la région des montagnes des Cardamomes. *JSS* 2(1): 19-47.
- Cabaton, Antoine. 1905. Dix dialectes indochinois recueillis par Prosper Odend'hal. Etude linguistique par Antoine Cabaton. *Journal Asiatique*, 10<sup>e</sup> série, 5: 265-344.
- Crawford, John. 1828 [2000]. *Journal of an Embassy from the Governor-General of India to the Courts of Siam and Cochin China*. London: Henri Colburn. [reprint 2000, New Delhi]
- Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *MKS* 15: 139-154.
- Edmondson, Jerold A. 1996. Voice qualities and inverse filtering in Chong. *MKS* 26: 107-116.
- Ellul, Jacques. 1968. Enregistrement magnétique de vocabulaire saoch, région de Veal Renh, Cambodge.
- Ferlus, Michel. 1981. A propos d'un allongement vocalique devant la finale r en khmer ancien. *ASEMI* 12(1-2): 101-109.
- Ferlus, Michel. 1992. Essai de phonétique historique du khmer (Du milieu du premier millénaire de notre ère à l'époque actuelle)", *MKS* 21: 57-89.
- Ferlus, Michel. 2004. The origin of tones in Viet-Muong. In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*. Edited by Somsonge Burusphat. Arizona State University: 297-313.
- Ferlus, Michel. 2005. L'intérêt linguistique des transcriptions chinoises concernant le Cambodge ancien (Fou-nan et Tchen-la). *Dix-neuvièmes Journées de Linguistique de l'Asie Orientale*, 30 juin - 1<sup>er</sup> juillet, 2005. Paris: CRLAO (EHESS-CNRS).
- Ferlus, Michel. 2009. What were the four Divisions of Middle Chinese ? *Diachronica* 26(2): 184-213.
- Filippi, Jean-Michel. 2000. *Lexique et analyse phonétique du Saoch*. Manuscrit.
- Filippi, Jean-Michel. 2008. *Recherches préliminaires sur les langues des minorités du Cambodge*. Publié par l'UNESCO et l'IDSSC, Phnom Penh, Editions Funan.
- Garnier, Francis. 1873 [1976]. *Voyage d'exploration en Indo-Chine effectué pendant les années 1866, 1867, et 1868...* Paris, Hachette. Vol. 2: 498-506.
- Gregerson, Kenneth J. 1976. Tongue-root and register in Mon-Khmer. *Austroasiatic Studies Part 1*, Oceanic Special Publication N° 13. Edited by Philip N. Jenner, Laurence C. Thompson & Stanley Starosta. The University Press of Hawaii: 323-369.
- Gubler, J. s. d. Liste de mots pears recueillis dans les Phnom Krevanh. Manuscrit.
- Guesdon, Père Joseph. 1930. *Dictionnaire cambodgien-français*, 2 vols. Paris: Plon.
- Harmand, J. 1878-79. Notes de voyage en Indo-Chine: les Kouys – Ponthey-Kakèk. *Annales d'Extrême-Orient* 1: 322-339.

- Haudricourt, André-Georges. 1965. Les mutations consonantiques des occlusives initiales en Mon-Khmer. *Bulletin de la Société de Linguistique de Paris* 60(1): 160-172.
- Headley, Robert K. 1977. A Pearic vocabulary. *MKS* 6: 69-149.
- Headley, Robert K. 1978. An English-Pearic vocabulary. *MKS* 7: 61-94.
- Headley, Robert K. 1985. Proto-Pearic and the classification of Pearic. In *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*. Edited by Surya Ratanakul, David Thomas, & Suwilai Premrirat. Bangkok, Mahidol University: 428-478.
- Henderson, Eugénie J.A. 1952. The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14(2): 149-174.
- Hoshino, Tatsuo. 2002. Wen Dan and its Neighbours: The Central Mekong Valley in the Seventh and Eighth Centuries. In *Breaking new Grounds in Lao History, Essays on the Seventh to Twentieth Centuries*. Edited by Mayoury Ngaosrivathana and Kennon Breazeale. Chiang Mai, Silkworms Books.
- Huffman, Franklin E. 1976. The register problems in fifteen Mon-Khmer languages. *Austroasiatic Studies Part 1*, Oceanic Special Publication N° 13. Edited by Philip N. Jenner, Laurence C. Thompson & Stanley Starosta. The University Press of Hawaii: 575-589.
- Huffman, Franklin E. 1985. The Phonology of Chong, A Mon-Khmer Language of Thailand. In *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*. Edited by Surya Ratanakul, David Thomas, & Suwilai Premrirat. Bangkok, Mahidol University: 355-388.
- Isara, Choosri. 2002. Dialects of Chong. *MKS* 32: 55-70.
- Isara, Choosri. 2004. Chaak and other Chong place names: linguistics remains of a Mon-Khmer people in Eastern Thailand. *MKS* 34: 67-77.
- Isara, Choosri. 2007. Investigating Contact-induced Language Change: Cases of Chung (Saoch) in Thailand and Cambodia. Thesis, Mahidol University.
- Isara, Choosri. 2009. Chung (Saoch) of Thailand and Cambodia: phonological and lexical comparisons. *MKS* 38: 69-85.
- Isarangura, Nai Noe. 1935. Vocabulary of Chawng words collected in Krat Province. *JSS* 28: 173-186.
- Jacob, Judith M. 1974. *A Concise Cambodian-English Dictionary*. London, Oxford University Press.
- Kanika Kenikanon & al. 1979. *Chong*. Bangkok, Bannakij. [in Thai]
- Kunwadee Phaetphithak. 1996. *Phaasaa Chong Muu Baan Khlong Saeng, Tambon Dan Chumphon, Amphoe Borai, Changwat Trat*. MA Thesis. Silpakorn University. [A description of the Chong language in Khlong Saeng village, Borrai district, Trat province]
- [Leclère, Adhémar. ~1900]. *Vocabulaire français-sâauch-cambodgien*. Manuscrit de l'Ecole Française d'Extrême-Orient, Paris. 65p. [identified by the spelling *sâauch* only used by Leclère, see *BSEI* 57, 1910]
- Leclère, Adhémar. 1910. Les Sâauch. *BSEI* 57: 91-114. La langue des Sâauch, 112-114.
- Malone, Dennis L. 2001. Language development in a minority language community: Report of the Chong Writers Workshop. *MKS* 31: 141-152.
- Martin, Marie A. 1974a. Remarques générales sur les dialectes pear. *Asie du Sud-Est et Monde Insulindien* 5(1): 25-37. Paris, Centre National de la Recherche Scientifique.
- Martin, Marie A. 1974b. Esquisse phonologique du sòmree. *ASEMI* 5(1): 97-106. Paris, Centre National de la Recherche Scientifique.
- Martin, Marie A. 1975. Les dialectes Pear dans leurs rapports avec les langues nationales. *JSS* 63(2): 86-95.
- Martin, Marie A. 1997. Les Khmers Daeum «Khmers de l'origine». Société montagnarde et exploitation de la forêt. De l'écologie à l'histoire. Paris: Presses de l'EFEO.
- Ménétrier E. 1926. *Monographie de la Circonscription résidentielle de Kampot*. Saigon: Editions d'Extrême-Orient. 89p. [saoch vocabulary p. 41-42]
- Morizon, René. 1936. Essai sur le dialecte des populations Pears des Cardamomes. Paris, Editions Inter. 65p.
- Moura, J. 1883. *Le royaume du Cambodge*. Paris, Ernest Leroux, 2 vols. [porr and samreh vocabulary, vol. 1: 440-447]
- Noppawan Thongkham. 2003. The phonology of Kasong at Khlong Saeng Village, Danchumphon Sub-District, Bo Rai District, Trat Province. MA Thesis, Mahidol University, Thailand.
- Pannetier, Adrien Louis Marie. s. d. *Notes sur les Pears, accompagné d'un vocabulaire*. Manuscript. Paris, EFEO. [not easily usable]
- Parkin, Robert. 1991. *A Guide to Austroasiatic Speakers and Their Languages*. Oceanic Linguistics Special Publication N° 23. Honolulu: University of Hawaii Press.
- Pavie, Auguste. 1881-82. Excursion dans le Cambodge et le Royaume de Siam, III. *Excursions et Reconnaissances* 4(10): 99-146. [voc. tchiong 114-116]

- Pelliot, Paul. 1902. Mémoires sur les coutumes du Cambodge. *BEFEO* 2(2): 123-177. Traduction du texte de Tcheou Ta-Kouan.
- Pelliot, Paul. 1903. Le Fou-nan. *BEFEO* 3(2): 248-303.
- Pelliot, Paul. 1904. Deux itinéraires de la Chine en Inde à la fin du VIII<sup>e</sup> siècle. *BEFEO* 4: 131-413.
- Pelliot, Paul. 1951. Mémoires sur les coutumes du Cambodge. *Œuvres posthumes de P. Pelliot*. Paris: Adrien Maisonneuve. Traduction revue et augmentée de 1902.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen lautlehre der Kharia-Sprache*. Wiesbaden, Otto Harrassowitz.
- Pornsawan Ploykaew. 2001. *Samre grammar*. Thesis. ILCRD, Mahidol University.
- Pornsawan Ploykaew. 2001. The phonology of Samre. *MKS* 31: 15-27.
- Pornsawan Ploykaew. 2005. The study of Samre's world view: the belief in sorcerer spirits. *MKS* 35: 171-182.
- Purtle, Dale. 1970. Manuscript vocabulary of Chong collected at Veal Renh, Kampot Province.
- Saifon Luamkhan. 1991. Phaasaa Chong Muu Baan Nam Khun I, Khlong Phlu, Amphur Makham, Changwat Chanthaburi. MA thesis. Silapakorn University.
- Schmidt, Pater Wilhelm. 1907, 1908. Les peuples mon-khmer, trait d'union entre les peuples de l'Asie centrale et de l'Australonésie. *BEFEO* 7 (1907): 213-263 & *BEFEO* 8 (1908): 1-35.
- Schmitt, J. 1904. Les chôngs. *Revue Indochinoise* II(3): 161-162.
- Seidenfaden, E. 1943. Review of R. Baradat, Les sãmre ou Peâr, population primitive de l'ouest du Cambodge, *BEFEO* 41: 1-150, 1941. *JSS* 34(1): 73-80.
- Sidwell, Paul. 1998. *A Reconstruction of Proto-Bahnaric*. Thesis, University of Melbourne.
- Sidwell, Paul. 2005. The Katuic Languages, Classification, Reconstruction and Comparative Lexicon. Lincom Europa.
- Sirikarn Charoentham. 1987. *Phaasaa Chong Muu Baan Thung Ta-In, Changwat Chanthaburi* [Chong Language, Thung Ta-In village, Chanthaburi province]. Thesis. Silapakorn University.
- Siripen Ungsitipoonporn. 2001. A phonological comparison between Khlong Phlu Chong and Wangkraphrae Chong. MA thesis. ILCRD, Mahidol University.
- Skeat, Walter William & Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsula*, two vols. London: Franck Cass & Co. Reprint 1966, New York: Barnes & Noble. [vocabularies from Doudart de Lagrée & Francis Garnier 1873: 498-517, and Moura 1883: 440-447]
- Sunee Kamnuansin. 2002. The syntax of an endangered language: A case study of Kasong. Paper presented to *The 8th International Conference on Thai Studies* (January 9-12, 2002). Nakhorn Pathom, Thailand.
- Sunee Kamnuansin. 2002. *Kasong syntax*. MA thesis. ILCRD, Mahidol University.
- Sunee Kamnuansin. 2003. Syntactic characteristics of Kasong: An endangered language of Thailand. *MKS* 33: 167-182.
- Surekha Suphanphaiboon. 1982. *The phonological system of Chong language*, Muban Takhianthong, Tambon Takhianthong, Amphoe Makham, Changwat Chanthaburi. M.A. Thesis, Srinakharinwirot Prasarnmit University, Bangkok. [in Thai]
- Surekha Suphanphaiboon. 1987. Kan Suksa Klum Chatiphan Chao Chong. Research paper presented to *The Social Science Association of Thailand*.
- Suwaphat Chanmekha. 2003. The Hidden Language: a case Study of Kasong Language Attitudes, Uses and Habilities in three Villages in Trat Province, Thailand. Thesis, Mahidol University. [in Thai]
- Suwilai Premsrirat. 2001. Austroasiatic languages and peoples. In *Annotated Bibliography for Ethnic Minorities in Thailand N°2*. Nakhon Pathom, ILCRD, Mahidol University. [unpublished ms in Thai]
- Suwilai Premsrirat. 2004. *Ethnolinguistic Maps of Thailand*. Ministry of Culture and Mahidol University.
- Suwilai Premsrirat. 2008. *Chong Dictionary*. Center for Revitalization of Endangered languages, Institute of Language and Culture for Rural Development.
- Tcheou Ta-Kuan. 1295/1296. See Paul Pelliot 1902.
- Theraphan L. Thongkum. 1984. Samre language. *Journal of Thammasat University* 15(1): 16-128. [in Thai]
- Theraphan L. Thongkum. 1988. Phonation types in Mon-Khmer languages. In *Vocal Physiology: Voice Production, Mechanisms and Function*. Edited by Osama Fujimura: 319-333.
- Theraphan L. Thongkum. 1991. An Instrumental Study of Chong Register. In *Austroasiatic Languages. Essays in honour of H.L. Shorto*. Edited by J.H.C.S. Davidson. School of Oriental and African Studies, University of London: 141-160.
- Thomas, David D. & Robert K. Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25: 398-418.

- Ueda, H. & Okada, T. 2003. Saoch vocabulary. Reports on minority languages in mainland Southeast Asia.
- Webber, Karl E. 1976. Ethnographic notes of the Chong population in Chanthaburi province southeast Thailand. *Thailand Research Project Diffusion of Innovation Coast Hinterland - Continuum Preliminary*. Office of Senior Research Fellow of Southeast Asia Institute of Heidelberg University of German Culture.

## Goal-marking in Munda with special reference to the Lower Munda languages

Arun Ghosh  
University of Burdwan

### Introduction

‘Goal’ is used here to subsume different semantic roles, viz. recipient (R), direction (D), location (L), beneficiary (B), causee (C) and experiencer (E) to test how these are marked and whether these are marked alike in the Munda languages. It has been argued in typological literature that the goal-marking morphemes are extremely elastic in nature. These morphemes are mostly polysemous in character as they encompass different semantic roles. (Rice and Kabata 2007:451) This is true synchronically as well as diachronically. Here in identifying different goal-marking morphemes and their extension pathways emphasis is given on the three south Munda languages, Remo, Gutob and Gta? which are classified as Lower Munda by S. Bhattacharya (1975) on some morphosyntactic criteria like a. lack of object and dative control in the verb; b. lack of marker for dual number; and c. use of genitive marker different from Upper Munda (Bhattacharya:1975), represented by Santali, Mundari, Ho, Korku, Kharia, Juang, Sora, Pareng and a host of minor Munda languages like Turi, Asuri, Birhor etc. There are two ways of encoding the patient (P) in Lower Munda: prefixation and suffixation depending on selection between the participant noun and pronoun. Marking also depends on animacy and selection between human and non-human participants.

In Remo prefixation is employed regardless of the participants being noun or pronoun. In Gutob and Gta? selection of the process depends on the selection between participant noun or pronoun-- prefixation is used when the participant P is pronoun and suffixation is used when the participant P is noun. In case of marking of the theme (T) the inanimate T is zero marked in all the three languages and if animate it receives the same marking as P. In Remo the recipient (R) receives the same marking regardless of its being noun or pronoun. In Gutob and Gta? it receives the same marking as P. The pronominal R in Gta? can also get double marking, that is, prefix and suffix both. Remo marks the direction (D), location (L), experiencer (E), causee (C) and beneficiary (B) with the same prefix assigned for P and R. In Gutob pronominal D, L, E, C and B are marked by the same prefix assigned for pronominal P and R and the nominal D, L, E, C and B are marked by the same suffix assigned for nominal P and R. Gta? marks E and C with the same prefix or suffix assigned for P and R but uses different markers for D, L, and B. As L, D, C, E and B in Remo and Gutob, and E and C in Gta? are assigned the same marker as R all are treated here as a single argument, viz. Goal. The dative-accusative marker performing the role of P and R is treated here as the basic signal and there from extension pathways are tried to be located. After discussing the goal-marking in Lower Munda languages the Upper Munda languages are taken up for consideration to see whether the same extension pathways hold in the whole Munda group.

### Theoretical Background

The idea of locating the extension pathways of the semantic roles stems from different readings like Blansitt (1988), Heine (1990), Rice and Kabata (2007), Malchukov, Haspelmath and Comrie (2007) and Kittilä (2008). To substantiate his Function Contiguity Hypothesis Blansitt located in the available data of

---

Ghosh, Arun. 2011. “Goal-marking in Munda with special reference to the Lower Munda languages.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.52-68.  
Copyright vested in the author.

Munduruku “a common function marker for dative, allative and locative.” (Blansitt 1988:180) His Function Contiguity Hypothesis was schematically represented as

Object- Dative- Allative- Locative (Ibid. 177)

or

Object = Dative = Allative = Locative (Rice & Kabata 2007: 463)

He generalized among others “if an adposition occurs as both object marker and allative marker, it also occurs as dative marker. If an adposition occurs as both dative marker and locative marker it also occurs as allative marker.” (Blansitt 1988:186)

Heine (1990) in his model of Dative extension observed that the non-cognate suffixes *-ke* and *-ro* of the two Nilo-Saharan languages, Ik and Kanuri respectively, “shared many of the same functions, including the marking of indirect objects, directional locatives, goals, benefactives, purposes, reasons, manner and time complements, as well as marking subordinate clauses and serving as a derivational suffix to mark adverbs (Heine 1990: 129).”<sup>32</sup> Rice & Kabata argued that direction markers tend to extend to recipient marking diachronically.<sup>33</sup> In their endeavour the main purpose was to investigate “the concomitant semantic roles and functions that the principal goal-marking morpheme in a language also marks, such as location, recipient, possessor, experiencer, purpose, etc., as well as more traditionally conceived morphological cases, such as dative, genitive, etc”. (Ibid.: 452) Contesting Blansitt's Function Contiguity Hypothesis (1988) and Heine's Dative Extension model they proposed a third model of the grammaticalization pattern of the Allative (2007:494) in which they proposed four principal semantic domains, SOCIAL, SPATIO-TEMPORAL, LOGICAL-CONTEXTUAL and MENTAL with Social leading to Recipient extending to Addressee and Benefactive, Spatio-temporal to Locative extending to Time and Ablative, Logical-contextual to Purposive extending to Reason and Infinitival and with Mental to Conceptual extending to Perceiver and Experiencer.<sup>34</sup> Malchukov, Haspelmath and Comrie while looking at the ditransitive constructions across languages found that verbs of physical transfer like ‘give’, ‘sell’, ‘lend’ etc. describe “a scene in which an agent participant causes an object to pass in the possession of animate receiver (= recipient)”. They also found that verbs denoting mental transfer like ‘show’ or ‘tell’ behave alike. That is, verbs having recipient and verbs denoting direction behave alike. In their view “the animate argument of ‘show’ and ‘tell’ is not a recipient in the narrow sense”, but can be regarded as an R-argument. Moreover they also found that benefactive constructions having prospective recipient are expressed like the ditransitive constructions. In their findings causee in the causative constructions is found to be functioning like the R, which led them to argue that “this is of course not an accident, because the meanings of transfer verbs contain a ‘cause’ element”. (Malchukov, Haspelmath and Comrie 2007: 2-3) Kittilä is in favour of neutralizing the differences between the semantic roles of recipient and goal in his discussion on differential goal marking (DRM). Observing that “both recipient and goal can be seen as end points of transfer” (Kittilä 2008: 248) he is in favour of lumping these roles together. He further argues that “there are languages which accord R a uniform marking regardless of animacy, which also justifies seeing these as different manifestations of a single semantic role”. (Kittilä 2008: 248)

Taking a cue from previous discussions and having a close look at the Munda data, especially those of the Lower Munda<sup>35</sup> languages, it has been proposed that in Munda all the three arguments, viz. recipient (R), direction (D) and beneficiary (B) can be considered as manifestations of a single semantic role, that is, goal, symbolized as R by Kittilä. Aside these three, ‘causee’, ‘permisee’, person permitted for receiving

---

<sup>32</sup> Sally Rice and Kaori Kabata, Cross-linguistic Grammaticalization Patterns of the ALLATIVE. *Linguistic Typology*, 11, 2007, p.464.

<sup>33</sup> See also Seppo Kittilä & Silvia Luraghi's Questionnaire for their on-going project “Differential marking of spatial relations: the case of direction with human landmarks”. (E-mail circulation).

<sup>34</sup> For a schematic presentation of Rice and Kabata's model see Sally and Kabata, 2007: 494, Figure 21.

<sup>35</sup> The name Lower Munda was given to three extreme south Munda languages, Remo, Gutob and Gta? by S. Bhattacharya in his article titled “Munda studies: A new classification of Munda” (*Indo-Iranian Journal*, v.xvii: 1, pp.97-101) where he isolated these languages from other Munda languages on some morphosyntactic criteria.

something or undertaking some action and the Experiencer<sup>36</sup> are also encoded in the same way as the R in Munda. Other functions are also taken into account as causee involves some kind of transfer of action from one person to the other; person permitted is the receiver of the permission and in the psychomatic verbs whether sensory, emotional or physical like ‘feel good’, ‘be angry’, ‘be hungry’ the logical subject of the sentence receives the R-marking. Moreover, as the roles discussed so far involve some kind of transfer of action and/or event we are also in favour of considering these roles together into a single role, designated hereafter as goal (R). In the following discussion we will primarily concentrate on the encoding of goal (R) in the Lower Munda languages, viz. Remo, Gutob and Gta? and compare it with other Munda languages, especially Santali, Mundari, Korku, Kharia, Juang, Sora and Birhor.

### Goal-marking in Lower Munda

The lower Munda languages, Remo, Gutob and Gta? although taken together, do not encode ‘goal’ in a uniform manner. The extension of semantic role of R varies to some extent. The following examples will illustrate how the R is encoded in the three Lower Munda languages, Remo, Gutob and Gta?. Let us take up Remo<sup>37</sup> data first:

Remo:

1.     gitiŋəŋ         *a-niŋ*                 muŋŋ   gimɛ   bɛɽ-ɔʔ  
       3SG             ACC/DAT-1SG         one   goat   give-PST  
       ‘He gave me a goat.’
2.     *a-nɔ*                 niŋ   surɔŋ                 bɛʔ-ti-ŋ  
       ACC/DAT-2SG     1SG   medicine             give-NPST-1SG  
       ‘I shall give you medicine.’
3.     ramo   *a-sonja*                 muŋŋ   gimɛ   bɛɽ-ɔʔ  
       Ramu ACC/DAT-Sonya     one   goat   give-PST  
       ‘Ramu gave Sonya a goat.’
4.     ɑɽi   *a-meri*                 muŋŋ   upoaɽ   bɛɽ-ɔʔ  
       Hadi ACC/DAT-Mary     one   gift   give-PST  
       ‘Hadi gave a gift to Mary.’

In the examples the R whether nominal or pronominal is uniformly marked with the ACC/DAT prefix *a-*. All the sentences are in ditransitive frame with a human R and non-human T. The non-human T is unmarked. Examples (3) and (4) have alternative forms in which the Rs ‘Mary’ and ‘Sonya’ can receive double marking, that is, accusative-dative prefix *a-* and address clitic<sup>38</sup> *-la*, (which can also be dispensed with) as in the examples (5) and (6):

5.     ɑɽi   *a-meri-la*                 surɔŋ                 bɛɽ-ɔʔ  
       Hadi ACC/DAT-Mary-CLIT     medicine             give-PST  
       ‘Hadi gave medicine to Mary.’
6.     ramo   *a-meri-la*                 muŋŋ   upoaɽ   bɛɽ-ɔʔ  
       Ramu ACC/DAT-Mary-CLIT     one   gift   give-PST  
       ‘Ramu gave a gift to Mary.’

<sup>36</sup> Subject receiving R-marking has been designated as Experiencer by Gregory D.S. Anderson (2008) and Leukas Neukom (2000) and as Dative subject by Manideepa Pattanayak (2008).

<sup>37</sup> Remo is the language of the Bondas distributed in and around the Bonda Hills in the Khairput block of the Malkangiri district of Orissa. There are at least two known dialects of the language—Plains Remo and Hill Remo. The people call their language *remɔsam*. Data presented here represent the Hill dialect.

<sup>38</sup> The clitic is basically used as an endearing element with the person addressed or referred. It is used with the subject noun (+human) as in *sobita-la maj-na kɔlɔm sunɔʔ bɛɽ-ɔ* ‘Sabita sold her pen.’ *selanla niŋ-pulaj ɖaʔ ruŋ-ɔʔ* ‘The girl brought water for me.’ It can also be dispensed with without disturbing the meaning or reference, as in *sobita-la maj-na kɔlɔm sunɔʔ bɛɽ-ɔʔ* and *selan niŋ-pulaj ɖaʔ ruŋ-ɔʔ*.

The goals of the verbs of mental transfer like ‘show’, ‘tell’, ‘bring’ and motion verbs like ‘throw’ receive the same encoding as the R. Examples (7)-(11) illustrate the argument:

7.     gitinəŋ         *a-niŋ*                     uʔsram         uʔ-ɔʔ         bibɛʔ  
        3SG             ACC/DAT-1SG         story           tell-PST         RED-give  
        saʔ-gi-ta  
        come-PRF-NPST  
        ‘He has come to tell me a story.’
8.     gitinəŋ         *a-niŋ*                     nɛ-ʃ     ruŋ-ɔʔ         bɛʔ-ɔʔ  
        3SG             ACC/DAT-1SG         1-child   bring-PST         give-PST  
        ‘He brought and gave me a child.’
9.     *a-jəŋ-ɖāj*                             sun- ɔʔ  
        ACC/DAT-mother-PB.KIN tell-PST  
        ‘He told his mother’.
10.    gitinəŋ         *a-niŋ*                     mujŋ   bire    tul-ɔʔ  
        3SG             ACC/DAT-1SG         One   stone   throw-PST  
        ‘He threw me a stone.’
11.    gitinəŋ         *a-niŋ*                     mujŋ   bəl    kandɛl-ɔʔ  
        3SG             ACC/DAT-1SG         One   ball    throw at-PST  
        ‘He threw a ball at me.’

In the examples from (7)-(11) the direction is encoded through the same marking as the R with the human goal. The ‘goal’ of the motion verbs like ‘come’, ‘enter’, ‘climb’, ‘go’ is encoded in the same way, that is with the accusative-dative prefix *a-*, as in the examples (12-15):

12.    *a-ɖio*                     wiy-α  
        ACC/DAT-house         go-IMP  
        ‘Go home.’                                     (Frank Fernandez 1967: 112)
13.    niŋ    (*a-ɖ io-boʔ*)<sup>39</sup>         gay-ti-ŋ  
        1SG   ACC/DAT-house         enter-NPST-1SG  
        ‘I enter/ shall enter the house.’             (Frank Fernandez 1967: 68)
14.    may   sak-seta         *a-semuʔ*                     ɖayk-ta  
        3SG   come-PART   ACC/DAT-tree         climb-NPST  
        ‘Having come he will climb the tree.’         (Sahu, Samantaray, Patel 1993:30)
15.    *a-niŋ*                     lɔ  
        ACC/DAT-1SG         come  
        ‘Come to me.’

In contrast to Gutob and Gtaʔ (see examples 50, 63-64) Remo marks the place-names and locations with the same accusative-dative marker, as in (16)-(17):

16.    niŋ    *a-mundlipaɖa*                     uj-ti-ŋ  
        1SG   ACC/DAT-Mudulipada         go-NPST-1SG  
        ‘I shall go to Mudulipada.’

---

<sup>39</sup> Here R-marking morpheme α- is optional. Similarly -boʔ can also be dispensed with in favour of α-.



17. nɔ a-af uj-tu-nɔ  
 2SG ACC/DAT-weekly market go-NPST-2SG  
 ‘You will go to the market.’

There is another way of encoding the D/L (direction/location) in Remo, that is, through the D/L-marking suffix *-bɔʔ* and that applies to both human and non-human goals. Consider examples (18)-(20):

18. *bajar-bɔʔ* uj-a mari gimɛ-se ruŋ  
 market-D/L go-IMP and goat-meat bring  
 ‘Go to the market and bring mutton.’
19. *ɖaktɔr-bɔʔ* uj-a ja a-pɛ surɔŋ bɛʔ-tɔ  
 doctor-D/L go-IMP who ACC/DAT-2PL medicine give-NPST  
 ‘Go to the doctor who will give you medicine.’
20. gitinɔŋ laʔa-bɔʔ uj-ga  
 3SG field-D/L go-PRF  
 ‘He went to the field.’
21. badɔl aʔi-bɔʔ uj-sũð-sega mamara-sega sun-ɔʔ  
 Badal Hadi-D/L go-CLIT-PART something-PART tell-PST  
 ‘Going to Hadi Badal told something’.

In these examples the direction-location (D/L) marker *-bɔʔ* encodes both human and non-human goals. In (18) and (20) *-bɔʔ* encodes the non-human goal and in (19) and (21) the human goal. Regular alternation of this suffix with the accusative-dative prefix as found marking the R clearly shows that R and D/L in Remo are treated alike. Examples (22)-(25) justify our argument:

22. niŋ *ɖio-boʔ* gay-t-iŋ  
 1SG house-D/L enter-NPST-1SG  
 VS  
 niŋ a-ɖio gay-t-iŋ  
 1SG ACC/DAT-house enter-NPST-1SG  
 ‘I enter the house.’ (Frank Fernandez 1967: 68)
23. oʔom *gusoʔ-bitre* ɖi-ta  
 arrow dog-in COP-NPST  
 VS  
 oʔom a-gusoʔ ɖi-ta  
 arrow ACC/DAT-dog COP-NPST  
 ‘The arrow is in the dog.’ (Frank Fernandez 1967: 68)
24. niŋ *korji-bagboʔ* layk-t-iŋ  
 1SG chair-D/L sit-NPST-1SG  
 VS  
 niŋ a-korji lajk-t-iŋ  
 1SG ACC/DAT-chair sit-NPST-1SG  
 ‘I sit on the chair.’ (Frank Fernandez 1967: 68)
25. gulaj-ne *konɖa-bɔʔ* onɖaj-ga  
 boys-PL hill-D/L go-PRF  
 VS  
 gulaj-ne a-konɖa onɖaj-ga  
 boys-PL ACC/DAT-hill go-PRF  
 ‘The boys went to the hill.’

The psychosomatic verbs<sup>40</sup> encode the subject as experiencer of transfer point of bodily or mental sensations with the same R-marker, that is *a-*, as in the examples (26)-28). It is the experiencer subject which acts as the transfer point of physical or psychological experience and not the experimenter object, quite in line with the argument put forward by Croft that the experiencer-subject verbs are purely stative and that ‘the experiencer is characterized as simply being in a mental state regarding the stimulus’.’ (Palmer 1994:27; Croft 1991:214-15). The experiencer subject here, however, does not control agreement in the verb.

26. *a-nij* kuru-sa?ga nij kiaŋ sum-o?-nij  
ACC/DAT-1SG hungry-PART 1SG rice eat-PST-1SG  
‘As I was hungry I ate rice.’
27. *a-nij* lu-lor-ɖusu?-gu-ta  
ACC/DAT-1SG RED-vomit-DESID-PRF-NPST  
‘I wish to vomit.’
28. *a-nij* si suso?p-ɖen-ta  
ACC/DAT-1SG fever RED-get-PROG-NPST  
‘I am getting fever.’

Beneficiary which ‘prototypically refers to entities, usually animates that are indirectly affected by the action of the verb’ (Palmer 1994:31) can optionally be encoded with the R marker, that is, the beneficiary in the benefactive constructions may be marked with the accusative-dative *a-* or may be marked with *a-* along with the clitic *pulaj* ‘for’, as in (29)- (31):

29. sobita *nij-pulaj* muj kələm səb-ɔ?  
Sabita 1SG-CLIT one pen buy-PST  
‘Sabita bought a pen for me.’
- 29a. sobita *a-nij* muj kələm səb-ɔ?  
Sabita ACC/DAT-1SG one pen buy-PST  
‘Sabita bought a pen for me.’
- 29b. sobita *a-nij-pulaj* muj kələm səb-ɔ?  
Sabita ACC/DAT-CLIT one pen buy-PST  
‘Sabita bought a pen for me.’
30. selan *nij-pulaj* ɖa? ruŋ-ɔ?  
3SG 1SG-CLIT water bring-PST  
‘She brought water for me.’
- 30a. selan *a-nij* ɖa? ruŋ-ɔ?  
3SG ACC/DAT-1SG water bring-PST  
‘She brought water for me.’
31. gitinəŋ *nij-pulaj* semu? sisep'-ɖen-ta  
3SG 1SG-CLIT tree RED-cut-PROG-NPST  
‘He is cutting the tree for me.’
- 31a. gitinəŋ *a-nij* semu? sisep'-ɖen-ta  
3SG ACC/DAT-1SG tree RED-cut-PROG-NPST  
‘He is cutting the tree for me.’

<sup>40</sup> For a detailed discussion of the psychosomatic verbs in Mundari see Osada (1992: 104-109).

- 31b. *gitiŋə*      *a-niŋ-puləj*      *semuʔ sisep'-dɛn-ta*  
 3SG      ACC/DAT-CLIT      tree RED-cut-PROG-NPST  
 'He is cutting the tree for me.'

The 'causee' in the causative constructions receives the same R-marking as the cause element also involves transfer of action from one person to the other. In Remo any or all of the verbs can be transformed into a causative regardless of transitivity with 'the simple meaning of causing someone to perform the relevant action'. (Palmer 1994: 215). Here also C receives R-marking. Consider (32) – (33):

32. *ramo a-sonja-la*      *kiaŋ ə-sum-əʔ*  
 Ramu ACC/DAT-Sonya-CLIT      rice CAUS-eat-PST  
 'Ramu caused Sonya to eat rice.'

33. *a- əʔəŋ-gu*      *ə-lemə*  
 ACC/DAT-child-PB.OFF      CAUS-sleep  
 'Make the child sleep.'

The extension pathways of R in Remo may be shown schematically as follows:

ACC-DAT--->RECIPIENT=DIRECTION=LOCATION=EXPERIENCE =BENEFICIARY = CAUSEE

A closely related language and a close neighbour of Remo Gutob<sup>41</sup> shows more or less the same typological pattern as found in Remo. Here too the R in the ditransitive frame, direction-location of the motion verb as also the Experiencer of the psychomatic verbs are encoded in the same way. While the pronominal goal is marked by the accusative-dative prefix *o-* the nominal goal is marked by the clitic – *ləj/puləj*. The D/L-marker in respect of both human and non-human goals is *-bo* which optionally alternates with the R-marker *-ləj/puləj*, meaning 'for', that is, the nominal R of the ditransitive frame optionally acts as goal. The beneficiary, the prospective or projected recipient is also found to be receiving the accusative-dative prefix *o-* along with the clitic *-ləj/puləj* in case of pronominal goal and the clitic *-ləj* in case of nominal goals. Consider (34) – (38) where the pronominal goal (both recipient and goal) in the ditransitive frame receives the accusative-dative prefix *o-*:

34. *dajakori o-niŋ*      *beʔ-tu*  
 kindly ACC/DAT-1SG      give-NPST  
 'Kindly give me.'
35. *pen o-niŋ*      *kodli beʔ-tu-pen-ki*  
 2PL ACC/DAT-1SG      banana give-NPST-2PL-Q  
 'Will you give me banana?'
36. *maj o-niŋ*      *sun-ə*  
 3SG ACC/DAT-1SG      tell-PST  
 'He told me.'
37. *maj o-niŋ*      *muj-ro bəl tiŋ-ə*  
 3SG ACC/DAT-1SG      one-CLAS ball throw-PST  
 'He threw me a ball.'
38. *sobita o-niŋ*      *muj-ro boj riŋ-ə*  
 Sabita ACC/ DAT-1SG      one-CLAS book bring-PST  
 'Sabita brought me a book.'

<sup>41</sup> Gutob is the language of the Munda speaking Gadabas distributed mainly in the Lamtaput block of the Koraput district of Orissa. Data analyzed here were collected in the Lamtaput block.

Here the pronominal goals (R) of the verbs of physical and mental transfer are marked by the accusative-dative prefix *o-* and the T is marked zero. Examples of the nominal goals are the following:

39. *aṭi meri-laj muj-ro upoar beṭ-ə*  
 Hadi Mary-CLIT one-CLAS gift give-PST  
 ‘Hadi gave a gift to Mary.’
40. *badəl sobita-nu-laj muj-ro boj riṅ-ə*  
 Badal Sabita-GEN-CLIT one-CLAS book bring-PST  
 ‘Badal brought a book to Sabita.’

*laj-pulaj* is sometimes found to be alternating with *-bo* as in the alternative sentence in (41a). Compare (41) and (41a):

41. *sobita sonja-nu-pulaj muj-ro boj riṅ-ə*  
 Sabita Sonya-GEN-CLIT one-CLAS book bring-PST  
 ‘Sabita brought a book to Sonya.’
- 41a. *sobita sonja-nu-bo muj-ro boj riṅ-ə*  
 Sabita Sonya-GEN-D/L one-CLAS book bring-PST  
 ‘Sabita brought a book to Sonya.’

The experiencer (E) is encoded with the accusative-dative marker *o-* in case of pronominal experiencer and *-laj* in case of nominal experiencer. As in Remo the experiencer does not control agreement in the verb. Consider (42) and (43):

42. *o-niṅ sos laḡaj-guni*  
 ACC/DAT-1SG thirst be affected-COP  
 ‘I am thirsty.’
43. *o-niṅ besi duk-ḡeṅ-guni*  
 ACC/DAT-1SG very be sad-PROG-COP  
 ‘I am very sad.’

Beneficiary is also marked by the accusative-dative marker *o-* and the clitic<sup>42</sup> *-laj*. Consider (44) and (45):

44. *o-niṅ-laj ḡa? riṅ-tu-nom-ki*  
 ACC/DAT-1SG-CLIT water bring-NPST-2SG-Q  
 ‘Will you bring water for me.’ (Asha Kiran Society 2002:14)
45. *nisani maapru-pulaj gisiṅ riṅ-nen*  
 Nisani god-CLIT fowl bring-3PL  
 ‘Let them bring fowl for the Nisani God.’

In the ditransitive frame with two arguments—R and T and both are human R is marked by *-bo* and T is marked by *-laj/pulaj*. In similar constructions with R and nonhuman T the R is encoded with *-laj* and the T is left unmarked. Consider (46) and (47):

46. *niṅ oḡon-laj joṅ-ḡej-nu-bo beḡ-o?-niṅ*  
 1SG child-ACC/DAT mother-P.B-KIN-GEN-LOC give-PST-1SG  
 ‘I gave the child to the mother.’

<sup>42</sup> *-laj* is assigned the accusative status as it is basically used for P-marking.

47. ramu *sonja-laj* muj-ro gime? beḍ-o?  
 Ramu Sonya-CLIT one-CLAS goat give-PST  
 ‘Ramu gave Sonya a goat.’ (Rajan & Rajan 2001: 32)

The causee is encoded with the same accusative-dative marker as in (48) and (49):

48. o-nom niṅ niṅ-nu *apuṅ-lay* əb-soi-tu-niṅ  
 ACC/DAT-2SG 1SG 1SG-GEN father-CLIT CAUS-see-NPST-1SG  
 ‘I will show my father to you.’ (Bhattacharya 1975: 164)
49. nej gutob-log *sahebo-pulaj* samo əb-gir-nej  
 2PL Gadaba-people British-CLIT language CAUS-learn-2PL  
 ɖu-tu  
 AUX-NPST  
 ‘We the Gadabas are teaching the Saheb (our) language.’

Unlike Remo Gutob does not encode place-names with the R-marker. Consider (50):

50. niṅ bier *mudlipaḍa* i-lom-niṅ  
 1SG tomorrow Mudulipada go-FUT-1SG  
 ‘I shall go to Mudulipada tomorrow.’

The extension pathways of R in Gutob as found in the examples (34)-(50) can therefore be schematized as follows:

ACC-DAT -->RECIPIENT=DIRECTION=LOCATION=EXPERIENCER BENEFICIARY = CAUSEE

The other member of the group Gta?<sup>43</sup> follows more or less the same pattern. Like Remo and Gutob there are two markers for encoding R, one for pronoun and the other for noun. While the nominal R is marked by *-ke*<sup>44</sup> the pronominal R is marked by the accusative-dative prefix *a-*. Sometimes the pronominal goal receives double marking, that is, pronominal R is marked by the accusative-dative prefix *a-* and the suffix *-ke*. The causee and the experiencer receive the same encoding, *a-* in case of pronouns and *-ke* in case of nouns. Pronominal goal with the verbs of physical transfer, mental transfer as also with motion verbs is marked with the accusative-dative prefix *a-*. Consider (51) – (54):

51. *a-niṅ* muj ʈaa biʔ-la  
 ACC/DAT-1SG one rupee give-IMP  
 ‘Give me one rupee.’
52. mε *a-niṅ* husra ani-te  
 3SG ACC/DAT-1SG story tell-PST  
 ‘He told me a story.’

Nominal goal with the same verbs is marked with *-ke*:

<sup>43</sup> Gta? is the language of the Didayis spoken in the Kudumulguma block of the Malkangiri district of Orissa. They themselves call their language *Gta?sa*. There are three known dialects of the language—Plains Gta?, Hill Gta? and Cut off area Gta?. My data are drawn from the cut-off area. S. Bhattacharya calls the language by their ethno-name Didayi.

<sup>44</sup> *-ke* may be a loan from Indo-Aryan as most of the eastern Indo-Aryan languages have this suffix used in dative-accusative. But if it is a loan from Indo-Aryan then through which language does it come? The most plausible answer would be Oriya, as it is the next neighbour to Oriya. But interestingly Oriya possesses the suffix in the form of *-ku*, not *-ke* which is found in Bangla. And Bangla is in the neighbourhood of Gta? in the southern part of Malkangiri district where more than 600,000 Bengalees are staying in the Dandakaranya project area

53. ramu *sonja-ke* mujŋ gimi biʔ-te  
 Ramu Sonya-ACC/DAT one goat give-PST  
 'Ramu gave a goat to Sonya.'
54. neŋ *ɖomru-ke* muj ciʔi biʔ-neŋ-ge  
 1SG Domru-ACC/DAT one letter send-1SG-PST  
 'I sent a letter to Domru.'

Like Remo and Gutob causee also receives the accusative-dative encoding. Consider (55)-(56):

55. ramu *ũjo-ke* a-ʃoŋ-te  
 Ramu baby-ACC/DAT CAUS-eat-PST  
 'Ramu fed the baby.'
56. mɛ *a-na* miʔ<sup>h</sup>a a-ʃoŋ-te  
 3SG ACC/DAT-2SG sweets CAUS-eat-PST  
 'He caused you to eat sweets.'

Experiencer as perceiver of physical and mental sensation is encoded with the accusative-dative marker as in (57)-(58):

57. *a-niŋ* besikeɖe sob-ɖiŋ-te  
 ACC/DAT-1SG very behungry-PROG-NPST  
 'I am very hungry.'
58. tæŋ kitoŋ *remwa-ke* bweʔ-tur-la *remwa-ke* gæ-ke  
 that god man-ACC/DAT spit-DS man-ACC/DAT itch-PST  
 'The god spit on the man and the man started to feel itchy.'  
 (Anderson 2008:697)

Example (58) is typical in the sense that along with the experiencer subject (the second italicized field in the sentence) the experiencer object (P) is marked with *-ke* (the first italicized field in the sentence). Examples (59) and (60) show double marking of R:

59. *a-me-ke*<sup>45</sup> ter gisiŋ biʔ  
 ACC/DAT-3SG-ACC/DAT this fowl give  
 'Give him this fowl.'  
 (Bhattacharya 1975b:166)
60. *a-mɛ-ke* biʔ-la  
 ACC/DAT-3SG-DAT give-IMP  
 'Give him.'  
 (Panda: 1989:13)

In a ditransitive frame with R and T and both human, R is marked with *-ke* and the T is left unmarked as in (61) and in the similar frame with human R and non-human T the R is marked with *-ke* and the T is left unmarked as in (62):

61. neŋ *jaŋ-ke* mujŋ *mujnom* biʔ-te  
 1SG mother-DAT one child give-PST  
 'I gave the child to the mother.'
62. neŋ *jaŋ-ke* muiŋ *gisiŋ* biʔ-te  
 1SG mother-DAT one fowl give-PST  
 'I gave mother a fowl.'

---

<sup>45</sup> As examples (59) and (60) are taken from Bhattacharya (1975) and Panda's (1994) data no pattern can be found regarding the distribution of *a-*, *-ke* and *a-.....ke*.

Unlike Remo and like Gutob Gta? does not overtly mark the place-names. Compare (63) and (64) with (65):

Gta?:

63. niŋ njigɔɔ      *mudulipaɔa* waj-e  
1SG tomorrow Mudulipada go-FUT  
'I shall go to Mudulipada tomorrow.'

Gutob:

64. niŋ bijer      *mudlipaɔa* i-lom-niŋ  
1SG tomorrow Mudulipada go-FUT-1SG  
'I shall go to Mudulipada tomorrow.'

Remo:

65. niŋ *a-munɔlipaɔa*      uj-ti-ŋ  
1SG ACC/DAT-Mudulipada go-NPST-1SG  
'I shall go to Mudulipada.'

In Gta? there is one locative-directional marker *-rini* which is not found to mark the R although there is enough evidence of R marking the direction.

The extension pathways of R in Gta? may be schematised as follows:

ACC-DAT----> RECIPIENT = DIRECTION = CAUSEE = EXPERIENCER

The pattern of extension of the semantic role of R as discussed so far is not the same in all the three Lower Munda languages. While it encompasses the roles of R, D, L, E, B, C in Remo and Gutob, it covers only R, D, C and E in Gta?, Beneficiary and Location falling outside the purview of R. Now let us consider the grammaticalization pattern of the R in other Munda languages outside Lower Munda.

### Goal-marking in Upper Munda:

The grammaticalization pattern of R as attested in other Munda languages, can be shown in (66)–(86):

Santali:

The recipient (66), direction (67-69) and causee (70) are marked by the applicative and pronominal incorporation in the verb in Santali:

66. daka    *ɛm-a-d-iŋ-a-e*  
rice    give-A-TM-1SG-FIN-3SG  
'He gave me rice.'
67. hɛc'-*a-d-iŋ-a-e*  
come-A-TM-1SG-FIN-3SG  
'He came to me.' (SK cf. LN. 49)
68. uni    *lɔi-a-ko-a-e*  
3SG    tell-A-3PL-FIN-3SG  
'He will tell them.'
69. dare-ŋ    *bɛŋget'-a-k'-kan-tahɛkan-a*  
tree-1SG    look at-A-INAN-COP-COP:PST-FIN  
'I was looking at the tree.' (Bo. Cf. LN.53)

70. *ɲɛl-oco-a-d-e-a-e*  
see-CAUS-A-TM-3SG-FM-3SG  
'He showed it to him.'

## Mundari:

The same process holds in Mundari as in (71-73):

71. *maŋɟi seta-ko-ɲ om-a-d-ko-a*  
food dog-PL-1Sg give-A-TM-3PL-FIN  
'I gave food to the dogs.'
72. *dasi-ko-e kaji-a-t'-ko-a*  
servant-PL-3SG tell-A-TM-3PI-FIN  
'He told his servants.' (Bhattacharya 1975: 150)
73. *diku-ɲ itu-a-d-ko-a*  
Hindu-1SG teach-A-TM-3PL-FIN  
'I have taught Hindi to them.' (Osada 1992: 95)

## Korku:

In Korku the recipient as in (74), direction as in (75) and causee as in (76) are marked the same suffix *-ke*:

74. *am-ke ijɲ mya-kama:y ghaliba*  
you-to(IO) I one-work-obj give up  
'I will give you a work.' (Nagaraja1999: 46)
75. *ijɲ Dic-ke ambesasa kule-c-lakken*  
I he-obj.mango-bring send-per-cont  
'I am sending him to bring mango.' (Nagaraja1999:46-47)
76. *dij' koro-ke ij-en ghaleij*  
that man-ACC/DAT 1SG-D/L show  
'Show me that man.' (Bhattacharya 1975:157)

## Birhor:

In Birhor the recipient receives the *-ke* marking along with pronominal incorporation in the verb, as in (77):

77. *oni sim ij-ke ago-ij-mi*  
that fowl 1SG-ACC/DAT bring-1SG-IMP  
'Bring that child to me.' (Bhattacharya 1975: 154)

## Kharia:

In Kharia the recipient as in (78), causee as in (79) and direction as ((80) are marked by the suffix *-te*:

78. *am am-a beʔa-nom-te ij-te tere*  
2SG 2SG-GEN son-2SG-ACC/DAT 1SG-ACC/DAT give  
'You give me your son.' (Bhattacharya 1975:158)
79. *ij ij-a apa-ij-te am-te ob-iyo-ijɲ'*  
1SG 1SG-GEN father-1SG-ACC 2SG-ACC CAUS-see-1SG  
'I will show my father to you.' (Bhattacharya 1975: 158)



80. *in ulaga-ʃhoŋ daru-te ɖep-naiŋ-gʻ*  
 1SG leaf-for tree-ACC/DAT climb-1SG-FUT  
 ‘I will climb the tree for leaves.’ (Bhattacharya 1975: 159)

Juang:

In Juang *-te* marks the recipient (81), causee (82)-(83) direction (84):

81. *ram ʃæmɔ-te ara iŋa baŋbaŋndɔ kete*  
 Ram Shyam-ACC/DAT Self House burn about  
*me-gata-yɔrɔ gam-ɔ*  
 NEG-say-INF say-PST  
 ‘Ram told Shyam not to tell anybody about his burning house.’  
 (Pattanayak 2008:212)
82. *aiŋ-te juaŋ-ka gata-ro-ki ab-soŋ-iŋ*  
 1SG-ACC/DAT Juang-GEN language-DEF-PL CAUS-teach-1SG  
 ‘Teach me the language of Juang.’ (Bhattacharya 1975:160)
83. *am aiŋ-te goble-kon-om-te m-am-ŋo-iŋ*  
 2SG 1SG-ACC/DAT nephew-2SG-ACC/DAT CAUS-see-1SG  
 ‘You will show your nephew to me.’ (Bhattacharya 1975: 160)
84. *ramɔ-a [aiŋ-a inja-te ɔŋɔr-te]*  
 Ram-GEN [1SG-GEN house-ACC/DAT go-INF ]  
*belɔ a-si-an*  
 time NEG-be-PST  
 ‘Ram does not have time to go to my house.’ (Pattanayak 2008: 70)

Sora:

Like the Lower Munda languages Sora also marks the recipient with the prefix *-a*, as in (85)-(86):

85. *kuni a-tarbaŋ-ji a-maŋɖra tiya*  
 that ACC/DAT-flower-PL ACC/DAT-man give  
 ‘Give those flowers to the man.’ (Bhattacharya 1975:162)
86. *kuni a-tarbaŋ-ji a-niŋ-ji tiya*  
 that ACC/DAT-flower-PL ACC/DAT-3PL give  
 ‘Give those flowers to them.’ (Bhattacharya 1975: 162)

Examples cited can be sub-grouped into two: examples (66) to (73) and (77), (82)-(83) on the one hand and (74) to (76), (78)-(79) and (85)-(86) on the other. They differ in respect of R-marking in the verb. In Santali (66-70) Mundari (71-73), Birhɔɾ (77) and Juang (82-83) the R is cross-referenced on the verb. But while Santali and Mundari take the applicative *a-* tagged on to the pronominal referent or to the TAM their close kin Birhɔɾ does not take any applicative marker, simple pronominal forms are added to the verbal predicate in the form of infix. In Korku (74)-(76), Kharia (78-79), Sora (85-86) and in Juang (81) there is no verbal cross-referencing. While in Kharia (78-80) and Juang (81 and 84) the R is marked by *-te* in the sentence Sora (85-86) marks R by the accusative-dative prefix *a-*. One interesting point about the second group (Kharia and Sora) is that the suffix *-te* and prefix *a-* employed for marking the R is also used for marking the T which may lead to ambiguity. To disambiguate the phenomenon the ordering of T and R may be taken into account. Kharia and Sora normally follow the S T R V pattern, though S R T V pattern for Kharia can not be ruled out.<sup>46</sup> Although supporting evidences are not available for all languages examples like (67), (70), (72), (77), (74), and (80) clearly show that recipient and direction/location converge and support our hypothesis that movement or transfer in respect of object or location can be subsumed under R

<sup>46</sup> This is confirmed by John Peterson in an informal talk.

(goal) in Munda. In some languages outside Lower Munda Beneficiary (B) is also marked in the same way as R. Consider (87)–(90):

Santali:

87. parkom bel-a-ko-m peṛa-ko hec'-en-a  
 Bedstead place-A-3PL-2SG visitor-PL come-PST:INTR- FIN  
 'Place a bedstead for them to sit on, visitors have come.' (Bo.cf. LN: 48)

Mundari:

88. daru-m mag-a-n-ta-n-a  
 tree-2SG cut-A-1SG-AM-ITM-FIN  
 'You are cutting the tree for me.' (Osada 1992: 93)

Korku:

89. Do Di-ku-ke co:ja-ma lija sege-ba De-Ten  
 and dem-pl-IO why-re cloth bring-np dem-abl  
 iṅ-ke-ka mya soy-lubu sa:li  
 I-IO-rest one good-cloth bring-imp  
 'And why do you bring new clothes for them rather bring one new cloth for me.'  
 (Nagaraja 1999:155)

Juang:

90. ara judi golo-bə ɔn-de aṅ-te  
 3SG If Keonjhar-DIR go-CON 1SG-ACC/DAT  
 paçolam-R-r-ene  
 shawl FUT-bring-FUT  
 'If he goes to Keonjhar he will bring a shawl for me.' (Pattanayak2008: 36)

At least in these three languages outside Lower Munda Beneficiary receives the R- marking.

The Experiencer also receives the R-marking outside Lower Munda. Consider (91)-(92) and (95):

Santali:

91. hola-e dak'-a-t'-le-a  
 yesterday-3SG rain-A-TM-1PL-FIN  
 'Yesterday we had rain.' [lit. rain affected us] (Bo. Cf. LN: 50)

Mundari:

92. sowan-ja-ʔ-n-a  
 smell-AM-TM-1SG-FIN  
 'I have sensed a smell.' (Osada 1992: 106)

Korku:

93. iṅ-en kapara kasu-lakken  
 I-dlc head ache-cont  
 'I am having headache.' (Nagaraja 1999: 100)
94. iṅ-en da dadam-nen  
 I-dlc water thirsty-pt  
 'I am thirsty.' (Nagaraja1999:100)

Juang:

95. *aiŋ-te*                      *emalō i-sere*  
 1SG-ACC/DAT      cold   be-PRF  
 'I have got cold.'  
 (Pattanayak 2008: 193)

Here in these examples as the syntactic subject is affected by either the event ('rain' as in Santali) or sensation ('smell' as in Mundari) it receives the R-marking. In Korku, however, the experiencer (E) receives a separate D/L marker which is not used for marking the recipient, causee or beneficiary. Along with direction, beneficiary, causee and experiencer applicative also encodes the person permitted or receiving permission in permissive construction in Santali as in (96):

Santali:

96. *kirij-oco-a-d-iŋ-a-e*  
 buy-C/P-A-TM-1SG-FIN-3SG  
 'He permitted me to buy.'

In the permissive construction (96) the argument receiving permission is also encoded with the applicative *a-* as the permittee is also an indirect recipient.

The grammaticalization pattern of R in the Upper Munda languages can be schematically represented in two blocs---- one represented by Santali and Mundari where R is marked by the applicative and the other represented by Korku, Birhor, Kharia, Juang and Sora where R is marked by the accusative-dative morpheme.

1. APPLICATIVE[Santali and Mundari]--->  
 RECIPIENT = DIRECTION = CAUSEE = BENEFICIARY =EXPERIENCER = (PERMISSEE)<sup>47</sup>

2. ACC/DAT [Korku, Birhor, Kharia, Juang and Sora]---> RECIPIENT=DIRECTION= CAUSEE=  
 BENEFICIARY=(EXPERIENCER)<sup>48</sup>

On the basis of the discussion made separately of Lower and Upper Munda so far the over all semantic-functional extension of R in Munda may be delineated in a linear model as follows:

**R→RECIPIENT=DIRECTION=LOCATION=CAUSEE=(BENEFICIARY)=EXPERIENCER  
 =(PERMISSEE)**

The model explicates that all the roles are not shared by all the Munda languages. The common extension is Recipient, Direction, Location, Cause and Experience. But while Gutob, Remo, Santali, Mundari, Korku, Juang have Benefactive extension other languages do not share this feature (as attested in the data). The Permissive construction is found only in Santali. Of the Lower Munda languages Remo and Gutob have elaborate extension process covering six roles—R, D, L, E, B, C uniformly. Gta? shows only four-way extension covering R, D, C, E. Outside Lower Munda Santali has elaborate pattern covering six roles—R, D, C, B, E, P. For paucity of data nothing conclusive can be said about Birhor and Sora, though the general tendency leads us to the assumption that the six-way extension path of R might have been the general tendency of proto-Munda.

## Conclusion

The paper examines encoding of goal (R) from a Munda perspective. The phenomenon comprises different types: one, whether the recipient of a ditransitive construction receives the same encoding as the goals of direction-location. It has been found from the available data that recipient and directional-locational

<sup>47</sup> The permissive construction is available in Santali, hence PERMISSEE is kept in paranthesis.

<sup>48</sup> Experiencer is found to be receiving R-marking in Kharia and Juang. Korku marks E with a direction-location marker which is not used to mark the recipient.

goals receive the same encoding, not only in Lower Munda but in other Munda languages as well. Two, whether beneficiary (B), causee (C) and permisee (P) receive the same marking as R. It has been found that all these roles receive the same marking as R. Three, whether coding of the syntactic subject coincides with R. It is also found to be partially true. The present study (to the best of my knowledge) is the first cross-Munda analysis of R (goal), raising many issues that deserve to be investigated in more detail. It is true that it may not be so easy to make generalization about R-marking in Munda in the light of the data presented. Some more data need to be incorporated for that. Still it is to be noted that the phenomena discussed give some idea about the general tendency of R-marking in the Munda languages.

#### Abbreviations

A	applicative	INAN	inanimate
ABL	ablative	INF	infinitive
ACC	accusative	IO	indirect object
AKS	Asha Kiran Society	ITM	intransitive marker
AM	aspect marker	LN	Leukas Neukom
AUX	auxiliary	NEG	negative
B	beneficiary/benefactive	NPST	non-past
Bo.	P.O.Bodding	OBJ.	object
CAUS	causative	P	patient
CLAS	classifier	PART	Participle
CLIT	clitic	PB.KIN	post base kinship marker
C/P	causative/ permissive	PB.OFF	post base marking offspring
CON	conditional	PL	plural
cont	continuous	PROG	progressive
COP	copula	PRF	perfect
D/L	direction/location	PST	past
DAT	dative	Q	question marker
DEF	Definitive	R	recipient (goal)
DESI	desiderative	RED	reduplicated
DIR	directional marker	R&R	Rajan and Rajan
DL	dual number	SG	singular
dlc	direction-location	SK	L.O. Skrefsrud
FIN	finite marker	T	theme
FUT	future	TAM	tense aspect marker
GEN	genitive	TM	tense marker
IMP	imperative	VC	verb clitic.

### Acknowledgement

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany has provided financial assistance for undertaking field survey in the southern part of Orissa for collecting data in Remo, Gutob and Gta?. I would like to thank Professor Bernard Comrie, Director, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany for arranging the financial assistance and for constant support in undertaking linguistic study of the dying Munda languages. I also owe much to him for going through the first two drafts of this paper. I owe much to Dr Seppo Kittilä of the University of Helsinki for his constant encouragement in building the theoretical framework of this paper and for going through the first draft of the paper. I thank Dr. Nicholas Evans of Australian National University and Dr. John Peterson of the University of Leipzig for their comments on earlier drafts of the paper. I also remember with gratitude the help I received from my informants, especially Lachimi Challan, Soma Kirsani, Badal Dhangramajhi, Jaya Bisoyi, Subhas Muduli, Jogu Bodnayek, Guru Kirsani and Niranjan Sisa during my field trip to Koraput and Malkangiri in July-August 2009, February-March and September-October 2010. Lacunae if there is any is of course mine.

### References

- Anderson, Gregory D S and K. David Harrison, 2008. Remo. In: Self (ed.) *The Munda Languages*. London: Routledge, pp. 557-632.
- Asha Kiran Society, 2002. *Gutob Language Learners' Guide*. Lamtaput, Koraput.
- Bhattacharya, Sudhibhushan, 1975. *Studies in Comparative Munda Linguistics*. Simla: Indian Institute of Advanced Studies.
- \_\_\_\_\_. 1975. Munda studies: A new classification of Munda. *Indo-Iranian Journal*, v. XVII: 1, pp.97-101.
- Blansitt, Edward L. Jr., 1988. Datives and Allatives. In: Michael Hammond, Judith Moravcsik & Jesica Wirth (eds.), *Studies in Syntactic Typology*. pp.173-191. Cambridge: CUP.
- Bodding, P.O. 1929 *Materials for a Santali Grammar, Part II*. Dumka: The Santal Mission of the Northern Churches.
- \_\_\_\_\_. 1925-1929 *Santal Folktales*. Oslo: H. Aschehoug & Co.
- \_\_\_\_\_. 1929-1936 *A Santal Dictionary*. Oslo: A. W. Broggers Boktrykkeri.
- Croft, W. 1991. *Syntactic Categories and grammatical relations*. Cambridge: Cambridge University Press.
- Fernandez, Frank, 1967. *A Grammatical Sketch of Remo: A Munda Language*. PhD thesis, University of North Carolina at Chapel Hill.
- Kittilä, Seppo, 2008. Animacy effects on differential Goal marking. *Linguistic Typology*, 12 pp.245-268.
- Lukas Neukom, 2000. Argument marking in Santali. *Mon-Khmer Studies*, 30, pp.95-113.
- Malchukov, Andrej, Martin Haspelmath, Bernard comrie, 2007. Ditransitive constructions: A typological overview. *First draft*.
- Matson, D. M., 1964. *A Grammatical Sketch of Juang*. PhD dissertation, Madison, University of Wisconsin.
- Nagaraja, K. S. *Korku Language Grammar, Text, And Vocabulary*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Osada, Toshiki, 1992. *A Reference Grammar of Mundari*. Tokyo: Tokyo University of Foreign Studies.
- Panda, Pramodakumari, 1989. *Didayi*. Bhubaneswar: Academy of Tribal Dialects and Culture.
- Pattanayak, M., 2008. Juang. In: Gregory D S Anderson (ed.), *The Munda Languages*. London: Routledge, pp. 508-556.
- Rajan, H. & Jamuna Rajan, 2001. *Grammar write up of Gutob-Gadaba*. Lamtaput: Asha Kiran Society.
- Rice, S. & K. Kabata, 2007. Crosslinguistic grammaticalization patterns of the ALLATIVE. *Linguistic Typology* 11, pp.451-514.
- Sahu, Samantaray and Patel, 1993. *Bonda*. Bhubaneswar: Academy of Tribal Dialects and culture.
- Skrefsrud, L.O. 1873. *A Grammar of the Santhal Language*. Benares: Medical Hall Press.

## A Synopsis of Mal Phonetics

**Theraphan L-Thongkum, Chommanad Intajamornrak,  
Kanitha Putthasatien, Yupaporn Huadsiri, and Supaporn Phalipat**  
Department of Linguistics, Chulalongkorn University

### Abstract

The consonants, vowels and pitches or tones of Mal (T'in), a Khmuic language of the Mon-Khmer language family, were acoustically studied using computer programmes and were statistically tested. The three places of articulation of initial stops in Mal can be identified using the Locus Equation (LE) method. Even though the distinct slopes and y-intercepts of the linear equation can characterize the place of articulation, it is more effective to use the slope values. The ratio of duration for long-to-short vowels is 2:1. The overall vowel space of short vowels is smaller than that of their long counterparts but the variation within the space of each short vowel is more dispersed. The phonation type of the initial sonorant has an influence on the F0 value of the following vowel. Some Mal varieties have become two-tone languages, /high/ and /low/. The five stages of tonal evolution induced by language contact with Khammueang are postulated.

### 1. Introduction

The Lua' or T'in are a Mon-Khmer speaking ethnic group living in Nan province, northern Thailand and the adjacent Sayaburi province of Lao PDR (see map in the appendix). In Nan province, the 34,600 Lua' speakers are scattered throughout ten districts: Bo Kluea, Pua, Chaloeam Prakhiat, Chiang Klang, Thung Chang, Santisuk, Mueang Nan, Mae Charim, Wiang Sa and Song Khwae (more detailed information can be found in L-Thongkum and Intajamornrak, 2009). According to Filbeck (1972), the Lua' language of Nan province consists of two major dialects, i.e. Mal and Pray, and each of these two dialects has a few sub-dialects or varieties. This classification has been confirmed by Singnoi (1988) and Jirananthanaporn (1993). However, the Lua' have never identified themselves as Mal or Pray. Perhaps, the ethnonyms and language names "Mal" and "Pray" are only known to anthropologists and linguists, not among the Lua' or the local people of Nan province. They always say "we are Lua' people and we prefer to be called Lua' by outsiders". They do not want to be KhonT'in (meaning natives or local inhabitants) because it has derogatory connotations, i.e. stupid, ignorant and uncivilized people.

In Nan, besides the Mal-Pray language, the other Mon-Khmer languages are also spoken, i.e. Khmu' and Mlabri. A number of cognate words can be found in Lua' (Mal and Pray), Khmu' and Mlabri, such as 'to snap (the fingers)': \*phlas (Proto-T'in), phlayh (Mal), phat (Pray), plíjh (Khmu' Rawk) and plalh (Mlabri); 'to wake (someone up)': \*phrəh (Proto-T'in), phyəh (Mal), phrəh (Pray), phrəh (Khmu' Rawk) and papuuuh (Mlabri).<sup>49</sup> These Mon-Khmer languages belong to the Khmuic branch of the Mon-Khmer language family (Filbeck, 1972 and 1978).

---

<sup>49</sup> These two examples are from L-Thongkum (2007). More information on the Mal, Pray, Khmu' and Mlabri lexicon, about 2,400 entries, is provided in L-Thongkum et al., for further comparative studies.

---

Theraphan L-Thongkum, Chommanad Intajamornrak, Kanitha Putthasatien, Yupaporn Huadsiri, and Supaporn Phalipat . 2011. "A Synopsis of Mal Phonetics." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.69-80.  
Copyright vested in the authors.

Mal has two major varieties, tonal and non-tonal. The tonal variety, spoken in Ta Luang and Yotdoiwatthana villages in the districts of Pua and Bo Kluea respectively, was selected for our acoustic study. The result of our literature survey shows that an acoustic study of Mal does not exist, even though Mal phonology has been studied synchronically and diachronically by Filbeck (1972, 1976, 1978, 1990); Singnoi (1988); L-Thongkum (2007) and L-Thongkum et al. (2007). Our acoustic studies of Mal relating to consonants, vowels and tones were made under the research project ‘Linguistic Diversity in Nan Province: A Foundation for Tourism Development’ funded by the Thailand Research Fund (TRF) for three years (2004-2007).

This paper is a synthesis of the research findings presented in three M.A. theses (Huadsiri, 2007; Phalipat, 2007 and Putthasatien, 2007) and a Ph.D. dissertation (Intajamornrak, 2009c). The objectives of our research on Mal phonetics are: to test whether the Locus Equation (LE) method can satisfactorily identify the places of articulation of initial stops; to analyse and compare the duration, formant frequency and vowel space of short and long vowels; and to investigate the acoustic characteristics of the two tones in order to postulate the stages of tonal evolution from the past to the future.

## 2. Phonological sketch of Yotdoiwatthana Mal

Mal has many initial consonant clusters due to the reduction of prefixes, infixes and sesqui-syllables (minor or unstressed syllables before major or stressed syllables), otherwise it can be said that the Mal phonological system is rather simple in comparison with those of its sister languages. The consonant, vowel and tone systems of Yotdoiwatthana Mal are shown below.

### Consonants

		<i>Labial</i>	<i>Alveolar</i>	<i>Palatal</i>	<i>Velar</i>	<i>Glottal</i>
<i>Plosive</i>	VI. unaspirated	p	t	c	k	ʔ
	VI. Aspirated	ph	th		kh	
	Vd.	b	d	ɟ	g	
<i>Nasal</i>	VI.	hm	hn	hɲ	hŋ	
	Vd.	m	n	ɲ	ŋ	
<i>Fricative</i>	VI.		s			h
<i>Aproximant</i>	VI.	hw	hl			
	Vd.	w	l	j		
	Pregottalised	ʔw				

### Consonant clusters

Cw	phw thw kw khw lw
Cl	pl phl bl kl khl
Cj	pj phj bj kj gj mj sj
NC (l/j)	mph mpl mphi nth ns ŋk ŋkh ŋg ŋkl ŋkj

### Vowels

<i>Monophthongs</i>	Front		Central		Back	
	Short	Long	Short	Long	Short	Long
High	i	i:	ɨ	ɨ:	u	u:
Mid	e	e:	ə	ə:	o	o:
Low	ɛ	ɛ:	a	a:	ɔ	ɔ:
<i>Diphthongs</i>	iə		(iə)		uə	
	i:ɨ e:ɨ e:ɨ e:ɨ e:ɨ a:ɨ a:ɨ o:ɨ c:ɨ c:ɨ i:ɨ e:ɨ					

**Tones**                    ˊ (high)                    ˋ (low)

<sup>50</sup> Proto-Tʻin (Proto-Lua) \*r has become \*ɨ in Yotdoiwatthana Mal, for example, \*phar > pháɨ ‘to fly’; \*thar > tháɨ ‘rope, cord, string’; \*khe:r > khé:ɨ ‘carambola’ and so forth.

### 3. Methodology

The data for the acoustic study of consonants, vowels and pitches or tones was collected at Ban Ta Luang in Amphoe Pua which is a district of the province of Nan, northern Thailand. All of the words used in the devised word lists were pronounced by three female speakers (30-50 years old) and directly recorded with a high-quality microphone onto a computer notebook using Adobe Audition. With regard to the study of pitch patterns which can reflex tonal evolution in Mal, three more Mal varieties spoken in Ban Kwet, Ban Phu Kok, Ban Yotdowatthana and a conservative variety of non-tonal Pray spoken in Ban Huai Lom were acoustically analysed and compared. An acoustic study of tones in the Pua variety of Khammueang pronounced by native speakers and bilingual speakers in Mal and Khammueang was done to help confirm our claims on the birth of tone in Mal.

In order to test whether the Locus Equation (LE) method can satisfactorily identify the three places of articulation of initial stops in Mal, i.e. bilabial, alveolar and velar, suitable word lists were devised. The 81 words used in this study had CVV(C) syllable structure, initial stops with the three places of articulation and three manners of articulation: p ph b t th d k kh g and nine vowels: i: e: ε: i: ə: a: u: o: ɔ:. The pronunciation of each speaker was recorded three times. The 729 test tokens (81 words x 3 times x 3 speakers) were acoustically analysed with Praat version 4.4.13. The second formant frequency (F2) at the burst of the obstructive airstream (F2 at burst) and the steady state of F2 at the 50% of vowel duration (F2 vowel) were measured in hertz. The locus equation was calculated from a linear equation and then plotted on a linear regression graph. The F2 values at the burst of the nine vowels were plotted along the y axis, while the F2 values in the steady state of the nine vowels were plotted along the x axis, in order to obtain the distinct slopes and y-intercepts of the linear equation which can characterise the three places of articulation. See the results in 4.1.

With regard to the acoustic analysis of vowels, nine short and long counterparts (18 vowels): i-i: e-e: ε-ε: i-i: ə-ə: a-a: u-u: o-o: and ɔ-ɔ:, were used. The duration and formant frequencies (F1, F2) of the nine pairs of vowels were analysed with Praat version 4.5.06 and statistically tested with a t-test (<0.05). The devised word list consists of 90 words, five words for each vowel. Three recordings were made for each of the three speakers. The 810 test tokens (90 words x 3 times x 3 speakers) were acoustically measured. The average values of F1 and F2, S.D., the vowel spaces and square units of the two sets of vowels (short vs. long) and the vowel space of each vowel showing its variation were calculated and then plotted using the three programmes: extractFeatures, Vowel plot and Polygon, written for us by Pathhawi Chanwaiwit. See the results in 4.2.

To test the hypotheses that the fundamental frequencies (F0) of vowels following voiceless initial sonorants (hm hl) are higher than those following voiced initial sonorants (m l) and that the F0 difference is statistically significant, a suitable word list for the tests was devised. The data from each of the three female speakers was recorded three times. The 720 test tokens (80 words x 3 times x 3 speakers) were acoustically measured with Praat version 4.4.04 and the significance of F0 difference was tested with a t-test (<0.05). The investigation of F0 behaviour was carried out in two steps. In the first step, the acoustic characteristics of the two tones (high vs. low) were analysed in order to obtain an overview of the tone shapes, then, the analysis of F0 at every 25% (0%-100%) of the normalised duration was done. In the second step, which was the main objective of our tone investigation, the same 720 test tokens mentioned previously were used again for investigating the fundamental frequencies of vowels following voiceless and voiced initial sonorants. The results of the measurement in real time at every 25 millisecond (msec.) of vowel duration were plotted with Microsoft Excel. See the results in 4.3.

To see clearly how the Yotdowatthana variety of Mal has become tonal, the pitch or F0 patterns occurring in native words having different types of initial consonant, final consonant and vowel height in the other three Mal varieties and a Pray variety were also acoustically studied. The pitch patterns of Tai (Khammueang) loanwords and those of native words (minimal pairs) were also analysed in order to test the hypothesis that tonal evolution in Mal was induced by an external factor, i.e. language contact with Khammueang spoken by the majority of people living in Nan province. Three word lists consisting of 265, 262 and 140 words were devised to suit the multipurposes of our acoustic investigation. For each purpose, three female speakers were used and three recordings for each speaker were made using Adobe Audition



version 2. All together, the F0 values of 6,003 test tokens were acoustically analysed with Praat version 4.5.24 and the significance of the findings was tested with t-test (<0.05). See the results in 4.4.

**4. Results**

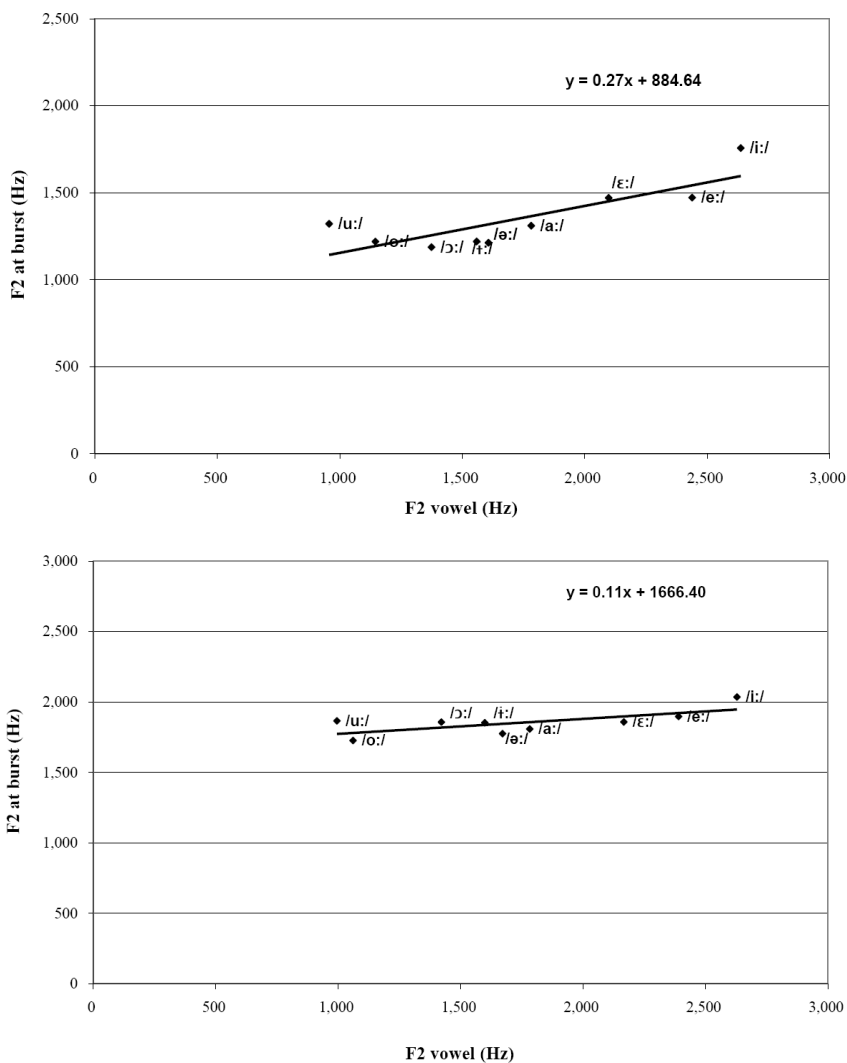
With regard to the acoustic characteristics of Mal consonants, vowels and tones, the following are our findings.

**4.1 Consonant**

The Locus Equation (LE), a method for identifying the places of articulation of initial stops, used by Sussman and Shore (1996) and Modarresi et al., (2005) was applied. The co-articulation of nine initial stops, i.e. the bilabials /p ph b/, alveolars /t th d/ andvelars /k kh g/ and nine long vowels, i.e. /i:/ e:/ ε:/ i:/ ə:/ a:/ u:/ o:/ ɔ:/ in Mal was investigated by means of the LE technique.

The second formant frequency at the burst of vowels (F2 at burst) and in the steady state of the vowels (F2 vowel) was measured and analysed.

The LE values were calculated from a linear equation plotted on a linear regression graph. The values of F2 at burst are shown in the y axis and those of the F2 vowel are shown in the x axis, see examples in *Figure 1*.



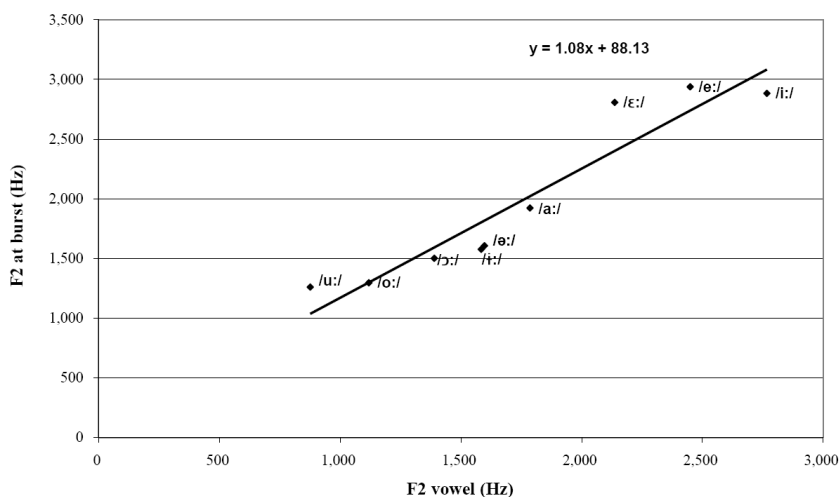


Figure 1 The locus equations of /b d g/

The distinct slopes and y-intercepts of the linear equation can characterise the places of articulation of the three sets of initial stops as shown in Figure 2. However, using the slope value seems to be more effective in identifying the place of articulation than the y-intercept value. The results confirm the hypothesis that the velar stop has the highest slope, while the lowest slope occurs in the alveolar stop. It is also affirmed by the F-test that the difference in the slope values is statistically significant. More details can be found in Huadsiri (2007, 2009).

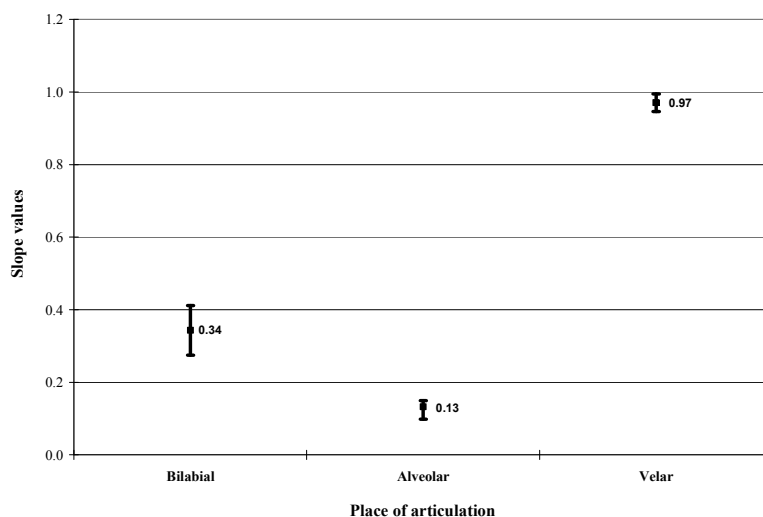


Figure 2 The mean slope values of initial stops

#### 4.2 Vowel

Mal has nine pairs of short and long vowels: /i-i:/, /e-e:/, /ɛ-ɛ:/, /ɨ-ɨ:/, /ə-ə:/, /a-a:/, /u-u:/, /o-o:/ and /ɔ-ɔ:/ (Details can be found in the phonological sketch section.) The duration and formant frequencies (F1 and F2) of these eighteen vowels were acoustically analysed. The average durations of short and long vowels are 212.05 msec. and 482.15 msec. respectively. Broadly speaking, the ratio of duration of long-to-short vowels is about 2:1. The difference between the average duration of short and long vowels is statistically significant. See Figure 3.

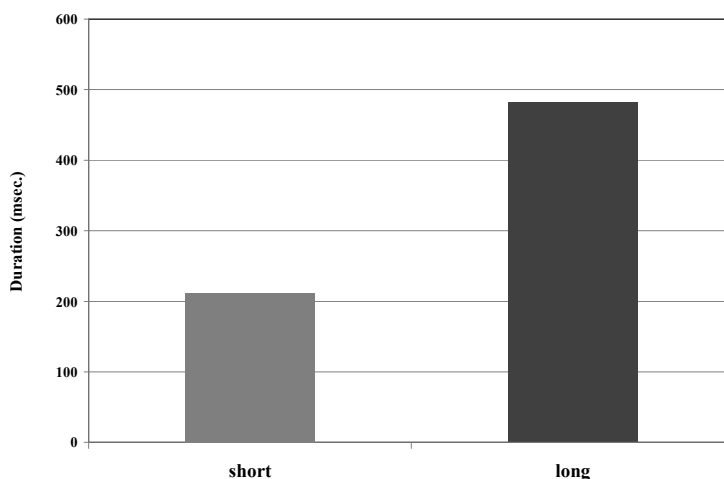


Figure 3 The average durations of the short and long vowels

Generally, the short front vowels /i e ε/ have a higher F1 but a lower F2 than those of their long counterparts /i: e: ε:/. This means that the short front vowels are lower and more centralised. For central vowels, /ɨ/ is lower than /i:/ since /ɨ/ has a higher F1, whereas /ə/ and /a/ are higher and more centralised than /ə:/ and /a:/ due to the fact that /ə/ and /a/ have a lower F1 but a higher F2. With regard to back vowels, the short vowels /u o ɔ/ have a higher F1 than their long counterparts /u: o: ɔ:/. This indicates that the short vowels are lower than the long vowels and that /ɔ/ is more peripheral than /ɔ:/, see Figure 4. The difference between the F1 and F2 of the short vowels and their long counterparts is statistically significant, except for the F1 of the pairs /ə-ə:/ and /a-a:/, and the F2 of the pairs /ε-ε:/, /i-i:/ and /a-a:/.

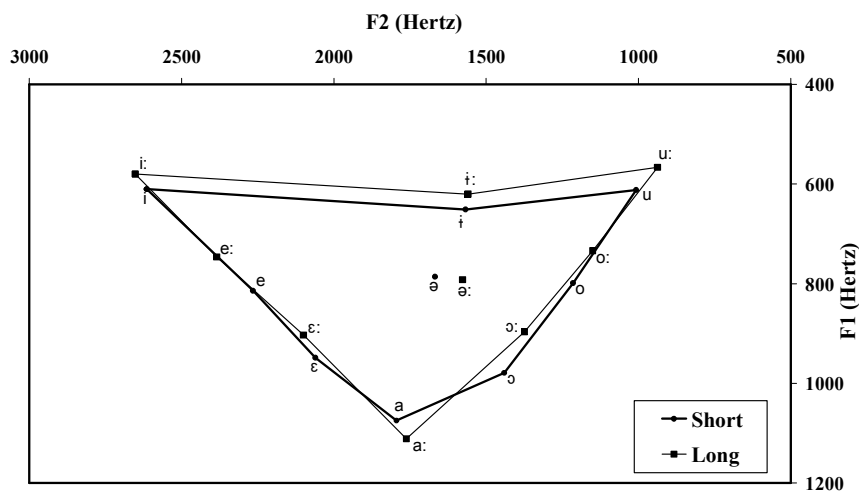


Figure 4 The overall vowel spaces of the short and long vowels

The overall space of the short vowel is smaller (394, 914.92 square units) than that of the long vowel (438, 329.98 square units) as can be seen in Figure 4. As is evident in Figures 5 and 6, each of the short vowels in Mal has a greater variation within its space than that of the long vowels. More details can be found in Phalipat (2007, 2009).

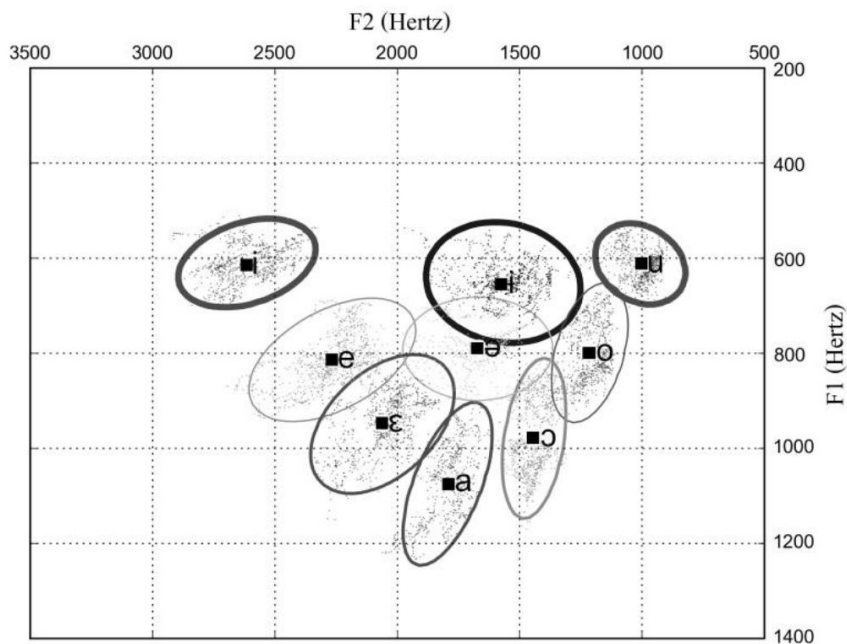


Figure 5 The vowel spaces of the short vowels

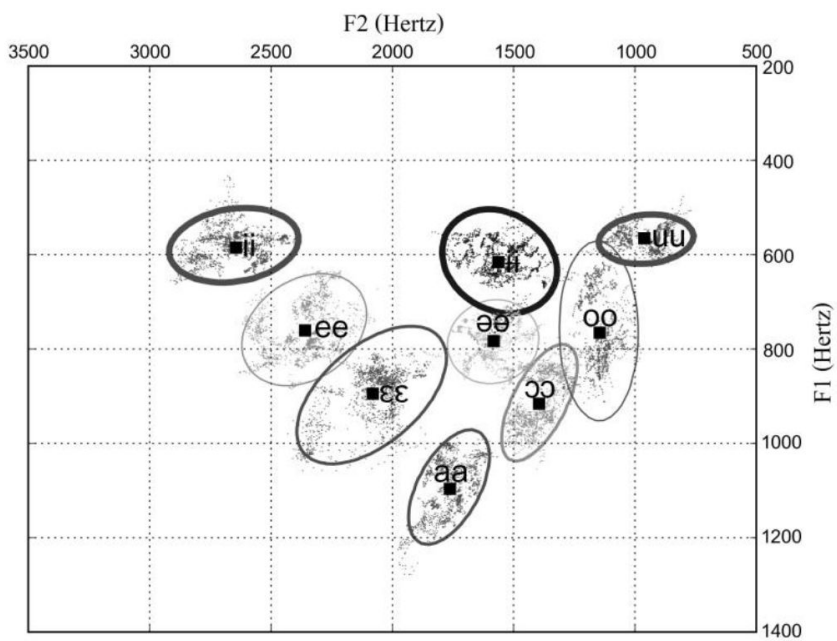
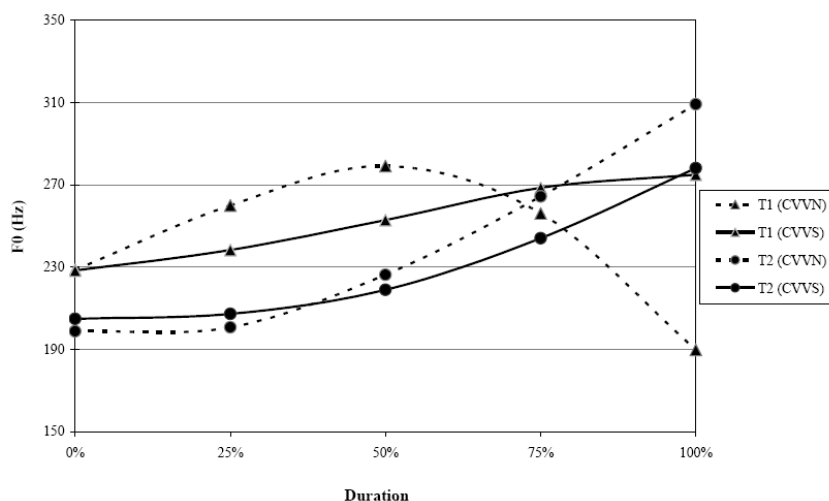


Figure 6 The vowel spaces of the long vowels

### 4.3 Tone

The acoustic characteristics of the two tones, i.e. /High/ and /Low/, were analysed. The fundamental frequency at every 25% (0%-100%) of normalised times was measured. The result of the F0 measurements indicates that the high tone has two phonetic shapes [high-falling] in the non-checked syllable and [high-level] in the checked one, while the low tone has only one characteristic, i.e. [low-rising] in both types of syllable as is shown in *Figure 7*.



*Figure 7* The phonetic characteristics of the two tones

Filbeck (1978) classifies the Lua' (T'in) language into three groups, i.e. A, B and C, using the criteria of lexical and phonological development. He says that Mal B has two tones, the rising tone and the non-rising tone. Some minimal pairs can be found in Filbeck (1972), for example, /kãan/ 'work' (loanword) - /kaan/ 'defeated' (loanword); /cãan/ 'be able' (loanword) - /caan/ 'to hire' (loanword) and so forth. With regard to pitch difference, there are three levels of pitch height: high, mid and low. These three pitch levels correlate with the three degrees of stress: primary stress with high pitch, secondary stress with mid pitch and unstressed with low pitch. There are three kinds of pitch contours: level, falling and rising. The high pitch with falling contour (high-falling) and the mid pitch with falling contour (mid-falling) occur in smooth or non-checked syllables having primary stress and secondary stress, respectively. The high-level pitch occurs in long checked syllables with primary stress, while the mid-level pitch occurs in those with secondary stress. Short checked syllables and syllables having a short vowel with a final /-h/ can have one of the two level pitches, high-level or low-level.

Singnoi (1988) and Jiranathanaporn (1993) agree with Filbeck (1972 and 1978) and confirm that Mal has two "phonemic pitches" or two tones, the rising tone and the non-rising tone. Singnoi (1988) summarises the shapes of the two tones as is shown in Table 1.

<i>Syllable structure</i>	<i>Phonetic characteristics</i>
Open and closed smooth syllable	High-falling, Low rising
Checked syllable	
Short	Level
Long	Level, Low-rising
Syllable with short vowel and /-h/	Level

*Table 1* The phonetic characteristics of the Mal tones  
(Adapted from Singnoi, 1988)

Our acoustic findings, to some extent, agree with Singnoi's description of the phonetic characteristics of the two "phonemic pitches" of the Mal variety spoken in Ban Wangsao in Amphoe Chiang Klang which is her research site. It is possible that in the Mal variety of Ban Yotdoiwatthana studied by our research team and the one described by Singnoi (1988), even though both of them have two tones, the tone shapes or the phonetic realisation of the two tones may be different. Moreover, since the two studies were done about twenty years apart, phonological variation and change would not be abnormal.

#### 4.4 Tonogenesis

Proto-Mal, reconstructed by Filbeck (1978), is non-tonal and most of the modern varieties of Mal are still non-tonal. What is the cause of tone birth in Mal?

The Lua' (T'in) or Mal-Pray were forced to move from the mountains and to settle in Pa Klang Refugee Camp located in the lowland area of Amphoe Pua during the communist infiltration about forty years ago. Later, due to a positive situation in the remote areas of Nan, they returned to the mountains. They became bilingual in Mal and Khammueang during their stay in the lowlands.

Filbeck (1972) states that the two processes of tonal development, i.e. by means of independent innovation and by means of contact with a tonal language (Thai) occurred in Mal (p.111). A number of minimal and near minimal pairs of Thai loanwords found in Mal suggests that the Mal tonal system (rising vs. non-rising) emerged through contact with Thai (p.115). However, the emergence of the rising tone seems to have been the result of an independent innovation (p.116).

In our opinion, the two tones in Mal should be analysed as high vs. low, not falling vs. rising or non-rising vs. rising. Khammueang has six tones: low rising (A1-2), mid-rising (A3-4), mid-level (B1-3), mid-falling (B4), high-level (C1-3) and high-falling (C4). The Mal two tones (high vs. low) are assigned to Khammueang loanwords, no matter what tones they are in the donor language. Khammueang words with mid-rising, high-level and high-falling tones, when borrowed into Mal, tend to have the Mal high tone and vice versa, the ones with low-rising, mid-level and mid falling tones, tend to have the Mal low tone.

Based on the results of our thorough investigation, we claim that language contact with Khammueang, a variety of northern Thai dialect spoken in Nan, was the cause of tone birth in Mal. The high-falling pitch seems to be typical in Mal native words, even in the non-tonal varieties. The tone shape [low-rising] has been borrowed from Khammueang by Mal. It is worthwhile pointing out that a large number of basic Khammueang words have either one of the two rising tones /low-rising/ (A1-2) and /mid-rising/ (A3-4). The rising contour seems to play an important role for Mal bilingual speakers, i.e. it helps induce and enhance the low-rising characteristic of the Mal low tone. The five stages of tonal evolution in Mal can be postulated as follows:

- \*Stage 1: [↘] / [┐] (native words)
- \*Stage 2: [↗] / [┐] (native words and loanwords)
- \*Stage 3: [↘] / [┐] (native words) [↘] / [┐] and [↙] (loanwords)
- \*Stage 4: [↘] / [┐] and [↙] (native words and loanwords)
- \*Stage 5: /High/ [↘] and /Low/ [↙]  
[┐]

At the present stage, Stage 5, the result of the acoustical measurement of the two tones suggests the idea that, at the next stage (Stage 6) sometime in the future, Mal may have three tones: /High/, /Mid/ and /Low/. Perhaps, the third tone could develop from the voicing of voiceless sonorants, e.g. /hm/ > /m/, /hl/ > /l/ and so on, as shown in *Figure 8*. More details can be found in Putthasatien (2007, 2009) and Intajamornrak (2007, 2008, 2009a, 2009b, 2009c).

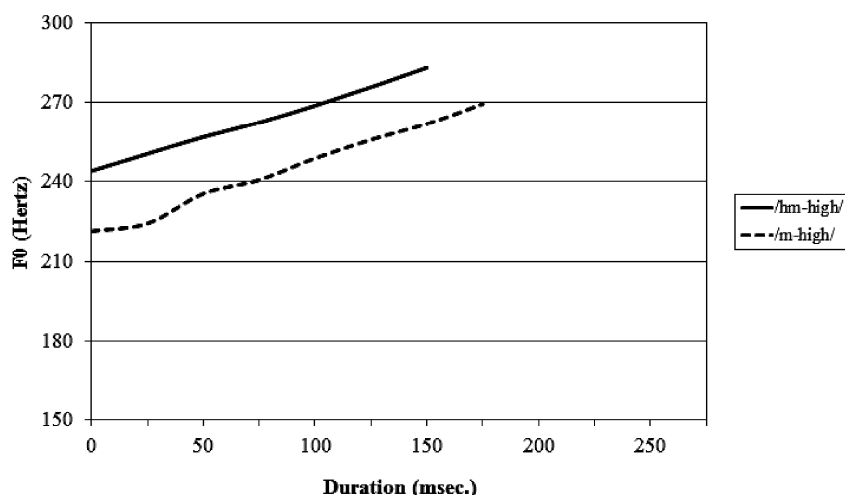


Figure 8 The average F0 values of vowels following voiceless and voiced bilabial nasals in words having high tone

## 5. Conclusion

The results of our acoustic analysis confirm the claim of Sussman et al., that the LE method can help identify the place of articulation of the three types of initial stop: bilabial, alveolar and velar.

The distinction between the average duration of short and long vowels with the ratio 2:1 is statistically significant. This finding confirms the phonological analysis that Mal has a distinctive vowel length. The smaller overall space of the short vowels shows their tendency to be more centralised than their long counterparts. Moreover, the variation within the space of each short vowel is more dispersed. This indicates the fact that the short vowels have more variants than the long vowels.

The birth of the /High/ and /Low/ tones was induced by language and cultural contact between Mal and northern Thai. Five stages of tonal evolution can be postulated. The results of our study seem to suggest that internal factors have had no role to play in Mal tonal evolution from the past to the present.

Vowels following voiceless initial sonorants have a higher F0 value or a higher pitch than those following voiced ones, and the F0 difference is significant. This internal factor could be a cause of tonal development in the future.

## Acknowledgements

We would like to express our sincere gratitude to the Thailand Research Fund (TRF) and Chulalongkorn University for funding our research project. Many thanks go to the local authorities of the districts of Pua and Boklua in Nan province and our Mal friends whose kind assistance and hospitality made our field research possible.

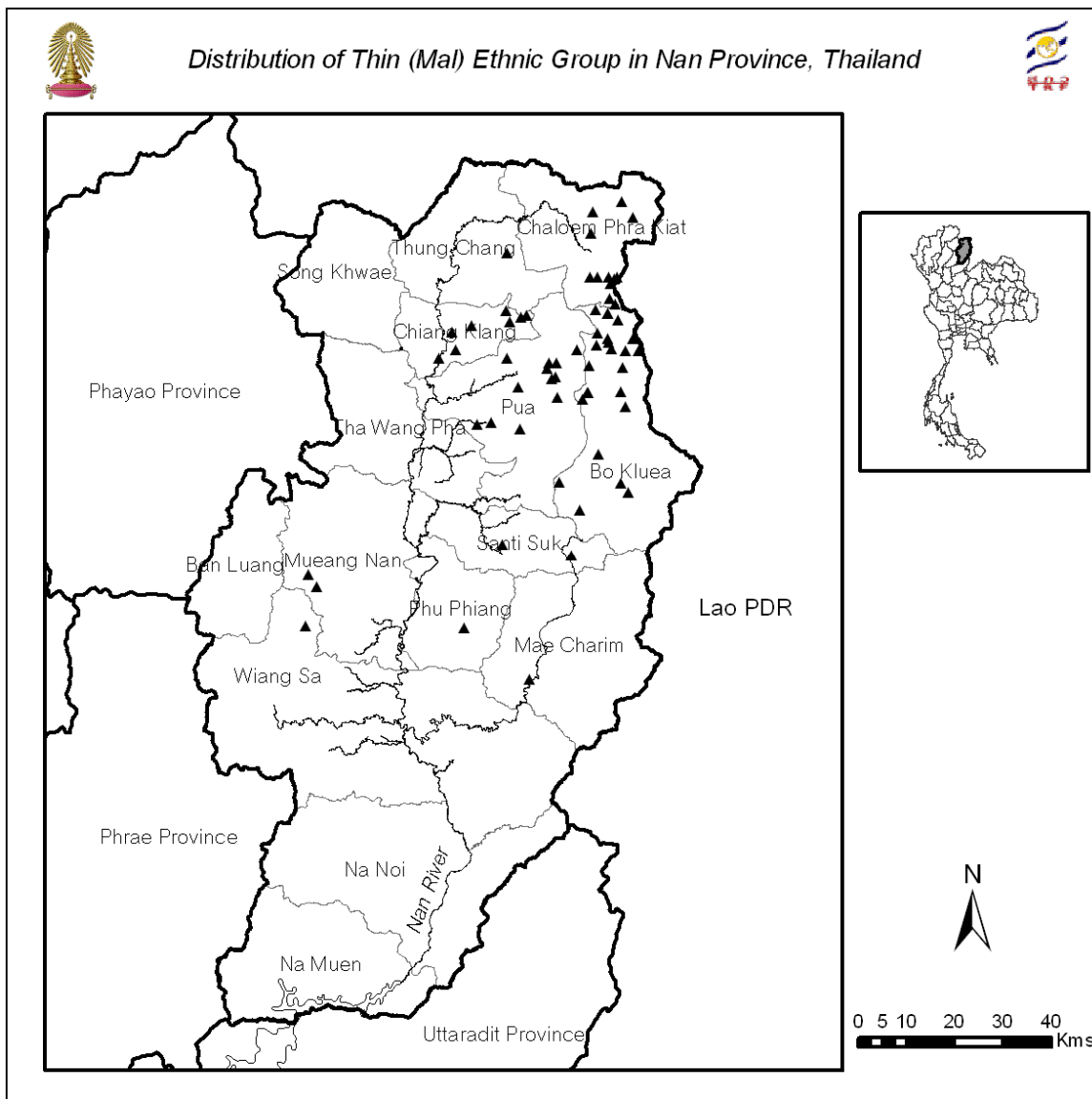
## References

- Filbeck, David. 1972. Tone in a dialect of T'in. *Anthropological linguistics* 14:111-118.
- \_\_\_\_\_. 1976. Mal (Thin). In William A. Smalley (ed.), *Phoneme and orthography: language planning in ten minority languages of Thailand*. Canberra: Australian National University.
- \_\_\_\_\_. 1978. *T'in: a historical study*. Pacific linguistics, series B. no. 49. Canberra: Australian National University.
- \_\_\_\_\_. 1990. Mal phonology revisited. *Mon-Khmer studies* 18-19: 1232-1261.

- Huadsiri, Yupaporn. 2007. Identifying the place of articulation of initial stops with the locus equation method: a case study of Hmong, Mien and Mal. Unpublished MA thesis, Department of Linguistics, Faculty of Arts, Chulalongkorn University. (in Thai)
- \_\_\_\_\_. 2009. An acoustic study of the place of articulation of initial stops in Hmong, Mien and Mal using locus equation (LE) method. In Minegishi, Makoto, Kingkarn Thepkanjana, Wirote Aroonmanakun and Mitsuaki Endo (eds.), *Proceedings of the Chulalongkorn-Japan linguistics symposium*. Tokyo: CbLLE Tokyo University of Foreign Studies. pp. 173-182.
- Intajamornrak, Chommanad. 2007. Tai loanwords in Mal: a minority language of Thailand. Paper presented at The 9<sup>th</sup> Annual Southeast Asian Studies Graduate Student Conference, organized by the Southeast Asian Program, Cornell University. 16 – 18 March, 2007. (To appear in *Mon-Khmer studies* 39)
- \_\_\_\_\_. 2008. Fundamental frequencies of Tai Yuan tones spoken by Lua' Mal speakers at Ta Luang village, Pua district, Nan province. Paper presented at the 41<sup>st</sup> International Conference on Sino-Tibetan Languages and Linguistics, organized by SOAS, University of London, England, 21 – 24 September, 2008. (To appear in *Manusya, special issue* 17)
- \_\_\_\_\_. 2009a. Acoustical investigation of pitch patterns in Ta Luang Mal: implications for tonal evolution. In Minegishi, Makoto, Kingkarn Thepkanjana, Wirote Aroonmanakun and Mitsuaki Endo (eds.), *Proceedings of the Chulalongkorn-Japan linguistics symposium*. Tokyo: CbLLE Tokyo University of Foreign Studies. pp. 203-212.
- \_\_\_\_\_. 2009b. The acoustic characteristics of lexical pitches in four varieties of the Mal language of Nan province: implications for tonal evolution. Paper presented at the RGJ-Ph.D. Congress X, organized by the Royal Golden Jubilee Ph.D. Program, Jomtien Beach Resort Hotel, Chonburi, 3 – 5 April, 2009.
- \_\_\_\_\_. 2009c. On becoming a tonal language of Mal-Pray spoken in Nan province. Unpublished Ph.D. dissertation, Department of Linguistics, Faculty of Arts, Chulalongkorn University. (in Thai)
- Jiranathanaporn, Supatra. 1993. Linguistic geography of minority groups in Nan province: a lexical study of Pray and Mal dialects. Phitsanulok: Naresuan University. (in Thai with English abstract)
- L-Thongkum, Theraphan. 2007. Final fricatives \*-s and \*-h in Proto-T' in lexicon. *Manusya, special issue* 13: 87-93.
- L-Thongkum, Theraphan, Phinnarat Akharawatthanakun, Chommanad Intajamornrak, Kanitha Putthasatien, Yupaporn Huadsiri and Supaporn Phalipat. 2007. Dictionary of the 15 languages spoken in Nan province. Unpublished research report submitted to the Thailand Research Fund (TRF), Bangkok.
- L-Thongkum, Theraphan and Chommanad Intajamornrak. 2009. Tonal evolution induced by language contact: a case study of the T' in (Lua') language of Nan province, Northern Thailand. *Mon-Khmer studies* 38: 57-68.
- Minegishi, Makoto, Kingkarn Thepkanjana, Wirote Aroonmanakun and Mitsuaki Endo (eds.). 2009. *Proceedings of the Chulalongkorn-Japan linguistics symposium*. Tokyo: CbLLE Tokyo University of Foreign Studies.
- Modarresi, Golnaz., Harvey M. Sussman, Bjorn Lindblom and Elizabeth Burlingame. 2005. Locus equation encoding of stop place: revisiting the voicing/VOT issue. *Journal of Phonetics* 33: 101-113.
- Phalipat, Supaporn. 2007. The relationship between duration and formant frequency of vowels: a case study of Hmong, Mien and Mal. Unpublished MA thesis, Department of Linguistics, Faculty of Arts, Chulalongkorn University. (in Thai)
- \_\_\_\_\_. 2009. A comparison of the acoustic characteristics of Hmong, Mien and Mal vowels. In Minegishi, Makoto, Kingkarn Thepkanjana, Wirote Aroonmanakun and Mitsuaki Endo (eds.), *Proceedings of the Chulalongkorn-Japan linguistics symposium*. Tokyo: CbLLE Tokyo University of Foreign Studies. pp. 183-194.
- Putthasatien, Kanitha. 2007. The behavior of fundamental frequencies influenced by voiceless and voiced initial sonorants: a case study of Hmong, Mien and Mal. Unpublished MA thesis, Department of Linguistics, Faculty of Arts, Chulalongkorn University. (in Thai)
- \_\_\_\_\_. 2009. The effect of voiceless and voiced initial sonorants on the fundamental frequencies (pitches) of the following vowels in Hmong, Mien and Mal. In Minegishi, Makoto, Kingkarn Thepkanjana, Wirote Aroonmanakun and Mitsuaki Endo (eds.), *Proceedings of the Chulalongkorn-Japan linguistics symposium*. Tokyo: CbLLE Tokyo University of Foreign Studies. pp. 195-202.
- Singnoi, Unchalee. 1988. A comparative study of Pray and Mal phonology. Unpublished MA thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Sussman, H.M., and Jadine Shore. 1996. Locus equations as phonetic descriptions of consonantal place of articulation. *Perception and Psychophysics* 58: 936-946.



Appendix



## Mon Converging Towards Thai Models: Evidence from a Mon Historical Text<sup>51</sup>

Patrick McCormick

University of Washington in Seattle

Language is central to the transmission of history, so that the actual form of language in a historical text can reveal the contacts and cultural practices of its speakers. A language, in turn, is a mirror of social circumstances and contact, of historical changes, of the history of social interactions and settlement, as well as the vagaries of political affiliation and dominance. In the writing of Southeast Asian histories, the Mons are considered to be one of the earliest groups to have arrived in the Mainland to have developed one of the earliest, most advanced civilizations after having come into contact with Indic civilization and Buddhism. This understanding has deep resonances outside the discipline of history, including in linguistics.

I take here examples of the language found in a Mon historical text to examine the intersection between Mon history and the study of the Mon language to question the common assumption that similarities between Mon and the neighboring languages of Thai and Burmese are always due to primary Mon influence. The evidence of the text suggests that Mon-speaking community in 19<sup>th</sup>-century Siam, surrounded by speakers of Thai, came to replicate Thai patterns of syntax and usage without necessarily borrowing many word or word forms from Thai. Reflecting the social contingencies of a contact situation, these developments are not surprising from a linguistic perspective, but from the perspective of Burmese or Thai history and historiography, interpreting Mon influence as anything but primary, or to see the Mon language as being worked upon by other languages, may be highly unexpected, troubling, or unacceptable.

The texts of the *Rājāvamsa Kathā*<sup>52</sup> are a collection of historical and literary narratives about a different period of Mon history. The longest and best-known component text, for which there are also Thai- and Burmese-language retellings, is the narrative of *Rājādhirāj*.<sup>53</sup> The narrative traces the rise of the illustrious ancestors of Rājādhirāj before turning to trace his career as a military hero, both against other Mon polities and against the Burman court, spanning roughly the late 13<sup>th</sup> to the latter part of the 14<sup>th</sup> centuries AD.<sup>54</sup> “*Rājāvamsa Kathā*” is one of the names that a Siamese Mon monk, Nai Candakantā, gave to a collection of prose historical narratives. He compiled several independent historical texts and printed them in book form in two volumes in 1912 and 1913 at Pāk Lat, a Siamese Mon village now on the outskirts of

---

<sup>51</sup> My thanks to Laurie Sears, Mary Callahan, Christoph Giebel, Charles Keyes, Jacques Leider, Mathias Jenny, and Paul Sidwell for reading and commenting on earlier versions of this chapter. Thanks also to Christian Bauer, Nicoletta Romeo, John Okell, and Justin Watkins for providing me feedback on presenting the data. Special thanks to my Mon informants in Burma and Thailand, who I do not name here because of local sensitivities.

<sup>52</sup> စန္ဒကန္တာ နှစ် “ရာဇဝံသကထာ(စာအုပ်)” ၊ ချုပ်ပက်လတ်၊ ဘာကြိုင်စိင်၊ ၁၉၁၂။ [Candakantā, Nai. *Rājāvamsa Kathā* (2 Vols). Pāk Lat, Siam: 1911-1912, in Mon]

<sup>53</sup> The relationship between the narrative in the three languages is not one of whole translation of one into the other, although individual passages at times may be very similar. I avoid the term “version,” with the implication of an “original,” and instead used the term “retelling,” which opens the possibility of transmission through oral recitation and performance.

<sup>54</sup> The Mon title is ဂလာန်ကြီးမြတ်တော်ဒေတော်သို့ချုပ်တံသာ *Galān Ktāw Bdah Tnow Datow Smim Ḍuñ Hamsā*, meaning, “On the Origin of the *Hamsāvati* Succession of Kings.” Unlike in the Burmese and Thai titles, the name “Rājādhirāj” is not present in the Mon, but I refer to it here as “the Rājādhirāj narrative” for the sake of convenience.

McCormick, Patrick. 2011. “Mon Converging Towards Thai Models.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAALA. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.81-97.

Copyright vested in the author.

Bangkok.<sup>55</sup> The language of the *Rājāvamsa Kathā*, particularly the narrative of *Rājādhirāj*, appears highly “Siamified.”<sup>56</sup> The narrative appears to have been either translated from Thai, or recorded as retold in a Siamified dialect of Mon from what was then Siam. This evidence in the Mon text of the close contact between the Thai and Mon languages has apparently escaped the notice of decades of Mon and Burma scholars, including those Burma Mon<sup>57</sup> scholars living in Thailand who have become fluent and literate in Thai.

I believe that the process of linguistic convergence, in which Siamese Mon has replicated Thai linguistic models, is the most likely explanation for most of the unusual features of the Mon language of the *Rājāvamsa Kathā*. Involved here are languages in contact, which implies speakers in contact who are embedded in a variety of social connections and interactions. There are not many sources on the Mons living in Siam before the 20<sup>th</sup> century. Mons settled scattered throughout what is now Central Thailand. The usual understanding<sup>58</sup> is that these Mons arrived from Burma in waves starting in the 17<sup>th</sup> century, although there is mention of Mon-speaking communities at Ayutthaya in earlier centuries.<sup>59</sup> Mons cultivated rice and worked as potters, while women engaged in trade and men could serve in the Siamese army in ethnically segregated regiments. There were also Mon women at the Thai court. Rice cultivators were the first to assimilate to the Siamese, whereas those living in more isolated communities, particularly if engaged in trade or an occupation that the Siamese did not engage in, tended to maintain their language much longer.<sup>60</sup> There appears to have been regular contact between Mons in both countries, particularly between religious institutions. We can thus establish a social contact in which some of the features of the language of the Siamese Mon developed, reflecting the contingencies of contact outside of Burma.

### Linguistic Evidence for Convergence<sup>61</sup>

I consider here some of the features of the language of the text of *Rājādhirāj* as evidence of extended contact with, and possible translation from, the Thai language. Of particular relevance are examples of Mon prose, using *native* words, reshaped and reworked along the lines of Thai linguistic models. We might say the text speaks Mon in Thai ways. Following this is a set of examples of this process of convergence and replication.

Mon scholars, when asked about the unusual language of the *Rājādhirāj* narrative, either deny that there is anything unusual, or think of the language as archaic or full of errors. Many Burma Mons familiar with the Mon language of Thailand think of it as preserving features now lost in Burma. While the Mon dialects of Thailand do in fact preserve some vocabulary no longer used in daily speech in Burma dialects, the modern spoken varieties display even greater convergence towards Thai models than does the language of the *Rājāvamsa Kathā*.

---

<sup>55</sup> The printed text has been available in Siam and Burma for close to one hundred years and has been the focus of scholarly attention.

<sup>56</sup> I use the term “Siamified” because this process started before there was the modern state of Thailand.

<sup>57</sup> That is, Mons from inside Burma, in opposition to the Siamese or Thai Mons, the ไทยรามัญ *Thai Rāmañ*.

<sup>58</sup> See Robert Halliday, “Immigrations of the Mons into Siam,” in Michael Smithies, ed., *The Mons: Collected Articles from the Journal of the Siam Society* (Bangkok: Siam Society, 1986). Also Suphorn Ocharoen สุภรณ์ โอเจริญ, *มอญในเมืองไทย* (กรุงเทพฯ: สำนักงานกองทุนสนับสนุนการวิจัย, 2541) [Suphorn Ocharoen, *The Mons in Thailand*, 1988]

<sup>59</sup> See Victor Lieberman, *Strange Parallels: Southeast Asia in Global Context, C. 800-1830*, Volume 1: Integration on the Mainland, vol. 1 (New York: Cambridge University Press 2003).

<sup>60</sup> See Brian Foster, *Commerce and Ethnic Differences: The Case of the Mons in Thailand*. (Athens: Ohio University Press, Center for International Studies, Southeast Asia Program, 1982). It is in some of these areas connected to the surrounding areas by road only since the 1980s where Mon is best maintained in Thailand.

<sup>61</sup> The first version of this chapter was conceived of without any knowledge of Bauer’s 1986 article on structural borrowing in Thai Mon. Bauer has taken a very similar approach, examining examples of modern spoken Thai Mon for evidence of contact with Thai. While I have taken the written language of the late 19<sup>th</sup> century as the basis of this examination, Bauer has taken the modern spoken language. Nevertheless, our findings are substantially similar. See Christian Bauer, “Structural Borrowing in Mon: Towards Language Death?” in *Journal of Language and Culture* vol. 6 no. 2 (Bangkok: Mahidol University) 1986.

Yet at the same time, there are patterns and expressions in Siamese Mon that may be only partially meaningful, or misleading, to speakers of Burma Mon, but which become clear when interpreted through the lens of a Thai expression or usage. My evaluations of the naturalness and markedness of the language has been strongly shaped by reading it together with my Mon language tutor, Nai Hawng Htaw.<sup>62</sup>

Below I consider examples of reanalysis, the introduction of new patterns, grammaticalization, and direct borrowings. Particular examples may represent more than one phenomenon. “Reanalysis” is a process in which speakers reinterpret Mon words and word forms to mirror those of Thai. The results of are varied: at times the reanalyzed form will exactly mirror the model language, while at others, hypercorrection or incomplete learning may result in forms that do not match exactly the model language. “Grammaticalization” is a very process in which words are reinterpreted to fulfill grammatical functions.<sup>63</sup> Many of the classifiers widely found in languages of Mainland Southeast Asia have their origins in nouns, and some of the so-called “directionals” or adpositions of Thai and Burmese have their origins in verbs and nouns. I also consider examples of Pāli being used in different ways between the languages of Burma and Thailand, and of a few direct Thai loanwords. I further examine examples of what I have called “translationese,” language that is still so close to the model or source language as to be unintelligible without reference to the model expression.

There is often a surprising congruity between Burma Mon and Burmese on the one hand, and a disjuncture between Burma Mon and Thai on the other. This congruity in fact highlights a conceptual stumbling block: many Mons think that Mon and Thai “should” be similar because of basic typological similarities, such as word order, which is shared between Mon and Thai but not with Burmese, but in fact Burmese and Burma Mon have come to share many patterns and even ordering of sentence elements.

### Formal Possessive Marker

Unlike Burma Mon, Thai and Burmese can formally mark possession with the use of so-called particles. While all three languages can use a strategy of juxtaposition, with either the possessor before the possessed (in Burmese), or possessed before the possessor (in Thai and Mon), this latter is the only strategy available in Burma Mon. In Thai, a common strategy is the grammaticalized use of the word *kʰɔːŋ* or “thing,” which has lost its original meaning, but indicates possession, as in:

Thai <sup>64</sup>	ขนม	ของ	เด็ก
	<i>kʰənǒm</i>	<i>kʰɔːŋ</i>	<i>dèk</i>
	snack	POSS	child

The above sentence, when rendered into Burma Mon, is:

Burma Mon	ကွက်	ကောန်
	<i>kwaiŋ</i>	<i>kon</i> <sup>65</sup>
	snack	son/daughter

In the *Rājādhirāj* narrative, we find a seemingly incongruous use of the word *krɔ̃p*, derived from Sanskrit *dravya*, which means “property, possessions; treasure.” In fact, however, this appears to be a replication and grammaticalization of the Thai use of *kʰɔːŋ*. Because of the possible reading of the Mon term

<sup>62</sup> This is a pseudonym.

<sup>63</sup> Following the definition of Payne, this is a process in which certain words take on a grammatical function, thus losing their original meaning. See Thomas E. Payne, *Describing Morphosyntax: A Guide for Field Linguists* (Cambridge, UK: Oxford University Press, 1997), p. 239 and p. 262.

<sup>64</sup> According to Huffman, Khmer has a usage exactly paralleling the Thai example. See Huffman “Thai and Cambodian—a Case of Syntactic Borrowing?” *Journal of American Oriental Society* 93, no. 4 (1973): 489-509.

<sup>65</sup> Burma Mons whom I have consulted reject the usage of this word as a possessive marker, although there is evidence that Burma Mons who have settled on the Thai side of the border have already replicated this Thai pattern. See Jenny 2010 (forthcoming). The Maticchon dictionary of Siamese Mon does in fact list *krɔ̃p* as having the Thai grammaticalized usage. See คณะกรรมการจัดทำพจนานุกรมมอญ-ไทย *พจนานุกรมมอญ-ไทย ฉบับมอญสยาม* (กรุงเทพฯ: มติชน, 2548), p. 143. [Committee for the Preparation of the Mon-Thai Dictionary, *Mon-Thai Dictionary, for Thai Mon*, (Bangkok: Maticchon, 2005)]

as “property,” in some contexts the incongruity is not as apparent as in others. In all the examples below, the normal Burma Mon was to express possession would be to drop the use of *kr̥p*.

p 301<sup>66</sup>

နရာမိလ္လာ	သို	ဗိုလ်ဗး	ကြော်ခတာ	စိင်ချေ	ၼု
Nar̥amilla	with	general-troop	family	elephant-horse	<b>THING</b>

ကေ။	အနိုင်အနိုင်	ကော်	တိတ်	ကျဲ	ပုံ	ပြင်	ကေ
cakao?, <sup>67</sup>	ʔənqin.əkʰaiŋ	cək	tət	klɛŋ	pəŋoʔ	pəraŋ	cakaoʔ
body,	strenuously	march	go.out	come	direct	side	body

တံဇိုင်	တု	ရာဇာဓိရာဇ်	ပုန်။
təm.cəŋ	təlaʔ	rəcəcətʰiʔrət	plɔn.
foot	lord`	Rājādhirāj	AGAIN

*Narāmilla, together with his troops, their families, and calvary, forced themselves on to march straight back to the presence of Lord Rājādhirāj.*

In this above example, to a Burma Mon, it sounds as though Narāmilla is also taking along his possessions or treasures.

The next example is noteworthy for reasons other than just the use of *kr̥p*. This excerpt is from a scene in which the speech of children or the mad is interpreted for portents of the future. Baññā Noy, the name of Rājādhirāj when he was younger, has sent some of his followers to hear the news at Hamsāvati, where some of his men have gone to a gate of the city to listen. We find words and phrases that are highly reminiscent of Thai, but cannot necessarily be put back into Thai word-for-word. This example, as in many that follow, exemplifies the linguistic slippage that can mark the language of *Rājādhirāj*.

We find the presence above of an otherwise unknown Sanskrit word *rājaśāstra*, here *rəcəsət*, following the Thai pronunciation *ราชศาสตร์*<sup>68</sup> *rā:tcʰsà:t*. In Thai usage, this is a law promulgated by the king in accordance with the principles of the *Dharmaśāstra*. Following the interpretational lens of Thai, at first glance the phrase *neh mə tɔh cəŋək* implies, “the person who becomes/will become great.” A more likely interpretation would be the common Thai expression, *ผู้เป็นใหญ่* *pʰi: pen yáy*, a common description of someone who is a “superior.”

pg 194

ချိန် အာ ဝ	တြင် ထသံင်ကိုင်	ဂံ	မိင်	ဗြ	ၼု
cʰɔŋ ʔa	pəðəa tərəŋ həsəŋ.kaiŋ	kɔʔ	mɔŋ	pəruʔ	<b>kr̥p</b>
when go	in door T.K.	GET	hear	sound	<b>THING</b>

ဒါရက	မ	ဟို
tɛəŋkəkaʔ	mə	hɔm
child	REL	say

<sup>66</sup> Unless otherwise indicated, page numbers are all from the Rājādhirāj section of Phra Candakantā’s *Rājāvamsa Kathā*.

<sup>67</sup> The three lines provide script, phonetic transcription, and gloss. Words that are capitalized have been grammaticalized. Abbreviations: DEIC *deictic*; DUR *durative* or progressive; FIN abbreviation for FINISH, acting as a conjunction; FOC for *focus* particle; HON = honorific; INT *intensifier*; IRR and REAL for *irrealis* and *realis*; NEG *negative*; REL *relative*; TOP *topic*; VOC = vocative; a period between two words reflects that they are equivalent to one word in the other language; = indicates that one unit contains a combination of fused meaningful units. Numbers followed by S or P indicate pronouns: 1P is the *first person plural*. In keeping with Burma Mon tradition, phonetic renderings reflect reading, formal pronunciations rather than colloquial.

<sup>68</sup> This word has come through Thai because of the treatment of the final consonants: if the Sanskrit had descended through Mon, it would likely not result in a final –t, but a final –s, realized as –h.

ဂး	သို့	ဇော	ကျင်	ဗဒ်	ရာသောတ်	ဂုံ	တင်စွေဟ်
<b>kəh</b>	səmoij	cəno:k	klɛŋ	hətɕh	rɛːcəsət	kɔ?	tə.cəneh
<b>SAY</b>	sming <sup>69</sup>	great	come	establish	rājaśastra	GET	attack

ဒိုင်တဲ	က	ညး	မ	ဒ	ဇော်၊	ညး	မ
tɕŋ.toəkao?	ŋɛh	mə	tɕh	cəno:k	ŋɛh	mə	
accept	with	person	REL	be	big.	person	REL

ဒ်	ဇော်	မ	ဂုံ	ဇ္ဇး။
tɕh	cəno:k	mə	kɔ?	cəŋh.
be	be.big	REL	GET	be.victorious

When they went to the Tasawng Kaing gate, they heard the sound of child saying, "The great sming who brings about the rājaśastra will attack and repulse the superior man. The superior man will be victorious."

For purposes of comparison, the following is an example of the use of *krɔp* following the normal Burma Mon meaning of "possession, treasure," in this case the latter, which reflects more closely the original Sanskrit meaning. This example is taken from the သုပတ်ရာဇာဝင်ဒတောင်သို့ *Slapat Rājāvaṅ Datow Smin*<sup>70</sup> or "(Treatise) on the History of the Kingly Lineage," sometimes referred to in English as "The History of Pegu."

p 17 (Slapat Rājāvaṅ Datow Smin)

သို့	အိန်	ဂ	ညာတ	ကေတ်	ဒမံ	ဓာတ်	ကျာ	ပိ	တေ	တဲ၊
səmoij	?in	kɔ?	ŋət	ket	mɔŋ	tʰət	kyaik	pəe?	te?	tuy,
King	Indra	GET	see	take <sup>71</sup>	LIVE	relic	Buddha	three	those	FIN

ပ	ပူဇသက	ဒြပ်	ရတ်	ဟ	တိ	ကို	ဗွတ်	တို့၊	ရ။
pa?	paocɛə.səka	<b>krɔp</b>	rət	hɔ?	təm	kə	pəno:t	təwɔ	ra?
do	pūjā	<b>treasure</b>	jewel	not	know	with	extent	limit	FOC

Indra, catching sight of the three Buddha relics, made limitless offerings of treasures and jewels.

**Quotation Particle**

A feature of Burma Mon syntax that sets it apart both from both Thai and Burmese is that it lacks a particle marking direct speech and other such complements. In Thai, the usual particle is *wā*: a grammaticalized verb meaning "say." The fact that *wā*: is used even with verbs meaning "say, tell," indicates that the original meaning has been "bleached" in the process of grammaticalization. In Burma Mon, a natural way for speech to be quoted is to put the speech first, followed by a verb of saying or hearing. We also commonly find examples of such verbs both preceding and following a quote.

We find in the text of *Rājādhiraṅ* repeated examples of *kəh*, a Literary Mon<sup>72</sup> verb meaning "say, tell," being used exactly following the Thai pattern of *wā*:. Consultation both with Burma Mon speakers and

<sup>69</sup> Title of ambiguous meaning, from "king" through "local leader" or "headman."

<sup>70</sup> From *Journal of the Burma Research Society* vol. 31 no.1 (Rangoon: Burma Research Society, 1923), p. 15. This text is often erroneously referred to as *Slapat Rajavaṅ Datow Smin Roṅ*, but the final word is an assertive focus particle. Following tradition, I have given titles in Indic transliteration, not transcription.

<sup>71</sup> According to Mathias Jenny (personal communication, March 2010), the usage of *V+ket* may be grammaticalized, in a manner not paralleled in either Thai or Burmese, to mean "to do something for one's own benefit," in opposition to *V+kn*, which means "to do something for someone else."

<sup>72</sup> As in many languages of Southeast Asia, there may be a sharp divide in the syntax, vocabulary, usage, and even phonology between written and spoken styles of the language. The Burmese linguistic tradition recognizes this distinction, with "spoken" language standing in contrast to "literary" language of writing and formal speech. Burma Mons have varying degrees of control over the literary form of the language, depending on education. Many areas of Literary Mon vocabulary have fallen out of use in the past century.

with contemporaneous texts from inside Burma has revealed that this usage appears to be unknown in spoken Burma Mon. A complication to using the evidence of this expression is that *kəh* is the second element in the Literary Mon phrase *həm kəh*, an elaborate expression meaning “to say” used in formal language.<sup>73</sup> Instances of uses outside of this literary collocation may provide stronger evidence for the Mon replication of the Thai model, as in:

p294

တ	့ညး	ရာဇာဓိရာဇ်	တ	ဂး	ပျီန	ဗာ	စေ့
təla?	ŋəh	rəʔcəʔətʰi?rət	təm	kəh	pənan	həmɣə	cih
lord		Rājādhirāj	know	SAY	army	burman	descend

ရီ	ချင်	ပြန်။
rɔ̃m	dʒɪŋ	prɔ̃n.
surround	town	Pye.

Lord Rājādhirāj knew **that** the Burmese army had come down and surrounded Prome.

The above example follows Thai word order, whereas in Burma Mon, the subordinate clause would precede the main verb, at least in Literary Mon. We may note that this Mon construction of placing the subordinate clause before the first is suspiciously similar to the natural word order of Burmese. A Burma Mon rendering of the above might be:

ပရူ	ပျီန ဗာ	စေ့ ရီ	ချင်	ပြန်
parao	pənan	həmɣə cih	rɔ̃m	dʒɪŋ prɔ̃n
circumstances	army	burma descend	surround	town Prome

ဂံ	တ့ညး	ရာဇာဓိရာဇ်	ဂံ	တီ	ကေတ်။
kəh,	təla?ŋəh	rəʔcəʔətʰi?rət	kəʔ	təm	ket.
TOP, lord	Rājādhirāj	GET	know	TAKE	

A final example from the *Slapat Rājāvañ Datow Smin* displays a more common Burma Mon strategy of enclosing speech with multiple, non-grammaticalized uses of *kəh*, which is highly reminiscent of the parataxis of oral texts. At times, it can be difficult to determine which instance of *kəh* goes with which instance of speech.

p 21 (Slapat Rājāvañ Datow Smin)

သုတ်ဗြ	မိင်	တဲ	သ့း	ရ။	ဒ်	ရ	သတ်	ဂံ
səmot.prəʔ	moɪŋ	tuy	sə.kəh	ra?	təh	ra?	saik	kəh
young.woman	hear	FIN	IRR.say	FOC	be	FOC	manner	TOP

အဲ	သ့း	လမျီု	ပဲ	ရတ	ပိ	ရ။	အဲ	ဒင်
?ə	səlah	ləmyɔ̃m	pədoə	rət	pəe?	ra?	?ə	tɔ̃ŋ
1S	free	in	life	jewel	three	FOC.	1S	accept

ဂစိုတ်	ရ	သုတ်ဗြ	ဂး။
kill	FOC	young.woman	say.
həcət	ra?	səmot.prəʔ	kəh.

Having heard that, the young woman said, “In that case, I donate my life to the Three Jewels. I will be killed,” said the young woman.

<sup>73</sup> An “elaborate expression,” found throughout the region, are made of compounds of words with the same or similar meaning to give a feeling of “weightiness” to the occasion or subject matter.

### Interrogative Strategies

There are many features of syntax that Mon shares with Burmese. Based on inscriptional evidence, it appears that sentence-final question markers have existed in both languages (in unrelated forms) for centuries. Burmese and Mon share a common interrogative pattern which stands in contrast to the Thai and Khmer patterns.<sup>74</sup> Burmese and Mon usually indicate yes-no questions with an “absolute” question particle, Burmese *là* and Mon *ha*.<sup>75</sup> Questions involving relative words, however, are marked with different particles, usually *lê* in Burmese and *rao* in Mon.<sup>76</sup> Thai, in contrast, uses a variety of expressions to ask absolute questions, including *ไหม* *māi*, often with an expectation of confirmation. Another strategy is *หรือ* *rǔi*, often followed by *เปล่า* *plàw* “or not?,” which has less an expectation of confirmation. The Thai-Khmer and the Burmese-Burma Mon strategies are not directly equivalent and so do not always correlate with each other.

It is therefore striking when we find the seemingly incongruous use of Mon sentence-final *ha* in mid-sentence or clause-initially. Sentences such as the following have led Burma Mon scholars to wonder whether there was some kind of misprint or elision in the text:

p 201

အဲ	ချိန်	လဝ်	လမျီု	ကု	မိ	ရ။
ʔoə	c <sup>h</sup> ɔn <sup>77</sup>	lə	ləmyɜm	kaoʔ	mjɪʔ	raʔ
1s	give.over	PUT	life	with	mother	FOC

လုဟ	ဝံ	မိ	ကေတ်တာ	လမျီု	အဲ	ဟာ ဟာ
ləmuh	wuʔ	mjɪʔ	ket.na	ləmyɜm	ʔoə	ha ha
now	this	mother	take	life	1S	INT INT

ဗူး	လမျီု	အဲ။
pəlh <sup>78</sup>	ləmyɜm	ʔoə
free	life	1S

(Baññā Noy, having read a letter, is now speaking to his wetnurse) *I have entrusted my life to you, my mother. Now will you take or free it?*

This sentence highlights the continued slippage between Thai Mon and Thai. The doubling of the *ha* may occur because the speakers of the replicating language do not precisely match the model form, or the form recorded is in flux before being established in place of the old usage. In the first Thai-Mon sentence, the first instance of *ha* may be a continuation of the old interrogative pattern, while the second instance has taken on the meaning of “or” following the Thai model. While sounding unnatural to the Burma Mon ear, the above sentence does not actually sound acceptable when translated directly into Thai:

เดี้ยว นี้	แม่	เอา	ชีวิต	ฉัน	หรือ	ปล่อย	ชีวิต	ฉัน*
dǎw.ní:	mā:	aw	c <sup>h</sup> i:wít	c <sup>h</sup> ǎn	rǔi	plò:y	c <sup>h</sup> i:wít	c <sup>h</sup> ǎn
Now	mother	take	life	1S	or	free	life	1S

The Thai word *rǔi* by itself means “or” and can be used to indicate alternatives. The Burmese and Burma Mon strategies to indicate alternatives, however, differs from the Thai. One natural strategy is to ask

<sup>74</sup> See Huffman 1973.

<sup>75</sup> Contemporary spoken Thai Mon often dispenses with the use of *rao* altogether, thus aligning the Thai Mon usage with Thai, although such dropping is not unknown in Burma Mon.

<sup>76</sup> Burmese has alternate forms for both absolute and relative questions, though these two are the most common.

<sup>77</sup> This word is the result of a typical sound change in Thai Mon – the Burma Mon equivalent is *p<sup>h</sup>ynn*, made up of causative *pə* + *cɜn* “donate.” Such sound changes sometimes have the effect of rendering common words into something exotic-sounding to Burma Mons, who suspect that Thai Mon has preserved “archaic” words.

<sup>78</sup> Despite the conjunct spelling, for the causative meaning, the initial must be realized as a sesquisyllable. Such spellings are common enough and may reflect the uncertainty of the scribes when dealing with conjunct versus non-conjunct forms. Central Thai script has no conjuncts, whereas most other Indic scripts do.



two parallel questions, each stating one alternative. The above second sentence, when rendered into Burma Mon, would be:

လှူဟ် မိ	ကေတ် လမျီု	အဲ	ဟာ၊	ဗလး	လမျီု	အဲ	ဟာ။
ləmɥh mi?	ket ləmyɜm	ʔoə	ha,	pələh	ləmyɜm	ʔoə	ha
now	mothertake	life	1S	INT,	free	life	1S INT

The last example sentence begins Thai and ends Burmese. The Thai honorific prefix **พร** *phrá?* appearing in the title of Min Gaung’s name, while at the end is the phrase *hə?seəŋ pəwa?*, highly reminiscent of the Burmese phrase မဟုတ်သည့်အလုပ် *mə hou? θj ʔəloo?*, a “bad deed” or “something inappropriate.” Of interest here is the last clause, beginning Thai-fashion with Mon *ha*. Nai Hawng Htaw wanted to move this *ha* to the end of the previous sentence, thinking that it might form a rhetorical question or might be a typesetting error. At least as likely an interpretation is to view the phrasing in light of Thai syntax:

p393

အဒိ	တူညး	ဖြ	မင်	င်	ကို	နွံ	လိက်	နင်	ထံ	ဂ်
hətɔh	təla?ɲəh	p <sup>h</sup> ra?	mɛaŋ	kɔŋ	kɔ	nɯm	lɔik	nɛəŋ	no?	kɔh
fact	lord	Phra	Mañ	Kamñ	GIVE	exist	letter	bring	this	TOP

ကသပ်	ဂြိုဟ်	နက်	လောန်၊	ဆဂး	ဒ်	ဇကရာဇ်	နွံ	ကို
kəsəp	krɔh	nɛək	lɔn.	c <sup>h</sup> əkɛh	tɔh	ekarət	nɯm	kɔ
thought	harm	deep	very.	but	be	king	exist	with

ဝေ သစ္စာ	ဟာ	ချပ်	အလိက်စွဲ	ဟ	သေင်	ပွ။
thə sətʃəa?	ha	c <sup>h</sup> əp	əlit.cənaɲ	hɜ?	seəŋ	pəwa?
law loyalty	INT	think	strategy	not	correct	deed.

*Bayin Min Gaung means great harm having this letter sent. But he must be loyal, being the King. Or is this an ill strategy?*

**Politeness Strategies**

Out of the three languages under consideration here, Thai has the most elaborated speech levels. The language has relied heavily on foreign, particularly Khmer and Pāli-Sanskrit vocabulary, taking many loanwords from these languages to provide special sets of terms used in reference to royalty and monks. In both Thai and Burmese, there are a variety of pronouns and particles available to indicate social distance between speaker and hearer or speaker and referent, or to indicated attitudes of the speaker towards the statement. Mon has the least elaborated honorific and pronoun system, with few distinctions made according to status or gender. In the texts of recent centuries, we find a few simple pronoun and vocabulary differences that index honorifics. Burma Mon has strategies to indicate politeness in making requests, suggestions, or commands. As in Burmese, these usually take the form of sentence-final particles and verbs used as “softening” strategies.

In Thai, a high-frequency strategy is the use of the verb **ขอ** *k<sup>h</sup>ɔː*; literally *ask for, request* that has been grammaticalized as a way to indicate politeness, often without a directly stated subject. A direct translation of Thai *k<sup>h</sup>ɔː* into Mon is *ʔat*, which in Burma Mon is used with nouns and is not used as a politeness strategy. Clauses introduced with *ʔat* are found in *Rājādhirāj*, and I observed that this usage was again confusing for Burma Mons, striking in their seeming unnaturalness to the Burma Mon speaker. The Burma Mon equivalent of this sentence would drop the *ʔat*.

p 199

အာတ	ကို	ဗို	လဗး	သရဲ	စိင်	ချေ
ʔat	kɔ	pɔ	ləpɛh	səray	coiŋ	c <sup>h</sup> eh
request	give	officer	soldier	hero	elephant	horse

ကို အဲဒိုက် ညံင် ရံ ဂပ် ညိ။  
 ko ?oə.doik ɲəŋ rum kəp ɲi?  
 GIVE 1S.vassal so.that enough suitable LITTLE.  
*Give me enough soldiers and mounts to me (so as to be suitable).*

The next sentence is another example of one which is redolent of Thai, but does not translate well directly back into Thai, especially since, as in the previous example, the combination of *?at ko* suggests Thai *ขอให้* *kʰə̌: hây*, “may (I)...” “let (me)...”. At the same, as with the above example under interrogative strategies, a direct back-translation leads to something unacceptable to Thai speakers.<sup>79</sup>

p 215  
**အာတ** ကို ဒိုက် သွံ ဒိုင် ကာ ခေဝ်ရော် ဂုဏ် နဲ  
**?at** ko doik səkʰə̌? tʰɲ ka kʰlɛ.rək kɯn nəy  
 request give vassal IRR.get accept affair repay gratitude lord  
 ဒို့ အိုတ် လမျီု ညိ။  
 tʰə̌ ?ot ləmyɜm ɲi?  
 until end life LITTLE  
*Let me repay my gratitude to you, my Lord, until the end of my life.*

To what extent was the above sentence acceptable to Thai Mons of the time? Did they find it immediately interpretable? Burma Mons have difficulty in understanding it and reject it on the grounds of vagueness and unnaturalness, with the result that such sentences are a frequent obstacle to reading the Mon-language *Rājādhirāj*. Here is a more typical Burma Mon example, from *Wottu Mi Doñ Keh Htaw*, “The Story of Golden-Nib Mi Dong” from 19<sup>th</sup>-century Burma.<sup>80</sup> The following features paratactic repetition of the verb “request.”

p 61 (Mi Doñ Keh Htaw)  
 န နဲ မ ပွဲ မြဲသ် ဝံ ဗောဓိသတ **အာတ** တ်  
 nɛə nɔə mə pətom məɛə.saik wu? pʰtʰisot **?at** tao  
 with way REL begin manner this, Bodhisattva **request** DUR  
 အခေါင် ပဲ သိုင် မ သိုက် လို့လှော် သွတ်  
 ?əkʰoŋ pəðə səmoij mɛə? sɔm.kɔ ləɲim.lənaik səmot  
 permisson in lord father together mourn true  
 ဗုတ် ပ တဲ ကွတ် **အာတ်** တ် လို့ ရ။  
 pənot pɔe? ɲəə kənat **?at** tao ləwɜ ra?  
 extent three day extent, **request** DUR alwaysFOC

*In this manner, beginning in this way, the Bodhisattva kept mournfully asking for permission [to go] from his father for three days, he kept asking [sic].*

A further politeness strategy, although not as widely found, involves the Mon word for *help*, although sometimes the Mon rendering of the Thai word is found. Thai speakers can make requests for the listener to do something on behalf of the speaker by placing *ช่วย* *cʰuay* “help” at the beginning of a sentence. Such a strategy does not exist in either Mon or Burmese, and so the presence of verbs meaning *help* at the beginning of sentences in the *Rājādhirāj* narrative is conspicuous. There are two strategies for rendering *cʰuay* into Mon: translating it into its Mon equivalent *rɜm bəiŋ*, or using *cʰə̌ə*, the closest phonetic rendering

<sup>79</sup> A direct translation such as *ขอให้ข้าจะได้รับการตอบแทนบุญคุณเจ้าจนหมดชีวิตหน่อย* does not appear acceptable, while *ข้าขอตอบแทนบุญคุณเจ้าจนหมดลมหายใจ* may be more so.  
<sup>80</sup> လောကသီရိ၊ နဲ “ဝတ္ထုမိညောင်ကျော်ထင်”၊ ညိုင်မတ်မို့၊ သုဝဏ္ဏပုံနှိပ်တိုက်။ Nai Lokasiri, ed., *Wotthu Mi Doñ Keh Htaw* [*The Story of Golden-Nib Mi Dong*], (Mawlamyaing: Thuwunna Press, ND).

of the Thai into Mon. The idea of “help” appears weakened or removed from the Thai expression, so that Siamese Mon speakers, when attempting to render the Thai into Mon may have preferred to introduce the Thai expression. Burma Mon readers tend to interpret this usage literally, in the sense of “give me help and ....”

The first example uses both the Mon and the Thai expressions, and is closed with another Thai-like usage of Mon *kəm* “also,” paralleling the positioning in Thai of *dūay* “also” in requests:

p 198

မိ	မိၤငံ	ရှဲ	ဟိၣ်	ကိုၣ်	ညိၣ်	ကိုၣ်။
mi?	paing	chəʊ	həm	kr	ni?	kəm
mother	<b>HELP HELP</b>		say	GIVE	LITTLE	also

*Mother, please tell him for me.*

The second sentence is more elaborate and features both native (sentence-final *ni?*) and borrowed strategies:

p 209

တလဂုဏ	ဖျဲ	မေတ္တာ	ရှဲ	အာ	ဟိၣ်	ကိုၣ်	ဗညာ	နဲ	ညိၣ်။
təla?kʉn	p <sup>h</sup> yeh	mɛt̩ta	chəʊ	?a	həm	kr	pəŋɛə	nɔə	ni?
monk,	throw	<i>mettā</i>	<b>HELP</b>	go	say	GIVE	Baññā	Noy	LITTLE.

*Your reverence, please<sup>81</sup> go tell Baññā Noy for me.*

**Grammaticalization**

Having already considered the cases of turning *thing* into a possessive marker and *say* into a quotation marker, I consider here two other cases of grammaticalization. The first is the use of the word *arrive* to indicate a complement of verbs of thinking and feeling, and the other is the use of the word *search* to indicate direction of motion or action towards humans. I have chosen these two for their relative frequency in the text of *Rājādhirāj*, but also because their interpretation may be deceptive to the Burma Mon reader. Depending on the context, these usages appear either superfluous in a sentence, or seem to have a literal meaning that is in contrast to their intended meaning.

**Arrive**

In Thai, the verb *t<sup>h</sup>ūŋ* can mean *arrive* but has a function of also marking direction towards the endpoint of something. At the same time, it can also be used with more abstract concepts and means something like *about, concerning*. The use of *t<sup>h</sup>ūŋ* is particularly common with the verbs like *think of, remember* and *say, talk about*. In the following sentence, we have a clear example of the Thai usage together with one of a few direct Thai loanwords: the term *k<sup>h</sup>it* means *think, plan, consider, wonder, be of the opinion*. In Burma Mon, the various meanings of *k<sup>h</sup>it* tend to be broken up into disparate expressions, some borrowed from Burmese. Because *k<sup>h</sup>it* is used quite frequently in Thai, it may be that speakers borrowed the Thai word into Mon to fill a perceived gap in their own language, and is not some kind of incomplete translation:

<sup>81</sup> The collocation *throw mettā* is used in both Thai (แม่เมตตา *p<sup>h</sup>è: mé:ttā:*) and Mon to mean *bless*, but here appears to be used in the sense of “please,” or perhaps “do the favor of...” This usage is described for modern spoken Thai Mon, where it is used as a polite expression, not having to do with monks:

အာတ်	ဖျဲ	မေတ္တာ	ကောန်	ကလောန်	ဏံ	ကိုၣ်	အဲၣ်ခိုက်	ညိၣ်။
?at	p <sup>h</sup> yeh	mɛt̩ta	klon	kəlon	na?	kr	?əə.dəik	ni?
request	throw	<i>mettā</i>	do	work	this	GIVE	1S.vassal	LITTLE.

“Please do this work for me.”

In Thai Mon, the pronoun *?əə.dəik* is not used only when talking to monks but as a general polite 1S pronoun. See ละออล เป็นเจริญ, *ระบบไวยากรณ์มอญ* (มหาวิทยาลัยศิลปากร, 2526), p. 138. [La-or Paencaroen. *Mon Grammatical Systems*. MA Thesis, Silapakorn University, 1983].

p 199

ကို	ဗညာ	နဲ	ဆိတ်	နိပ်	ကျေ	တ	မေ	ထ
ko	paŋɰə	noə	kʰít	ɲp	cəkaoʔ	təlaʔ	mɛ	thao
GIVE	Baññā	Noy	think	ARRIVE	body	Lord	Me	Thao

*Let Baññā Noy think of Lady Me Thao; let Baññā Noy miss Lady Me Thao.*

To the Burma Mon, the above sentence may sound something like, “Let Baññā Noy think arriving at Tala Me Thao herself,” although Burma Mons are not familiar with the Thai loanword. A Burma Mon rendering of this sentence would be rather different, something on the order of the following, although the Mon rendering cannot cover the same range of meanings as the Thai:

ကို	ဗညာ	နဲ	ဂံ	ပိသ္တာ	တ	မေ	ထ
ko	paŋɰə	noə	kɰʔ	pɰh.səna	təlaʔ	mɛ	thao
GIVE	Baññā	Noy <sup>82</sup>	GET	miss	Lord	Me	Thao

*Let Baññā Noy miss Lady Me Thao.*

### Search

In Thai, another word used in a similar way is the *หา* *hǎ*: which has been grammaticalized to indicate directionality towards humans. The verb literally means *look for*, but in many contexts, the idea of actual searching is absent. This usage is often coupled with the verbs *come* and *go*.<sup>83</sup> In Burma Mon, as in Burmese, there are ways of indicating directionality towards a human goal. In the following sentence, we find a clear example of Thai Mon having replicated the Thai pattern, with the native Mon *klay*, *search*, *look for* being pressed into service. For a Burma Mon speaker, this usage is discordant because of the Burma Mon speaker’s desire to interpret the meaning literally:

p 210

ကို	ဗညာ	နဲ	တိတ်	ကျဲ	နဲ	အဲ	ရ။
ko	paŋɰə	noə	tet	klɰŋ	klɰy	ʔoə	raʔ
GIVE	Baññā	Noy	exit	come	search	1S	FOC

*Let Baññā Noy come to me, Have Baññā Noy come to me.*

In Burma Mon, this sentence could be rendered:

ကို	ဗညာ	နဲ	တိတ်	ကျဲ	ဇေရင်	အဲ	ရ။
ko	paŋɰə	noə	tet	klɰŋ	cəɰɰŋ	ʔoə	raʔ
GIVE	Baññā	Noy	exit	come	vicinity	1S	FOC

*Let Baññā Noy come to me, Have Baññā Noy come to me.*

Here is an example of *klay* from *Wotthu Mi Doñ Keh Htaw*, which shows a more typically Burma Mon usage of the word:

p 115 (Mi Doñ Keh Htaw)

လု	အ	ကံ	ဥဒ္ဓုန်	သိုင်	တုံ
lup	ʔa	klɔʔ	ʔucɰn	səmoɪŋ	tuy,
enter	go	orchard	garden	king	FIN

<sup>82</sup> When rendered in Mon, Thai names tend to be rendered following the Indic values of their letters. Modern Mon has neutralized the distinction between graphic OY and AY, both now *oə*, so that spellings can be interchangeable. I suspect this is a representation of Thai *น้อย* *nó:y*, “small” or “junior.”

<sup>83</sup> There is some similarity between the Thai and English in such sentences as, “Come find me in the library tomorrow,” where there is not an implication of an actual search.

ကွဲ	စွဲ	စ	သတ်ဆ	ရိုဟ်	ဇုတ်	ကြွက်	ပွတ်။
klay	bəh	ʃiəʔ	sət.cʰuʔ	rʒh	cʊt	krək	pəŋəh.
search.for	pluck	eat	fruit	root	bone	mango	jackfruit

*They went into the garden of the king, searched for, plucked and ate the measly mangoes and jackfruits.*

**Changing “see” into “think.”**

A final set of expressions, admittedly chosen out of a very wide range of choices, have to do with a Thai usage covering a broad range of meanings. These include ideas of thinking, planning, considering, and agreeing, all based on on the Thai verb *hěn*, or *see* in English. The usage takes on a variety of meanings depending on the complementizer it is coupled with. These include *hěn* “see; plan; consider”; *hěn wā*: “be of the opinion; think”; and *hěn dūay* “agree, be of an accord.”

The narrative of *Rājādhirāj* features examples of the direct equivalents of each of these expressions, at most only marginally intelligible in Burma Mon. The Burma Mon equivalent of Thai *hěn* is *ṇat*, which can only mean “see” and does not match the semantic range of the Thai. The expression *hěn wā*., literally *see say*, the second element is the grammaticalized quotation particle discussed above. The Mon rendering of *ṇat kəh* appears meaningless to Burma Mon speakers. The reader frequently meets these usages throughout *Rājādhirāj*, but because of their idiomatic, extended meanings, they posed significant hurdles towards comprehension for the Burma Mon speakers that I consulted. This was especially the case with the Mon *ṇat km*, rendering *hěn duāy*, which sounds like *see also*. The expression *see say* is one of the most frequent in the text, in large sections occurring on nearly every page.

p 202

ညာတ်	ဂး	ဒး	အိုင်ဒုင်	လဝ်	ကွာ	ရ။
ṇat	kəh	təh	ʔəiŋ.tʃŋ	lɔ	kla	raʔ
see	SAY	HIT	endure	set-down	before	FOC

*(I) think we will have to endure it for the time being.*

In Burma Mon, this might be rendered as:

ဒး	အိုင်ဒုင်	လဝ်	ကွာ	ရ။	အဲ	ထိုင်။
təh	ʔəiŋ.tʃŋ	lɔ	kla	raʔ,	ʔəə	tʰeəŋ
HIT	endure	set-down	before	FOC,	1S	think

We may note that the above Burma Mon sentence not only employs a Mon rendering of the Burmese ထိုင် *tʰi*, English *think, have an opinion* but even replicates Burmese word order by placing the quotation before the verb, as is typical of SOV word-order languages like Burmese.

**Miscellaneous Syntactic Examples**

The following are several less common examples of expressions that follow a Thai model, including changes in word order and calquing. For example, Thai and Siamese Mon agree in the ordering of collocations, with the head following the verb. Burma Mon and Burmese, however, have the head before the verb. The order verb-head may be prototypical for SVO languages like Thai and Mon, so it is possible that Burma Mon has changed under pressure from Burmese, or that Siamese Mon has preserved an older feature, or redeveloped it through contact with Thai. In the narrative, we find apparently “reversed” expressions:

အိုတ်	စိုတ်
ʔət	cət
use.up	mind
<i>be fed up, be uninterested</i>	

This is rendered in Burma Mon as *cvt ʔvt*.<sup>84</sup> The apparent reversal is frequently found in collocations involving the metaphor of “mind,” the figurative seat of emotions in Mon, Thai, and Burmese. Here we have a Thai expression calqued into Mon:

p 204

လူ	ချုပ်	အောက်	စိတ်	ပဲ	ဗိုလ်	ပို	ရ။
ləpaʔ	čʰəp	ʔon	cnt	pəðə	pɜ.pənan	poy	raʔ
do.not	think	few	mind	in	forces	3P	FOC

*Do not feel insignificant because of our forces.*

To a Burma Mon, this sentence sounds something like, “Do not feel sad because of our forces,” an interpretation that does not fit easily with the context. Rather, the sentence appears to be a rendering of the Thai collocation **น้อยใจ** *nó.y cay* literally, *few mind*, meaning rather *feel small, feel inferior*. This interpretation fits better with the context. A Burma Mon rendering might be the following, which following common Literary Mon usage, reverses the two clauses in a Burmese-like manner:

ပဲ	ဗိုလ်	ပို	ဂံ	လ	စိတ်	အောက်	ရ။
pəðə	pɜ.pənan	poy	kəh,	ləpaʔ	cnt	ɔt	raʔ
in	forces	3P	TOP,	do.not	mind	small	FOC

**Calque of the Thai Collocation **ขอโทษ** *kʰə: tʰə:t***

The Thai expression to express *ask for forgiveness*, also used as a polite expression meaning *I’m sorry*, is made up the collocation of **ขอ** *kʰə:*, *request* as discussed above, and the Sanskrit **โทษ** *tʰə:t*, meaning *crime, offense; sin, fault*. Burma Mon speakers use strategies meaning absolution and forgiveness which are not exactly equivalent to the Thai. There is no Burma Mon expression directly equivalent to the Thai, so that the following sentence sounds to a Burma Mon speaker as though a sin is being requested. It makes sense only when interpreted through a Thai lens:

p 391

အောက်	ဒုတ်	သို့	ဗိုလ်	ဒပ်။
ʔat	tʰh	səmoij	pɜ.nay	təp
request	fault	<i>smiñ</i>	leader	army

*....(they went in and) asked forgiveness of the leader of the army*

**Thai Pāli**

The central position of Pāli both to religion and intellectual thought in Theravada Buddhist Southeast Asia is well-known, yet the specifics between how languages of the region make use of Pāli vocabulary varies in detail. Just as the usage of Latin and Greek words varies between the languages of Europe, so do the meaning and usage of individual Pāli and Sanskrit words vary between Thai, Burmese, and Mon. Compared with Thai and Khmer, there are fewer Sanskrit loanwords in the languages of Burma, reflecting the narrower role of Sanskrit learning compared to Pāli, at least in recent centuries. Throughout the texts of the *Rājāvamsa Kathā*, there are a number of Pāli and Sanskrit words commonly used in Thai but unknown or used with a different meaning in Burma Mon and Burmese.<sup>85</sup> With few exceptions, these words are rendered in a Pāli or Palicized form in Mon. Below is a short list of some of the more notable examples. I have given the meaning of the Sanskritic vocabulary according to the Thai usage. The Mon pronunciations are often educated guesses, as the words would be pronounced in Literary Mon.

<sup>84</sup> This particular expression has no direct Thai equivalent. It is, however, exactly parallel to the Burmese expression of the same meaning, စိတ်ကုန် *sei? kōv*.

<sup>85</sup> As far as I have learned, the usage of Pāli and Sanskrit does not vary much between Burmese and Mon, with the exception of some old loanwords into Mon.

Mon <sup>86</sup>	Pāli/Sanskrit	Thai Usage	Comment
တေဇနုဘာဝ <i>teceṇaṇup<sup>h</sup>ṅaḥwṅa?</i>	tejānubhāva	เดชาณาภาพ <i>king's power, influence</i>	No such specialized usage appears to exist.
ပဒေသ <i>paṭesa?</i>	pradeśa	ประเทศ <i>country</i>	The Pāli word means <i>variety</i> in Burma. In Thai Mon, this term replaces ချင် <i>dṣi</i> “city” in the collocation ကွန်ချင် <i>kwan dṣi</i> “country,” creating ကွန်ပဒေသ <i>kwan paṭesa?</i>
ပယောဇနု <i>paḃaocanṅa?</i>	prayojana (Skt)	ประโยชน์ <i>use, benefit</i>	The Pāli ဂုဏ် <i>kun</i> from Pāli <i>guṇa</i> is used with similar meaning.
ပဉ္စာ <i>paṅṅa</i>	pañhā	ปัญหา <i>problem</i>	Mon can use a Sanskritic form, ပြဿနာ <i>prattṅṅa</i> or ပြဿနာ <i>praṅṅa, praṅṅa</i> .
ဝိဝိ <i>wi<sup>h</sup>?<sup>h</sup>i</i>	vidhī	วิธี <i>way, method; customary usage</i>	The Thai usage covers two meanings not connected in Mon. For <i>method</i> , နဲ့ <i>no</i> or နဲ့ဝ် <i>no</i> is used. For <i>custom, usage</i> , အခိုင်ကွဲ <i>ṅak<sup>h</sup>aik kṅa</i> can be used.
သမယု <i>samoaya?</i>	samaya	สมัย <i>age, period</i>	ခေတ် <i>k<sup>h</sup>et</i> derived from Pāli <i>khetta</i> is used.
သဝါဏ္ဏခေါတ် <i>sawannak<sup>h</sup>ot</i>	svargata (Skt)	สวรรคต <i>“go to heaven,” ie, die</i>	Instead of “Pālifying” the spelling to something like - * <i>saggagata</i> , the Thai pronunciation has been rendered in Mon
သောဝဏီ <i>saowṅwe</i>	sauvanīya (Skt)	เสาวนีย์ <i>queen's command</i>	No such specialized usage appears to exist.
သွေဟ် <i>sneh</i>	sneha (Skt)	เสน่ห์ <i>charm, appeal</i>	A Mon word of this shape means <i>landing place</i> . There is no exact equivalent to the Thai term.
အနုညာတ <i>ṅanunata?</i>	anuññāta	อนุญาต <i>(give) permission</i>	In Burma, this rare word means <i>nonentity</i> . In Thai, it is used as a verb, for which Mon uses ကိုအခေါင် <i>ko ṅak<sup>h</sup>oṅ</i> . <sup>87</sup>
က္ခိဗလ <i>ṅitt<sup>h</sup>iḃalṅa?</i>	iddhibala	อิทธิพล <i>power, force</i>	Mon uses various words inherited from Pāli, including အာဏာ <i>ṅana</i> , ဩဇာ <i>ṅaocṅa</i> , or က္ကသ <i>ṅiṅsa?</i> conn. Pāli <i>īsa</i> ; or native အဝို <i>ṅawṅ</i> .
အာသဲ <i>ṅasoa</i>	āśaya (Skt)	อาศัย <i> dwell, live, reside</i>	A variety of native expressions, မိုင် <i>mṅi</i> , literary တ် <i>tao</i> , or ပဒတန် <i>paṅ ḥatṅn</i> or ပဒတ် <i>paṅ ḥatao</i> , can all be used.

Table 1: Mon Words Following the Thai Use of Pāli-Sanskrit

### Thai Loanwords

The introduction of loanwords – actual material – from one language into another is not necessarily as common as might be expected in contact situations, especially where there is long-term bilingualism. Interestingly, other than the Pāli words considered above, there appear to be few direct loanwords from Thai. Some of the notable example of those that do occur include *เอง* *e:ṅ* meaning *(one)self*, which is borrowed into Mon as အိုင်, probably *ṅoiṅ*. We find a curious usage of the word *ဝဲ* *boə*, which in Mon means *bean, pea*, but from the context appears to be the Thai word *บัว* *bua*, *rice-gall midge*.

<sup>86</sup> Some of the Thai Mon pronunciations are conjectural.

<sup>87</sup> The second element is derived from an earlier pronunciation of Burmese, now အခိုင် *ṅak<sup>h</sup>wi* but once something like *ṅak<sup>h</sup>wṅi*.

p 219

သို့	မ	ရူ	ဒုံ	ဆ	ကောန်	အမတ်။	တုပ်	ပွာ	ညှင်
səmoɪŋ	mɛəʔ	rū	tɔ̃h	ɕʰaʔ	kon	ʔəmat	tup	pəma	ŋɔ̃ŋ
<i>Smiñ</i>	Father	Rū	be	only	child	minister	be.like	example	like

သောင်	ဝဲ။	ပွဲ	သောင်	ပဝ်	ပြိုင်	ကို	သို့	သွပ်
səneəŋ	ʔəw.	plah	səneəŋ	pə	praɪŋ	kɔ̃	səmoɪŋ	səpə
wing	<i>boa</i>	extend	wing	fly	compete	with	king	garuda

လှောင်	ကော	မှသို့တ်	ကို	မူး	ဂုံ	ကော	မာန်	က	ဟာ။
lɔ̃ŋ	kləʔ	masəmit	kɔ̃m	məkɛh	kɔ̃h	kləʔ	mən	kɔ̃m	ha.
sail	cross	sea	also	if	TOP,	cross	can	also	INT.

*Smiñ Ma Rū is only the son of a minister, like the wings of the boa. If he were to extend his wings and fly in competition with Garuda to cross the sea, would he really be able to?*

The sentence is meaningful only when we take the word *boa* with its Thai meaning of a small insect. The following sentence illustrates an isolated instance of the use of the word မှဲ့ *mai*, which in Mon means *corrupted, rotten*, rendering the sentence unintelligible:

p 382

မဝ	ရဲ	ကယျာဝ်ဆွာ	မိင်	သတ်	ဂုံ	တု	ပိုယ	ဇက်	ပုန်
mɛəŋ	rəy	kəyɔ.cʰəwa	mɔɪŋ	saik	kɔ̃h	tuy	poy	cɛək	pənan
Mañ	Rāy	Kayawchwa	hear	way	that	FIN	3P	march	army

ကျင်	အလောန်	ထံ	ဂုံ	မှဲ	ဂး	ဗဇီ	စိင်	ကု
klɔ̃ŋ	ʔələŋ	noʔ	kɔ̃h	məy	kɛh	pəcɔ̃	coɪŋ	kaəʔ
come	occasion	this	DEIC	intend	SAY	cause.fight	elephant	with

သို့	မည်ညှိ။
səmoɪŋ	mɔ̃ŋ ŋɔ̃ʔ
<i>smiñ</i>	Mon LITTLE.

Taking the word at its Thai meaning, หมายถึง *mă:y* in English, mean, intend, have the meaning, expect, the whole sentence makes sense: Min Ye Kyawzwa heard what [his astrologer] had to say [and said,] ‘Our coming this time was meant to be to engage with the Mon sming in elephant battle.’

**Translation or Retelling?**

As some of the examples already considered have suggested, many sentences in the Mon-language *Rājādhirāj* do not make for good Literary Mon as understood in Burma. Yet these sentences do not form smooth Thai when translated word-for-word. An assumption of many scholars when dealing with literary or historical narratives found in more than one language tradition is that one must be a translation from the other. A survey of the narrative in *Rājādhirāj* in its Burmese, Mon, and Thai retellings does not suggest a complete direct translation between any of them, or at least, between any of the texts extant today. The overall grammatical and syntactic coherence of the Thai Mon of the *Rājādhirāj* narrative, especially when viewed through the lens of convergence, suggests that much of it has been recorded in a Thai Mon dialect reflecting a noticeable degree of convergence towards Thai models.<sup>88</sup> Nevertheless, there are sentences and expressions

<sup>88</sup> The modern Thai Mon dialects display an even higher degree of convergence towards Thai. In addition to Bauer’s work cited above, see also for example จำปา เยื้องเจริญ และ จำลอง สารพัดนึก, *แบบเรียนภาษามอญ*, (กรุงเทพฯ: ภาควิชาภาษาตะวันออก คณะโบราณ มหาวิทยาลัยศิลปากร, 2528 [1985]) [Champa Yeuangcharoen & Chamlong Saraphatneuk, *Learning Mon* (Bangkok: Southeast Asian Language Department, Faculty of Antiquities, Silpakorn University, 1985)]; and พระมหาจรรยาจอกสมุทร, *การศึกษากการใช้คำและเรียนคำในมอญ* (มหาวิทยาลัยศิลปากร, 2539 [1996]) [Phramaha Jaroon Chork Samud, *Study of the Use of Words in Mon*, (Bangkok: Silpakorn University, 1996)].



in the *Rājādhirāj* that appear to be unintelligible. These could be examples of direct, though incomplete, translation; examples of grammatical developments in Thai Mon that have neither a Burma Mon analogue nor a Thai model; or examples of expressions that were intelligible at a certain time in a particular context, but are not any longer.

The first example is quite startling, particularly because there is a similar, though not exact, sentence in the main, most common Thai-language retelling of *Rājādhirāj*. The meaning of the Thai sentence is, “It is up to you, my Lord,” although the Thai Mon sentence is wholly unintelligible. The heart of the expression is the formal Thai *สุดแต่* *sùt tæ̀:*, meaning *be up to (one’s wish), depend on (someone’s wish)*. The literal meaning of the expression is *ending from*, from which it has taken on its extended meaning. The first element, *sùt*, has been used in the Mon sentence. Perhaps the expression as it stands had become common in Siamese Mon at the time. Alternately, the Mon speakers who recorded this text may have been unfamiliar with the Thai and were unable to render an adequate translation. In this passage, one person is telling another that he will work on her behalf, to which she replies:

p 198  
 သုတ် ဆ တု အဲ ရှ။  
*sut* c<sup>h</sup>a? t̄əla? ʔəə ra?  
*sut* only lord 1S FOC

The Mon sentence, which appears in any case incomplete, is wholly unintelligible without reference to the Thai. The native Mon *sut* means “silk” or “sūtra,” suggesting something like “Only silk as much as My Lord.” The following Thai sentence, however, makes the apparently intended meaning clear:

p 92 Thai-language *Rājādhirāj* of Hon<sup>89</sup>  
 สุดแต่ พระเจ้า พี่ คิด เกิด  
*sut tæ̀:* p<sup>h</sup>rǎʔ.cāw p<sup>h</sup>i: k<sup>h</sup>ít t<sup>h</sup>ɔ̄t  
 up.to Lord older.sibling think EXHORTIVE  
*As you like; It’s up to you, my Lord.*

A page later is another obscure passage, one that does not appear to have a direct Thai equivalent in Hon’s printing of *Rājādhirāj* in Thai:

p 199  
 သိတ် အာ ဥပ္ပို ညးဂံ ရာ အိုတ် ရှ။  
*k<sup>h</sup>ít* ʔa ʔuppay ɲə̄h.kə̄h rao ʔət̄ ra?  
 think go stratagem who even use.up FOC

Parsing the sentence to render *ɲə̄h.kə̄h* as “who,” the usual meaning of the collocation, does little to elucidate the meaning when put into Thai, which is ungrammatical. If we reparse the expression *ɲə̄h kə̄h*, which could then mean *that person*, another meaning is possible:

คิด จะ ไป ุบาย เขา นั้น ไม่ หมด แล้ว  
*k<sup>h</sup>ít* cà pay ʔùba:y k<sup>h</sup>ǎw nán mây mòt lǎ:w  
 think IRR go stratagem 3S that, not use.up PERF

This second Thai sentence then means something like, *There is no end to his thinking of stratagems*, which fits the context. In any case, the sentence is not wholly acceptable either when read as it is in Mon or when it is recast in Thai.

---

<sup>89</sup> See เจ้าพระยาพระคลัง, *ราชวิธีราช ของ เจ้าพระยาพระคลัง (หน)* (กรุงเทพฯ: สำนักพิมพ์ศิลปปบบรรณาคารพิมพ์จำหน่าย, 2423 [1880], [Chao Phraya Phrakhlang (Hon), *Rājādhirāj of Chao Phraya Phrakhlang (Hon)*, (Bangkok: Silapabannakhan Press, 2001 reprint).

### Conclusion: Intersections Between Historiography and Linguistics

Local Mon scholars must often work to make the *Rājādhirāj* narrative meaningful. This work is necessary in part because the reality of modern-day Burma is far removed from the context in which the texts were created. More widely, ideas about history have influenced lines of linguistic inquiry. A persistent theme is that the Mons and speakers of Austro-Asiatic languages are “originary” because they were among the first peoples to appear in the historical record. Many scholars take this understanding of the Mons as the underlying explanation for similarities between Mon and other languages. The logic of this argument may be hard to escape – if the Mon language of the Mon *Rājādhirāj* has been influenced by Thai, then Thai itself was influenced by Mon. The risk is that such an assumption may render other interpretations invisible. Much work has yet to be done to determine the extents and limits of Mon-Khmer primacy. Directionality and the contingencies in the convergence between Mon and Burmese for at least the past millennium remains to be carefully studied. A survey of Mon literature since the 17<sup>th</sup> century suggests an increasing harmonization from Mon towards Burmese. Recognizing the contingencies of specific situations may go far to render visible the linguistic and social context that the language of the *Rājādhirāj* narrative reveals—a widespread bilingualism among Mon speakers surrounded by Thai speakers for several centuries, resulting in the convergence of Siamese Mon towards Thai models.

### REFERENCES

- Aikhenvald, Alexandra Y. “Grammars in Contact: A Cross-Linguistic Perspective,” in A.Y. Aikhenvald & R.M.W. Dixon, eds. *Grammars in Contact: A Cross-Linguistic Typology*. Oxford University Press, 2007: pp. 1-66.
- . “Mechanisms of Change in Areal Diffusion: New Morphology and Language Contact.” *Journal of Linguistics* vol. 39, (2003): pp. 1-29.
- Aikhenvald, A.Y. & R.M.W. Dixon. “Dependencies between Grammatical Systems.” *Language* vol. 74, no. 1 (1998): pp. 56-80.
- . “Introduction,” in A.Y. Aikhenvald & R.M.W. Dixon, eds. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford University Press, 2001: pp. 1-26.
- Bauer, Christian. “Morphology and Syntax of Spoken Mon.” PhD Dissertation, School of Oriental and African Studies, University of London, 1982.
- . “Structural Borrowing in Mon. Towards Language Death?” *Journal of Language and Culture* 6, no. 2 (1986).
- Halliday, Robert. *A Mon-English Dictionary*. Rangoon: Mon Cultural Section, Ministry of Union Culture, Government of the Union of Burma, 1955.
- Heine, Bernd & Tania Kuteva. *Language Contact and Grammatical Change*. Cambridge University Press, 2005.
- Huffman, Franklin. “Burmese Mon, Thai Mon, and Nyah Kur: A Synchronic Comparison,” *Mon-Khmer Studies* vol. 16-17, (1990): pp. 31-84.
- . “Thai and Cambodian - a Case of Syntactic Borrowing?” *Journal of American Oriental Society* vol. 93, no. 4 (1973): pp. 489-509.
- Jenny, Mathias. *The Verb System of Mon*. Arbeiten Des Seminars Für Allgemeine Sprachwissenschaft. Zürich: Universität Zürich, 2005.
- Matras, Yaron. *Language Contact, Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press, 2009.
- McCormick, Patrick A. *Mon Histories: Between Translation and Retelling*. PhD Dissertation, History Department, University of Washington, 2010.
- Shorto, Harry L. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press, 1962.

## **A Study of Language Use and Literacy Practices to Inform Local Language Literature Development among Khmu in Thailand**

**Michelle M. Miller**  
Payap University Linguistics Institute  
SIL International

**Timothy M. Miller**  
SIL International

### **Introduction to language development in Khmu**

The Khmu are a major ethnolinguistic group of northern Southeast Asia. The largest proportion of the estimated 700,000 Khmu resides in Laos, and smaller communities are found in Vietnam, Thailand, and China (Gordon 2005; National statistics center of the LAO PDR 2007). Khmu is a Mon-Khmer language and encompasses a complex web of ethnic and dialect groupings (Chazee 1999; Suwilai 1987).

Khmu was first documented in the late 19<sup>th</sup> century by French scholars in Luang Prabang, Laos (Preisig 1994). Systematic linguistic analysis, documentation, and efforts toward language development began among Khmu in Laos in the mid-1950's with the work of William Smalley (Smalley 1961), and have been continued by numerous other scholars (Lindell 1974; Lindell *et al* 1981; Svantesson *et al* 1994; Suwilai 1987). A small body of Khmu literature has been produced using both Lao- and Roman- script orthographies<sup>90</sup> that are based on an 'Eastern' Khmu dialect.<sup>91</sup> (Preisig 1990; Suksavang *et al* 1994; Suksavang and Preisig 1998).

During the past 30 years there has also been extensive linguistic research conducted among Khmu speakers in Thailand (Suwilai 1987, 1998, 2001; Cholthissa 1988) and there have been experimental efforts to apply this research toward the development orthographies (Suwilai 1990; Supatra 1988; Cooper 1998, 1999), but currently no writing system has community endorsement nor is in wide use among Khmu communities in Thailand for the development of Khmu language literature.

### **Research purpose: Informing Khmu language development in Thailand**

In 2008, we (Timothy and Michelle Miller) were in contact with members of Khmu communities in Thailand who expressed interest in developing an orthography representative of the dialects spoken in Thailand. There already seemed to be adequate linguistic research upon which to base orthography development, but there was no current documentation of sociolinguistic or literacy data to inform both the process and products of a language development effort.

Consequently, in July and August of 2008, we undertook a study of a cluster of Khmu communities in Nan province, Thailand to gather data in four areas:

---

<sup>90</sup> An 'orthography' is "a standardized system for writing a particular language. The notion includes a prescribed system of spelling and punctuation" (Benson and Kosonen 2010:2 cf Crystal 1999:244).

<sup>91</sup> Also referred to as 'Southern', particularly in earlier literature. This refers to the Khmu Uu (or Khmu Cueang) and related groups who use the negative particle 'am' and are found in Luang Phrabang, Phongsaly, Xiengkhouang, Houaphan, Vientiane, Bolikhamxay and Sayabouly.

---

Miller, Michelle M. and Timothy M. Miller. 2011. "A Study of Language Use and Literacy Practices to Inform Local Language Literature Development among Khmu in Thailand." In Srichampa and Sidwell (eds.) *Austroasiatic Studies: papers from ICAALA. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.98-111.  
Copyright vested in the author.

1. Language and culture maintenance and shift
  - To inform strategies for safeguarding Khmu language and culture
2. Standard Thai literacy skills and practices
  - To inform instructional approaches and the design of relevant materials
3. Attitudes toward language development
  - To inform community participation in a potential language development process
4. Perceptions of educators working in Khmu communities regarding the adequacy of Khmu children's Thai language skills for successful classroom learning
  - To inform plans for a potential local language instructional approach in the school setting

We acknowledge that the scope of our study was limited to only five of the estimated 30 Khmu villages in Thailand, but from informal contact we have had with other Khmu villages we believe that what we learned is representative of at least some of the other communities, and we hope that this study will stimulate a broadened research and documentation effort.

In the remainder of this paper we will briefly introduce the research setting, summarize key findings from our study, and present several recommendations.

### The research setting

Thailand is home to an estimated 10,000 to 30,000 ethnic Khmu (Gordon 2005; Aguetant 1996). Most of the Khmu villages are situated along the border with Laos in Chiang Rai and Nan provinces, though there are small pockets of Khmu in Lampang and Uthaitani provinces (Suwilai 2002b).

The Khmu village settlements within Thailand are fairly scattered and have been established for varying lengths of time. Many villages in the Lao border regions trace their histories to migrations from Laos within the last 50 years, while other villages have been established for much longer periods of time (Suwilai 2002; Supatra 1988).

We chose five villages in Chon Daen sub-district of Nan province as the site of this research because this sub-district has the highest concentration of Khmu speakers in Thailand. Khmu in these villages identify themselves as part of the *Tmooy Thruel* clan and speakers of a common dialect (Miller & Miller 2009). Their dialect may be classified as part of the Western Khmu dialect grouping which includes Khmu referred to as *Khmu Rook* in Laos (Svantesson & House 2006; Chazee 1999).

Village name	Thai name	Local Khmu name (IPA)	No. of households	Population
Nam Paan	บ.น้ำปาน	/ʔom pa:n/	65	450
Huay Klaep	บ.ห้วยเกลบ	/ʔom kle:p/	71	378
Nam Lu	บ.น้ำหลู	/ʔom lúʔ/	74	453
Huay Moi	บ.ห้วยมอย	/ʔom mɔ:j/	24	244
Ban Mai Chaidaen	บ.ใหม่ชายแดน	/ʔom sê:n/	39	209
Total			273	1734

Table 1: Name and population of Khmu villages in Chon Daen sub-district, 2008 (Songkhwa District Office registrar, July 2008).

### Research Methods

Data for this research project was gathered during a 6-week period in 2008 using questionnaires, interviews and personal observation. Nearly 70 respondents participated in this study, and informed consent was given by all respondents.

## Research Findings

### 1. Language maintenance or shift

It has been well documented that extensive language shift to a regional language or to Standard Thai (the *de facto* national and official language) is taking place among many of the ethnic minority language communities in Thailand (ILCRD 2008; Vail 2006; Tehan & Nahas 2009). Fourteen of the nation's 74 documented languages are currently deemed "severely endangered" by the Institute of Language and Culture for Rural Development (ILCRD), Mahidol University, Thailand (ILCRD 2008) and many others can likely be classified as "unsafe" to use the terminology of Brenzinger *et al* (2003). Previous to this study the language use within Khmu communities in Thailand had not been documented for 20 years.

There are a several analytical frameworks that have been used to identify particular points of vulnerability with respect to language vitality and to prioritize strategies for language 'safeguarding' or revitalization, which include:

- Joshua Fishman's Graded Intergenerational Disruption Scale (Fishman 1991)
- UNESCO's proposed nine factors on language vitality and endangerment (Brenzinger, Matthias *et al* 2003)
- Landweer's eight Ethnolinguistic Vitality Factors (Landweer 2000)

In our analysis we will present five key factors that seem particularly salient in understanding the situation of the Khmu.<sup>92</sup> The indicators we will discuss come primarily from Fishman (indicator 1) and Landweer (indicators 2-5 below). We are using the Landweer (2000) framework because it brings the issue of economic security into particular focus.

- 1) Intergenerational transmission of the mother tongue
- 2) Population and group dynamics
- 3) Domains of language use
- 4) Access to a stable and acceptable income base within the locale
- 5) Social outlook regarding and within the speech community

#### 1.1. Intergenerational transmission of the mother tongue

Khmu children are currently learning Khmu in the home, and it is the primary language of communication between all three generations within most households. All of the respondents to our survey (36 of 36) reported that they can speak Khmu, *Muang* (the Nan-Phrae province variety of *Kammuang* or Northern Thai) and Standard Thai (ST). Ninety-four percent of respondents reported that Khmu is their strongest language.

The data below provide further documentation of how language is being used in the home:

- Between spouses: Khmu was the language spoken between husband and wife in over 80% of the homes surveyed. In the remaining 20% respondents report that *Muang* or ST is used because one of the marriage partners is not Khmu.
- With parents: All respondents reported that they speak Khmu with their parents.
- Parents with their children: 89% of respondents reported speaking Khmu with their children.
- With siblings: 97% reported speaking Khmu with brothers and sisters.
- First language of children in the village: 89% reported that Khmu is the first language children speak and 6% reported that Khmu is learned simultaneously with the mother tongue of a non-Khmu parent. Thus, 95% said that Khmu is the first language of the children.
- The teachers at village Early Childhood Development Centers reported that the children "cannot speak any Standard Thai" or "speak just a few words, such as 'hello' or 'eat rice'" when first

---

<sup>92</sup> Note that a more complete analysis is presented in Miller & Miller 2009.

entering these state-sponsored daycare programs. This was further evidence that Khmu is indeed the language spoken at home.

According to Joshua Fishman (1991) transmission of the language in the home is “the threshold level” for language maintenance, the level at which small languages continue to survive and even thrive (cf. Lewis 1996:8; Fishman 1991:92). We thus assess this finding as a clearly positive indicator of Khmu language maintenance.

## 1.2. Population and group dynamics

According to Songkhwa district records, the absolute populations of the villages that participated in this study have been stable and have even shown growth during the past 20 years (Miller & Miller 2009; Supatra 1988). There are only a few non-ethnic Khmu living in the villages because of marriage to a Khmu spouse.

Some significant changes are taking place, however, in the group dynamics of the Chon Daen sub-district villages. There has been a dramatic increase in the number of young people, particularly women, who are seeking work outside the village after completing their education. (Most young people typically complete secondary grade 3—or *mathayom 3*). Of female respondents 35 and older, *none* had ever worked outside the village, but 78% of women ages 15 to 34 had worked in an urban center for a period of several months or longer, some on multiple occasions. It should be noted that our sample population included only people currently living in the village.

Male respondents reported that employment outside the village was not a recent development, as all male respondents in both older and younger groups had worked outside the village on one or more occasions.

People most commonly reported finding employment as factory laborers, gas station attendants, construction laborers, and drivers. Forty percent of respondents reported working outside the village on multiple occasions for time periods ranging from six months to two years. Some reported seeking work on a seasonal basis following the agricultural cycle, while others sought more permanent work and come home occasionally during holidays. It should be added that not everyone who goes to the city to work finds the experience to be financially profitable. Several respondents noted that women generally have more success saving money from their urban work experience than the men who tend to spend more money on alcohol.

The increase in the number of Khmu youth seeking work in urban areas has led to changes in marriage patterns. Village leaders observed that very few young people are currently marrying other Khmu. Other respondents' estimates of the number of Khmu youth marrying non-Khmu ranged between 50 and 90 percent.

The question of what effect this will have on the language of the children appears to depend on where the children are raised. We do not have clear data indicating how many young people who marry outsiders return to the village, but our data describes several scenarios. One respondent noted that sometimes marriages fail and the Khmu spouse, generally the wife, will return to the village to raise the child. We met a number of young mothers in this situation. Several respondents explained that sometimes the children of mixed marriages are left in the care of grandparents in the village while the parents work in the city. In these cases the children speak Khmu as their mother tongue. One respondent noted that a number of young women had married young men from northeastern Thailand and had moved to the Khmu village because no suitable farmland was available in the spouse's home area.

When we asked both Khmu village leaders and individual respondents about community attitudes toward marriage with non-Khmu, responses were similar. Among the survey respondents, none expressed an unfavorable opinion toward youth marrying non-Khmu, with the exception of one respondent who reported a taboo forbidding Khmu to marry a person of Htin ethnicity. Most respondents expressed the opinion that, “It depends on the couple to decide who they will marry,” and many expressed that the key issue should be love between the man and woman, not ethnicity. Interestingly, two respondents said they hoped to marry someone from Bangkok in order to raise their social status.

There is evidence in our data that this current ambivalence toward marriage with non-Khmu is a more recent trend, and that in the past marrying a fellow Khmu was an important value. One respondent said that in the past it was an important cultural value to marry a Khmu, but that now it does not matter. *Kamnan* (sub-district chief) Min explained his perception of how attitudes and practices toward marriage have changed, and expressed concerns that parents have related to intermarriage:

*Now only a very small percentage of Khmu marry fellow Khmu from our villages. In the past few years only a small percentage of Khmu young people has married within the Khmu tribe. In the past, people might arrange a marriage between neighbors or distant relatives from the time their children were young. In doing so parents could ensure that the person their child was marrying was diligent and from a good family. Now they have no way of helping ensure that their children's marriage is stable. They do not know whether the person might have AIDS or something else. When some in-laws (outsiders married to village children) come to visit the village, their in-laws will not speak to them. There is a feeling that they have stolen their child away to live elsewhere, and there are bad feelings about this in some cases (Kamnan Min).*

Regarding connectedness with the broader population of Khmu, the residents in Chon Daen sub-district villages reported only minimal contact with other Khmu communities in Thailand outside Nan province. For most people, contact with Khmu from outside the local district was reportedly rare. Chon Daen residents do have fairly regular contact with Khmu from Laos, though, at a weekly border market.

Landweer's (2000:12) language vitality scale regarding population and group dynamics brings into focus the influence of the migration of outsiders on the ethnolinguistic vitality of local language communities. However, in this Khmu context, while there is some in-migration of outsiders, a more relevant appraisal of changes in population and group dynamics is linked to the increasing number of young people, particularly women, moving out of the village to seek employment. There are currently only a few non-Khmu who have migrated into the Khmu villages, and many of these have learned to speak Khmu.

Given this observation we have modified Landweer's scale as follows. Note that a score of (4) indicates high ethnolinguistic vitality potential while a score of (1) indicates low potential.

- (4) Marriage between members of the same vernacular; children raised as monolingual speakers of the vernacular
- (3) Intermarriage practiced by some members of the speech community but children are raised by parents or guardians within the vernacular community to speak vernacular and possibly another language
- (2) Intermarriage practiced; one or more parents maintain work outside vernacular community and children raised for a period of time in household of grandparents speaking vernacular
- (1) Intermarriage practiced and family established outside vernacular community; speak an LWC (language of wider communication or regional trade language) as their mother tongue

Given the data gathered, we assigned a score of 2 on this scale for population and group dynamics.

### 1.3. Domains of language use

All respondents in our study were multilingual, speaking Khmu, *Muang*, and Standard Thai, though we observed that the oldest respondents showed more limited standard Thai skills. Khmu was reported to be the language used in the village between residents (even some non-Khmu spouses were reported to learn to speak Khmu), and *Muang* for local commerce. Standard Thai is used in the domains of education, most broadcast media, official functions, and is the only language of literacy.

Our research revealed that Standard Thai likely has increasing influence in the community and home, domains in which Khmu has been the primary language used, due to two recent developments.

Firstly, Khmu pre-school children are being exposed to Standard Thai much earlier and more intensively as a result of the opening of state-funded local daycare centers four years ago. Virtually all children between the ages of two and five attend. The teachers in four of the five daycare centers are Khmu, but are being encouraged by their sub-district administrator to speak Thai with the children.

Secondly, within the last five years television ownership has dramatically increased. According to village leaders, five years ago almost no one owned a television, but now satellite television has become available and our survey indicated that nearly 80 percent of respondents own a television and watch daily.

Below is Landweer's (2000:9) ethnolinguistic vitality scale, which relates domains in which the local language is used with potential for language maintenance:

- (4) Home, cultural events, social events and other domains
- (3) Home, cultural events, social events
- (2) Home, cultural events where the vernacular is used, but is mixed with an outside lingua franca or other local language(s)
- (1) Home, where the vernacular is used but is mixed with an outside lingua franca or other local language(s).

We assess the relative strength of Khmu language vitality in Chon daen to be a '3' on this scale given that in 'other domains' such as education, broadcast media, and commerce either Standard Thai or Muang is the language used.

#### 1.4. Access to a stable and acceptable income base within the locale

Landweer (2000) states that, "One of the most common factors influencing a community to shift from one language to another is that adequate work environments using their mother tongue do not exist for their children" (Landweer 2000:17).

The livelihood of the Khmu in Chon Daen is based primarily on subsistence hillside rice farming and on gathering forest products. Some Khmu families in Chon Daen raise cows, pigs and or chickens on a small scale as part of their livelihood base, and a few families have small fish farms. Nearly all of the respondents to our survey questionnaire (89%) report they are involved in farming. The respondents who reported 'student' as their occupation (8%) help their parents on the family farm when they are free, bringing total number of respondents involved in farming to nearly 100%.

Village leaders described two difficulties Khmu community members currently experience in trying to provide for the needs of their families in the locale.

Firstly, the Khmu in Chon Daen do not have legal entitlement to the land they farm. The inability to own land in the village locale causes stress and frustration, because without land rights Khmu have no economic security, no means to acquire additional land needed to adequately support their families, and no source of collateral to apply for loans. Within the last decade some of the land used by the Khmu was even reclaimed as part of a national reforestation initiative. This particular land had been used for grazing cattle, thus the loss of this land forced most residents to sell off their cattle. Cattle ownership had served as a sort of emergency savings fund for many Khmu families.

A second challenge the Khmu face relates to the lack of a stable means of earning cash in the village locale. Corn is being developed as a cash crop, but the market price is volatile. Many families accrued further debt on their investment in corn crops in the year this study was conducted due to the low market price, the costs of fertilizers and loss caused by rats and wild pigs. One leader lamented that *all* of the families in his village had accrued large debt associated with the costs of agriculture and there was no way to earn cash locally to repay these debts.

Landweer (2000:17) has developed a scale linking ethnolinguistic vitality to the availability of a stable means of income generation in the local setting. This scale emphasizes the importance of having a



means of access to stable income bases in an environment that uses the local language. (A higher score of the below scale indicates greater potential for language maintenance.)

- (4) Stable and acceptable economic base where the vernacular is the code of choice
- (3) Adequate dual economy where the language used is dictated by choice of economic base
- (2) Marginal subsistence economy requiring augmentation of the traditional means of subsistence with non-vernacular, cash-based economic schemes
- (1) Dependence on an economy requiring use of a non-vernacular

From the data we have gathered, our assessment is that the Khmu speech community in Chon Daen rates a 1.5 on this scale. The issues discussed here have led to increasing debt and a growing dependency on finding outside employment. Khmu families face an economic situation that depends on the use of Standard Thai or a regional language such as Muang in order to secure wage labor in urban centers.

### 1.5. Social outlook regarding and within the speech community

According to Landweer (2000), the primary focus of inquiry regarding this vitality factor is, “Is there internal and/or external recognition of the language community as separate and unique within the broader society?” (Landweer 2000:15). Landweer explains that “a strong ethnic identity can influence language choice. In other words, the perception a group has of itself can be supportive or can undermine the value associated with their language and ultimately their own use of the language” (Landweer 2000:14).

Cultural elements that the Khmu in Chon Daen sub-district described as unique and important to their identity included their language, traditional agricultural and healing ceremonies, rice-wine production, and stories of the ancient cultural hero ‘Ceuang’. Khmu-style farming, trapping and gathering from the forest were also mentioned as valued and uniquely Khmu.

Village residents are keenly aware that the Khmu are losing their distinctiveness, however. The village leader of Ban Mai Chaidan shared his insights as to why this is happening:

*If you look at the various tribes [found in Thailand—he mentioned the Karen, Lahu, , Hmong, Mien, Htin and Khmu], the Htin, Khmu and Khmu Lue seem to be most flexible...discarding their customs and culture in order to follow the ways of others... Khmu men no longer have traditional clothes...our women don't really wear the traditional clothes.*

Several Khmu described the conflicting feelings and the stress that comes from living between traditional and encroaching modern life-styles. The leader of Huay Klaep village articulated his outlook on the times during an interview as follows:

*We used to be a very small group that did not interact much with broader society and its influences, but in the past 20 years the values related to materialism have crept in and led to changes. In the past, girls did not really go out from the village to work. We used to simply eat foods from the forest, but now people, especially teens, see the comforts and luxuries of the outer society and want them; they do not want the old ways.*

*If society changes fast and we are not able to adjust ourselves rapidly enough, it will be hard for us. Take car ownership, for example. If we do not have the resources to buy a car and borrow money as is done in other parts of society, the payments are a burden and hardship.*

*Modernizing brings benefits though. For example, in previous times we did not have tin or tile for roofs, but we used leaves. Making a leaf-roof was a lot of work and after two to three years it would leak and we would have to re-thatch. We have adopted this modern development and benefitted. So, if we don't adopt these newer ways...it can bring suffering as well (Village leader Sanit Saorungtoy).*

We used two questions in our survey to probe community perceptions regarding attitudes of young people toward Khmu language and culture. In response to the question, “*Are the young people proud of*

being Khmu?” 75% responded affirmatively, while several were unsure and only one interviewee responded that young people were not proud of being Khmu.

A second question asked, “Have you ever observed situations in which Khmu young people chose not to speak Khmu to each other?” Respondents related observations that when Khmu youth were in Bangkok, outside the village, in a situation where outsiders are present or when at school they tended to speak Thai or *Muang*. Several respondents felt that Khmu was not spoken because young people were embarrassed to speak their mother tongue. Several respondents did note that in the village Khmu youth speak Khmu together. One youth in his early twenties said that young people use a lot of slang that is really a mixture of languages.

From these responses we conclude that Khmu youth still have a level of appreciation for their language and culture, but are aware of and probably somewhat embarrassed by the lower social status of their language when with speakers from a language and culture with higher prestige. Among Khmu peers in the village, Khmu seems to be the language that is spoken.

As for the perspective of northern Thai outsiders in the surrounding area regarding the distinctives of the Khmu, those we talked to knew little about the Khmu except that Khmu are not followers of Buddhism but practice a “spirit religion.”

Ethnolinguistic vitality factor	Significant changes or observations reported	Overall future language maintenance impact ( <i>positive, negative, or neutral</i> )	Landweer vitality scale rating where applicable
1. Intergenerational language transmission	<ul style="list-style-type: none"> <li>▪ Children speaking K in the home</li> <li>▪ K the language of communication between 3 generations within households</li> </ul>	positive	Not applicable
2. Population and group dynamics	<ul style="list-style-type: none"> <li>▪ Dramatic increase in young women working outside village</li> <li>▪ Few K with K marriages</li> <li>▪ Attitude toward mixed marriages increasingly neutral</li> </ul>	negative	2
	<ul style="list-style-type: none"> <li>▪ Few non-Khmu immigrants in the villages</li> </ul>	neutral	
3. Domains of language use	<ul style="list-style-type: none"> <li>▪ K the primary language of homes in village (children’s first language)</li> <li>▪ K the primary language used in daily village life</li> <li>▪ K the ‘most fluent’ language for nearly all respondents</li> </ul>	positive	3
	<ul style="list-style-type: none"> <li>▪ ST used exclusively in education; urban workplace</li> <li>▪ Earlier exposure to ST at day care for 2-3 year olds</li> </ul>	negative	
	<ul style="list-style-type: none"> <li>▪ M used for local commerce</li> </ul>	neutral	
4. Access to a stable and acceptable income base within the locale	<ul style="list-style-type: none"> <li>▪ Lack of land rights</li> <li>▪ Market for cash crops unstable</li> <li>▪ Subsistence rice farming; large debt</li> <li>▪ Increased demand for consumer goods through broadcast media exposure</li> </ul>	negative	1.5
5. Social outlook regarding and within the speech community	<ul style="list-style-type: none"> <li>▪ Positive attitude toward culture/language and local language development</li> </ul>	positive	2
	<ul style="list-style-type: none"> <li>▪ Disappearing visible cultural distinctives; ‘flexibility’</li> </ul>	negative	
	<ul style="list-style-type: none"> <li>▪ Youth desire and pursue modern lifestyle</li> </ul>	negative	

Table 2: Summary of ethnolinguistic vitality factors

Abbreviations: Standard Thai—ST; Muang—M; Khmu—K.

Landweer's (2000:15) scale for rating a speech community's social outlook in terms of its influence on language maintenance or shift is presented below:

- (4) Strong internal identity, high status or notoriety conferred by outsiders, with cultural markers present
- (3) Strong internal identity, neutral status or notoriety conferred by outsiders, with cultural markers present
- (2) Weak internal identity, neutral status or notoriety conferred by outsiders, with some cultural markers present
- (1) Weak internal identity, negative status or notoriety conferred by outsiders, with few if any cultural markers present

Based on the data we gathered, we have assigned a score of 2 for this ethnolinguistic factor.

### **Summary of factors indicating potential for language maintenance or shift in Chon Daen Khmu villages**

Transmission of the language to children in the home is indeed the most positive indicator of ethnolinguistic vitality. Overall, however, language and culture maintenance does seem to be threatened in some significant ways. Table 2 below summarizes the ethnolinguistic vitality factors discussed and describes the nature of the impact they may have on future language maintenance. Regarding the Landweer vitality scale score, recall that '4' indicates high potential for language maintenance and '1' low potential.

### **2. Standard Thai literacy skills and practices**

In our literacy survey we gathered data related to literacy skills in Standard Thai, and use of print and broadcast media in order to inform potential local language development activities that may include the production of literacy instructional materials, local language literature, and non-print media products.

#### **2.1. Literacy skills**

For the purpose of this study, we used the fairly broad definition of 'literate' put forth by UNESCO: "A person is literate who can, with understanding, both read and write a short statement on his or her everyday life" (UNESCO 2006). We did not attempt to administer a literacy skills test, but used a self-reporting evaluation instrument to gather data. Respondents to our literacy questionnaire were asked to describe their current reading and writing abilities by choosing between the four skill levels described in Tables 3 and 4 below. Note that reading and writing were treated as separate skills.

<b>Reading skill level choice</b>	<b>Definition of skill level</b>
1. Cannot read	I cannot read any words.
2. Poor	I can read only a few words.
3. Fair	I can read most words and sentences in books and newspapers; I read somewhat slowly.
4. Very good	I can read anything I want such as a newspaper, government forms, agriculture or health materials without difficulty and fairly quickly.

*Table 3: Skill level choices for reading skills*

<b>Writing skill level choice</b>	<b>Definition of skill level</b>
1. Cannot write	I cannot write any words.
2. Poor	I can write my name and a few other words.
3. Fair	I can write simple sentences, but have trouble spelling many words.
4. Very good	I can write a letter, story, report and fill in a form without difficulty.

*Table 4: Skill level choices for writing skills*

The respondents were grouped into two categories based on their self-assessments as follows:

1. **‘Literate’** if reported their ability to read AND write as being “Very Good” or “Fair”
2. **‘Not Literate’** if reported that they “Cannot Read/Cannot Write” AND/OR “Read Poorly/Write Poorly”

Among younger cohort (ages 15-34), all of the respondents (18 of 18) rated their reading and writing skills as either “Fair” or “Very good” and were therefore classified as ‘Literate’, though the women rated their ability levels higher than the men overall.

Our findings were quite different among older cohort (ages 35 and older). Only four of 18 (or 22%), were classified as ‘Literate’ and only one of these was a woman.

## 2.2. Literacy practices

When we investigated the question, “*For what purposes do Khmu community members use literacy skills?*” we learned that generally, most Khmu in Chon Daen do not depend on written materials to access information, neither do many people read for entertainment or pleasure. There are several possible reasons for this:

- 1) The majority of the respondents over 35 years old were farmers and the skills required to maintain their livelihood were ‘learned by doing’ and not dependent on literacy.
- 2) Access to newspapers and other types of reading material is quite limited. There were no reported village libraries or reading rooms.
- 3) Other media forms are increasingly available and appealing. Nearly 80 percent of respondents watch television daily and 70 percent report listening to the radio daily.

## 3. Attitudes toward language development

The prospect of developing a way to read and write Khmu was met with many enthusiastically positive responses from village leaders, parents, community members and even Thai teachers and administrators in the local schools. Several noted that other ethnic minority groups around them had developed writing systems and literature in their languages. However, some of the respondents were skeptical as to whether it would be possible to develop a way to write the unique sounds of their Khmu language.

The most commonly articulated perceived benefit of developing a writing system was that doing so would help preserve the Khmu language and culture. Other perceived benefits were also mentioned and included the following:

- *We could write and tell people our true feelings and others would know what we mean right away.*
- *People could write secrets to each other.*
- *Khmu would be a ‘complete’ language.*
- *People who are not Khmu could learn Khmu.*
- *We could write the words on paper for the Khmu spirit ceremonies.*
- *If Khmu people could produce a written history of our people, written in our own language, we would have a better understanding of who we are. We would have more confidence to try to pull ourselves up [from a lower socio-economic status].*

## 4. Perceptions of educators working in Khmu communities regarding the adequacy of Khmu children’s Thai language skills for successful classroom learning

Because our study revealed that nearly all Khmu children in Chon Daen sub-district are monolingual speakers of Khmu at the time they enter school, and there is a growing concern in Southeast Asia for the welfare of students who must cope with learning through a language that is not their mother tongue (Kosonen

& Young 2009), we thought it important to learn about the educational experience of Khmu children, more specifically asking the following questions:

- *Is there a perception among Thai educators working in Khmu communities that a language barrier adversely affects the Khmu students' ability to learn academic content?*
- *Do local educators see value in incorporating a mother-tongue learning component into the local curriculum?*

We interviewed teachers and administrators at two schools, serving over 300 students, the large majority of whom were Khmu. Key findings from our interviews include the following:

- Teachers do experience significant difficulties communicating with the students in grades Kindergarten through primary grade 2 (*prathom 2*).
- Most teachers believe that a lack of Standard Thai speaking and listening skills adversely affects the students' educational outcomes, but several added that a language barrier is not the only factor that may cause some children to struggle. One teacher who had taught young children in other minority language communities suggested that a mother-tongue first curriculum would help the children learn better during the early years of school.
- Limited data on national exam scores (O-NET) and grade repetition rates indicate that Khmu students are well behind many of their national peers regarding mastery of academic content (Miller & Miller 2009).
- The headmaster and several teachers expressed support for teaching Khmu culture and language as one component of the school curriculum. The headmaster said he would implement a curriculum immediately if he had one, but lacks the expertise and financial resources to develop such a local language curriculum component.
- Regarding informal language use by the students during breaks from the classroom, all the teachers we interviewed reported that primary school children (through primary 6) speak Khmu with each other.

### **5. Conclusions: Moving forward with language development**

The purpose of this study was to inform a potential language development effort among the Khmu in Thailand. In this section we will make specific recommendations regarding Khmu language development based on these research findings. It should be noted that since this research was conducted, leaders from these Khmu communities have taken the initiative to pursue orthography development that employs the Standard Thai script. A recent partnership has been established between these communities, Mahidol University's Institute for Cultures and Languages in Asia, SIL, and the Thailand Research Fund in order to conduct community-based research leading to local literature development.

#### **5.1. Now is the time to pursue Khmu language development**

The Khmu are facing a rapidly narrowing window of opportunity in which to safeguard their language. There is currently a vital, multi-generational community of speakers and yet the past decade has brought some significant changes that are weakening the chain of language transmission to the youngest generation. Fishman (1991), Crystal (2000), and Brenzinger *et al* (2003) all emphasize the foundational role of orthography development, language documentation, local language literature production and facilitation of teaching the language in school in safeguarding and strengthening threatened languages.

#### **5.2. Develop a network of Khmu in Thailand**

Because this study revealed that Khmu communities are fairly isolated from one another, it will be important to develop a closer network so that the Khmu community at large in Thailand can contribute to and benefit from resources that are developed. A community newsletter, website, and Khmu language radio programming could help facilitate the building of closer ties between Khmu communities.

### 5.3. Facilitate Khmu literacy skills development

Our literacy survey indicated that there are both literate and non-literate adults in the community who want to learn to read Khmu. Different literacy instructional approaches will need to be developed to meet the needs of these two groups.

For those who can read Thai, self-study materials that employ a transfer skills approach (Waters 1998) would likely be sufficient to facilitate acquisition of Khmu language literacy skills.

For children or adults who do not have literacy skills in any language, literacy instruction will require the participation of trained teachers and the development of materials intended to teach initial literacy concepts and skills.

### 5.4. Develop relevant print and non-print Khmu language media

Research revealed that both the use and availability of reading material in these communities was minimal. Thus a local language development program should include the development of literature that is perceived to be interesting and relevant, as well as a means of making this literature readily accessible.

Further community research should be done to identify story-tellers, singers, musicians and those with other special skills and knowledge so that these cultural forms and resources can be captured in both print media and audio-visual formats such as video and photo journals.

Networking with local organizations (such as the local public health center) could facilitate cooperative efforts to produce resources in the Khmu language that can have practical benefits in meeting communication needs.

### 5.5. Support local language instruction for Khmu children in the formal education system

There is growing support in Thailand for using local languages in education. Most commonly, local language and local culture appreciation are taught as a school subject, though there is also cautious support for a local language-based multilingual educational approach in some contexts. For the Khmu children of Chon Daen, both of these classroom applications would be appropriate.

The teachers in schools serving Khmu children gave clear testimony that in the early grades the Khmu children have difficulty understanding Standard Thai, the language of classroom instruction. An educational program that allows the students in the early grades (kindergarten and early primary) to *begin* learning in their mother tongue and gradually bridge into Standard Thai-based instruction would likely increase students' comprehension of academic material and improve educational outcomes (Kosonen and Young 2009). Continuing to incorporate the teaching of Khmu as a subject, then, for older students would be important step toward safeguarding the Khmu language by fostering appreciation for their local language and promoting its use in new domains (Crystal 2000).

### 5.6. Develop new avenues for local income generation

Consult with outside experts in agriculture and microenterprise to pursue developing means of local income generation that are more viable and stable than current cash crop ventures. The movement of young people out of Khmu communities and into urban areas will likely increase if new avenues of local income generation are not developed.

## **6. Recommendations for further research**

Sociolinguistic research and documentation needs to be extended to include other Khmu communities in Thailand to determine whether their situation is similar to that of the Khmu in the communities that participated in this study.

## REFERENCES

- Aguettant, Joseph L. 1996. Impact of population registration on hilltribe development in Thailand. *Asia-Pacific Population Journal* 11:4. 47-72. <http://www.unisicap.org/esid/psis/population/journal/1996/v11n4a3.htm> (Accessed 17 June 2004).
- Benson, Carol and Kimmo Kosonen. 2010. Language-in-education policy and practice in Southeast Asia in light of the findings from Ethiopia. In K. Heugh & T. Skutnabb-Kangas (eds.). *Multilingual Education Works*. New Delhi: Orient Black Swan (Forthcoming).
- Brenzinger, Matthias *et al.* 2003. Language vitality and endangerment: By way of introduction. Paris: UNESCO Ad Hoc Expert Group on Endangered Languages. [http://www.linguae-celticae.org/dateien/UNESCO\\_Language\\_vitality\\_and\\_endangerment.pdf](http://www.linguae-celticae.org/dateien/UNESCO_Language_vitality_and_endangerment.pdf). (Accessed 19 October 2009).
- Chazee, Laurent. 1999. *The peoples of Laos*. Bangkok: White Lotus Press.
- Cholthisa Ong-arj. 1988. A phonology of Khmu Lue at Nan province. Salaya: Institute of Language and Culture for Rural Development, Mahidol University M.A. thesis.
- Cooper, Michael and Susan Cooper. 1998. A preliminary phonological analysis of Huay Jo Khmu (Technical Paper No. 38). Chiang Mai: Payap Research and Development Institute and the Summer Institute of Linguistics.
- Cooper, Michael and Susan Cooper. 1999. Tri-lingual phrasebook: Khmu -Thai -English (Technical Paper No. 41). Chiang Mai: Payap Research and Development Institute and the Summer Institute of Linguistics.
- Crystal, David. 2000. *Language death*. Cambridge: The Cambridge University Press.
- Fishman, J.A. 1991. *Reversing language shift*. Clevedon: Multilingual Matters.
- Gordon, Raymond G., Jr (ed.). 2005. *Ethnologue: Languages of the world*. 15th Edition. Dallas, TX: SIL International. [http://www.ethnologue.com/show\\_country.asp?name=TH](http://www.ethnologue.com/show_country.asp?name=TH) (Accessed 30 September 2009).
- IILCRD Research. 2008. *Endangered languages and cultures*. Institute of Language and Culture for Rural Development. Salaya, Thailand: Mahidol University. <http://www.lc.mahidol.ac.th/en/research-groups-endangered.php> (Accessed 15 June 2009).
- Kosonen, Kimmo and Catherine Young (eds). 2009. *Mother tongue as a bridge language of instruction: Policies and experiences in Southeast Asia*. Bangkok: SEAMEO.
- Landweer, M. Lynn. 2000. Indicators of ethnolinguistic vitality. *Dallas: Notes on Sociolinguistics* 5:1.5-22.
- Lewis, M. Paul. 1996. Language maintenance in seven K'iche' communities. *SIL Electronic Working Papers* 1996-002, July 1996. [http://www.sil.org/silewp/1996/002/silewp\\_1996-002.html](http://www.sil.org/silewp/1996/002/silewp_1996-002.html)
- Lindell, Kristina. 1974. A vocabulary of the Yuan dialect of-the Kammu language. *Acta Orientalia* 36.191-207.
- Lindell, Kristina, Jan-Olof Svantesson and Damrong Tayanin. 1981. Phonology of Kammu dialects. *Cahiers de Linguistique Asie Orientale* 9.45-71.
- Miller, Michelle and Timothy Miller. 2009. *A sociolinguistic and literacy study of Khmu communities in Thailand: toward the process and products of language development*. Chiang Mai: Payap University Linguistics Institute.
- National Statistics Center of the Lao PDR. 2007. <http://www.nsc.gov.la> (Accessed 19 October 2009).
- Preisig, Elisabeth. 1990. *The Kmhmu' orthography: Dialect of Xieng Khouang, Luang Prabang and Sam Neua*. Unpublished manuscript.
- Smalley, William. 1961. *Outline of Khmu structure. An essay of the American Oriental Society No 2*, ed. by Edward H. Schafer. New Haven: American Oriental Society.
- \_\_\_\_\_. 1994. *Linguistic diversity and national unity: Language ecology in Thailand*. Chicago: University of Chicago Press.
- Suksavang Simana, Somseng Sayavong and Elisabeth Preisig. 1994. *Kmhmu'-Lao-French-English dictionary*. Vientiane, Lao P.D.R.: Ministry of Information and Culture, Institute of Research on Culture.
- Suksavang Simana and Elizabeth Preisig (English). 1998. *Kmhmu' livelihood: farming the forest*. Vientiane, Lao PDR: Institute for Cultural Research, Ministry of Information and Culture.
- Supatra Jiranathanaporn. 1988. *An experiment using Thai orthography with Khmu phonology at Ban Nam Pan in Nan*. Thailand: Mahidol University M.A. thesis.
- Suwilai Premsrirat. 1987. *Khmu, a minority language of Thailand*. Pacific Linguistics: Series A, No.75. Department of Linguistics, Research School of Pacific Studies, The Australian National University.

- \_\_\_\_\_. 1990. Khmu traditional medicine and Khmu-Thai medical conversation. Salaya, Thailand: Institute of Language and Culture for Rural Development, Mahidol University.
- \_\_\_\_\_. 1993. Thai-Khmu-English dictionary. Salaya, Thailand: Mahidol University, Institute of Language and Culture for Rural Development.
- \_\_\_\_\_. 1998. Phonological variation and change in the Khmu dialects of northern Thailand. *Mon-Khmer Studies (MKS)* 29.57-69.
- \_\_\_\_\_. 2001. Tonogenesis in Khmu dialects of Southeast Asia. *Mon-Khmer Studies (MKS)* 31.47-56.
- \_\_\_\_\_. 2002. Thesaurus of Khmu dialects in Southeast Asia. Salaya, Thailand: Mahidol University, Institute of Language and Culture for Rural Development.
- Svantesson, Jan-Olof, Damrong Tayanin and Kristina Lindell. 1994. Khmu-Lao dictionary. Vientiane, Lao PDR: Ministry of Information and Culture.
- Svantesson, Jan-Olof and David House. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology* 23.309-333.
- Tehan, Thomas M. and Ramzi W. Nahas. 2009. Mpi present and future: Reversing language shift. *Mon-Khmer Studies (MKS)* 38.87-104.
- UNESCO 2006. Education for All Global Monitoring Report 2006. [http://www.unesco.org/education/GMR2006/full/chapt6\\_eng.pdf](http://www.unesco.org/education/GMR2006/full/chapt6_eng.pdf). (Accessed 1 March 2009).
- Vail, Peter. 2006. Can a language of a million speakers be endangered? Language shift and apathy among Northern Khmer speakers in Thailand. *International Journal of the Sociology of Language* 178.135-147.



## Multiple uses of the pronoun *de:* in Kmhmu'

R. Anne Osborne  
SIL International

Kmhmu' personal pronouns have a singular, dual and plural distinction, with gender differentiation in the second and third person singular forms. As well as these forms there is also an unspecified pronoun *de:*. Unlike the other personal pronouns, *de:* has no semantic constraints with respect to person, gender or number. It exhibits a range of meanings depending on its context. From the speaker's point of view, what motivates the choice of *de:* over one of the specific pronouns? How do the hearers know who *de:* is referring to? Are the contextual factors involved in interpreting the meaning of *de:* purely pragmatic, or can syntax help? This paper explores the syntactic and pragmatic constraints that govern the use and interpretation of *de:*.

The data for this paper comes from the Kmhmu' spoken in Vientiane province of the Lao PDR. Many people have helped with collecting, translating and analysing texts. In particular I would like to thank Ajarn Suksavang Simana', Elisabeth Preisig, Ajarn Sosavanh Silaphet, Mrs Tan Ounpachanh and Miss Pang Vilay.

Dr Suwilai (1987:33) and Suksavang et al. (1994:166) describe *de:* as a reflexive pronoun. Van den Berg (1988:5) says it does not have a reflexive meaning, but occurs in certain dependent clauses where it is co-referential with the subject of the main clause. Suksavang et al. (1994:166) also report a co-referential meaning and observe that *de:* can 'replace all other pronouns', and can also mean 'alone'. Building on these findings, this paper presents four structurally defined uses of *de:*, one of which has multiple pragmatic functions.

### Syntactically Defined Uses of *de:*

These four uses of *de:* occur in distinctive syntactic contexts. Where there is an antecedent reference in a clause, *de:* has a co-referential function. Where there is no antecedent, *de:* is interpreted as a first person meaning that is unspecified for number. As the final element in a clause, *de:* means 'alone'. When it occurs in the verb complex, *de:* combines with a verbal element to add aspectual or adverbial information to the main verb.

### Co-referential function

In a well-formed sentence *de:* is used anaphorically for any pronominal co-reference within a clause and across some clause boundaries. This syntactic context requires an antecedent reference for *de:*. In (1), *de:* is co-referent with the antecedent first person singular pronoun *ʔoʔ*.

- (1)  
*lɛʔ ʔoʔ<sub>i</sub> jɛt jɔʔ maʔ'kin de:<sub>i</sub>/ʔoʔ<sub>i</sub>*  
and 1sg stay with aunt CO-REF/ 1sg  
'And I stayed with my aunt'<sup>93</sup>.

---

<sup>93</sup> In the Kmhmu' possessive, possessors follow the possessed entity.

---

Osborne, Anne R. 2011. "Multiple Uses of the Pronoun *de:* in Kmhmu'." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.112-123.  
Copyright vested in the author.

Although use of the specific pronoun is not ungrammatical in this context, it is not considered to be ‘good Kmhmu’.

Co-reference often occurs in a possessive and may be used in inalienable possession such as kin terms (see (1)), and body parts, as shown in (2) in the possessive *hir' ηiam de:* ‘his heart/mind’.

- (2)  
*gə:*<sub>i</sub> *gɔ?* *taŋ* *hir' ηiam* *de:*<sub>i</sub> *lə?* *lə?*  
 3sgm so\_then set\_up heart CO-REF good good  
 ‘So he made up his mind firmly.’

In alienable possession also, *de:* may refer to the possessor, as shown in (3), where the possessor in *ʔom de:* ‘her water’ is co-referent with the subject *na:* ‘3sgf’.

- (3)  
*na:*<sub>i</sub> *gɔ?* *gɔ:k* *ʔom* *de:*<sub>i</sub>  
 3sgf so\_then carry\_on\_shoulder water CO-REF  
 ‘Then she carried her water on (her) shoulder...’

When *de:* has a co-referent function, the antecedent is usually the subject of the clause, as shown in (4) where *de:* is co-referent with *ja?* *gi'na:j* ‘that woman’.

- (4)  
*ja?* *gi'na:j* *ʔan* *kɔ:n'ηim*<sub>j</sub> *de?* *ŋuak* *de:*<sub>j</sub><sup>\*j</sup>  
 woman that give young\_person get hat CO-REF  
 ‘That woman gave the young person her hat.’

A non-subject antecedent is also possible, though unusual. An example is shown in (5) where *de:* is co-referent with *kɔ:n'ηim* ‘young person’, the non-subject argument in the clause.

- (5)  
*ja?* *gi'na:j* *ʔan* *kɔ:n'ηim*<sub>j</sub> *hak* *de?* *ŋuak* *de:*<sub>j</sub><sup>\*i</sup>  
 woman that give young\_person nevertheless get hat CO-REF  
 ‘That woman gave *the* young person his/her hat.’

In order for *de:* to be understood as referring to the non-subject argument, the conjunction *hak* ‘nevertheless’ must be inserted in the clause. Examples (4) and (5) demonstrate the effect of *hak* ‘nevertheless’ in changing the co-reference relation.

The domain of co-reference for *de:* is a clause. The clause boundary includes embedded clauses within a matrix clause, subordinate adverbial clauses and tail-head linkage clauses in a narrative. It does not extend across co-ordinate clause boundaries.

An example of an embedded clause is seen in (6), where *de:* is the subject of the complement clause *de: ci ha:n k<sup>h</sup>i'ʔni?* ‘I will die now’, and is co-referent with the subject of the matrix clause *ʔo?* ‘1sg’.

- (6)  
*ʔo?*<sub>i</sub> *ŋɔ?* *de:*<sub>i</sub> *ci* *ha:n* *k<sup>h</sup>i'ʔni?*  
 1sg fear CO-REF IRR die now  
 ‘I am afraid (that) I will die now.’

Co-reference with the object of the matrix clause requires use of a specific pronoun, as shown in (7) where *gə:* ‘3sgm’ is co-referent with *kɔ:n* ‘child’ the object in the matrix clause.

- (7)  
*ʔo?*<sub>i</sub> *ʔmɔ:k* *kɔ:n*<sub>j</sub> *ʔan* *gə:*<sub>j</sub><sup>\*i</sup> *jɔh*  
 1sg tell child IMP 3sgm go  
 ‘I told the child to go.’

If *de:* is used in this context, as shown in (8), the sentence is ungrammatical because it has no meaningful interpretation.

- (8)  
 \*  $\text{ʔoʔ}_i$   $\text{ʔmɔ:k}$   $\text{kɔ:n}_j$   $\text{ʔan}$   $\text{de:}_{*i/j}$   $\text{jɔh}$   
 1sg tell child IMP CO-REF go  
 ‘I told the child?’

An example of co-reference in an adverbial clause is seen in (9), where *de:* is the subject of the clause *p<sup>h</sup>ia de: ti: ʔah ti: ra'ma:ŋ k<sup>h</sup>ian li:n ɲaŋ* ‘so that he would become richer than before’ and co-referent with the subject in the main clause *ta:j gə:* ‘his older brother’.

- (9)  
*ta:j* *gə:* *de?* *suan* *kap* *tu:t t<sup>h</sup>ru:l* *p<sup>h</sup>ia*  
 older\_sibling 3sgm get garden with fig\_tree PURP  
  
*de:* *ti:* *ʔah* *ti:* *ra'ma:ŋ* *k<sup>h</sup>ian* *li:n* *ɲaŋ*  
 CO-REF IRR have IRR rich AUG more\_than old  
 ‘His older brother got the garden and the fig tree, so that he would become richer than before.’

An example of co-reference in a tail-head linkage is seen in (10). In the second sentence, when the clause is repeated, typically the subject is a null reference, and *de:* is co-referent with it.<sup>94</sup>

- (10)  
*le?*  $\text{ʔoʔ}_i$  *jet* *jɔ?* *ma?*<sup>h</sup>*kin*  $\text{de:}_i$   
 and 1sg stay with aunt CO-REF  
 ‘And I stayed with my aunt.’  
  
 $\emptyset_i$  *jet* *jɔ?* *ma?*<sup>h</sup>*kin*  $\text{de:}_i$  *ma?*<sup>h</sup>*kin*  $\text{ʔoʔ}_i$  *ʔam* *rak*  
 1sg stay with aunt CO-REF aunt 1sg NEG love  
 ‘(I) stayed with my aunt, (but) my aunt did not love (me).’

Co-reference does not extend across co-ordinate clause boundaries, but rather an NP or specific pronoun is required to identify the same referent. This is seen in (11), where *de:* is used as a co-reference in the first clause, but in the 2<sup>nd</sup> clause the specific first person singular pronoun *ʔo?* must be used to identify the same referent.

- (11)  
 $\text{ʔoʔ}_i$  *jet* *jɔ?* *ma?*<sup>h</sup>*kin*  $\text{de:}_i$   
 1sg stay with aunt CO-REF  
  
 $\text{ʔnaj}$  *ma?*<sup>h</sup>*kin*  $\text{ʔo?/*de:}$  *ʔam* *rak*  
 but aunt 1sg/CO-REF NEG love  
 ‘I stayed with my aunt, but my aunt didn't love (me).’

In *Kmhmu'*, clauses may be juxtaposed without any overt marking of the hierarchical relationship between them. The scope of the co-referential function of *de:* has implications for identifying clause boundaries and the nature of the relationship between clauses. Where *de:* continues to be co-referential across clause boundaries, it is indicative of a subordinate hierarchical relationship between clauses. Where it does not have scope over adjacent clauses, this marks a co-ordinate relationship.

<sup>94</sup> If this construction is analysed without a null subject, then a tail-head linkage is a case where the antecedent of *de:* is outside the clause.

**Unspecified first person meaning**

In the second of the four structurally defined uses of *de:*, where there is no syntactic antecedent, *de:* points to a referent in the speech situation or the cultural context. It evokes a set of referents that includes the speaker and is not specified as to number.

An example of this unspecified first person meaning is seen in (12) where *de:* occurs in the second sentence. There is no antecedent to *de:* in this clause or the preceding one and *de:* is interpreted as having the non-specific meaning ‘one’. This would include the speaker and anyone who might climb the mountain.

(12)

*gɔʔ daʔ ʔo: mok niʔ*  
so\_then far oh! mountain this  
‘Oh! (it) is far away, this mountain.’

*mia de: ga: kin'dru:m mok daʔ ʔniŋ*  
when unspecified climb underneath mountain at up\_there

*pa'ma:n sa:m si: ʔak niʔ lɛʔ*  
about three four Clf\_kms this PRT

‘When one climbs the lower slopes of the mountain up there, (it is) about three (or) four kilometres.’

In another context where there is no antecedent, *de:* combines with the noun *gon* ‘person’ in the construction *gon de:* ‘we people’. An example is shown in (13).

(13)

*ʔan ra'wa:j niʔ tuʔ pok meʔ gə: gɔʔ pok*  
COND tiger this want bite INDEF 3sgn so\_then bite

*mah gon de: tɛ:ŋ məh jɔʔ gə: ʔam bian*  
eat person unspecified do INDEF with 3sgn NEG can

‘If the tiger wanted to attack anyone, to eat anyone, then it attacked (and) ate (them); we people could not do anything to it.’

In this context the noun *gon* ‘person, people’ would identify a global generic referent, i.e. the whole human race. Combined with *de:* it refers to people, including the speaker, within the wider relevant community i.e., those who were in the area affected by the tiger.<sup>95</sup>

It is not clear whether *gon de:* is a compound or a noun phrase. The only element which may be inserted between *gon* and *de:* is the possession marker *deʔ*, which then changes the meaning from ‘we people’ to ‘our people’. Other elements, such as an adjective, relative clause, demonstrative, or indefinite pronoun may not be inserted between *gon* and *de:*, as shown in (14).

(14)

*gon deʔ de: tɛ:ŋ məh jɔʔ gə: ʔam bian*  
person POSS unspecified do INDEF with 3sgn NEG can

‘...our people could not do anything to it.’

*\*gon ɲam de: tɛ:ŋ məh jɔʔ gə: ʔam bian*  
person big unspecified do INDEF with 3sgn NEG can

‘...we big people could not do anything to it.’

<sup>95</sup> *gon de:* is less restrictive than *gon ʔiʔ* ‘we (specific group) people’, which includes the speaker and their family or a specific group identified in the discourse context.

\**gon jet k'he:t ni? de: te:ŋ məh jɔ?*  
 person live district this unspecified do INDEF with

*gə: ʔam bian*  
 3sgn NEG can

'...we people (who) live in this district could not do anything to it.'

\**gon ni? de: te:ŋ məh jɔ? gə: ʔam bian*  
 person this unspecified do INDEF with 3sgn NEG can

'...we this people could not do anything to it.'

Thus it seems that *gon de:* may be a compound. *Kmhmu'* speakers think of the construction as two words and it is difficult to come to a definitive answer on this.

In other speech situations, *de:* identifies a more specific first person referent, either singular or plural. An example of a plural referent is shown in (15). There is no antecedent reference for *de:* in this sentence. It is interpreted using cultural knowledge as a first person plural meaning, the family of the speaker, i.e. *kun de:* 'our village'.

(15)

*ŋɔ:r ru:ŋ taŋ da? wiaŋ rɔ:t da? kun*  
 road rough starting at Vientiane arriving at village

*de: gi'ni?*  
 unspecified that\_one

'The road was rough starting at Vientiane (and) arriving at our village.'

An example of a singular referent is shown in (16). Again there is no antecedent reference for *de:* in this sentence. It is interpreted using the context of the discourse as a first person singular meaning.

(16)

*ho:t joŋ ʔo? law sah ti: ʔam lə? bə?*  
 and\_then father 1sg say COMP IRR NEG good eat

*ʔom'sa: jɔ:r sah de: sir'ma?*  
 tea because unspecified have\_fever

'And then my father said, "(It) would not be good to drink, tea", because I had a fever.'

### **Alone**

A third syntactic context provides another interpretation of *de:*. When *de:* occurs as the final element in a clause, it can mean 'alone' or 'by oneself'<sup>96</sup>. This is seen in (17).

(17)

*bat gi: gə: gɔ? jɔh de:*  
 turn this\_one 3sgm so\_then go oneself

'Then (at) this time he went alone.'

This meaning is made clearer to the audience by the optional insertion of the counter expectation marker *hak* 'nevertheless', as shown in (18).

<sup>96</sup> There is no sense of an emphatic meaning here, where a referent is re-iterated to emphasise that it was this person and not another who went, such as in English "He himself went". Nor is there a reflexive meaning, which typically involves a single referent being the actor and undergoer of the same action, e.g. "He saw himself (in the mirror)". It is closer to the English use of the reflexive pronoun in a preposition phrase e.g. "by himself".

(18)

*bat gi: gə: gɔʔ hak jɔh de:*  
 turn this\_one 3sgm so\_then nevertheless go oneself  
 ‘Then (at) this time he went alone.’

Where the subject is plural, this sense of *de:* can also be used, see (19).

(19)

*plɔ:j ʔan nɔ: (hak) jɔh de:*  
 release CAUS 3pl nevertheless go oneself  
 ‘(We) let them go alone.’

This usage is seen less frequently than the co-referential or unspecified first person meanings.

### **Particle in the verb complex**

The fourth syntactic context in which *de:* is found is in the verb complex. In this context, *de:* appears to have an argument re-instantiation function rather than a referential function. Some speakers do understand it as co-referential with the subject of the clause, but others see it as having no meaning of its own, but merely necessary for a well-formed sentence. Therefore this use is included in this paper not so much because *de:* here is acting as a pronoun, but for the sake of completeness of the characterisation of grammatical contexts in which it may be found.

In a Kmhmu' clause the verb complex is identified by its position in relation to the conjunction *gɔʔ* ‘so\_then’, which, when it occurs, always follows the subject and precedes the verb complex. When *de:* appears in this context with *gɔʔ*, it always follows *gɔʔ* and so is part of the verb complex. Within the verb complex, *de:* combines with a verbal element and precedes the main verb, adding components of meaning to it. Five different verbal elements are seen with *de:*, giving aspectual and adverbial information about the action or state of the main verb.

### ***bian de:* PAST**

When used as a main verb *bian* means ‘to achieve, to accomplish, to reach’. In combination with *de:* in the verb complex, it signifies past action of the following main verb; something which has been achieved or accomplished. In (20), *bian de:* marks the construction of the rice field as an event accomplished in the past.

(20)

*mo:j dia hi'ʔih ʔiʔ bian de: ʔəh ʔeʔ*  
 one Clf\_times PRT 1pl PST construct rice\_field  
 ‘One time ok, we made an upland rice field. (or ...we got to make...)’

When used in this context *de:* can be omitted without any loss of meaning, as shown in (21).

(21)

*ʔiʔ bian ʔəh ʔeʔ*  
 1pl PST construct rice\_field  
 ‘We made an upland rice field. (or ...we got to make...)’

Negation of *bian de:* indicates that an action was not achieved, as shown in (22), where *ʔam bian de: ʔiak ʔom* means ‘didn't get to drink water’.

(22)

*te:ŋ nɛ:w məʔ gɔʔ ʔam bian de: ʔiak ʔom*  
 do whatever so\_then NEG PST drink water  
 ‘Whatever (he) did (I) didn't get to drink water.’

***taŋ de:* RESULT**

When used as a main verb *taŋ* means ‘to set up, to arrange, to commence’. In combination with *de:* in the verb complex, it signifies an action resulting from the event in the previous clause. In an example from a reported narrative about a bear hunt shown in (23), one of the hunters is attacked by the bear. His friend goes to his rescue and takes out his knife and advances on the bear. As a result of this action the bear releases the man. This is marked by *taŋ de:* in the result clause *gə: taŋ de: plɔ:j* ‘then it released (him)’.

(23)

*gə: gəj tɔ:t mi:t da? gwaŋ da? gi:*  
3sgm but\_then draw\_out knife at waist at this\_one  
‘But then he drew out a knife at (his) waist here.’

*jɔh tɔrɔ: gə: taŋ de: plɔ:j*  
go approach 3sgn then release  
‘(As he) went (and) approached, then (as a result) it released (him).’

The conjunction *gɔ?* can be inserted after the subject and before *taŋ de:* as shown in (24).

(24)

*gə: gɔ? taŋ de: plɔ:j*  
3sgn so\_then then release  
‘So then (as a result) it released (him).’

It is possible to omit *de:* with no obvious change in meaning, as shown in (25).

(25)

*gə: taŋ plɔ:j*  
3sgn then release  
‘Then it released (him).’

When *de:* is omitted, however, the inclusion of *gɔ?* produces an ungrammatical sentence, as shown in (26).

(26)

\**gə: gɔ? taŋ plɔ:j*  
3sgn so\_then then release  
‘So then it released (him).’

Negation of *taŋ de:* is not possible, probably for semantic reasons.

***pʰɔ: de:* ‘almost, just about’**

When used as a main verb *pʰɔ:* means ‘to be enough’. In combination with *de:* in the verb complex, it adds an aspectual meaning to the main verb and signifies a state that is about to be achieved or has almost come about. In (27) *pʰɔ: de:* is used to add aspectual meaning to the adjectival state verb *bah* ‘to be light’.

(27)

*ʔi? rəh jɔh da? rɛ? ŋian bri? pʰɔ: de: bah*  
1pl get\_up go at rice\_field when environment almost light  
‘We got up (and) went to the rice field when it was almost light.’

Omission of *de:* from this construction is ungrammatical, as shown in (28).

(28)

\**bri? pʰɔ: bah*  
environment almost light  
‘It was almost light.’

Insertion of the conjunction *gɔʔ* is grammatical, as shown in (29).

(29)  
*briʔ*            *gɔʔ*            *pʰɔ: de:*    *bah*  
 environment so\_then almost light  
 ‘It was almost light.’

Negation of *pʰɔ: de:* is not possible, perhaps due to semantic constraints.

***ɱian de:* ‘as if’**

When used as a main verb *ɱian* means ‘to be like, to be the same as’. In combination with *de:* in the verb complex, it signifies an event or state which appears to be true, as shown in (30).

(30)  
*ur*    *gi:*            *ɱian de:*    *ti:*    *lam*  
 stew this\_one as\_if IRR delicious  
 ‘This stew looks as if (it) will be delicious.’

Omission of *de:* produces a grammatical sentence with no change of meaning, as seen in (31).

(31)  
*ur*    *gi:*            *ɱian*    *ti:*    *lam*  
 stew this\_one as\_if IRR delicious  
 ‘This stew looks as if (it) will be delicious.’

The indefinite pronoun *məʔ* ‘someone, something’ can be substituted for *de:* with no change in meaning as shown in (32).

(32)  
*ur*    *gi:*            *ɱian məʔ*    *ti:*    *lam*  
 stew this\_one as\_if IRR delicious  
 ‘This stew looks as if (it) will be delicious.’

The fact that the indefinite pronoun can substitute for *de:* gives some support to the idea that *de:* retains some of its referential functions in this context.

***ʔɔ:r jɔʔ de:* ‘together, with one another’**

When used as a main verb *ʔɔ:r* means ‘to lead, to accompany’ and *jɔʔ* means ‘together’. In combination with *de:* in the verb complex, *ʔɔ:r jɔʔ de:* generally precedes an action verb and signifies that a group of people performed an action together, adding the meaning of accompaniment to the main verb. In (33) *ʔɔ:r jɔʔ de:* precedes the main verb *hɔ:j kʰraʃ* ‘to laugh at’ and indicates that the other people all joined together to laugh at them.

(33)  
*leʔ*    *briaŋ*            *gɔʔ*            *ʔɔ:r jɔʔ de:*    *hɔ:j kʰraʃ*    *si'na:*  
 and other\_people so\_then together laugh\_at 3du  
 ‘And other people together laughed at them.’

Omission of *de:* is grammatical and brings no change of meaning, as seen in (34).

(34)  
*leʔ*    *briaŋ*            *gɔʔ*            *ʔɔ:r jɔʔ*    *hɔ:j kʰraʃ*    *si'na:*  
 and other\_people so\_then together laugh\_at 3du  
 ‘And other people together laughed at them.’



**Pragmatically Defined Uses of *de:***

In the syntactic context where there is no antecedent reference and *de:* is interpreted as having an unspecified first person meaning, it may also communicate other meanings depending on the pragmatic context. These meanings include marking events off the storyline of a narrative, mitigating the force of a speech act and communicating friendship or intimacy.

**Marking events off the storyline**

In a personal narrative, *de:* can be used with the first person meaning to mark events off the storyline such as the author's comments on these events, or actions that are out of chronological sequence such as a flashback.

An example of marking author comments is shown in (35). The author tells of an experience where she is offered an ice-cream for the first time in her life. Her father asks her would she like one and she says yes. In relating her response, an event on the storyline, she uses the specific pronoun *ʔoʔ* '1sg'; *ho:t ʔoʔ law bəʔ* 'Then I said, "(I) do,"'. In the subsequent 4 clauses she goes on to comment on her experience. In these clauses she uses the unspecified pronoun *de:* repeatedly; *de: ʔam gə:j bəʔ* 'I had never eaten (it)', *de: law bəʔ* 'I said (I) did (want to eat it)', *de: di:m* 'I believed'. When she resumes the events of the story, she reverts to the specific pronoun *ʔoʔ* '1sg'; *ʔoʔ gəʔ glɛ:t glɛ:t jəh* 'I licked (the ice-cream), licked (and) went along'.

(35)

*ho:t ʔoʔ law bəʔ ʔnaj de: ʔam gə:j bəʔ ʔmo:j*  
and\_then 1sg say eat but unspecified NEG ever eat not\_one

*bat ʔam gə:j bəʔ ʔmo:j dia*  
Clf\_turns NEG ever eat not\_one Clf\_times

'Then I said, "(I) do," but I had never eaten (it) once; (I) had never eaten (it) even once.'

*de: law bəʔ de: di:m sah gə: ti: lam*  
unspecified say eat unspecified believe COMP 3sgn IRR delicious

'I said (I) did (want to eat it), I believed that it would be delicious.'

*ʔoʔ gəʔ glɛ:t glɛ:t jəh ho:t jəŋ ʔoʔ ŋa:ŋ jəh ka:l*  
1sg so\_then lick lick DIR and\_then father 1sg walk DIR before

'I licked (the ice-cream), licked (and) went along, and my father walked along in front.'

In (36) there is an example of marking events that are off the storyline because they are out of chronological sequence i.e. a flashback. In the first sentence the storyteller uses the specific first person singular pronoun *ʔoʔ* to refer to herself in an event on the storyline, *na:jmɔ: gəʔ ləŋ jəŋ ʔoʔ* 'The doctor then scolded my father'. As she relates events from a flashback in the next two sentences, she uses *de:* to refer to herself; *de: sir'maʔ* 'I had a fever...', *de: gəʔ ŋɔ:ŋ jəh laʔ* 'I still went out to play', *de: sir'maʔ* 'I had a fever', *de: məh kɔ:n ŋɛʔ* 'I was a small child'. When she returns to the storyline events, she resumes using the specific pronoun *ʔoʔ*; *seh ʔoʔ deʔ ʔom tʰi'le:* '(They) put saline into me'.

(36)

*na:jmɔ: gəʔ ləŋ jəŋ ʔoʔ deʔ sah jəŋ ʔoʔ ʔiaʔ*  
doctor so\_then scold father 1sg get COMP father 1sg stupid

'The doctor then scolded my father that my father was stupid.'

*jɔ:r sah ŋa:m ŋɔ:ŋ jɛt daʔ kuŋ de: sir'maʔ si'gi:*  
because when yet stay at village unspecified have\_fever today

*pim'gi: ʔam sir'maʔ sir'maʔ be:p tɪr'wa:ŋ mi:*  
tomorrow NEG have\_fever have\_fever type interval day

'Because when (we) were still staying in the village I had a fever one day, the next day (I) didn't have a fever; (I) had a fever alternate days.'

*si'gi:*        *de:*        *gɔʔ*        *ŋɔ:ŋ*        *jɔh*        *laʔ*        *pim'gi:*        *de:*  
today        unspecified        so\_then        yet        go        play        tomorrow        unspecified

*sir'maʔ*        *jɔ:r*        *de:*        *məh*        *kɔ:n*        *ŋɛʔ*        *sir'maʔ*        *riaj*        *le:w*  
have\_fever        because        unspecified        be        child        small fever        cease already

*gaj*        *jɔh*        *laʔ*        *nɛ:w*        *ʔnɛʔ*  
but\_then        go        play        type        that

‘Today I still went out to play, tomorrow I had a fever, because I was a small child, (when) the fever had ceased (I) went out to play, like that.’

*məh*        *diŋ'niʔ*        *nɔ:*        *gɔʔ*        *law*        *sah*        *məh*        *sir'maʔ pa'ʔa:t*  
be        like\_that        3pl        so\_then        say        COMP        be        malaria  
‘So (because it) was like that, they said (it) was malaria.’

*seh*        *ʔoʔ*        *deʔ*        *ʔom tʰi'le:*        *ton*        *bian*        *hok*        *daj*  
put\_in        1sg        get        saline        until        achieve        six        Clf\_bags  
‘(They) put saline into me, until (it) reached six bagfuls.’

The default reference pattern for events on the storyline with the same subject as the previous clause is a zero reference (Osborne 2009:95). The unusual referencing pattern seen in (35) and (36) with a switch from the specific pronoun to repeated use of the unspecified pronoun *de:* is a device for marking clauses which are off the mainline of events.

### **Mitigating effect**

In this same syntactic context where there is no antecedent reference, but in a different pragmatic context, the unspecified first person meaning of *de:* may be used to signal mitigation of the emotive force of a speech act.

An example of this is seen in (37), which occurs later in the same story as (35). Having tried the ice-cream and found it unpleasant, the author throws it away. When her father discovers this he scolds her and addresses her using the pronoun *de:* to soften the force of his speech.

(37)  
*baŋ*        *ʔan*        *de:*        *bəʔ*        *de:*        *laʔ*        *pit*        *gə:*  
3pl        give        generic        eat        generic        PRT        discard        3sgn  
‘Someone gives one (some) to eat, (and) one throws it away.’

In all previous conversations with her the father has used the specific pronoun *ba:* ‘2sgf’ to address her. The change to *de:* signals a mitigating of the accusatory force of the second person pronoun. Because the use of *de:* includes the speaker, it necessarily lessens the accusatory impact of the rebuke. A similar effect in English might be achieved by using first person plural instead of second person singular, e.g. ‘When someone gives us an ice-cream, we don’t throw it away’.

### **Signalling friendship and intimacy**

Another conversational use of *de:*, where there is no antecedent reference, is as a vocative to signal friendship.

The pragmatic context for this use is specific to female speakers who are close friends and age mates. It does not substitute for a specific pronoun, but rather for the name of the other person and may be interpreted as meaning ‘friend’, as shown in (38).

(38)  
*de:*        *ʔaʔ*        *jɔh*        *laʔ*        *ta'la:t*        *jɔh*  
friend        1du        go        play        market        go  
‘Friend, (let) us go to the market, ok?’

In another pragmatic context, that of close male-female relationships, *de:* is used as a form of intimate address. Either partner will address the other as *de:* and refer to themselves as *briaŋ*, which in most contexts means 'other people'. An example of this is shown in (39), where a young woman speaks to her fiancé on the phone of how she misses him. She uses *de:* to address him and refers to herself as *briaŋ*.

(39)

<i>de:</i>	<i>gɔʔ</i>	<i>sir'ʔe:ŋ</i>	<i>briaŋ</i>	<i>ʔoh</i>
unspecified	so_then	think_about	other_people	PRT
<i>briaŋ</i>	<i>sir'e:ŋ</i>	<i>de:</i>		
other_people	think_about	unspecified		

'Are you thinking of me? I am thinking of you (all the time).'

### Relationship Between the Uses of *de:*

There is a semantic core that is common to at least three of the four syntactically distinct uses of *de:* described in this paper. That is the element of co-referentiality. In the 'Co-referential function', the 'Unspecified first person meaning' and 'Alone', *de:* identifies a referent that has already been supplied either syntactically or from the speech situation or cultural context. Thus these forms are clearly semantically related and are probably one word with different uses rather than homophonous forms.

The exact function of the fourth use, 'Particle in the verb complex', is not yet clearly defined. Unlike the other three uses which are associated with reference, it is associated with the world of events. Thus it is not yet clear whether this is another use of the same word or merely a homophonous form.

Of the three related uses, there is no synchronic evidence to point to which use was historically prior to the others. In uses 1 and 3, 'Co-referential function' and 'Alone', the co-referent is syntactically available, while in use 2, 'Unspecified first person meaning', it is not. One could characterise the 'Unspecified first person meaning' as the more general interpretation which holds whenever the conditions for the other two syntactically constrained uses are not met. Also, this use has a number of wider pragmatic interpretations and perhaps on these grounds it might be hypothesised to be prior to the other uses. Shorto (2006) reconstructs a proto-Mon-Khmer \**de?* 'reflexive pronoun' which has reflexes in Palaungic, Bahnaric and Nicobarese in addition to Kmhmu'. Further research in the documented history of related languages may provide evidence for prior occurrence of the syntactic co-referent usage or the unspecified usage. What is clear from this study is that speakers have specific ways to prevent ambiguity in meaning between the various uses of *de:*. It is these syntactic and pragmatic constraints that are the focus of this paper.

### Conclusion

Four syntactically defined uses of the unspecified pronoun *de:* have been characterised. Where there is an antecedent reference, *de:* has a co-referential function with scope within a main clause, over an embedded clause, an adverbial clause or a tail-head linkage clause, but not over a co-ordinate clause. The scope of the co-referential function of *de:* has implications for identifying clause boundaries and the nature of the relationship between clauses.

Where there is no antecedent reference, *de:* has an unspecified first person meaning which is identified from the speech situation or cultural context. In certain pragmatic contexts, this unspecified first person meaning of *de:* may be interpreted as marking events off the storyline of a narrative discourse, mitigating the emotive force of a speech act, or signalling friendship or intimacy.

When *de:* is the final element in a clause, it means 'alone'. In the fourth syntactic environment, as part of the verb complex, *de:* occurs with other verbal elements to add aspectual or adverbial components of meaning to the main verb.

Although this function of *de:* in the verb complex and its relation to the other uses of *de:* is not yet clearly defined, the first three uses of *de:* share the common semantic core of co-referentiality and thus are

taken to be different uses of a single word. Their use is governed by syntactic and pragmatic factors that allow speakers to unambiguously identify referents and also communicate speaker attitude.

### References

- Berg, René van den, with the assistance of Mark Taylor. 1988. *Some Khmu' particles: A preliminary analysis*. Anduze, ms.
- Osborne, R. A. 2009. *Who's who in Kmhmu': Referring expressions and participant identification in selected Kmhmu' narrative texts*. Chiang Mai: PYU thesis. [http://www.sil.org/silepubs/Pubs/52451/Osborne\\_WhosWhoKmhmu\\_52451.pdf](http://www.sil.org/silepubs/Pubs/52451/Osborne_WhosWhoKmhmu_52451.pdf)
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics 579.
- Suksavang Simana', Somseng Sayavong & Elisabeth Preisig. 1994. *Kmhmu' – Lao – French – English dictionary*. Vientiane: Ministry of Information and Culture, Institute of Research on Culture.
- Suwilai Preamsirat. 1987. *A Khmu grammar*. Pacific Linguistics Series A – No. 75, Papers in South-east Asian Linguistics 10. 1-143.

## About *ʔaɔj* in Contemporary Khmer.

Denis Paillard

Laboratoire de linguistique formelle  
Université Paris Diderot

‘GIVE-verbs’<sup>97</sup> have given rise to numerous studies in various languages (cf. Newman: 1996, 1997 for a bibliography). In most languages, GIVE verbs have other uses than that described by Newman as the “act of giving”. Heine & Kuteva (2002) mention the following uses: “benefactive”, “dative”, “causative”, “purpose”. In some languages, GIVE-verbs are used as a passive form (cf. in particular Paris: 1998, Yap & Iwasaki: 1998, 2003). It should be noted that depending on the languages, GIVE verbs have widely differing uses, which suggests that the semantic core of the word cannot be reduced to that of transfer. Just compare for instance the uses of *give* in English and *donner* in French<sup>98</sup>.

Our study of the uses and values of *ʔaɔj* in Khmer will definitely not consider the transfer value as the basic one. Grounding on an inventory as complete as possible of all these values in contemporary Khmer, we bring forward a characterization of the semantic identity of *ʔaɔj* which makes it possible to account for the range of all its uses, showing what they have in common and in what they differ. Such a characterization will neither resort to the notion of grammaticalization nor that of desemantisation. On this point, our analysis differs from other studies dealing with GIVE-verbs in South-East Asian languages, such as Khmer (Bisang, 1996) Thai (Rangkupan, 2005, Thepkanjana & Uehara, 2008) or Mon (Jenny (2006).

### 1. Inventory of the uses of *ʔaɔj*

Our starting point is not a semantic one (we do not base on the “transfer” value) but a syntactic one, namely the various syntactic constructions where *ʔaɔj* is to be found. We present ten groups of uses of *ʔaɔj* ; each type of use is illustrated through a series of representative data and corresponds to specific syntactic properties. The terms used to refer to the various uses are purely conventional and are presented with the corresponding syntactic patterns.

#### 1.1. ‘Transfer’: $N_1 \text{ ʔaɔj } N_2 N_3 / N_1 \text{ ʔaɔj } N_2 V N_3$

- 1a) *koat ʔaɔj ba:j cʰkae maɔŋ pɔnman*  
3sg ʔaɔj rice dog hour how many  
“At what time does he give the dogs their meal?”
- 1b) *koat ʔaɔj ba:j cʰkae si: maɔŋ pɔnman*  
3sg ʔaɔj rice dog eat hour how many  
“At what time he will give the dogs their meal?” (they are starving)

---

<sup>97</sup> We use here GIVE as a generic term referring to verbs which in various languages can express the notion of transfer. This does not in any case mean that those verbs are synonyms of *give* in English.

<sup>98</sup> From a diachronic point of view, cf. the text by E. Benveniste about the notion of gift in *Vocabulaire des institutions indo-européennes* (1993). Benveniste shows that the bases giving rise to *give* and *donner* correspond to quite different representations. About *yaru* in Japanese, let's mention Oguma (2006).

---

Paillard, Denis. 2011. “About *ʔaɔj* in Contemporary Khmer.” In Srichampa and Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.124-137.  
Copyright vested in the author.

- 2a) *campush ka:ɲiʒ samnaŋ koat ʔaɔj ø kam<sup>9</sup>ka:*  
 as regards work construction 3sg ʔaɔj ø worker  
*muʒj t<sup>h</sup>ɲaj pram rɔ:j riʒl*  
 one day five hundred riel  
 “As regards the construction works, he gives the workers 5000 riels a day” (workers salary)
- 2b) *campush ka:ɲiʒ samnaŋ koat ʔaɔj tʂi kam<sup>9</sup>ka:*  
 as regards work construction 3sg ʔaɔj go worker  
*muʒj t<sup>h</sup>ɲaj pram rɔ:j riʒl*  
 one day five hundred riel  
 “As regards the construction works, he spends 5000 riels a day for the workers”  
 (manager’s expenses for the salaries)

As shown by examples (1) and (2), it is possible to have *ʔaɔj* either alone or combined with another verb (namely *tʂi* ‘go’ in most cases). With *ʔaɔj* alone, the Ns stand for arguments ((‘agent’, ‘object’, ‘beneficiary’) of *ʔaɔj* and the interpretation is that of transfer. With *ʔaɔj* combined with another verb, the Ns coming after *ʔaɔj* have a different status: those Ns are arguments of the second verb, whereas the preceding N only is an argument of *ʔaɔj*. The interpretation is different in (2b) the workers are not the beneficiaries but the recipients of the transfer: “money – 5000 riels – go to – workers» (the aim is to reckon the amount of money allotted to the labour costs).

The difference between (1a) and (1b) is not as important: in (1a) what is at stake is the everyday usual process of feeding the dogs, whereas (1b) refers to starving dogs waiting for the food: the sequence on the right is to a certain extent independent of that on the left.

Comparing those two constructions shows that the demarcation between *ʔaɔj* expressing the transfer and *ʔaɔj* as a ‘causative’ is quite blurred. With the transfer value, *ʔaɔj* commutes with *cuun* (with *cuun* the beneficiary has the highest social or parental status). But this local synonymy should not conceal the semantic differences between the two verbs: with *cuun*, the focus is on the relation between the two characters through their social differences. *Cuun* has not as many uses as *ʔaɔj*. Contrary to *ʔaɔj*, *cuun* can combine with the “causative” prefix *baŋ*. This difference can already be found in old Khmer

**Benefactive<sup>99</sup>:**  $N_1 V (N_2) ʔaɔj N_3 / N_1 V (N_2) ʔaɔj tʂi N_3$

This second use of *ʔaɔj* is closely related to the transfer value. The main difference lies in the presence of a verb on the left of the sequence: the process referred to by this verb is presented as completed for the benefit of a character appearing in the sequence on the right. From this point of view, the function of *ʔaɔj* is to point out who benefits by this process.

- 3) *sɔm cɔh t<sup>h</sup>laj ʔaɔj k<sup>h</sup>nɔm*  
 ask descend cost ʔaɔj 1sg  
 “Reduce the price for me”
- 4) *jɔk cnaŋ t<sup>h</sup>ɔm nuh mɔk ʔaɔj ma? p<sup>h</sup>aŋ*  
 take pot big dem come ʔaɔj mother part.  
 “Bring me this big pot” (the mother to her daughter)
- 5a) *koat tʂɲ siʒp<sup>h</sup>əi ʔaɔj (cuun) mda:j koat*  
 3sg buy book ʔaɔj cuun mother 3sg  
 “He bought a book for his mother”

<sup>99</sup> We take over this terminology from the literature dealing with GIVE-verbs.

- 5b) *koat tɔŋ sɨpʰəi វាជ (cuun) tɔi mda:j koat*  
 3sg buy book វាជ cuun go mother 3sg  
 “He bought a book, intended for his mother”

The benefactive is the second case where វាជ can commute with *cuun*. As shown by (5b), វាជ as well as *cuun* can be followed by the verb *tɔi* ‘go’. As for transfer, the presence of *tɔi* imparts a stronger autonomy to the sequence coming after វាជ: buying the book, on the one hand; allocating the book to the mother on the other hand.

#### Delegative: $N_1 V (N_2) វាជ N_3$

The delegative means that someone is making something instead of somebody else: in (6) ‘I’ drive the UNESCO manager’s car instead of him. The presence of a verb in the left sequence must be pointed out; the sequence on the right is limited to a single N whose relation to the process referred to by the V in the left sequence is the same as that involving the syntactic subject. The latter is presented as substituting for the former, which leads to considering one and the same process for the two agents.

- 6) *kʰnɔm bəʔk la:n វាជ nɨʒjuːʔ ju.nnesko*  
 1sg drive car វាជ manager UNESCO  
 “I drive the UNESCO manager’s car as his driver”
- 7a) *bakpraɛ វាជ ភាវាបាត និហ វាជ kʰnɔm pʰaŋ*  
 translate text dem. វាជ 1sg part.  
 “Translate this text for me!”
- 7b) *kʰnɔm bakpraɛ វាជ ភាវាបាត និហ វាជ koat*  
 1sg translate text dem. វាជ 3sg  
 “I translate this text instead of him (delegative) / for him (benefactive)”

As shown by the double interpretation made possible by (7b), the benefactive and the delegative seem pretty similar. To a certain extent at least, for on the other hand, they remain quite distinct: introducing *tɔi* after វាជ is possible only with the “benefactive” interpretation. The impossibility to have *tɔi* or any other V in the right sequence with the delegative can be accounted for by the implicit presence of a V identical to that in the left sequence, in accordance with the delegative interpretation

#### 1.4. $N_1 ku3 វាជ V$

*ku3*<sup>100</sup> specifies the preceding term as the very best item liable to fill such or such place of argument of the verb coming after *ku3*. Between *ku3* and the V it is possible to insert វាជ.

- 8) *bahcʰnaɔt lɔ:ʔ nih kʰmizn kaʔnaʔpaʔ na: ku3 វាជ cap វា:ram sah*  
 election time dem. neg. exist party indef. ku3 វាជ interest at all  
 “Regarding these elections, there isn’t one single party worthy of attention”
- 9a) *ba:ŋ miznɔp ci:wət baŋ kamsat mɛ:n*  
 older sibling name life older sibling miserable true  
*ku3 វាជ វា:nət nah*  
 ku3 វាជ take pity very  
 “Meanup, you have an unhappy life inspiring me compassion”
- 9b) *ba:ŋ miznɔp ci:wət baŋ kamsat mɛ:n ku3 វា:nət nah*  
 older sibling name life older sibling miserable true ku3 take pity very  
 “Meanup, it is true that you have an unhappy life, one feels compelled to compassion for you (one =

<sup>100</sup> *ku3* originates in a verb (‘suit’); it keeps this predicative value in some expressions such as *ku3 pɔm ku3* (*ku3 neg ku3*) ‘is it all right or not?’ and *ku3 វាជ រិ ku3 pɔm វាជ* ‘should we give it to him or not?’

one kind hearted)”

The difference between (9a) (presence of *ʔaɔj*) and (9b) lies in the fact that the relation between the N standing as *kuz* scope and the verb is actualized, whereas with *kuz* alone the relation is only potential, the N being in this case semantically closely related to the V place of argument. This is confirmed by the impossibility for *kuz* to be used alone in (10): the red complexion of the face has nothing to do *a priori* with being fear inspiring; the redness as an argument of “fear inspiring” actually comes from *ʔaɔj*.

- 10) *pe:l koat kʰɔŋ mək koat laŋ kraham kuz ʔaɔj kʰla:c*  
 time 3sg be angry face 3sg ascend red kuz ʔaɔj fear  
 “When he gets angry, his face turns red to the point of getting fear inspiring”

### 1.5 Jussive: V ʔaɔj PRED

*ʔaɔj* can appear after a V expressing an injunction. The sequence that follows is a predicate specifying the process expressed by V or one of the arguments. The alternative  $\emptyset / ʔaɔj$  corresponds to a difference of interpretation: without *ʔaɔj* the term plainly specifies the way the process is carried out; with *ʔaɔj*, the qualification by the postposed predicate is presented as a target.

- 11a) *nəi ø sɲɪzm (sdap ke: nɪjɪzj sɔn)*  
 remain  $\emptyset$  still (listen people speak first)  
 “Keep still! (listen to the others first)”
- 11b) *nəi ʔaɔj sɲɪzm (lec tu:k ʔejlej haɔj)*  
 remain ʔaɔj still (sink boat at this time PART.)  
 “Keep still! Or you’ll make the boat sink!”
- 12a) *haet ʔej ka: co:lcɔt nɪjɪzj kʰlaŋ ʔaŋcɔŋ*  
 reason what PART. like speak loud so  
 “But why do you enjoy speaking so loudly!”
- 12b) *sɔm nɪjɪzj ʔaɔj kʰlaŋ ciɔŋ nɔŋ ba:n te:*  
 part.politeness speak ʔaɔj loud more deict. can PART.  
 “Could you speak louder?”
- 13a) *nɪjɪzj muɔj muɔj ba:n te:*  
 speak one one can PART.  
 “Could you speak slowly?”
- 13b) \* *nɪjɪzj ʔaɔj muɔj muɔj ba:n te:*  
 speak ʔaɔj one one can PART.

In (13b), the term *muɔj muɔj* specifying ‘speak is not a predicate (‘slowly’ is the reduplication of the numeral ‘one’), which accounts for the impossibility to have *ʔaɔj* :

- 14a) *nɪjɪzj kʰlej kʰlej tɔi ʔah pe:l haɔj*  
 speak short short go finish time PART.  
 “Make it short we have no time left!”
- 14b) *nɪjɪzj ʔaɔj kʰlej tɔi...*  
 speak ʔaɔj short go...  
 “To sum up/to conclude...”

Should also be noted the difference of interpretation between (14b) and (14a) where the postposed predicate is reduplicated: the reduplication means that the validation of the predicate depends but on the interlocutor.



As shown by the series (15), this construction with ʔaɔj introducing a specification of the process is not possible with an assertion, a fact which confirms that with ʔaɔj this qualification is presented as a target. When ʔaɔj is replaced by *ba:n* ('get), the target is presented as actually being reached (cf. (15c)).

15a) *rət ʔaɔj liən*  
run ʔaɔj fast  
"Run fast"

15b) *koat rət ø (\*ʔaɔj) liən*  
3sg run ø \*ʔaɔj fast  
"He runs fast" (ʔaɔj is impossible in an assertion)

15c) *koat rət ba:n liən*  
3sg run acquire fast  
"He managed to run fast"

Finally, it must be noted that this construction can be met in a causative construction (second ʔaɔj):

16) *jək kambət nuh tɛi ʔaɔj pa: samliɛŋ ʔaɔj mət tɛi*  
take knife dem. go ʔaɔj dad sharpen ʔaɔj sharp part.  
"Bring this knife to your father, for him to sharpen it"

In this use, it is not possible to have the modal negation *kəm* unlike what is to be found with the permissive (1.6) and the causative (1.7).

**Permissive: ʔaɔj N<sub>2</sub> V / N<sub>1</sub>ʔaɔj N<sub>2</sub> V (N<sub>3</sub>)**

*Permissive* refers to constructions where ʔaɔj is followed by a sequence N<sub>2</sub> V (N<sub>3</sub>) where N<sub>2</sub> refers to a human being and V to a process aimed at by N<sub>2</sub>, but whose achievement depends *a priori* on the subject of ʔaɔj. The latter can be either the locutor (ex. (17a)) or a N or a Pronoun (ex. (17b)). This use is possible in an injunction as well as in an assertion.

17a) *ʔaɔj kʰnəm tɛi pʰa:ŋ*  
ʔaɔj 1sg go also  
"Let me come with you!"

17b) *koat ʔaɔj Dara: tɛi*  
3sg ʔaɔj Dara go  
"He let Dara come with him"

**Causative: ʔaɔj N<sub>2</sub> V / N<sub>1</sub>ʔaɔj N<sub>2</sub> V**

The causative is framed in the same construction as the permissive, the difference lying in the fact that nothing is said about the relation linking *a priori* N<sub>1</sub> and the process expressed by the V in the sequence coming after ʔaɔj.

18a) *pe:l koat mək dal ʔaɔj koat cam kʰnəm bantec*  
time 3sg come arrive ʔaɔj 3sg wait 1sg a little  
"When he gets there, tell him to wait for me a little!"

18b) *koat ʔaɔj wiz cam kʰnəm bantec*  
3sg ʔaɔj 3sg wait 1sg a little  
"He told him to wait for me for a while"

19) *nijɔj ʔaɔj kʰnəm sdap pʰa:ŋ*  
tell ʔaɔj 1sg listen PART.  
"Tell me about it for me!"

1.8. P1 *ʔaɔj* P2:  $N_1 V \text{ } ʔaɔj \text{ } N_2 V (Z) / N_1 V N_2 \text{ } ʔaɔj \text{ } V (Z)$

*ʔaɔj* is used as a relator between the propositions P1 and P2. The first clause P1 refers to an event bringing about the event introduced by the second clause. The verbs liable to appear in P1 are ‘causative’ verbs or *verba dicendi*. It is possible to find in Gorgoniev (1966) and in Bisang (1992) a representative list of these verbs. It must be noted that for a part of these verbs, a construction without *ʔaɔj* is also possible. Besides, in some cases, the  $N_2$  subject of the verb in P2 can stand before *ʔaɔj*. The examples hereafter help to understand the difference in the interpretation of these three constructions.

20a) *kʰɲɔm noam koat mək Ø mɔ:l ptʰɛah*  
1sg lead 3sg come Ø look house

“I bring him along to see the house”: either 1) “for him to see my house and know how it looks like”; or 2) “for him to know where I’m living”.

20b) *kʰɲɔm noam koat mək ʔaɔj mɔ:l ptʰɛah*  
1sg lead 3sg come ʔaɔj look house

“I bring him along to see the house”: either 1) the house is for sale; or 2) someone is needed as a caretaker for the house.

In (20a) (with Ø) the two events are part of the same whole event associating the two subjects, whereas in (20b) the two events (therefore the two subjects) are autonomous and independent one from the other: the V in P2 is aimed as a target involving the  $N_2$  subject of V2, which is confirmed by the impossibility to have *ʔaɔj* in (21) and, conversely that of Ø in (22a):

21) *sʔaek pe:l kʰɲɔm noam kən Ø (\*ʔaɔj)*  
to tomorrow time 1sg lead child Ø

*tʂi sa:la: cam kʰɲɔm chist cɔ:l lep*  
go school wait 1sg. existence enter play

“To tomorrow, I’ll avail on the occasion of taking my son to school to come and see you”

22a) *kʰɲɔm ciz neak noam wiz ʔaɔj (\*Ø) skoal caowa:j kʰaet*  
1sg. be person lead 3sg. ʔaɔj know chief province

“I’m the one who did introduce him to the province governor”

22b) *kʰɲɔm ciz neak noam ʔaɔj wiz skoal caowa:j kʰaet*  
1sg. be person lead ʔaɔj 3sg. know chief province

“(he wanted to be introduced to the province governor and) I am the one who managed a meeting”

In (22b) where the  $N_2$  subject of V2 comes after *ʔaɔj*, the meeting with the governor is presented as a goal for  $N_2$ , independently of who made it possible for this meeting to take place: P1 is therefore subordinated to the achievement of P2. A difference similar to that noted between (22a) and (22b) can be found again considering (23a-b); in (23a), the event expressed by P1 is at stake: what matters is not mainly knowing where (P2) takes place but rather knowing who made it possible for (P2) to happen (the event expressed by P2 comes under a negative judgment). On the reverse, in (23b) P1 is totally subordinated to P2: knowing the location is an appraised information, whereas knowing who made it possible for P2 to be achieved is a secondary matter.

23a) *neak na: noam wiz ʔaɔj skoal kanlaeɲ nih*  
person indef. lead 3sg. ʔaɔj know place dem.

“Who made him know this place?” (he wasn’t supposed to know this place)

23b) *neak na: noam ʔaɔj wiz skoal kanlaeɲ nih*  
person indef. lead ʔaɔj 3sg. know place dem.

“Who made him know this place?” (it’s a good thing)

There being no interdependence between the subjects of the two events when N2 comes after វាជ is confirmed by (24) and (25) where P2 refers to a detrimental event. In this case, the N subject of the V in P2 necessarily comes after វាជ:

- 24) *kʰnɔm so:mtɔh kʰnɔm noam វាជ baŋ piʔbaʔ daɔjsaʔ kʰnɔm*  
 1sg. beg forgiveness 1sg. lead វាជ 2sg. difficult because 1sg.  
 “I beg your pardon for getting you into a difficult situation”
- 25) *វាឡើ ត្បែរ: វាឡើ វាឡើ noam វាជ ke: មនៈល ឡើ វាជ គ្រុនៈ*  
 2sg. do so 2sg. get វាជ 3sg. look easy get family  
 “What you have done makes our family open to contempt” (A father rebukes his son who got arrested by the police for a theft)

(25) means that the neighbours were only waiting for an occasion to mock at the family (preconstruction of P2), and it’s the son who gave such an occasion.

We present two other series confirming the difference of interpretation due to the position of the N<sub>2</sub> in the sequence:

26) A mother is complaining about his son who doesn’t want to study:

- 26a) *វា: ម្តេជ នង បាឺ ក្បែរ មន បាឺកាម វិឺ*  
 anaph. one deict. if 1sg. neg. force 3sg.  
*វាជ រិន តៈ វិឺ វាត រិន តៈ*  
 វាជ study part. 3sg. neg. study part.  
 “This one, if I don’t force him to study, he won’t do it (he refuses to study, he has to be made to do it)”

- 26b) *វា: ម្តេជ នង បាឺ ក្បែរ មន បាឺកាម វាជ*  
 anaph. one deict. if 1sg. neg. force វាជ  
*វិឺ រិន តៈ វិឺ វាត រិន តៈ*  
 3sg. study part. 3sg. neg. study part.  
 “This one, if I don’t force him to study, he won’t do it (he will never consider doing it on his own initiative, he has to be prompted to do it)”

In (26a), the presence of the subject of P2 preceding វាជ emphasizes the interdependence between mother and son as regards studying, whereas in (26b) (subject of P2 coming after វាជ) ‘studying’ is presented as an aim the mother endeavours to reach. It can also be observed that វាជ can hardly be removed.

- 27a) *caowa:j sre:j វាឡើ មន បា:n បាឺកាប វាជ វាឡើ*  
 boss woman 2sg. neg. get order វាជ 2sg.  
*ត្បែរ: តៈ វាឡើ ត្បែរ: ក្បែរ វាឡើ តា*  
 do part. 2sg. do body 2sg. restr.  
 “Your boss did not tell you to do it, you did it on your own ?!”

- 27b) *caowa:j sre:j វាឡើ មន បា:n បាឺកាប វាឡើ វាជ*  
 boss woman 2sg. neg. get order 2sg. វាជ  
*ត្បែរ: តៈ កាត វាជ វាឡើ ផ្សេង ត្បែរ: តា*  
 do part. 3sg. វាជ person else do part.  
 “Your boss did not ask you to do it, she asked somebody else to do it”

- 27c) *caowa:j sre:j ក្បែរ បាឺកាប មនៈជ ក្បែរ វាជ*  
 boss woman 1sg. order mother 1sg. វាជ  
*ក្បែរ រក លុយ វាជ កាត*  
 1sg. seek money វាជ 3sg.  
 “My boss ordered my mother to ask me to go and fetch money for her (the boss)”

In (27a) and (27b), the verb *baŋkoap* ‘order’ (P1) comes under a negation: with the anteposition of N<sub>2</sub> (ex. 27a), the independence of P2 is reinforced: P1 is not the event which brings P2 about. With the postposition of N<sub>2</sub> (ex. 27b), what comes under the negation is the identity of the agent of P2. Finally, in (27c) where we find both a N object of *baŋkoap* ‘order’ and a postposed N, the N object takes over from the agent of *baŋkoap* as regards the validation of the process referred to by P2.

When the V of P1 is a strict causative verb, such as *bandal* ‘cause’, N<sub>2</sub> postposition is only possible: owing to its semantic value, the verb *bandal* cannot express a relation between the two subjects:

- 28) *koat k<sup>h</sup>mizn t<sup>h</sup>wə: ʔej dael mizn pra:jaɔj dal kɾəmħon te:*  
 3SG. NEG. do INDEF. REL. have interest reach Firm PART.  
*koat bandal ʔaɔj taε kɾəmħon k<sup>h</sup>at lɔj nɔŋ*  
 3SG. cause ʔaɔj only firm lose Money DEICT.  
 “He doesn’t do anything useful for the firm, all he does is getting it into losses!”

We now present a last series with *t<sup>h</sup>wə:* (‘do’) as a causative verb and with an inanimate N as the subject of the V of P2.

- 29a) *t<sup>h</sup>wə: la:n ʔaɔj (\*Ø) c<sup>h</sup>eh sen cam tsi*  
 make car ʔaɔj \*Ø work first wait go  
 “Fix the car before you leave!” (manage to make it work before leaving)
- 29b) *nɛak na: t<sup>h</sup>wə: la:n ʔaɔj (\*Ø) c<sup>h</sup>eh*  
 person indéf. make car ʔaɔj \*Ø work  
 “Who fixed the car?” (I thought there was no one to fix it)
- 29c) *nɛak na: t<sup>h</sup>wə: ʔaɔj (\*Ø) la:n c<sup>h</sup>eh*  
 person indéf. make ʔaɔj \*Ø car work  
 “Who made the car work?” (I don’t want it to be put in working order)
- 29d) *pe:l k<sup>h</sup>ɲom t<sup>h</sup>wə: Ø (\*ʔaɔj) la:n c<sup>h</sup>eh haɔj*  
 time 1sg. make Ø \*ʔaɔj car work part.  
*kɲom ka: riɜpcam dam ba:j*  
 1sg. then prepare cook rice  
 “After getting the car fixed, I cooked the rice”

In (29), the variation (with an equal lexicon) involves a. the place of the subject of P2 (*la:n* ‘car’) before or after *ʔaɔj*; b. the presence or not of *ʔaɔj* between P1 and P2. In (29a) and (29b), *la:n* must necessarily stand before *ʔaɔj*: P1 is first. In (29c) *la:n* must necessarily stand after *ʔaɔj*: stating that “the car is working” (P2) is what leads to wondering who is responsible for fixing it (cf. the comments given on the above 22b et 23b examples). The impossibility to have *ʔaɔj* in (29d) is due to the fact that the event ‘fix the car’ is part of a succession of events given in order and therefore taken as a whole. In other words, entering a succession makes it impossible for P2 to be dissociated from P1

The construction **P1 ʔaɔj P2** shows various degrees of autonomy of P2 as regards P1, depending on the position of the subject of the V in P2. When N2 stands before *ʔaɔj*, the event expressed by P1 is prominent (the autonomy of P2 is quite relative). When N2 comes after *ʔaɔj*, P2 is presented as a goal to be reached (or that has been reached), and P1 is subordinated to the achievement of this goal.

### 1.9. Optative: *ʔaɔj+taε p* and *kɔm+ʔaɔj+taε p*

In this case *ʔaɔj* is at the initial position<sup>101</sup> and is followed by *taε* (‘only’). *ʔaɔj* combined with *taε*, as a marker of restriction, means that **p** is the only thing taken into account as regards the evaluation / interpretation of a situation contextually introduced. This situation takes into account a set of possible options,

<sup>101</sup> It can be preceded by the modal negation *kɔm*, cf. the hereafter examples (33) and (34).

and the selection of **p** as the only one actually selected is presented as kept away from any kind of control, whether it be virtual or actualized. This accounts for the fact that nothing can be found on the left of ʔaɔj, not even a subject / agent. Given all that can be associated with this set, the only thing deemed relevant is **p**, whether actualized or not.

30) *wie sɔkcɔt twɜː ta:m ʔaɛŋ tɛaŋʔah ø ʔaɔj taɛ ʔaɛŋ nijsɔj lɔʔaː nɔŋ wiɜ*  
 3sg. agree do follow 2sg. all ø ʔaɔj taɛ 2sg. speak well with 3sg.  
 “He agrees to do anything you want, on the only condition that you speak kindly to him”

31) *ʔaɔj taɛ baːn tɕi ʔaɔj kʰŋɔm twɜː ʔɛj kaː kʰŋɔm twɜː daɛ*  
 ʔaɔj taɛ acquire go ʔaɔj 1sg. do what part. 1sg. do also  
 “On the only condition that I’m free to go away, I can be asked anything, I will not refuse to fulfil it”

32) *kʰmeːŋ nɔŋ sralaŋ ʔəw wie ʔaɔj taɛ wie*  
 child Deict. love father 3sg. ʔaɔj taɛ 3sg.  
*kʰəŋ ʔəw wie wie ɲɔŋəŋ pʰliɜm*  
 see Father 3sg. 3sg. smile immediately  
 “This child does love his father: catching sight of his father is enough for him to smile”

In (32), what is selected is the complex sequence formed by “see his father – smile right after».

Examples of ʔaɔj taɛ with the modal negation *kɔm*:

33) *wie sɔkcɔt twɜː ta:m ʔaɛŋ tɛaŋʔah kɔm*  
 3sg. agree do following 2sg. everything kɔm  
*ʔaɔj taɛ ʔaɛŋ nijsɔj ʔaːkraʔ nɔŋ wiɜ*  
 ʔaɔj taɛ 2sg. speak nastily with 3sg.  
 “He agrees to do anything you want on the only condition that you do not speak to him nastily”

34) *wie krɔantɛ jɔːk ʔaɛŋ tɕi kʰɔm mɜn ʔɛj teː kɔm ʔaɔj*  
 3sg. only take 2sg. go put in jail neg. indéf part. kɔm ʔaɔj  
*taɛ wie waj tʰaɛmtɜt nɔh tok ciɜ samnaːŋ tɕi*  
 taɛ 3sg. beat addition deict. consider be luck part.  
 “He (the cop) put you into police custody, no use making a fuss about it, the important thing is that he didn’t rough you up, you should consider yourself lucky!”

In (33) and (34) *kɔm* indicates that neg + p (‘not speak badly) is the only important thing.

Let’s point out that in some cases, either ʔaɔj taɛ or *kɔm ʔaɔj taɛ* can be found:

34a) *kʰmeːŋ nɔŋ mɜn dɜŋ ciɜ jaːŋmeːc wie ʔaɔj*  
 child deict neg know be how 3sg ʔaɔj  
*taɛ wie kʰəŋ ʔəw wie wie jɔm pʰliɜm*  
 taɛ 3sg see father 3sg 3sg cry immediately  
 “I don’t know what is the matter with this child: as soon as he sees his father, he starts crying”

34b) *kʰmeːŋ nɔŋ mɜn dɜŋ ciɜ jaːŋmeːc wie kɔm ʔaɔj*  
 child deict neg know be how 3sg kɔm ʔaɔj  
*taɛ wie kʰəŋ ʔəw wie wie jɔm pʰliɜm*  
 taɛ 3sg see father 3sg 3sg cry immediately  
 “I don’t know what is the matter with this child: seeing his father is enough to start crying right away”

With ʔaɔj taɛ, considering “see his father - cry” is nothing but a mere statement, grounding the remark “I don’t know what is the matter with this child”. The presence of *kɔm* means that the reference value is not “see his father - start crying”, which underlines the contradiction between “see his father” and “start

crying”. This is confirmed by the fact that in (32) only *ʔaɔj tae* is possible, *kɔm ʔaɔj tae* being impossible (the sequence which makes the scope for *ʔaɔj tae* is appraised).

### 1.10. *tʰa ʔaɔj* and *sdej ʔaɔj*: ‘criticize’, ‘rebuke’: [ $N_2 V$ ] $N_1$ *tʰa ʔaɔj* / *sdej ʔaɔj* $N_2$

Combined with two *verba dicendi* *tʰa:* and *sdej*, *ʔaɔj* mean ‘criticize’ (*tʰa:*), ‘rebuke’ (*sdej*). This interpretation is framed by the following rules a. the N coming after *ʔaɔj* refers to the person being criticized or rebuked; b. the sequence expressing the event giving rise to the criticism or the rebuke is present in the left context, but resuming it is impossible after *ʔaɔj*.

35) *mɔn nɔŋ tʰa: (\*sdej) ʔaɔj ke: mɔl kʰluɔn ʔaɔŋ sɔn*  
 before part. *tʰa:* \**sdej* *ʔaɔj* people look oneself 2sg first  
 “Before criticizing the others, just look at yourself!” (Someone criticizes the people around him about what they wear)

36a) *baɜ wɜ mɔ:k jɪt kɔm tʰa: ʔaɔj wɜ*  
 if 3sg come late neg.mod. *tʰa:* *ʔaɔj* 3sg  
 “If he’s late, don’t be critical!”

36b) *baɜ wɜ mɔ:k jɪt kɔm sdej ʔaɔj wɜ*  
 if 3sg come late neg.mod. *sdej* *ʔaɔj* 3sg  
 “If he’s late, don’t rebuke him!”

37) mother asks her husband not to rebuke their son for being out every night

37a) *pe:l wɜ mɔk kɔm sdej ʔaɔj wɜ cam*  
 time 3sg. come neg.mod. *sdej* *ʔaɔj* 3sg. wait  
*sʔaek nijsj pɔŋjɔl wɜ ta:m samruɔj tɔi*  
 tomorrow speak explain 3sg. following quiet go  
 “When he comes back, don’t shout at him ! Wait until tomorrow to have a quiet explanation with him!”

37b) *pe:l wɜ mɔk kɔm tʰa: ʔaɔj wɜ cam*  
 time 3sg. come neg.mod. *tʰa:* *ʔaɔj* 3sg. wait  
*ʔaek nijsj pɔŋjɔl wɜ ta:m samruɔj tɔi*  
 tomorrow speak explain 3sg. following quiet go  
 “When he comes back, don’t criticize him! Wait until tomorrow to have a quiet explanation with him!”

38) *baɜ wɜ mɔ:k jɪt kɔm tʰa: (\*sdej) ʔej ʔaɔj wɜ*  
 if 3sg come late neg.mod. *tʰa:* \**sdej* indef *ʔaɔj* 3sg  
 “If he’s late, don’t tell him anything!”

39) *ʔaɛŋ sdej (\*tʰa:) ʔaɔj koat riɔŋ ʔej pɔt tɔi*  
 2sg. *sdej* *tʰa:* *ʔaɔj* 3sg. matter indef. be true go  
 “What are you rebuking him about exactly?”

The differences of interpretation between *tʰa: ʔaɔj* and *sdej ʔaɔj* can be grounded on the semantic core of the two verbs: *tʰa:* deals with words and ways of saying things, whereas *sdej* deals with events, as shown by a comparison between (38) and (39): in (38) the indefinite *ʔej* refers to the class of possible words, and *tʰa:* only is possible; whereas in (39) *riɔŋ* ‘story’ refers to an event (which has to be identified) and *sdej* only is possible<sup>102</sup>.

The negative dimension of the verbs (“criticize”, “rebuke”, “shout at”) can be explained by the fact that the implication of  $N_2$  in an event giving rise to a negative (verbal) reaction (“be late”, “be out every

<sup>102</sup> This difference has to be backed out by a systematic study of these two verbs

night”) gets N<sub>1</sub> to say something about it: a reproachable behaviour leads to N<sub>1</sub> seizing verbal power on N<sub>2</sub>: he is putting him on (personal) trial.

## 2. A synthesis of the ten classes of uses

This presentation of ten large classes of uses of *្រាជ* has made it possible to show that each use can be characterized by a set of specific syntactic properties grounding the interpretation of the utterance with *្រាជ*. This presentation has also made it possible to bring out a characteristic feature at work in all the uses of *្រាជ* : the partial autonomy of the sequence coming after *្រាជ*. As a rule, this autonomy of the sequence following *្រាជ* comes together with the - possible or necessary - presence in this sequence of a V different from *្រាជ* . Owing to this verb, the sequence can refer to an event which is part of the complex event expressed by the whole utterance. As regards the presence of a V in the sequence before *្រាជ*, it should be noted that this is the case with the following uses: benefactive, delegative, jussive, P1 *្រាជ* P2 and criticize/rebuke.

We hereafter resume the ten types of uses in reference to the autonomy of the sequence coming after *្រាជ*.

**Transfer** is characterized by the possibility to introduce a verb after *្រាជ*. The presence of a V (normally *ត្រិ* ‘to go’), which is not compulsory, means that the transfer is not considered as the mere passage of an entity from a subject S1 to a subject S2 (S1 being the active part in this passage). With a postposed V, putting this entity in relation with S2 is considered as prior, and taking S1 into account is subordinated to this relation.

**Benefactive** comes with a V standing both on the left and of the right of *្រាជ*, The one on the right is not compulsory contrary to the transfer case. Benefactive means that the process referred to by the verb standing before *្រាជ* is not considered only from the point of view of the subject agent (S1) of this process, but from the point of view of another subject S2 standing on the right of *្រាជ*. The presence of *ត្រិ* on the right of *្រាជ* reinforces the subordination of the process achieved by S1 to the interest of S2.

**Permissive and causative**: the compulsory presence of a verb after *្រាជ* means that the subject-agent of *្រាជ* is taken into account as it allows the event referred to by the V to be possible or achieved.

**P1 ្រាជ P2** is characterized by the compulsory presence of a V both on the left (P1) and on the right (P2) of *្រាជ*. The V of P1 has a causative value and is understood as the achievement of the event referred to by the V in P2. However, when the subject of the V in P2 comes back to the left of *្រាជ*, the event referred to by P1 gets its own autonomy as an event.

**Jussive** and *kue*: alternative  $\emptyset$  / *្រាជ*. The mere possibility of a construction without *្រាជ* means that the event is referred to mainly by the verb other than *្រាជ*; This V is on the left in the jussive case, on the right in the *kue* case. In the jussive case, the predicate on the right of *្រាជ* is not interpreted as a mere determination of the process, but as a goal to be reached - the validation of the process being considered regarding the achievement of this goal.

**Delegative**: the impossibility to have a V in the sequence coming after *្រាជ* and the compulsory presence of a V before does not make an exception of the delegative. As shown above, the validation of V by a subject S1 is not at stake for itself: it matters only in regard with the substitution of S1 to a subject S2 standing on the right of *្រាជ*. The delegative implies a verb, but this verb is not explicit, being the same as in the left sequence, or already present in the left context.

**Optative**: no sequence at all before *្រាជ* : the only thing which matters is the (actual or aimed at) achievement of the only event referred to by the sequence coming after *្រាជ*.

**Criticize – rebuke**: the impossibility to have a V coming after *្រាជ* lies on a mechanism comparable to that described in the delegative case, but for one difference: the V which is not explicited on the right of *្រាជ* is present in the left context and not directly in the sequence before *្រាជ*. But the V has not the same interpretation in the two positions. In the left context, it refers to an event (which is considered as having a negative value); in the utterance with *្រាជ* preceded by the verbs *ត្រា*: and *ស្រេជ*, it corresponds to something

that is said (meaning a predication): the agent of the process on the left is resumed by the N coming after  $\lambda a\lambda j$  as a subject of whom is said that he is responsible for a process.

### 3. Characterization of $\lambda a\lambda j$ .

As indicated before, it is not possible to account for these ten uses of  $\lambda a\lambda j$  basing on a central value called transfer. It appears that  $\lambda a\lambda j$  is not just a verb like any other, but justifies its definition as a “metapredicate”, with the following characterization: its function is to put in relation two events **E1** and **E2**, the first one being introduced as the trigger of the second one:

$$E1 \lambda a\lambda j E2$$

The relation thus established between **E1** and **E2** gives  $\lambda a\lambda j$  a causative dimension, but this relation should not be reduced to a mere causal relation. In 2. the autonomy of the sequence coming after  $\lambda a\lambda j$  has been pointed out; in the above notation, this sequence corresponds to **E2**. This autonomy of **E2** leads to the assumption that **E2** comes first, and that **E1** is only taken as the trigger of **E2**. This primacy of **E2** means that **E2** is introduced independently of **E1**, even though the realization of **E2** interacts with that of **E1**.

An event can be minimally defined as involving a subject (**S**) and a predicate (**p**). Our hypothesis on the semantic core of  $\lambda a\lambda j$  can therefore be schematized as follows:

$$S_1 \quad p_1 \quad (E1) \quad \lambda a\lambda j^{103} \quad S_2 \quad p_2 \quad (E2).$$

This semantic function is at work in all the uses of  $\lambda a\lambda j$ . All these uses correspond to various modes of realization coming from the specific units embodying the sequences **S<sub>1</sub> p<sub>1</sub>** and **S<sub>2</sub> p<sub>2</sub>**. The sequence corresponding to the use **P1  $\lambda a\lambda j$  P2** gives the largest extension: **S<sub>1</sub> p<sub>1</sub>** and **S<sub>2</sub> p<sub>2</sub>** correspond each to a clause formed on the pattern **N V (XY)**. The transfer value corresponds to the minimal extension,  $\lambda a\lambda j$  being the only verb, **S<sub>1</sub> p<sub>1</sub>** and **S<sub>2</sub> p<sub>2</sub>** reducing to Ns.

From this viewpoint, describing such or such use of  $\lambda a\lambda j$  consists in interpreting the different constituents of the utterance as framed by the **S<sub>1</sub> p<sub>1</sub>  $\lambda a\lambda j$  S<sub>2</sub> p<sub>2</sub>** pattern. We hereafter present a table showing the organization of the ten uses of  $\lambda a\lambda j$  with an indication of the element of the abstract pattern realized by each constituent<sup>104</sup>. When an element of the abstract pattern is not realized, we use the  $\emptyset$  symbol. An element of the abstract pattern can be realized by more than one constituent. When the materialization of an element of the abstract pattern is optional, we put the corresponding element into brackets.

Uses	<b>S<sub>1</sub></b>	<b>p<sub>1</sub></b>	$\lambda a\lambda j$	<b>S<sub>2</sub></b>	<b>P<sub>2</sub></b>
<b>Transfer</b>	N	$\emptyset$	$\lambda a\lambda j$	N	N
	N	$\emptyset$	$\lambda a\lambda j$	N	V + N
<b>Benefactive</b>	N	V + N	$\lambda a\lambda j$	$\emptyset$	N
	N	V + N	$\lambda a\lambda j$	$\emptyset$	V + N
<b>Delegative</b>	N	V + (N)	$\lambda a\lambda j$	N	$\emptyset$
ku3	N	ku3	$\lambda a\lambda j$	(N)	V
<b>Jussive</b>	the addressee	V	$\lambda a\lambda j$	$\emptyset$	PRED
<b>Permissive</b>	(N)	$\emptyset$	$\lambda a\lambda j$	N	V
<b>Causative</b>	(N)	$\emptyset$	$\lambda a\lambda j$	N	V
<b>P1 <math>\lambda a\lambda j</math> P2</b>	N	V	$\lambda a\lambda j \lambda a\lambda j$	N	V
	N	V + N (= S <sub>2</sub> )		$\emptyset$	V

<sup>103</sup> Keeping  $\lambda a\lambda j$  in the notation of its semantic characterization comes from the fact that it works as what we called a “metapredicate”: it does not express an event as such, but plays a central role in the complex event corresponding to the relation established between **S<sub>1</sub> p<sub>1</sub> (E1)** and **S<sub>2</sub> p<sub>2</sub> (E2)**.

<sup>104</sup> In this table, we only mention the constituents in direct relation with the abstract pattern **S<sub>1</sub> p<sub>1</sub> (E1)  $\lambda a\lambda j$  S<sub>2</sub> p<sub>2</sub> (E2)**.



<b>Optative</b>	Ø	Ø	វាជ	N	V
<b>Criticize</b> /	N	tḥa:	វាជ	N	Ø
<b>Rebuke</b>	N	sdej		N	Ø

### Conclusion

This approach of វាជ leads to stating that the various uses and values of this verb are always constructional. It makes it impossible to consider one of the values (that of transfer in the present case) as more basic than the others. It breaks the widely spread idea according to which the lexical units “encodes” entities or events of the world (a central hypothesis in Newman’s works on GIVE). It brings forward the unity and coherence of វាជ in its various uses.

The semantic identity of វាជ is to be found in everyone of its uses, through variations coming from the other constituents of the utterance. The characterization we have put forward appears as a *schematic form*. This means that the interaction between វាជ and some of the items of the context is double: a. as a scheme, it organizes the elements of the context, framing them in a given pattern; b. as a form, it gets its substantial value (its content) from the lexical units embodying this abstract form in a given construction.

### Abbreviation

DEICT	deictic or demonstrative
RELAT	relative
INDEF	indefinite
PART	particle
NEG	negation
1SG	1 <sup>st</sup> person of singular personal pronoun
2SG	2 <sup>st</sup> person of singular personal pronoun
3SG	3 <sup>st</sup> person of singular personal pronoun

### References

- Benveniste, E. (1993) *Vocabulaire des Institutions Indo-Européennes* 1, Paris: Editions de Minuit.
- Bisang, W. (1992). *Das Verb im Chinesischem, Hmong, Vietnamesischen, Thai und Khmer*. Tübingen:Gunter Narr Verlag.
- Gorgoniev, Iou (1966) *Grammatika khmerskogo jazyka*, Moskva
- Heine B. & Kuteva T. (2002) *World lexicon of grammaticalization*, Cambridge University Press.
- Jenny, M. (2005), *The verb system of Mon*, Arbeiten des Seminars für Allgemeine Sprachwissenschaft, Nr. 19, Universität Zürich.
- Khin S. (2002) *La grammaire du khmer moderne*. Paris: Editions You Feng.
- Newman, J. (1996) *Give: a cognitive linguistic study*, De Gruyter
- Newman, J. (ed.) (1997) *The linguistics of giving*. Studies in Typological Linguistics, 36, John Benjamins.
- Oguma, K. (2006) “Identité et variation du verbe *yaru* en japonais”, in: *Constructions verbales et production de sens*, Besançon: Presses de l’Université de Franche Comté, pp. 307 – 321.
- Paris, M. C. (1998) “Syntaxe et sémantique de quatre marqueurs de transitivité en chinois standard: *ba*, *bei*, *jiao* et *rang*”, *La transitivité*, Lille, Presses universitaires du Septentrion, Coll. UL3, p. 355-370.
- Rangkupan, S. (2005) “The Syntax and Semantics of GIVE – Complex Constructions in Thai”, Paper presented at the 2005 International Course and Conference on Role and Reference Grammar, Institute of Linguistics, Academia Sinica, Taipei, Taiwan.
- Thepkhanjana, K. & Uehara, S. (2008), “The verb of giving in Thai and Mandarin Chinese as a case of polysemy: a comparative study”, *Language Sciences* 30, pp. 621 – 651.
- Yap & Iwasaki (1998) “A typological and semantic analysis of the relationship between causatives and passives in the periphrastic ‘give’ construction of some East and Southeast Asian languages”. Paper presented at the Eighth annual meeting of the Southeast Asian Linguistics Society.

Yap & Iwasaki (2003) "From causatives to passives: A passage in some East and Southeast Asian languages", *Cognitive Linguistics and Non-Indo-European languages*, Ed. by Casad & Palmer, Mouton de Gruyter, Berlin – New York, p. 419 – 445.

# Pearic, a Dying Branch of Austroasiatic Languages and Its Struggle for Survival

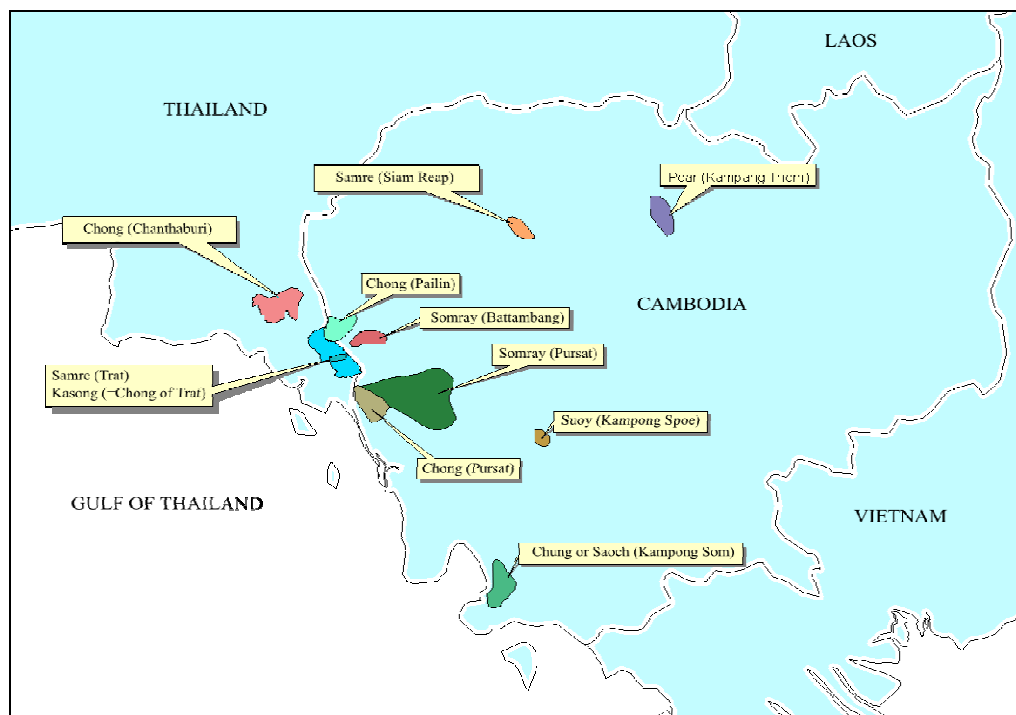
**Suwilai Premsrirat**

Center for Documentation and Revitalization of Endangered Languages  
Research Institute of Languages and Cultures of Asia (RILCA)  
Mahidol University, Thailand

## 1. Pearic languages in the Austroasiatic Language Family

Austroasiatic languages are the substratum of indigenous languages in mainland Southeast Asia. 160 Austroasiatic languages can be found in the large area of mainland Southeast Asia. According to Diffloth (1974), Austroasiatic languages are divided into two main sub-families; Munda and Mon-Khmer. The latter consists of Northern, Southern and Eastern Mon-Khmer, of which Pearic is a branch. Other branches in Eastern Mon-Khmer are the Khmeric, Bahnaric and Katuic.

Pearic belongs to the Eastern Mon-Khmer division of the Austroasiatic language family. There are seven languages in the Pearic branch: Chong, Chung, Kasong, Samre, Suoi, Somray and Pear. Some are found along the Eastern border of Thailand and across the border to Cambodia. Even though there are record numbers of Pearic languages and speakers in various locations throughout Cambodia (as shown in Map1 and Map2), they are not easily found these days.



Map 1: Pearic location in Eastern Thailand and Cambodia (Isara Choosri, 2007)

Suwilai Premsrirat. 2011. "Pearic, a Dying Branch of Austroasiatic Languages and Its Struggle for Survival." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAALA. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.138-153.  
Copyright vested in the author.

## 2. Pearic languages in Thailand

Of the more than 70 languages of Thailand, 23 are Austroasiatic languages, and among those are four small Pearic languages: Chong, Kasong, Samre, and Chung (Suwilai et al. 2004). They are found mainly in the Eastern region near the Cambodian border, except for the Chung people who are found in Karnchanaburi in Western Thailand, where their ancestors were brought as prisoners of war in A.D.1830.

### 2.1 Pearic languages as seriously endangered languages

The twenty first century is an age of rapid change, and bio-cultural diversity is being threatened. Ninety percent of world languages are estimated to face extinction by the end of this century, and these are mainly ethnic minority languages (Krauss 1992). For the indigenous Austroasiatic Language Family, the whole branch of Pearic languages is seriously endangered (Suwilai 2007). The changes in the ecology of language are caused by the global economy, global culture, and global communication. Powerful mass media channels are streamed right into the home, even in remote areas and they mainly use dominant languages. This is exacerbated by policies in national language and education that only promote the use of the national/official language for formal occasions, as the medium of instruction in schools, and as the language of the mass media. Also, increased options for job prospects for the young require the use of a widely-used language. As such, the younger generation does not see the value of, and even has a negative attitude toward its own ethnic languages. In turn, ethnic languages become stigmatized. As young people speak the official language more and more, they speak their ethnic languages less and less, or even stopped speaking them altogether. At the moment, in Thailand there are 15 seriously endangered languages; nine are Austroasiatic languages, and four of them are Pearic languages. Three Pearic languages, such as Samre, Kasong and Chung (Sa-oc), are at the last stage of endangerment prior to extinction, whereas Chong is at the second from the last stage (Suwilai, 2007).

According to Fishman’s GIDS, all languages of Thailand are on the weak side. They are at different stages of endangerment; Kasong, Samre and Chung (Sa-Oc) are at the last stage of endangerment before extinction, because there are only a few fluent speakers remaining. Chong, Lawa (Gong), and So (Thavung) are at stage 7 because only the older generation still uses the language enthusiastically among themselves, but not the younger generation. The languages in stage 6 are Urak Lawoi, Moklen / Moken, Mpi and Bisu which are used in the home and community, but not enthusiastically. Mlabri and Maniq (Sakai) are still used in daily life, but the number of speakers is small. There are not more than 300 speakers for each of these groups. Lavua and Nyah Kur are larger groups, but the children still use their ethnic language in some communities. However, since Lavua, Nyah Kur, Chong and So (Thavung) are now undergoing a language revitalization program, and are being taught as subjects in schools in the area, they may be placed at stage 5.

<b>Weak Side</b>	Stage 8	So few fluent speakers that community needs to re-establish language norms; often requires outside experts (e.g., linguists). [ <i>Kasong, Samre and Chung (Sa-oc)</i> ]
	Stage 7	Older generation uses language enthusiastically but children are not learning it. [ <i>Chong, Lawa (Gong), So (Thavung)</i> ]
	Stage 6	Language and identity socialization of children takes place in home and community., [ <i>Urak Lawoi, Moklen, Mlabri, Maniq (Sakai), Mpi and Bisu</i> ]
	Stage 5	Language socialization involves extensive literacy, usually including non-formal L1 schooling or teaching L1 as a subject “Local Studies” in school [ <i>Chong, Lavua, Nyahkur, and So (Thavung)</i> ].
<b>Strong Side</b>	Stage 4	L1 used in children’s formal education in conjunction with national or official language.
	Stage 3	L1 used in workplaces of larger society, beyond normal L1 boundaries.
	Stage 2	Lower governmental services and local mass media are open to L1.
	Stage 1	L1 used at upper governmental level.

Table 1: Eight stages of language endangerment according to Fishman’s GIDS (Adapted from Suwilai Preamsirat, 2007)

The Pearic speaking people, especially, the Chong, Kasong, and Samre, are believed to be the indigenous people of eastern Thailand and the adjacent area in Cambodia, which was a part of the Ancient Khmer Empire. At the moment, the majority of the Chong are found in the Khao Khitchakut district and some are in the Pongnamron district, Chanthaburi. Although Chong descendants are found in a large area in Chanthaburi, at the moment fewer than 2,000 people still speak some Chong, and the ability among people of different age groups varies widely. Though Elderly people still speak the language among themselves, they have been heavily influenced by Thai vocabulary. Most Chong descendants under thirty years of age do not speak Chong; Thai is their first language. Only young Chong people can speak the language. Good Chong speakers may not number more than 200 individuals. They are found in Khao Khitchakut area in Chanthaburi, only 3 elderly speakers are found in Pongnamron, Chanthaburi. Chong is therefore classified as a language at stage 7 of Fishman's GIDS for threatened languages, where the language is used by the older but not younger generation. A language at this stage needs to multiply the use of the language in the younger generation (Fishman 1991).

As for Kasong, Samre and Chung, the situation is worse. Kasong speakers are found in only three villages (Khloangsaeng, Danchumphon, and Padao). Even though there was a record of more than 3,000 Kasong people (Chong people of Trat) throughout the Borai subdistrict, Trat province; there are now not more than 50 Kasong speakers and not more than ten good speakers. The Samre are found in two villages, with about 30 speakers and only a few good speakers. It is expected to become extinct in the near future (Pornsawan 2001). Samre was also recorded as spoken in Sanamchaikhet, in Chachoengsaw province. Samre descendants remain there, but from my last trip to this area more than ten years ago, the people had already lost their ethnic language. They could recall only few words. The abbot in a Buddhist temple said that he used to hear people speak the language when he was young, but had not heard it for a long time. There might be some Samre speakers in the Cambodian forest, but no investigation has yet been undertaken. Both Kasong and Samre are known by Thai people as Chong. Chung (Sa-oc) is closely related to Chong. *Sa-oc* 'skin disease' is a Cambodian Khmer exonym (Diffloth pers. Comm.) used for referring to these people in their ancient empire. This name is known in the area as "Auat". Fewer than 50 speakers are found in Srisawat district, Kanchanaburi province. Only a few good speakers can be found. Their ancestors were prisoners of war in the early Bangkok period, but most Chung descendants now speak mostly Thai. Some who live in the nearby area can speak Khmu. A village of Chung-speaking people has been found in Kampong Som in Cambodia (Isara 2009).

Samre, Kasong, and Chung (Sa-oc) can also be classified at stage 8 of Fishman's GIDS, where there is a social isolation of the few remaining speakers from the minority language. A language at this stage needs to be recorded for possible later reconstruction. Pornsawan (2001) wrote a doctoral dissertation on Samre grammar, where Samre phonology and syntax are documented. As for the Kasong, three M.A. students in the Linguistics Department at Mahidol University worked on Kasong phonology (Noppawan 2003), Kasong syntax (Sunee 2002), and the Kasong's language attitudes (Suwapat 2003). Even though these language groups are also found in Cambodia, according to Gerard Diffloth (pers. comm.) the whole branch of Pearic languages is endangered.

## **2.2 Phonological, lexical and syntactical characteristics of Pearic as a group of endangered languages**

(1) The shift of the Pearic lexicon toward the dominant language in the area is quite common. A large percentage of words are Thai loan words. There are also Khmer loan words in case of Chung. Even though some of the basic Pearic vocabularies are still retained, it is obvious that loan words from the dominant languages are prevalent and spread beyond nouns and verbs into all parts of the lexicon, including closed classes of grammatical words. More than half of the lexical inventory is influenced by the dominant language. Therefore it is now urgent for lexical documentation and text materials or discourse collection, so that knowledge systems and local wisdom inherited in the Pearic languages can be preserved as much as possible before they are all lost forever. The following examples demonstrate the basic Pearic vocabulary that is still retained.

	CHONG	CHUNG	KASONG	SAMRE
<u>Number</u>				
One	<i>mɔːj</i>	<i>mɔːj</i>	<i>môj</i>	<i>môj</i>
Two	<i>phâ:j</i>	<i>prâ:</i>	<i>pâ:</i>	<i>pâ:r</i>
Three	<i>phê:w</i>	<i>phê:w</i>	<i>phê:</i>	<i>phê:</i>
Four	<i>phô:n</i>	<i>phô:n</i>	<i>phô:n</i>	<i>pḥu:n</i>
<u>Kinship Terms</u>				
Mother	<i>mɛ:</i>	<i>mɛ:</i>	<i>mij</i>	<i>mijŋmɛ:ʔ</i>
Father	<i>ʔu:ŋ</i>	<i>ʔu:ŋ</i>	<i>k<sup>h</sup>u:ŋ</i>	<i>k<sup>h</sup>u:ŋ</i>
Husband	<i>kalɔːŋ</i>	<i>k<sup>h</sup>əlɔːŋ</i>	<i>klɔːŋ</i>	<i>klɔəŋ</i>
Offspring (children)	<i>k<sup>h</sup>e:n</i>	<i>k<sup>h</sup>e:n</i>	<i>k<sup>h</sup>e:n</i>	<i>k<sup>h</sup>i:n</i>
<u>Natural phenomena</u>				
Water	<i>t<sup>h</sup>ɑ:k</i>	<i>t<sup>h</sup>ək</i>	<i>tɑ:k</i>	<i>tɑ:k</i>
Fire	<i>p<sup>h</sup>ɛ:w</i>	<i>plɪ:w</i>	<i>plɛ:w</i>	<i>plɪ:w</i>
Mountain	<i>k<sup>h</sup>ənɔːŋ</i>	<i>nɔːŋ</i>	<i>nɔːŋ</i>	<i>nuəŋ</i>
Rain	<i>kamɑʔ</i>	<i>kəmɛ:</i>	<i>kamâ:</i>	<i>kamâ:</i>
Sun	<i>t<sup>h</sup>əŋiʔ</i>	<i>t<sup>h</sup>əŋiʔ:</i>	<i>t<sup>h</sup>anŋiʔ:</i>	<i>saniʔ</i>

Table 2: Pearic basic vocabularies still retained

Pearic basic vocabularies still retain numbers, kinship, natural phenomena, plants, animals, basic verbs and nouns. Some examples are shown in Figure 3. However, the examples of Chong sentences below illustrate that there are heavy influences from Thai. Even basic words such as nouns, verbs, pronouns, grammatical words, etc. have been borrowed from Thai (Thai loan words are in bold)

*ba:n luŋ haŋ tɔ:k<sup>R1</sup> p<sup>h</sup>ɑ:j<sup>R2</sup> se.<sup>R1</sup>*

house Uncle- Hang sells twenty)

‘LungHang’s shop sells (it) for twenty baht.’

*k<sup>h</sup>ɔ.<sup>R3</sup> naj mɛ:nam ʔi:n<sup>R1</sup> ra:j<sup>R3</sup> tuə*  
crocodile in river have ten clas.

‘There are ten crocodiles in the river.’

*p<sup>h</sup>əj<sup>R3</sup> c<sup>h</sup>ala:t kwa: dak<sup>R1</sup> na*  
it clever than they fp.

‘It is smarter than the others.’

*mu: p<sup>h</sup>əj<sup>R1</sup> c<sup>h</sup>ɔ:p wiwa:t sa.<sup>R1</sup>*  
they like argue each other

‘They like to argue with each other.’

*p<sup>h</sup>iən k<sup>h</sup>a:k<sup>h</sup>a:j mɔŋ<sup>R1</sup> sa.<sup>R1</sup>*  
friend trade together

‘They do business together.’

*mu: p<sup>h</sup>əj<sup>R1</sup> si: ha: k<sup>h</sup>on*  
 group it four five class.  
 ‘Four or five of them.’

(2) There have been great changes in syntactic structure. Grammatical words, auxiliaries, final particles and conjunctions are heavily borrowed from Thai. Only a few affixations are found and none are productive. The syntax of all Pearic languages is becoming more like Thai though some Mon-Khmer characteristics can still be observed.

*p<sup>h</sup>əj<sup>R1</sup> ɽih<sup>R1</sup> ho:c<sup>R1</sup> ɽih<sup>R1</sup>*  
 it not die not  
 ‘It does not die.’

2.1) The negative construction (Neg) -Verb-Neg in Chong and Chung is still retained, though the gradual change to the Thai negative construction can also be observed. The last example sentence below has the same negative pattern as that in Thai.

*k<sup>h</sup>ah<sup>R1</sup> naŋsi: ɽih<sup>R1</sup> jɔɽ<sup>R1</sup>*                      *ɽi:ɽ<sup>R1</sup> təŋ ce:w<sup>R1</sup> ɽih<sup>R1</sup>*  
 know book **not** fp.                              I must go **not**  
 ‘(I am) illiterate.’                                      ‘I must not go.’

*k<sup>h</sup>it ɽih<sup>R1</sup> ɽɔ:k*  
 think **not** out)  
 ‘(He) cannot think of (a way to do).’

Examples above illustrate negative constructions in Chong and Chung that are different from Thai but similar to Khmer. This confirms the settlement of Pearic people with the Khmer and probably dates from the ancient Khmer empire.

2.2) The Noun Phrase / Noun compound in Kasong and Chong illustrates the original construction and the alternative construction influenced by Thai.

English Gloss	KASONG		
	Original	Alternative	Thai
Younger and elder sibling	<i>mót khlij</i>	<i>khlij mó:t</i>	<i>phí: nɔ:ŋ</i>
	younger elder (sibling)	elder younger (sibling)	elder younger (sibling)
Husband and wife	<i>saŋin klɔ:ŋ</i>	<i>klɔ:ŋ saŋin</i>	<i>phuə miə</i>
	wife husband	husband wife	husband wife
Wife and children	<i>saŋin khen</i>	<i>khen saŋin</i>	<i>lû:k miə</i>
	wife child	child wife	child wife
English Gloss	CHONG		
	Original	Alternative	Thai
Face and eyes	<i>mət<sup>R3</sup> ŋa:j<sup>R3</sup></i>	<i>ŋa:j<sup>R3</sup> mət<sup>R3</sup></i>	<i>nâ: ta:</i>
	eye face	face eye	face eye
Younger and older sibling	<i>bo:t<sup>R1</sup> líŋ<sup>R1</sup></i>	<i>líŋ<sup>R1</sup> bo:t<sup>R1</sup></i>	<i>p<sup>h</sup>ɽ: nɔ:ŋ</i>
	younger elder (sibling)	elder younger (sibling)	elder younger (sibling)
Mother and father	<i>me:<sup>R3</sup> ɽu:ŋ<sup>R1</sup></i>	<i>ɽu:ŋ<sup>R1</sup> me:<sup>R3</sup></i>	<i>p<sup>h</sup>ɔ: mɛ:</i>
	mother father	father mother	father mother

Table 3: Noun phrase / Noun compound constructions in Kasong and Chong

2.3 Affixes in Kasong and Chong

Chong locative and causative prefixes and instrumental infix are found, but none are productive.

CHONG						
Affixes	Function	Example				
/pa-/	Locative	<i>dij<sup>R1</sup></i>	‘on’	→	<i>padij<sup>R1</sup></i>	‘above’
		<i>mu:n<sup>R1</sup></i>	‘behind’	→	<i>pamu:n<sup>R1</sup></i>	‘behind’
		<i>reʔ<sup>R1</sup></i>	‘in’	→	<i>pareʔ<sup>R1</sup></i>	‘inside’
		<i>t<sup>h</sup>ε:w<sup>R3</sup></i>	‘other’	→	<i>pat<sup>h</sup>ε:w<sup>R3</sup></i>	‘elsewhere’
/ma-/	Causative	<i>ho:c<sup>R1</sup></i>	‘die’	→	<i>maho:c<sup>R1</sup></i>	‘to kill’

KASONG						
Affixes	Function	Example				
/-n-/	Instrumental	<i>khét</i>	‘to comb’	→	<i>khanét</i>	‘comb’
		<i>ké:w</i>	‘to harvest’	→	<i>khané:w</i>	‘sickle’
		<i>pók</i>	‘to wrap’	→	<i>panók</i>	‘package’
		<i>kók</i>	‘to carry on the shoulder’	→	<i>kánók</i>	‘shoulder pole’

Table 4: Non-productive Kasong and Chong affixes

(3) The variation and change in phonetics and phonology of Pearic languages as a result of influence from Thai is obvious. There is a great variation in the pronunciation of elderly and younger speakers in both segmental and supra segmental phonemes. There has been a trend that the young are losing not only their ethnic vocabulary, but also outstanding contrastive registers. Some still show the contrast, but very weakly. Register complex and tonogenesis found in Pearic languages at the moment have been obviously influenced by the dominant Thai. Chong and Chung are considered register languages, whereas Samre is considered as a tonal language and Kasong is in transition with both 2 contrastive registers and 2 contrastive tones. Examples are shown below.

	R1 (Mid clear)	R2 (High creaky / Glottal constriction)	R3 (Low breathy)	R4 (Low breathy followed by high glottal constriction)
Chong	kəta:k ‘peanut’	kəta:ʔk ‘tongue’	t <sup>h</sup> ə:k ‘water’	
	klə:ŋ ‘bone’	k <sup>h</sup> ələw:ŋ ‘อุกกระพรวน’	kələw:ŋ ‘husband, male’	kələw:ŋ ‘temporary wood bridge’
Chung	ta:k ‘bean, peanut’	ta:ʔk ‘tongue’	t <sup>h</sup> ə:k ‘water’	
			mləw:ŋ ‘eel’	mləw:ŋ ‘salty’
Kasong	<b>R1 (mid)</b>	<b>R2 (high-falling)</b>	<b>R3 (operated low)</b>	<b>R4 (operated, mid, high falling)</b>
	tak ‘big’	kətâ:k ‘tongue’	tâ:k ‘water’	tâ:k ‘wet’
	klə:ŋ ‘bone’		kləw:ŋ ‘husband’	
	kəw:ŋ ‘น้ำใส’	kəw:ŋ ‘green frog’		kəw:ŋ ‘long’
	lale: ‘rotten worm’	p <sup>h</sup> ê: ‘three’	pê ‘water’	pê: ‘watch’
		c <sup>h</sup> ə: ‘dog’		cə: ‘sour’



	Mid	Low	Mid-low	
Samre	klɔəŋ ‘bone’	klɔ̀əŋ ‘husband’	klɔ̄əŋ ‘log’	--
	pɔ:c ‘ด้วง (กระป๋อง)’	kapɔ̀:c ‘turn upside down’	pɔ̄:c ‘to dip water’	--
	pat ‘to lick’	pà:t ‘to slice’	pâ:t ‘to walk pass’	--
	sɔəŋ ‘to dance’	sɔ̀əŋ ‘to smell’	sɔ̄əŋ ‘to tell’	--

Table 5: Register Complex and Tonogenesis in Pearic languages

The changes to and decline of the Pearic languages is obvious. But what is being done? Linguists have been documenting and describing various aspects of the Pearic languages as much as possible as recorded in Figure 8 below.

Documentation and Description	Survey Pearic languages and Stage of endangerment			
	CHONG	CHUNG (Sa-oc)	KASONG	SAMRE
	Stage 7	Stage 8		
Sociolinguistic/ Attitude Survey or Language vitality	✓	--	✓	--
- Phonology	✓	✓	✓	✓
- Register/ Tone/ In Transition	✓	✓	✓	✓
- Syntax	✓	✓	✓	✓
- Lexicon/ Dictionary/ Wordlist	✓	✓	✓	✓
- Dialects	✓	✓	--	--
- Songs / Poems	✓	--	--	--
- Folktales	✓	✓	✓	✓
- Orthography	✓	--	✓	--

Table 6: Pearic Language Documentation and Description

Apart from language documentation and description carried out by linguists, the language speakers themselves can also document their own language and culture from the needed perspective of native speakers of the language. The Lexicon/ Dictionary Compiling is urgent for all endangered languages. It can be done by both linguists and language speakers. The first Chong dictionary was carried out by the community in 2002; though it is more like a list of words and phrases. The words were mainly collected by the Chong from Khlong Phluu village. Then in 2009, linguists published a Chong-Thai-English Dictionary containing three dialects. The two Chong dialects, Northern and Southern Chong, are from the Khao Khitchakut area and one dialect from Pong Namron (Eastern Chong dialect).

The Chong still have enthusiastic groups of people who want to keep their language alive, especially people of middle age who are active and see the value of their own mother language; they actively participate in the Chong revitalization program. However, other Pearic languages such as the Kasong, Samre and Chung do not have such enthusiasm. The Kasong and Samre have only a few elderly speakers. They are found in the border area of Trat province. It is not likely that these languages will survive following the death of the

remaining elderly speakers. This prognosis can also be applied to the Chong which are in Srisaswat district, Karnchanaburi province in Western Thailand.

### 3. Pearic speakers's struggle for the survival of their language

Apart from the documentation and description of Pearic languages by linguists, there are the reactions from grassroots communities for language revitalization programs with guided cooperation from linguists. Language Revitalization Programs are an attempt to add new linguistic forms or social functions to embattled minority languages with the aim of increasing the languages' uses or users" (King, 2001, p. 23)

According to Crystal's Six Postulates of Language Revitalization, an endangered language will progress if its speakers: 1) increase their prestige in the dominant community; 2) increase their wealth relative to the dominant community; 3) increase their legitimate power in the eyes of the dominant community; 4) have a strong presence in the educational system; 5) can write their language down; and 6) can make use of electronic technology. The Chong comprise the first endangered linguistic group that has undergone the language revitalization program.

The Chong Language Revitalization Program (CLRP) began with cooperation between Chong elders and Mahidol linguists with strong community motivation and commitment. Financial support was received from the Thailand Research Fund (TRF) as community-based research conducted by the speakers themselves. The technical support as well as psychological and emotional support was received from the Research Institute for Languages and Cultures of Asia (formerly Institute of Language and Culture for Rural Development (Mahidol University-Linguists and education experts).

The Chong Language Revitalization Program (CLRP) is composed of orthography development, literature production, curriculum development, teaching Chong as a subject in school (by native language speakers), and a Chong community learning center for public at large.

The orthography development component is a complex process of developing a writing system for a previously an unwritten language. Native speakers have to be actively involved in the orthography development process with support from the linguists. The process involves selection of a script writing system or alphabet for the standardization of the systems and vocabulary expansion to produce literature and reading materials.

#### 3.1 Chong language revitalization program (CLRP), the first attempt

The Chong Orthography was based on linguistic research. The Thai script has been selected for use. Three criteria are considered *Linguistic factors* (simple, phonological adequacy) *socio- psycholinguistic factors*, and *technical factors*. Orthography is an important tool for recording the Chong language and local knowledge, and the writing system is used for teaching younger generation and as a symbol of ethnic identity.

Community involvement is indispensable to the success of the project. The newly developed orthography has to be accepted by the community. The community should actively participate in the development process from the very beginning; selecting the script, looking for the different or outstanding features of the target language, looking for minimal pairs, selecting the symbol for each sound, looking for examples and consonant / vowel / tone or register. After that the tentative orthography has to be tested to see how readable and acceptable it is to the community. Then, an alphabet chart with keywords and pictures is able to be produced. The Chong people are very proud of their work. The orthography represents their Chong identity as members of a unique ethnic community of Thailand.



Figure 2: Thai alphabet Chart

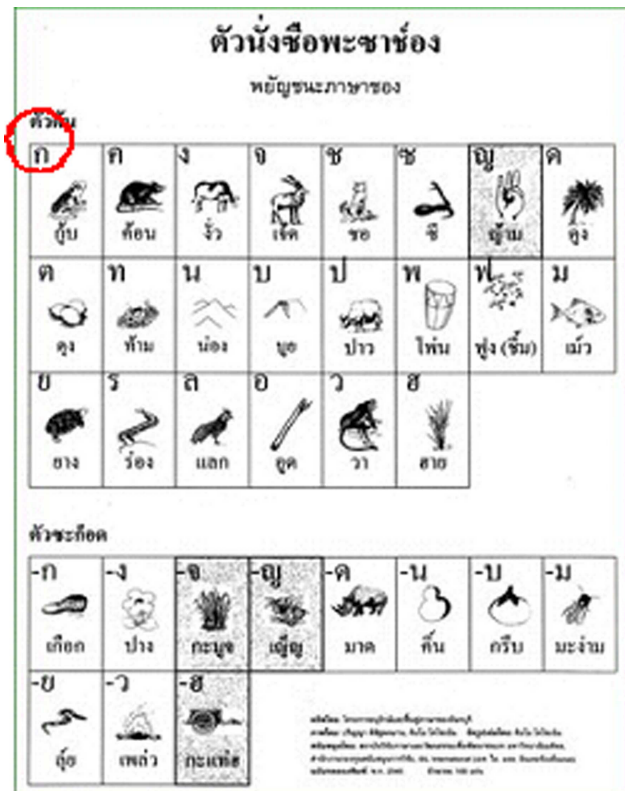


Figure 3: Chong alphabet chart

















ตัววันนยุค							
 วา	 ยาง	 ปาว	 ครา	 กะว้าย	 ร๋อง	 มะง่าม	 กะล่าง
 ค้อน	 ญ่าม	 ท้าม	 ซู้จ	 เม่ว	 จู้ย	 ร๋อย	 คิ่น

Figure 4: Chong registers

Using one word for one symbol, the practical orthography using a Thai-based script is simple. Once the tool for writing has been developed, the language speakers can start writing stories (at different difficulty levels), editing (story writing, language), illustrating and book binding. Reading materials in the local language have therefore been produced by various members of the community.

Since it is a dream of the Chong to have their language taught in the school system, the curriculum for teaching Chong as a subject has been developed. In this way, Chong students are able to use Chong as well as Thai (official language) in an educational atmosphere. They are proud of their part in Chong Language Revitalization Program (CLRP). The preparation of teaching materials is in accord with the cultural calendar and localized content.

For teaching Chong in school, teachers were selected from among the people in the community. The teacher selection criteria are based on proper pronunciation of Chong, dedication to the CLRP, and acceptance by the Chong community.

Apart from language classes in school, field trips are organized for the students to the community forest to learn about plants, animals, beliefs, food items etc., that are important to the local culture. The cultural activities are organized for community at large on cultural days, as well as for students to learn at the Chong Community Learning Center. For example, they learn to cook typical Chong food, as well as Chong dessert.

The initial success of CLRP, which is the first language revitalization program in Thailand, is quite encouraging. It is strongly empowering for the community. This program has revitalized not only the Chong people’s language, but also their self-confidence and self-esteem. As a way of promoting minority language education in school, it has become a model for other struggling, endangered groups such as the Nyah Kur, So (Thavung), Lavua etc. Additionally, the Chong students have been able to pass the National Standard Test in Thai and Math for the first time. This project has also contributed to a reconsideration of education policy. It is actually the first cooperative activity between linguists, education experts and the community to carry out a revitalization program for an endangered language.

### 3.2 The last breath to revitalize Kasong, a language in the last stage of endangerment before extinction.

As for, the Kasong, despite very few good speakers, the people would like to relearn their language with help from linguists; whereas Chung and Samre have no hope for survival after the death of the elderly speakers. They have neither the energy nor the enthusiasm of the young to preserve their language. However, there is a woman over 50 years of age who is still a fluent Kasong speaker, who is very much interested in collecting Kasong language and local wisdom. She wants to teach the language to the younger generation and become the key person for the Kasong revitalization project.

### 3.3 Pearic Languages, documentation and revitalization attempts

Language Activities	Revitalization	Peaic Languages and Stage of Endangerment			
		CHONG (7)	CHUNG (Sa-oc) (8)	KASONG (8)	SAMRE (8)
1. Community study / Language Situation survey / language vitality / Attitude Survey		✓	--	✓	--
2. Orthography Development		✓	--	✓	--
3. Literature/book in Vernacular Language		✓	--	--	--
4. Curriculum Development for teaching Ethnic Language (in School)		✓	--	--	--
5. Learning-Teaching-Reading Materials - Primer - TPR lessons		✓	--	--	--
6. Community Learning Center (outside school)		✓	--	--	--

### 4. Conclusion

The entire Pearic branch of AA is dying. Even though there are record numbers of Pearic speakers in various locations in Cambodia, they are not easily found these days. Kasong, Samre and Chung that are found in Thailand are at the last stage of endangerment (according to Fishman's GIDS) have very little hope of survival after the death of the last few elderly speakers. Chong which at the stage before last still have more and enthusiastic speakers who want to preserve their language and have joined hands with Mahidol linguists to reverse the situation. The lexical and syntactical characteristics of Pearic languages which are heavily influenced by the dominant (Thai) language as well as the register complex and tonogenesis found in Chong, Chung, Kasong, and Samre have been documented. While Chong and Chung are register languages, Samre is a tone language, and Kasong is at transitional stage of developing tones. Even though the Chong language revitalization process is considered to have started twenty years too late, the teaching of Chong as a subject in the formal school system, as part of the Chong language revitalization, has been rather successful and become a model for other languages with the same problem. On the other hand, for the Kasong, which is at the last stage, their descendants are trying to relearn and document their ethnic language as much as possible. Only a miracle can help preserve this language beyond the current generation even though they have developed a writing system as a tool for learning and documenting their language. In general, the Pearic speaking people have gradually given up their way of life and live as the dominant group does. The Chong children grow up with limited exposure to Chong language at home but learn various aspects of Chong language and culture by attending classes and other organized events. These languages are likely to become extinct by the end of this century. What can we do from the larger society? Apart from – language revitalizing, studying and documenting the language before it's lost forever, we can help with the development a national language policy that supports the use of indigenous languages in public, in school, and in mass media in their own area alongside the official / national language and international languages) for the sustainability of the preservation activities.

## References

- Diffloth, Gérard. (1984). *Davavavati Old Mon Languages and Nyah Kur*. Bangkok: Chulalongkorn University Press.
- Fishman, Joshua A. (1991). *Reversing Language Shift*. Clevedon, England: Multilingual Matters.
- Hinton, L. and K. Hale. (eds.). 2001. *The Green Book of Language Revitalization in Practice*. San Diego, CA: Academic Press.
- Isara Choosri. (2002). Mapping dialects of Chong in Chanthaburi Province, Thailand: An application of geographical information system (GIS). M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Isara Choosri. (2007). *Investigating contact-induced language change: case of Chung (Saoch) in Thailand and Cambodia*. Ph.D. dissertation. Research Institute for Languages and Cultures of Asia, Mahidol University.
- Krauss, Michael. (1992). "The world's languages in crisis." *Language* 68(1): 4-10.
- Malone, Dennis. (2003). "Developing curriculum materials for endangered language education: Lessons from the field." *International Journal of Bilingual Education and Bilingualism* 6(5):332-348.
- Matisoff, J. A. (1991). "Endangered languages of mainland Southeast Asia." In K.H. Robins and E.M. Uhlenbeck (Ed.s). *Endangered Languages, 189-228*. Oxford University Press.
- Nattamon Rojanakul. (2009). *Chong syntax*. M.A. thesis, Research Institute for Languages and Cultures of Asia, Mahidol University.
- Noppawan Thongkham. (2003). *The phonology of Kasong at Khlong Saeng village*. M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Pornsawan Ploykaew. (2001). *Samre Grammar*. Ph.D. dissertation. Institute of Language and Culture for Rural Development, Mahidol University.
- Siripen Ungsitipoonporn. (2001). *A phonological comparison between Khlong Phlu Chong and Wang Kraphrae Chong*. M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Sirirat Srisombat. (2009). *A community learning center and the revitalization of the Chong ethnic identity*. M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University. (in Thai)
- Smally, William. (1994). *Linguistic Diversity and National Unity: Language Ecology in Thailand*. Chicago: Chicago University Press.
- Sunee Kamnuansin. 2002. *Kasong syntax*. M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Suwaphat Chanmekha. (2003). *The hidden language: a case study of Kasong language attitudes, uses and abilities in three villages in Trat province*. M.A. thesis, Institute of Language and Culture for Rural Development, Mahidol University.
- Suwilai Premsrirat. (1996). "Phonological characteristics of So (Thavung), a Vietic language of Thailand." *MKS* 26:161-178.
- Suwilai Premsrirat. 2000. "Developing writing system for the Chong language". *JLC* 19.2:5-18 (in Thai).
- Suwilai Premsrirat. (2001). "The future of Nyah Kur". In Robert S. Bauer, ed. *Collected Papers on Southeast Asian and Pacific Linguistics*. pp. 155-165. Canberra: The Australian National University.
- Suwilai Premsrirat and Dennis Malone. (2005). "Language development and language revitalization in Asia." *Mon-Khmer Studies*. 35:101-120.
- Suwilai Premsrirat. (2006). *Thailand: Language Situation*. Encyclopedia of Language and Linguistics. 2nd Edition. Edited by Keith Brown. Oxford: Elsevier.
- Suwilai Premsrirat. 2007. "Endangered Languages of Thailand". *International Journal of Sociology of Language*. 186(2007): 75-93.
- Suwilai Premsrirat, Isara Choosri and Ekapong Suwankases (2001). *Ethnolinguistic mapping of Thailand*. Salaya: Mahidol University.



## Appendix



*Photograph 1-2: Pa Cin and Nai Chian, fluent speakers of Northern Chong dialect*



*Photograp 3-4: A Northern Chong speaking grandmother and a grandchild who cannot speak the language (Left). Mr. Kasem, the last fluent speaker of Southern Chong dialect (Right).*



*Photograp 5-6: The Chong speakers working with professional linguists (Left). The last three speakers of the Eastern Chong Dialect (Right)*



*Photograp 7-8: The well-known Kamnan Chern, the first Chong project leader (Left) The middle age key persons of Chong revitalization program (Right).*



Photograp 9-10: Two of the last few speakers of Samre and Kasong (Left). A Kasong- speaking grandmother and her grandson who cannot speak the language (Right).



Photograp 11-12: Chung Speakers in Cambodia (Left). Chung Speakers in Thailand (Right).



Photograp 13-14: Thai – based Chong language development



Photograp 15 – 17: Writing stories and producing teaching materials in Chong.



Photograph 18: An example of a Big Book in Chong.





Photograph 19: Curriculum development for teaching Chong as a subject in school.



Photograph 20 - 23: Learning to read and write in Chong.



Photograph 24 - 25: Learning more Chong vocabulary during the trip to the forest.



Photograph 26: Thai – based Kasong writing system.



Photograph 27 - 28: Developing Kasong writing systems using Thai alphabet.



Photograph 29 - 30: Pa Somsri, the last Kasong active speaker.



Photograph 31 - 34: Elderly Kasong speaker and younger Kasong descendants in their try to transfer knowledge.

# Aspects of Ho Phonetics and Phonology

Anna Pucilowski\*  
University of Oregon

## 1 Introduction

The phonetics and phonology of Ho, a North Munda language of central-eastern India, have not been well described to date. This paper is a more comprehensive description of that system than previous works (e.g., Deeney 2002, Anderson et al. 2008). We will see that some of the more interesting features of Ho include the phonetic prenasalization of word initial stops, as well as vowel harmony. In the rest of this section I briefly introduce Ho. Section 2 covers the consonants and section 3 describes the vowels. In section 4, I discuss syllable structure and section 5 looks at suprasegmental features, including vowel harmony.

Ho is spoken in the East Singhbhum district of Jharkhand and the Mayurbhanj and Keonjhar districts of the state of Orissa, India. There are approximately 1,500,000 Ho speakers (Lewis 2009). Ho is very closely related to Mundari, to the extent that some researchers have called Ho and Mundari dialects of the same language rather than separate languages (Pinnow 1959, cited in Osada 2008). According to Anderson et al., (2008) there is about 80-85% similarity between the two languages, at least for the Mayabhanj dialect (Osada 2008:161).

There are two known dialects of Ho. The Mayurbhanj dialect of Orissa has been much less studied than the Chaibasa dialect of Jharkhand. According to Anderson et al., there is some variation between the dialects in the vowel harmony (2008:199) and in the pronunciation of certain consonants (2008:201). More differences may come to light as the Mayurbhanj dialect is studied further.

### 1.1 Speakers for this study

The data for this sketch come from Ho speakers in Jharkhand, who are mostly from the Chaibasa area but living in Ranchi, Jharkhand's capital. The samples examined for this paper come primarily from one speaker. He is a university educated 30 year old who speaks Ho, Hindi, Mundari, Santali, Bengali, Oriya and English. The elicitation sessions were conducted primarily in English, translating into Hindi as necessary.

### 1.2 Methods

We collected approximately 2500 words over a three month period. The words were collected for a Talking Dictionary Project. Words were recorded once with their English translation as .wav files onto an Olympus LS-10 sound recorder at a sampling rate of 44,100Hz using the recorder's internal microphone.

---

\* The data for this paper were collected during fieldwork made possible by Living Tongues Institute for Endangered Languages. Special thanks also to Ganesh Murmu, Bhubneshwar Sawaiyan and the Ho students from the Department of Tribal and Regional Languages at Ranchi University.

---

Pucilowski, Anna. 2011. "Aspects of Ho Phonetics and Phonology." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.154-174.  
Copyright vested in the author.

## 2 Consonants

The phonemic consonant inventory of Ho is shown in table 1<sup>105</sup>. There are either 21 or 22 phonemic consonants, depending on the analysis of the palatal nasal (see below). The phoneme /n/ has allophones [n], [ɲ] and [ŋ]. In addition, the palatal and velar nasals, included in angle brackets, represent just one underlying phonemic consonant. This is discussed further in section 2.3.

	Labial	Dental	Retroflex	Palatal	Velar	Glottal
Stops	p, b	t, d	ʈ, ɖ	c, ɟ	k, g	ʔ
Fricatives		s				h
Nasals	m	n		<ɲ	ŋ >	
Flaps		r	ɽ			
Lateral		l				
Glides	w		j			

Table 1: Consonant Inventory of Ho

In the rest of this section I will discuss the consonants according to manner of articulation.

### 2.1 Stops

As seen in table 1, Ho has five stop points of articulation, including the retroflex stops, with voiced and voiceless counterparts. There is also a glottal stop.

Table 2 shows average VOT for the voiced and voiceless variants of three stops in Ho. I restricted measurements to an #\_a environment which meant that there were not enough tokens of /t/ and /d/ to measure (retroflex stops only appear in borrowed words). Additionally, I omitted tokens of /c/ and /ɟ/ because they are phonetically too fricative-like to measure VOT.

stop (no. of tokens)	average VOT time (ms)	range
k (10)	49	38 – 64
p (6)	30	17 – 43
t (8)	25	14 – 23
g (10)	-68	-46 – -48
b (10)	-89	-54 – -119
d (10)	-78	-47 – -112

Table 2: Average VOT for Ho stops

We can see from the table that VOT for unaspirated voiceless stops is not long, with the velar stop having the longest average duration at 49ms. We may treat Ho voiceless stops as short-lag stops (stops with VOT of 0-35ms), with the VOT of /k/ somewhat longer. The short-lag voiceless stops may be one of the factors in the prenasalization of voiced stops word initially (see in section 2.1.1).

Word initially and medially, there is always a phonemic contrast between voiced and voiceless stops. The following examples of minimal pairs or near minimal pairs illustrate contrast between voiced and voiceless stops.

<b>p : b</b>			
/puraʔ/	‘much, many’	/buraʔ/	‘to draw or ladle out’
/capa/	‘draw picture’	/caba/	‘finish’

<sup>105</sup>I use IPA transcriptions in table 1 and throughout the paper

		<b>t : d</b>		
/tul/	‘carry, in arms’		/dul/	‘pour’
/ata/	‘roast’		/ada/	‘know’
		<b>ʈ : ɖ</b>		
/tu:/	‘squirrel’		/ɖur/	‘quail’
/kuʈa/	‘straw’		/kuɖa/	‘roseapple’
		<b>c : ɟ</b>		
/ci/	‘or’		/ɟi/	‘soul’
/raca/	‘courtyard’		/reɟa/	‘coolie’
		<b>k : g</b>		
/kunʈi/	‘terraced, upland field’		/gunʈi/	‘cow’
/hake/	‘axe’		/haga/	‘brother’

We can also see contrast between the palatal and velar stops, as the following words show:

		<b>c : k</b>		
/cimin/	‘how much, many’		/kimin/	‘the wife of one’s son’
/-ici/	‘caus suffix’		/ikir/	‘deep’
		<b>ɟ : g</b>		
/ɟom/	‘eat’		/gom/	‘wheat’
/saɟaw/	‘decorate’		/sagaʔ/	‘type of grass seed’

Thus far we have seen that the voiced and voiceless pairs of stops are contrastive in initial and intervocalic position for each place of articulation. There are, however, a couple of instances of free variation in point of articulation. According to Anderson et al. (2008), there is individual variation among Mayurbhanj Ho speakers between [d] and [ɟ], and [t] and [c], so that /cimɪn/ ‘how many’ may also be pronounced [timɪn]. Similarly, ‘how do you say’ may be said as [tilekepe kadije] or [cilekepe kaɟije] (Anderson et al. 2008:201). The only examples they give of these consonants occur before [i] so it may be that /i/ is causing the palatalization in some words and therefore that /t/ and /d/ might be the underlying consonants.

The following examples demonstrate initial and medial contrast between the dental and retroflex stops, both voiced and voiceless.

		<b>t : ʈ</b>		
/mata/	‘ripe’		/maʈa/	‘whey’
/tu:ʔ/	‘mulberry’		/ʈu:/	‘squirrel’
		<b>d : ɖ</b>		
/duku/	‘sorrow’		/ɖuki/	‘urine’
/didi/	‘a vulture’		/ɖiɖi/	‘to stand on tiptoes’

There is however, free variation between /t/ and /ʈ/ and /d/ and /ɖ/ in some words, as in the following examples.

- (1) [tai] ~ [tʰai] ‘stay, remain’
- [data] ~ [dʰata] ‘tooth’
- [dandʒi] ~ [dʰandʒi] ‘a small handle, connecting device’

There are minimal pairs establishing the phonemic status of /t/ and /tʰ/.

	<b>t : tʰ</b>		
/buʈa/	‘tree trunk’	/buʈa/	‘old man’
/ʃuʈi/	‘defile’	/ʃuʈi/	‘friend, companion’
/goʈa/	‘whole’	/goʈa/	‘ground, land’

Again, Anderson et al. (2008:202) note that it is also possible to get some free variation between /t/ and /d/, and give the following examples from Deeney (2005).

- (2) [peʈe] ~ [peʈe] ‘pluck twig or small branch with one or both hands’
- [ka:ʈob] ~ [ka:ʈob] ‘crab’
- [poʈa] ~ [poʈa] ‘intestines’ (Deeney 2005)

The fact that /t/ can vary freely with both /t/ and /tʰ/ in some instances leads Anderson et al. (2008) to question the phonemic status of /t/. Given that there are minimal pairs, we must posit it as a phoneme at this stage. It seems that there are more instances of free variation intervocalically than initially. In fact, there is only one instance of variation initially ([tai] ~ [tʰai] ‘stay, live’). Retroflex sounds do however, only occur in borrowed words only so it may be the case that they can vary with non-retroflex sounds.

These instances of free variation are interesting and need further research. These may be instances of regional variation. In the case of dental/palatal variation before /i/, it may be a case of palatalization, and so not really free.

### 2.1.1 Prenasalization of voiced stops/phonetic variation in stops

There is a tendency for the voiced stops {b, d, g, ʒ, ɖ} to be prenasalized, e.g., /bandor/ → [ᵐbandor] ‘monkey’ in word initial position. The nasal that appears before the stop is homorganic and not normally equal to other nasals in intensity or length. Figure 1 shows a spectrogram of a prenasalized voiced stop in [ᵐbanda] ‘pond, tank’, where we can observe the nasal band that precedes the stop. We can see that the prenasalized stop is comparable in intensity to the second nasal portion ([n]) of this word. Not all pre-nasal portions were as intense as in this instance.

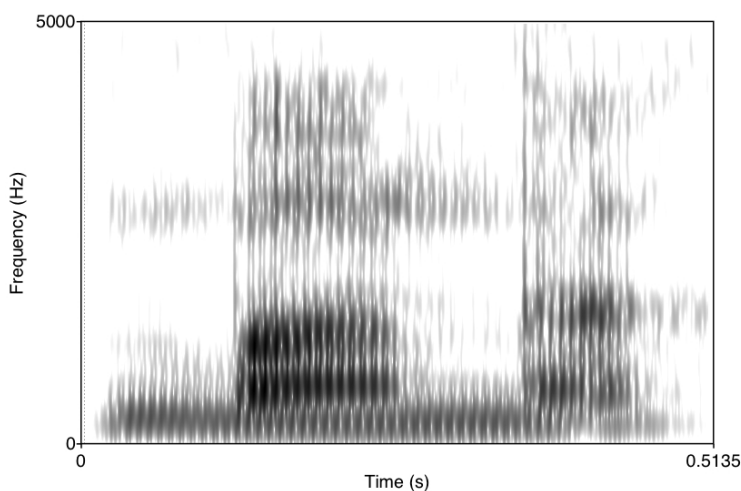


Figure 1: Spectrogram of [ᵐbanda] ‘tank, pond’

Prenasalization does not happen in every instance of a voiced stop. However, it can be observed in all of my speakers to some extent. Table 3 shows the proportion of words containing an initial voiced stop that were prenasalized.

	b	d	g	ʒ	ɖ
prenasalized	91% (247)	78% (134)	90% (165)	58% (107)	77% (20)
not prenasalized	9% (23)	22% (38)	10% (19)	42% (78)	23% (6)
total no. of tokens	279	172	184	185	26

Table 3: Percentage of prenasalized initial voiced stops in Ho words

I coded all of the word initial voiced stops in my data as either prenasalized or not prenasalized based on auditory impression. The tokens come from five speakers, three males and two females, all under the age of 30. We can see from the table that voiced stops in initial position are prenasalized more often than not. The instances of /ʒ/ being prenasalized are much fewer. This is most likely due to its more fricative-like nature. Cross-linguistically fricatives can be pre-nasalized, however, it is less common.

Two of the speakers (one male, one female) had fewer instances of prenasalization than the other three. However, they still produced prenasalized stops in some instances. It is possible that the prenasalization of initial voiced stops varies depending on social factors such as region and education. Given that all my speakers were under 30, future research must address whether the prenasalization is more common among younger speakers and therefore an indication of language change or whether all Ho speakers produce prenasalized stops and it is a stable phonetic feature.

The prenasalization of voiced stops is not always an indication of phonological change in progress. From an articulatory point of view, a prenasalized stop is just one maneuver which speakers can use to facilitate voicing during the stop closure and it serves to highlight the voicing in the stop (Johnson 2005:139, see also Ohala 1997).

It is difficult to see from the data whether the voiced stops are pre-nasalized when they occur medially. Ohala predicts that inter-vocalic voiced stops may spirantize (1997:95) and we do have some examples of that with some speakers in Ho, at least with the bilabial stop.

- (3) [jibon] ~ [jiβon] 'life'  
[babata] ~ [baβata] 'scabies, itch'

Whether intervocalic voiced stops spirantize regularly in Ho needs to be further investigated.

### 2.1.2 Word final stops

We now turn to stops in word final position. Voiceless stops do not appear in word final position, except in some borrowed words, e.g., /kek/ 'cake', /sut/ 'suit, salwar kameez' and /biskut/ 'biscuit', /kop/ 'cup' (from English) and /camac/ 'spoon' (from Hindi). I have no examples of word final /t/.

Voiced stops /b/ and /d/ appear phonemically word finally but there appears to be some phonetic variation in how they are pronounced. They are frequently unreleased or they are preglottalized and accompanied by a nasal release.

- (4) /lad/ → [lad̚] or [laʔd̚<sup>n</sup>] 'bread'  
/porob/ → [porob̚] or [po<sup>h</sup>roʔb̚<sup>m</sup>] 'feast, festival'

In Ho, both monosyllabic and polysyllabic words can have an audible nasal release in citation form. However, in connected speech, the nasal release can only be detected when it occurs at the end of an intonational phrase, elsewhere the stop is merely unreleased.

The voiced stops /d/, /g/ and /ʒ/ do not appear word finally except in borrowed words, e.g., /ne:g/ 'a ritual, rite' (from Hindi). However, there is also a glottal stop phoneme that appears word finally. Historically,

this is thought to be an allophone of /g/ (Anderson et al. 2008:200) but this is not clear synchronically.<sup>106</sup> The glottal stop is always followed by an echo vowel which has the same quality as the preceding vowel.

- (5) /daʔ/ → [daʔa] ‘water’  
/setaʔ/ → [se'taʔa] ‘morning’

I represent these words phonemically with the glottal stop, e.g., /daʔ/ ‘dog’, not /dag/. Words with a final glottal stop are contrastive with words that end in a simple vowel, either long or short, as in the following examples.

- (6) /iju/ ‘shout, call out’      /ijuʔ/ ‘fall from a height’  
/ʃo:/ ‘fruit’                      /ʃoʔ/ ‘sweep’

As in the examples with the nasal release, the echo vowel can be heard in the citation form of polysyllabic words, such as /setaʔ/ ‘morning’ as well as in monosyllabic words. Given that the echo vowel is entirely predictable, it is not normally written in phonemic transcriptions.

## 2.2 Fricatives

There are only two fricatives in Ho: /s/ and /h/. /s/ can appear in onset and coda position, although most of the instances of syllable final /s/ seem to be in borrowed words (/ʃinis/ ‘thing’ being an exception).

- (7) /sarkam/ ‘leaf’  
/rasi/ ‘juice’

/h/ can only appear in onset position.

- (8) /hisi/ ‘twenty’  
/hende/ ‘black’

In some words, it seems that [s] is in free variation with [ʃ], e.g., [sandʒi] or [ʃandʒi] ‘rooster’, [hɔʔoso] or [hɔʔoʃo] ‘goose’. This could be particular to certain speakers however. The bilabial fricative allophone was mentioned above.

## 2.3 Nasals

Five nasal consonant sounds are attested in Ho, but they do not all have phonemic status. The bilabial and dental nasals are the only nasals that can appear in all positions, as shown in table 4.

m		n	
/mana/	‘forbid’	/nama/	‘new’
/hambal/	‘heavy’	/hende/	‘black’
/nimin/	‘contents’	/guni/	‘wise’
/ikum/	‘kneel’	/isin/	‘cook’

Table 4: /m/ and /n/ in Ho

There are three other nasals that probably represent two phonemic consonants. First, the retroflex nasal [ŋ] has a very restricted distribution and is best analyzed simply as an allophone of /n/. It only appears intervocalically in very few instances, mainly loan words, e.g., [duŋa] ‘resin of a sal tree used to make incense’ (Anderson et al. 2008:202). Otherwise, [ŋ] only appears before /d/ and /t/ so might therefore be treated as an allophone of /n/.

- (9) [ˈmaŋdʒi]      ‘food’  
[ˈgaŋtʃi]      ‘small bell’

<sup>106</sup>In Mundari, Osada also regards the glottal stop as an allophone of both /g/ and /j/ (Osada 2008:102). According to him, /g/ appears after /a/ and /j/ after the other vowels.



Secondly, the velar and palatal nasals seem to be in complementary distribution, suggesting that they represent just one phoneme underlyingly. Neither appears in onset position or intervocalically. I do not, however, analyze all instances of [ŋ] as deriving from the same phoneme. First, we see examples of the velar nasal appearing before the velar consonants /k/ and /g/.

- (10) [ˈsiŋɡi]      ‘sun’  
       [ˈtʌŋku]     ‘pit (of fruit)’

Given that this is entirely predictable, we can treat this [ŋ] as an allophone of /n/. These instances of [ŋ] are normally represented as “n” in both Devanagari and romanized writing, which is further evidence of [ŋ]’s allophonic status here.

However, there is also an [ŋ] which appears word finally and thus seems to contrast with /m/ and /n/ while it is in complementary distribution with the palatal nasal [ɲ], as we see in the following table.

n		ɲ		ŋ	
[isin]	‘cook’	[tisiɲ]	‘today’		
[heben]	‘bitter’	[seteɲ]	‘spring, fountain’		
[oɾan]	‘bathe, wash oneself’	[aɲ]	‘I’	[eraŋ]	‘scold’
[ʃibon]	‘life’			[holoŋ]	‘flour’
[susun]	‘dance’			[sasaŋ]	‘tumeric’

Table 5: Alveolar, palatal and velar nasals in word final position

As we might expect, the palatal nasal appears mostly after the front vowels {i, e} and the velar nasal normally appears after the back vowels {u, o, a}:

ɲ		ŋ	
[betɲ]	‘wait’	[unuŋ]	‘play’
[setɲ]	‘spring, fountain’	[roŋ]	‘color’
[biɲ]	‘snake’	[sasaŋ]	‘yellow’
[tisiɲ]	‘today’	[gonoŋ]	‘orphan’

Table 6: Palatal and velar nasals in Ho

However, there are some exceptions to this tendency. Anderson et al. (2008) claim that both [ŋ] and [ɲ] can follow [u] (2008:202). Indeed, the only minimal pairs appear to be following [u].

- (11) [apuɲ] ‘father - vocative form’      [apuŋ] ‘my father’  
       [ruɲ] ‘to husk’                              [ruŋ] ‘sensation of having a limb asleep’

They claim that this is evidence of phonologically “front vs. back [u]” (Anderson et al. 2008:202). I measured the formants of [u] for [ruŋ] ‘to husk’ and [ruɲ] ‘sensation of having a limb asleep’ as spoken by my consultant and there was no significant difference between them. The F2 of the [u] in [ruŋ] was 1101Hz and in [ruɲ] it was 1191Hz. The palatal nasal had a slightly higher F2, and therefore fronter vowel but we need more tokens of [u] plus the palatal and velar nasals before we can draw any conclusions about a front vs. back [u].

In addition, there is one minimal pair involving [a]:

- (12) [aɲ]            ‘dawn’                              [aŋ]    ‘I’

Anderson et al. record [germoŋ] ‘a fleeting smile’ as another exception, because it has a palatal nasal following the back vowel [o]. They also note a tendency for speakers of Mayurbhanj Ho to have [iŋ] generally rather than the [i] of Chaibasa Ho, e.g., [tisiŋ] ‘today’ rather than [tisiŋ] (2008:202).

An additional question is how the nasal consonants interact with vowel harmony (see section 5.1). A low vowel [a] frequently raises to [e], when it appears after a high vowel. We might then ask whether a following [ŋ] changes to [ɲ] after a higher fronter vowel. In the only example I have where that rule might apply, the nasal appears to stay the same. /dijaŋ/ ‘rice beer’ becomes [dijeŋ] phonetically, and the nasal does not seem to change. In another example, [iminaŋ] ‘enough’, we expect the final [a] to raise to [e]. It would be interesting to see if the nasal consonant in turn changes. Unfortunately, in the examples we have, it is always followed by =ge, an emphatic clitic. The velar stop [g] seems to prevent the nasal from changing.

I here make some tentative suggestions about the number of nasal phonemes in Ho. We can posit three nasal phonemes: /n/, /m/ and maybe /ŋ/. /n/ has three allophones: [n], [ɲ], which appears before retroflex stops and [ŋ], which appears before velar stops. I also posit /ŋ/ as a phoneme with two allophones, [ŋ] and [ɲ] that appear word finally. [ŋ] occurs after the back vowels {u, o, a} and [ɲ] after the front vowels {i, e}. The phoneme could equally be /ɲ/.

## 2.4 Liquids

There are three liquid consonants in Ho: /l, r, ɽ/. The /l/ is clear in all positions and the /r/ phoneme is a flap. Both /l/ and /r/ can appear in all positions and they are contrastive.

- |      |         |         |         |              |
|------|---------|---------|---------|--------------|
| (13) | /kamaɽ/ | ‘lotus’ | /kamaɽ/ | ‘blacksmith’ |
|      | /lo:/   | ‘burn’  | /ro:/   | ‘dry’        |
|      | /ʒalom/ | ‘net’   | /ʒarom/ | ‘egg’        |

The retroflex flap /ɽ/ primarily appears intervocally and very rarely finally. It contrasts phonemically with /r/.

- |      |         |                  |        |                              |
|------|---------|------------------|--------|------------------------------|
| (14) | /gaɽi/  | ‘cart’           | /gari/ | ‘rake together’ ‘small shed’ |
|      | /haɽa/  | ‘bullock, steer’ | /hara/ | ‘grow’                       |
|      | /gaɽa/  | ‘river’          | /gara/ | ‘cement, mud paste’          |
|      | /sugaɽ/ | ‘handsome’       |        |                              |

Although /ɽ/ is contrastive with /r/, it is sometimes in free variation with /r/, as noted above in section 2.1. It is not clear whether this variation is intra-speaker or attributable to a particular regional dialect.

## 2.5 Glides

There are just two glides in Ho; /w, j/. Neither occurs word initially (although note that the Ho script is sometimes called *Warang Chiti*) and only very rarely in final position, mostly in loan words. The labial glide /w/ has a more restricted distribution than /j/. It only seems to occur after the back vowels {u, o} and before /a/, as in (2.5):

- |      |        |             |
|------|--------|-------------|
| (15) | /towa/ | ‘milk’      |
|      | /guwa/ | ‘betel nut’ |

The palatal glide /j/ seems to appear between all combinations of vowels, although I only have one example of it appearing between a back vowel and /a/ (where we see /w/), /hōjar/ ‘father-in-law’.

- |      |         |             |
|------|---------|-------------|
| (16) | /tuju/  | ‘jackal’    |
|      | /tajom/ | ‘after’     |
|      | /hojo/  | ‘air’       |
|      | /dijaŋ/ | ‘rice beer’ |

Both glide consonants are inserted to ease pronunciation in certain contexts, as we will see in section 2.6.3 on epenthesis.

## 2.6 Other Phonetic Processes

### 2.6.1 Free variation

/l/ are /n/ are contrastive phonemes in Ho, as we can see in the following examples:

- (17) /neka/ ‘like this’      /leka/ ‘count’  
 /panti/ ‘row’              /paltji/ ‘overturn’

However, we can also see some examples of free variation between /l/ and /n/, as in (2.6.1).

- (18) [nel] ~ [lel] ‘see’  
 [nili] ~ [lili] ‘honey bee’ (also [lele])

If this variation is always between [n-l] and [l-l], it could be the result of long-distance assimilation (or dissimilation). It might also be regional, rather than within-speaker. My consultants suggested that speakers who use the [l]-variant are from Chaibasa town.

Another type of variation that is closer to true free variation is between /ɟ/ and /j/. This type of variation can be heard within the same speaker and was also noted by Anderson et al. for Mayurbhanj Ho (Anderson et al. 2008:203. These examples of j/ɟ variation in /ɟarom/ ‘egg’ are spoken within phrases of each other, by the same person.

- (19) /..mendo      agu-ke-ɟ-a      biɟ      ɟarom.  
 ...but              bring-T/A-TR-FIN      snake egg.
- Endo biɟ      ɟarom-a?      agu-le-ɟ-redo.../*  
 So      snake egg-POSS      bring-T/A-TR-WHEN/IF  
 ‘...but he brought a snake egg. So when he brought the snake egg...’  
 (Girl-snake, lines 17-18)

### 2.6.2 Metathesis

We saw in the previous section that there is sometimes free variation with [n] and [l] in Ho. As well as this free variation, we also see metathesis with these consonants, as in example (20), where either [n] or [l] switches places with [r].

- (20) [rulbiɟ] ~ [nurbɟ] ‘milk-snake’ (Anderson 2008:204)

In the following example of metathesis (21) we see consonant variation between [k] and [r] and also that the order of the consonants switches, so that [s] precedes [k] in [maskal], but follows [r] in [marsal].

- (21) [maskal] ~ [marsal] ‘shine’ (of flame) (Anderson 2008:205)

Other lexical variants where metathesis has applied include:

- (22) [lapaŋ-lapaŋ] ~ [laŋgab-laŋgab] ‘out of breath’  
 [siɟimi] ~ [simiɟi] ‘bean’

In the first example above, we notice that the bilabial stop is voiceless intervocally, but voiced word finally. Also note that when the velar nasal appears word medially, it must be syllable final rather than syllable initial and so requires a homorganic stop to begin the next syllable.

Anderson et al. also note that metathesis can sometimes be obscured by infixation. For example, the reciprocal infix -p- can produce [bepɟa] ~ [tepeba] ‘meet’ from /beɟa/ ‘reach’ (Anderson 2008:204).

### 2.6.3 Epenthesis

There are various words in which either /j/ or /w/ is inserted in order to ease pronunciation. This is most frequently between morphemes, as in the following examples.

(23) [maʃasor] ‘master’ (Anderson et al. 2008:204)

(24) /ako/ ‘3.pl’ + /aʔ/ ‘poss’ → [akowaʔ]

In general, /j/ follows a front vowel and /w/ follows back vowels. After /a/, neither glide is necessary. According to Deeney, the glides that are inserted as epenthesis are normally written by Hos when writing Devanagari and he also uses them in his romanized script (Deeney 2002:xix).

## 3 Vowels

Ho has five contrastive vowel positions, but with two additional contrastive features: length and nasalization. Vowels can be short or long and nasalized or oral. Vowels can also be glottalized, as described in section 2.1 as a feature of certain word final glottal stops. Ho’s vowel harmony system will be described in section 5.1.

### 3.1 Short Vowels

As mentioned, Ho has a five vowel system; /i, e, a, o u/. Table 7 shows example words to demonstrate that all five oral vowels are contrastive in Ho.

/miʃai/	‘sweetmeats’	/ciʃi/	‘a letter’	/bita/	‘length between the tip of the thumb and the tip of one finger’
/meta/	‘to say to’	/cetan/	‘above’	/beʃa/	‘to arrive, reach’
/mata/	‘to ripen’	/caʃaʔ/	‘to split, crack open’	/bage/	‘leave, abandon’
/moʃa/	‘thick, fat’	/coka/	‘frog’	/boka/	‘a stupid person’
/muʃa/	‘nose’	/cuʃu/	‘mouse’	/buʃa/	‘tree trunk’

Table 7: Contrastive short vowels in Ho

The vowel plot in figure 2 shows measurements of F1 and F2 for 10 tokens of each vowel. The vowels all occurred in the first syllable of words of the type CV.CV (stress normally falls on the first syllable of this type of word, see section 5.2). They all appear between obstruents, mostly stops, but also some fricatives. Formant measurements were taken at the mid-point of the vowel. All tokens were recorded by the same male speaker.

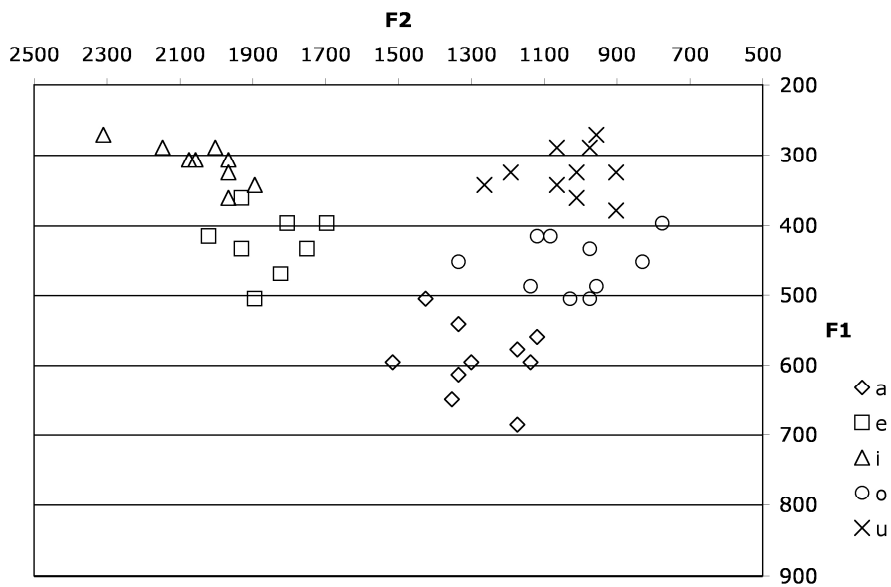


Figure 2: Short vowels in Ho

We can see from the vowel plot that the vowels in Ho have a similar distribution to what we might expect from a language with a 5-vowel system. Table 8 shows the mean formant values for the short vowels from figure 2 in Ho.

	F1 (Hz)	F2 (Hz)
i	311.3	2042.92
e	426.79	1856.29
a	592.29	1286.73
o	455.2	1021.58
u	325.33	1034.22

Table 8: Mean formant values for F1 and F2 in short vowels in Ho

There are two salient instances of allophonic variation of vowels in Ho, as well as some instances of vowel neutralization. The first is [e] as a raised variant of /a/, as shown in (25). The [e] allophone occurs only word finally in unstressed position after the high vowels {i, u}, as a type of harmony. Anderson et al. notice this for Mayurbhanj Ho and it is also a feature of Chaibasa Ho.

- (25) /kula/ → ['kule] 'tiger'  
 /luga/ → ['luge] 'nest'  
 /misa/ → ['mise] 'once'

but,

- /gaɽa/ → ['gaɽa] 'river'  
 /boʃa/ → ['boʃa] 'load, bundle'  
 /seta/ → ['seta] 'dog'

This is also discussed in section 5.1 on vowel harmony.

The second type of allophony in vowels is the nasalization of vowels when they precede a nasal consonant, as in the following examples.

- (26) /enga mindʒi/ → [ɛŋga mĩŋdʒi] ‘ewe’
- (27) /hilaŋ/ → [hilɛŋ] ‘disgust’

As we will see in section 3.3, nasalization is also a phonemic feature of Ho vowels.

There are also examples of vowel neutralization. A schwa vowel [ə] can occur in unstressed position, mostly word internally.

- (28) /kakala/ → [ˈkakəla] ‘to shout’
- /tisiŋ/ → [təˈsiŋ] ‘today’
- /dudulum/ → [dudəˈlum] ‘pigeon’

This is not particular to any vowel and it raises questions about stress patterns in Ho. It seems that Ho has three types of vowels: stressed, unstressed and reduced, and unstressed and unreduced (as in the first syllable of [dudəlum] ‘pigeon’). This leads us to ask whether Ho is a stress-timed language. At this stage it is not clear but future research will look at stress patterns and their relationship to vowel reduction, while also paying special attention to the interaction of stress with Ho’s complex morphology.

### 3.2 Long Vowels

All five Ho vowels can be either short or long. Figure 3 shows a vowel plot for 10 tokens of each long vowel. There are many fewer long vowels in my data than short vowels, so these tokens come from a variety of phonetic environments, including both open and closed syllables. There are also some tokens from another speaker, although still male.

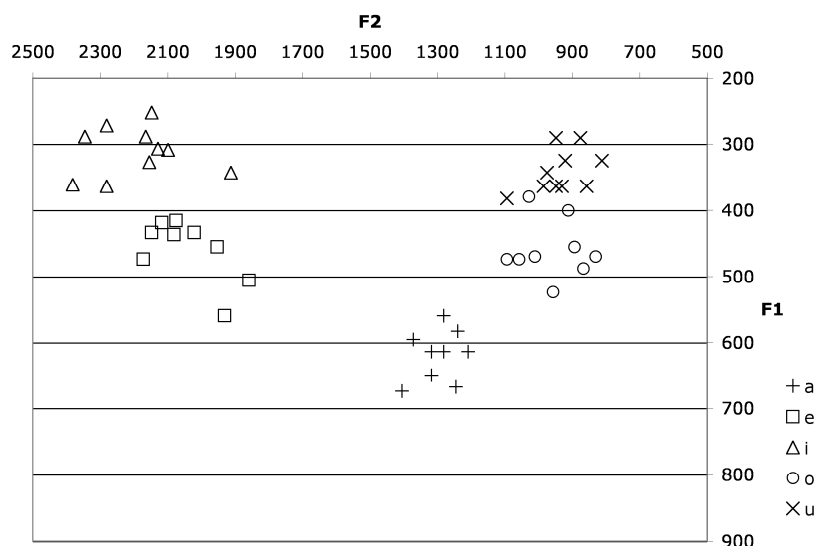


Figure 3: Long Vowels in Ho

There is some debate about whether vowel length is phonemic in Ho. Deeney (2002:xiii) claims that it is indeed phonemic. And, as we see from table 9, there are several minimal pairs that suggest this is true in the Chaibasa dialect.

[kani]	‘a pointed edge’	[ka:ni]	‘story’
[med]	‘eye’	[me:d]	‘iron’
[agu]	‘bring’	[a:gu]	‘lower’
[cera]	‘diarrhea’	[ce:ra]	‘beautiful’
[gom]	‘wheat’	[go:m]	‘to accompany someone’

Table 9: Long vs. short vowel minimal pairs in Ho

Zide (1991:537) and Anderson et al. (2008) state that vowel length is not phonemic in Ho. Zide claims that vowel length is in fact geminate. There is no synchronic morpheme boundary in the long vowels so there does not seem any advantage to positing that they are geminate rather than long.

Table 10 below shows the average length of long vowels versus short vowels in Ho, measured in milliseconds. I measured 10 tokens each of both long and short vowels. As stated above, there were fewer long vowels, thus the long vowels are from both open and closed syllables of varying word lengths, while the short vowels are all in the first syllable of words with CV.CV shape. We can see from the table that, on average, long vowels are more than twice as long as short ones in Ho.

	Short		Long	
	range	mean	range	mean
i	51-101 ms	71 ms	102-316 ms	203 ms
e	55-102 ms	71 ms	152-323 ms	212 ms
a	54-106 ms	80 ms	140-280 ms	202 ms
o	64-98 ms	83 ms	145-274 ms	217 ms
u	44-91 ms	75 ms	131-352 ms	203 ms

Table 10: Comparison of Length of Short and Long Vowels in Ho

Historically, at least some of the long vowels in Ho seem to result from the loss of an [ɾ]. If we compare Mundari and Ho vocabulary (table 11), we can see that in some words where Mundari has an /ɾ/, Ho has lost the consonant and is left with a long vowel.

Mundari	Ho	Gloss
[hoɾo]	[ho:]	‘man’
[duɾum]	[du:m]	‘sleep’
[sepeɾed]	[sepe:d]	‘young man’
[rẽɾẽʔ]	[rẽ:ʔ]	‘joy in the company of others’

Table 11: Comparison of Ho and Mundari vocabulary, from Deeney (2002:132-133)

Given that long vowels are distinguished in production, and the fact that we have minimal pairs for every vowel, we must posit that vowel length is now a phonemic feature of Ho. The relatively low frequency of long vowels cannot bear on their phonemic status.

### 3.3 Vowel Nasalization

In section 3.1, we saw that vowels are normally nasalized when they precede a nasal consonant. Nasalization is also a phonemic feature of vowels in Ho. Some examples of nasalized vowels contrasting with oral vowels are shown in table 12.

/ija/	‘grandmother’	/jĩja:/	‘humid, moisture’
/-rejo/	‘even if, although’ (verbal suffix)	/rẽjo/	‘creak, squeak’
/bal/	‘to burn a hole into’	/bāl/	‘flood’

Table 12: Nasal and oral vowels in Ho

Long vowels can also be nasalized, however there are fewer examples of these. Some are shown in (29).

- (29) /sĩ:ʔ/ ‘foul smell, stench’  
 /ẽ:ʔ/ ‘to erase, extinguish’  
 /rã:sa/ ‘joy, delight’  
 /dõ:si/ ‘thirty’  
 /ũ:r/ ‘leather, hide’

To summarize thus far, vowels in Ho can contrast in both length and nasalization so that both long and short vowels can be nasal or oral. Glottalization is also considered a feature of Ho vowels. This was discussed in section 2.1 as it is a result of word final stops.

### 3.4 Diphthongs

There are some instances of two vowel sequences in Ho that might be called diphthongs.

- (30) /bai/ ‘work, build’  
 /hau/ ‘red ant’  
 /jumbui/ ‘glutton’

The vowels in these examples maintain a smooth transition between the targets, and are longer than short vowels. I measured the lengths of the diphthongs and the results are presented in table 13.

Diphthong	Mean length (ms)	Range	No. of tokens
/ai/	223ms	126-292	11
/au/	216ms	150-330	6
/ui/	206ms	154-305	11

Table 13: Mean length of three diphthongs in Ho

We can see from the table that the average length of these two vowel sequences is about the same as the long vowels in Ho (see table 10), although, like the long vowels, there is a lot of variation. The variation seems to be dependent on whether the vowels occur in a closed or open syllable.

According to Deeney, Ho does not have diphthongs because two juxtaposed vowels normally retain their independent sounds. He argues that these must be treated as independent vowel sounds because each can be lengthened independently of the other, e.g., /bai/ ‘make’ plus the inanimate marker, *-i*, gives us [bai:], while adding the habitual marker lengthens the first vowel and gives us [ba:i] (Deeney 2002: xvii).

The vowel sequences above in (30) where there is a smooth transition between the vowels contrast with instances where both vowels retain their individual sounds, and rather than a smooth transition, we see a hiatus and sometimes a glottal stop between them. The following words demonstrate this:

- (31) /toroe/ ‘ashes’ → [toroʔe]  
 /moro/ ‘acid, sour’ → [moroʔe]  
 /aeʔ/ ‘3.sg’ → [aʔeʔ<sup>c</sup>]

Whether we call the two vowel sequences in (30) diphthongs or two-vowel sequences must depend on our definition of a diphthong. However, we must note that they are phonetically different from the examples in (31) which have a definite hiatus.

## 4 Syllable Structure

In this section we will see the basic syllabic structure of Ho words, as well as the phonotactic restrictions on those syllables.



#### 4.1 Basic Patterns

The patterns for Ho syllables in words are shown in table 14.

Monosyllabic		
VC	/ũr/	‘hide, leather, animal skin’
VVC	/a:ʔ/	‘herb’
CV	/nu/	‘drink’
CVV	/ʔu:/	‘squirrel’
	/bai/	‘work’
CVC	/cur/	‘surround, gush’
CVVC	/ju.r/	‘smooth’
Disyllabic		
V.CV	/u.ku/	‘hide’
V.CVC	/a.buŋ/	‘wash hands and feet’
CV.V	/go.e/	‘wilt, die’
CV.CV	/ca.pi/	‘wash’
CV.CVC	/de.raŋ/	‘maybe’
VC.CV	/en.ka/	‘like that’
VC.CVC	/aʔ.kar/	‘sense’
CVC.CV	/kun.ʔu/	‘wooden post’
CVC.CVC	/ban.dor/	‘monkey’
Trisyllabic		
V.CV.CV	/a.do.wa/	‘husked without boiling’
V.CV.CVC	/e.pa.raŋ/	‘quarrel’
CV.CV.CV	/sa.sa.ti/	‘torment’
CV.CV.V	/ku.la.e/	‘rabbit, hare’
CV.CV.CVC	/ko.lo.waʔ/	‘to give a loan’
CV.CVC.CV	/go.pon.de/	‘quarrel, recip.’
CV.CVC.CVC	/ka.ram.caʔ/	‘fox’
VC.CV.CV	/an.gu.ʔi/	‘finger’
CVC.CV.CV	/sin.du.ri/	‘vermillion’

Table 14: Syllable Structure of Ho words

VV represents both long vowels and diphthongs in this table. I have only included examples of monosyllabic words with VV sequences. However, long vowels and diphthongs also appear in all positions in multi-syllabic words.

In Ho, both CVV and CVC syllables are treated as heavy, while only CV and V syllables are light. This is evidenced by stress assignment (see discussion in section 5.2). Ho also has superheavy syllables: long vowels and diphthongs can appear in closed as well as open syllables.

- (32) [‘a:ndi] ‘marry, marriage’  
 [ca:s] ‘cultivate, farm’  
 [kon.ʔaid] ‘put branches into an earthen pot to support something being steamed’  
 [‘cau.li] ‘uncooked rice’

Table 15 is a summary of the phonotactic restrictions on the consonants in Ho.

	Onset	Coda		Onset	Coda
p	3	only borrowings	s	3	only borrowings
b	3	3	h	3	7
t	3	only borrowings	m	3	3
d	3	only borrowings	n	3	3
ʈ	3	3	ŋ	7	3
ɖ	3	3	ɲ	7	3
c	3	only borrowings	ɳ	7	3
ʃ	3	only borrowings	r	3	3
k	3	only borrowings	ɽ	7	3
g	3	only borrowings	l	3	3
ʔ	7	3	w	3 (only V-V)	only borrowings
			j	3 (only V-V)	only borrowings

Table 15: Phonotactic restrictions on Ho words

Note that table 15 only contains information about whether a consonant can occur in onset or coda position or not. The phonemic status of various sounds is discussed in the relevant sections.

The glide consonants, /j/ and /w/, can appear in syllable initial position, however they do not occur in word initial position, e.g., /tuju/ ‘jackal’, /guwa/ ‘betel nut’. The only other sonorants that can appear in word and syllable initial position are the nasals /m/ and /n/, as well as /r/ and /l/.

We should also note that sonorants seem to be preferred in coda position. There are some exceptions to this, and we see a few words with either /ɖ/, /ʈ/ or /b/ in syllable final position, as in (4.1).

- (33) /eɖka/ ‘wicked, bad’  
/tʈɛka/ ‘witchcraft’

## 5 Suprasegmental Features

### 5.1 Vowel Harmony

Ho has vowel harmony based on height. The high vowels /i/ and /u/ do not occur with the mid vowels /e/ and /o/ in a single word, with some exceptions. According to Deeney (2002:xviii), it is always the case that the mid vowels raise to harmonize with high vowels rather than vice versa.

Borrowed words from other languages, such as Hindi and English, are usually adjusted to conform to the vowel harmony rules.

- (34) *pothi* ‘book’ (Hindi) becomes [puti] in Ho

- (35) *police* becomes [pulis] in Ho

There are only a few rare exceptions where high vowels and mid vowels are combined within a single word. One very common one is [esu] ‘very’ (although note that it has a variant [isu] that does follow the harmonic pattern).

The second type of exception is due to the harmony that occurs when /a/ follows a high vowel and raises to [e]. The only low vowel in Ho, /a/, might be called a neutral vowel because it can occur with any of

the other four vowels in a single word. However, /a/ does seem to raise or centralize to [e] or [ə] in some words with /i/ or /u/ (see also section 3.1).

Figure 4 is a vowel plot showing the formant values for the raised /a/ vowel, alongside the other vowel measurements from section 3. I measured twelve tokens of /a/. All occurred finally after a high vowel in a word with CV.CV-type structure. All the tokens were spoken by the same male speaker.

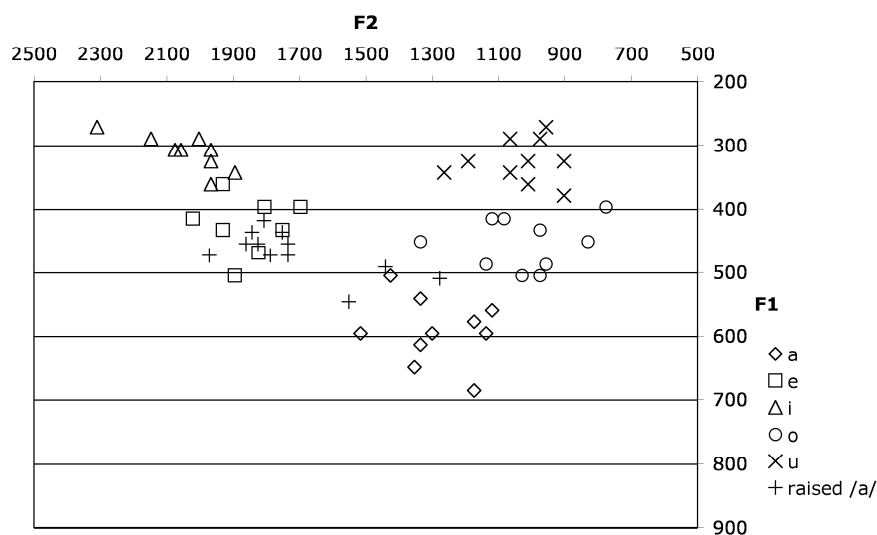


Figure 4: Raised /a/ vowel in Ho

We can see from the plot that in most cases the raised /a/ has a more [e]-quality, and indeed overlaps with many of the /e/ tokens. Although there are some instances where it is more like [ə].

Vowel harmony extends across certain morpheme boundaries in Ho so that some affixes will harmonize with the vowel in the stem. We see that vowel harmony in Ho is normally progressive, as in examples (36) and (37) but there are some examples of regressive harmony, as in example (35) above and (38) below.

(36) /hasu/ ‘sick’ +/o?/ ‘passive’ → [hasu:ʔ]

(37) /abu/ ‘1.PL.INCL’ +/leka/ ‘like, as’ → [abu-like]

(38) /en/ ‘that’ +/kij/ ‘dual’ → [inkij]

The aspect marker /-ke/ follows the verb root(s) in the clause and conveys something like perfective aspect. When /-ke/ follows a root with a high vowel, it usually becomes [-ki] as we see in examples (40) and (41).

(39) [jom-ke-q-a] ‘eat-T/A-TR-FIN’  
 [men-ke-q-a] ‘say-T/A-TR-FIN’  
 [cike-ke-q-a] ‘do-T/A-TR-FIN’ (/cika/ ‘do’)

(40) [tai-ki-n-a] ‘stay-T/A-ITR-FIN’

(41) [atom-idi-ki-q-redo] ‘remove-take-T/A-TR-WHEN/IF (snake bite, line 23)

Further evidence that /a/ is a neutral vowel is shown in the third example in (39). Here we see that the final /a/ of /cika/ raises to [e] and moreover, it serves to prevent the harmony from spreading from the stem vowel to the affixes. We can call this /a/ an opaque neutral vowel (van der Hulst & van de Weijer 1995).

Vowel harmony in Ho appears to be sensitive to morphological structure so that some affixes are more likely to undergo harmony than others. As we see in example (41) above, vowel harmony does not always spread across the whole word: *atom* does not become [atum] and *-redo* does not become [rido]. In terms of vowel harmony, *atom-idi* ‘remove-take’ is a compound word, and therefore less likely to undergo harmony (van der Hulst & van de Weijer 1995:501). The subordinating morpheme *-redo* is not part of a compound, but it may be considered less “close” than tense/aspect markers such as /-ke/. Van der Hulst and van de Weijer predict that derivational affixes are less likely to undergo harmony than inflectional affixes and it does indeed seem that in Ho, harmony is sensitive to morphological structure.

Other affixes that appear to undergo harmony include /-tan/, a progressive marker (42), /-(e)jan/, an intransitive past tense marker (43) and /-le/ a tense/aspect marker (44).

- (42) [nam-**tan**-a] ‘get-PROG-FIN’, but [lel-kɪŋ-**ten**-a] ‘look-DL-PROG-FIN’  
(creation myth, lines 23, 57)
- (43) [hara-**ejan**-a=kɪŋ] ‘grow-PST.ITR-FIN=dl’, but [hapanum-**ijen**-a=kɪŋ]  
‘young.woman-PST.ITR-FIN=DL’  
(creation myth, line 5)
- (44) [sen-**le**-n-redo] ‘go-T/A-ITR-WHEN/IF’,  
but [agu-**li**-ɖ-a] ‘bring-T/A-TR-FIN’  
(girl-snake story, lines 14, 16)

As well as tense aspect markers, some of the pronominal person clitics also undergo harmony. Pronominal person clitics indicate the subject and can attach as clitics to the word immediately preceding the verb.

- (45) [ciknaʔ=**laŋ**] ‘what=1.DL.INCL’, but [maŋɖj=**leŋ**] ‘food-1.DL.INCL’  
(creation myth, line 13)
- (46) [aben=**ben**] ‘2.DL.2.DL’, but [budi=**bin**] ‘intelligence=2.DL’  
(creation myth, lines 30, 35)

There are also some examples of disharmonicities. For example, the plural morpheme /-ko/ sometimes harmonizes to [-ku] when that word contains a high vowel (48).

- (47) [bare-ko] ‘brother-PL’  
[jo-ko] ‘fruit-PL’
- (48) [jilu-ku] ‘meat-PL’  
[maŋɖj-ku] ‘food-PL’

However, this does not seem to be consistent. In the following examples, /-ko/ follows a high vowel but does not harmonize.

- (49) [somɣaw-budi-ko] ‘intelligence-PL’
- (50) [dudəlum-ko] ‘pigeon-PL’

It seems that the words that do not harmonize with /-ko/ are either compound words (49) or longer words (50).

Although the pronominal person clitics above seem to harmonize, other clitics are not affected by harmony. For example, the focus clitic /=do/ never seems to raise to [du] after a high vowel.

- (51) [munu=do] not [munu=du] ‘beginning-FOC’  
 [muni=do] not [muni=du] ‘hermit-FOC’

Postpositions do not harmonize:

- (52) [buru-te] ‘forest-ABL’  
 (53) [disum-re] ‘world-LOC’

The evidence suggests that vowel harmony in Ho is affected by morphological structure. At this stage, we do not know all the morphemes that harmonize but we might predict that there is some reflection of morphological “closeness” so that the more inflectional affixes such as tense/aspect and person markers will harmonize, while the more derivational-like morphemes, the postpositions, as well as most clitics do not harmonize.

Vowel harmony in Ho is a little studied area of research. We can say that the mid vowels harmonize with the high vowels, while the low vowel /a/ raises to [e] or [ə]. Given that vowel harmony is in the direction of the high vowels, we see examples of both progressive and regressive harmony. Further research will reveal all of the morphemes that normally undergo harmonization.

### 5.2 Rhythm and Intonation

As is the case with most Munda languages, Ho does not appear to be a tone language. Korku, another North Munda language, is the only Munda language known to have tonal contrast (Zide 1966).

Very little is known about rhythm and intonation in Ho. The findings I present here for Ho must be seen as preliminary. The following hypotheses are based on my subjective impression about which was the prominent syllable in a given word.

Syllable weight seems to correlate with the prominent syllable in Ho. As discussed in section 4, we can say that both CVV and CVC syllables count as heavy in Ho and these syllables are normally emphasized.

- (54) [sin.du.ri] ‘vermillion’  
 [kun.tu] ‘wooden post’  
 [ce:.ra] ‘pretty’  
 [ra.’cam] ‘cut with scissors’

Ho also has superheavy syllables and these always attract emphasis:

- (55) [kon.’ɖaid] ‘put branches into an earthen pot’

When a word contains two or more syllables of equal weight, emphasis falls on the first syllable, as in the following examples.

- (56) [mu.ni] ‘hermit’  
 [sa.du] ‘holy man’  
 [ra.ca] ‘courtyard’  
 [ca.ka.ɽã] ‘snacks served with rice beer’  
 [ge.ɟe.ra] ‘obstinate’

This is less clear in words where we have two or more heavy syllables, e.g., CVC.CVC. In some instances, it seems like the above rule applies, i.e., emphasis falls on the first syllable:

- (57) [saŋ.gar] ‘hunt’  
 [pam.pal] ‘butterfly, moth’

However, in some instances of a homorganic nasal-stop sequence, the final syllable can also be prominent:

- (58) [ban.ˈcaw] ‘save, salvation’  
 [ban.ˈdor] ‘monkey’

This may be related to the phenomenon of prenasalized stops (see section 2.1.1). The nasal, in these cases, is acting as part of the onset of the second syllable rather than the coda of the first syllable. This creates a heavier second syllable, which thereby attracts stress. Further evidence for this is the fact that /bacaw/ is given as an alternative for /bancaw/ ‘save, salvation’ (Deeney 2005:22) (both from Hindi *bachana*).

Although more research on rhythm is needed, we can hypothesize that syllable weight determines word level prominence in Ho so that heavy syllables are emphasized. Where there are two or more syllables of equal weight, the first syllable is stressed. Exceptions to this generalization are homorganic nasal-stop sequences which occur word medially, creating a heavier following syllable which then carries stress. Future research will compare word level prominence with rhythm and patterns of emphasis at the phrase and sentential levels.

## 6 Conclusions

This paper has described the basic phonetics and phonology of the Chaibasa dialect of the Ho language. We have seen that Ho has many intriguing features, including prenasalized voiced stops, which are described here for the first time. Ho’s vowel system is also interesting. I have argued that vowel length is indeed phonemic, based on minimal pairs and phonetic analysis. The vowel harmony system was described in section 5.1 and is shown to be of a familiar type, based on vowel height, but sensitive to morphological closeness.

More research is needed in many areas, including the prenasalized stops as well as details of the suprasegmental features of vowel harmony and stress assignment. Especially valuable to the study of Ho phonetics would be sociolinguistic data so that we can see how Ho is changing, as well as the influence of regional languages, such as Hindi and Oriya, which is presumed to be great.

## References

- Anderson, G. D. S. (ed.) 2008. *The Munda Languages*, New York: Routledge.
- Anderson, G. D. S., Osada, T. & Harrison, K. D. 2008. ‘Ho and the other Khewarian Languages’, in G. D. S. Anderson (ed.), *The Munda Languages*, New York: Routledge, pp. 195–250.
- Deeney, J. 2002. *Ho Grammar and Vocabulary*, (2<sup>nd</sup> edition). Ranchi: Xavier Ho Publications.
- Deeney, J. 2005. *Ho-English Dictionary*, (2<sup>nd</sup> edition). Ranchi: Xavier Ho Publications.
- Johnson, K. 2005. *Acoustic and Auditory Phonetics*, Malden, MA: Blackwell.
- Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the World*, Sixteenth edition. Dallas: SIL International.  
 Online version  
 <<http://www.ethnologue.com>>
- Ohala, J. 1997. Aerodynamics of phonology, Proceedings of the 4th Seoul International Conference on Linguistics [SICOL] 11-15 Aug 1997.
- Osada, T. 2008. Mundari, in G. D. S. Anderson (ed.), *The Munda Languages*, New York: Routledge.
- van der Hulst, H. & van de Weijer, J. 1995. ‘Vowel Harmony’, in J. Goldsmith (ed.), *The Handbook of Phonological Theory*, Cambridge, MA: Blackwell, pp. 495–534.
- Zide, N. 1966. ‘Korku low tone and proto-Korku-Kherwarian vowel system’, in N. Zide (ed.), *Studies in Comparative Austroasiatic Linguistics*, The Hague: Mouton, pp. 214–229.
- Zide, N. 1991. ‘Lexicography of other languages of the Indian subcontinent: The Munda languages’, in F. J. Haumann, O. Reichmann, H. E. Wiegard L. Zgusta (eds), *An International Encyclopedia of Lexicography*, Vol. 3, Berlin and New York: Walter de Gruyter, pp. 2533–2547.

## Notes on Glottal Constriction in Gorum

Felix Rau  
Leiden University

### Introduction\*

In this paper I take a look at the phonology of glottal constriction in Gorum, a South Munda language of the Austroasiatic stock spoken in the Eastern Ghats in India. The common denominator of the phenomena I focus on in the following is glottalization, i.e. stricture at glottal level, and can be represented by the feature [+constricted glottis]. In the present study I emphasize the role of the syllable and other prosodic units. Through this focus I take up one aspect of an older approach by Aze (1971, 1974), which is now generally ignored. Additionally, I will highlight the comparative and historical implications of this analysis.

The three phenomena I am interested in are the following: the glottal stop /ʔ/, as in (1), a series of preglottalized voiced obstruents /ʔb, ʔd, ʔj, (ʔg)/, such as /ʔd/ in (2), and vowels with creaky voice articulation, i.e. /a̰ ḛ ḭ o̰ ṵ/, as /a̰/ in (3).

(1) ɖaʔ [ɖaʔ] ‘water’

(2) ɖaʔd [ɖaʔd<sup>n</sup>] ‘for’

(3) a̰ [a̰] ‘husking pit’

The glottal stop and creaky voice are purely glottal phenomena and differ from one another mainly in the degree of glottal constriction and its timing relative to the vowel. The glottal stop is a complete obstruction of the airflow at glottal level and is perceptually clearly delimited from the adjacent vowel. Creaky voice involves a lesser degree of glottal constriction and extends over the whole duration of the vowel and a following sonorant in the rhyme, if present. Perceptually, creaky voice is a property of the vowel, as it cannot be separated from it. The glottalization in preglottalized obstruents, on the other hand, is part of a complex phenomenon and occurs at the boundary between vowel and obstruent, parallel to the oral closing gesture. The glottalization is here only one aspect of the phoneme.

Most other Munda languages only have the glottal stop and the (pre-)glottalized obstruents, although Juray (Zide 1982) and Sora (Donegan p.c.) also have creaky phonation as a variant of the glottal stop. In most Munda languages, the (pre-)glottalized obstruents are considered allophones of the non-glottalized obstruents, as they occur only in syllable-, stem-, or morpheme-final position. This is the case, for example, in Santali (Ghosh 2008) and Kharia (Peterson 2008). Gorum, however, seems to be unique among Munda languages in having all three phenomena. Also, as I will argue, these phenomena are phonemic and involve one feature of glottalization, whose proper domain is the syllable.

---

\* I would like to thank all ICAAL 4 participants and especially Patricia Donegan and Gerard Diffloth for their helpful comments and stimulating discussions. Additionally I am also indebted to Juliette Huber for commenting on earlier versions of this paper. Some of the data used in this paper were collected on field trips partially funded by the Leiden University Fund.

---

Rau, Felix. 2011. “Notes on Glottal Constriction in Gorum.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.174-183.  
Copyright vested in the author.

### Previous Accounts

Aze (1971, 1974) and Zide (1963, 1982) analyze the status of these three phenomena in very different, in fact incompatible, ways. Aze (1971, 1974) subsumes the three under a single prosodic phoneme, an analysis that allows him to reduce Gorum syllable structure to (C)V(N) and (C)V?(N), where /ʔ/ represents glottalization of the syllable. However, his analysis cannot distinguish between creaky voice phonation and the glottal stop. Furthermore, it fails to distinguish a combination of either one of these with a nasal from the homorganic preglottalized obstruent. As such, Aze's analysis cannot account for a set of minimal pairs and hence has to be regarded as insufficient.

Zide (1963, 1982), on the other hand, recognizes three distinct phenomena, assuming that creaky voice and the glottal stop are two distinct phonemic segments, while preglottalized obstruents are a non-phonemic variant of voiced obstruents. The treatment of creaky voice as a segment is phonologically unfortunate. Also, while the glottalization of obstruents might very well have been a non-phonemic process in earlier stages of Gorum, this analysis seems not to be appropriate to account for the present state, nor does it seem to be psychologically adequate.

In my own analysis, all three glottal phenomena involve one feature [+constricted glottis] or glottalization, which is connected to the syllable. In so far, thus, I follow Aze. However, in my view all three are distinct and phonemic, so that in this respect I am more in agreement with Zide. Historically, I believe she is right in assuming that the preglottalized obstruents are not phonemic. However, synchronic evidence shows that they have become so, due to the heavy influx of Indo-Aryan loan vocabulary, so that preglottalized obstruents contrast with non-glottalized ones in coda-/stem-final position.

### The Phonemic Status of Glottal Phenomena

The fundamental differences in the previous accounts illustrate the difficult phonemic status of the three phenomena. It is in fact difficult to demonstrate their distinctiveness, since they are largely confined to mutually exclusive contexts; also, the phonetic differences can be very subtle.

Each of the three phenomena contrasts with its absence, i.e. [+constricted glottis] contrasts with [–constricted glottis]. The examples in (4)–(9) demonstrate this. The difference between (8) and (9), however, is a secondary one between native and loan vocabulary, as (9) is a loan from Telugu, probably via Desia Oriya. Synchronically, it is nevertheless real and does not seem to be different from the other two contrasts. Examples (8) and (9) are especially relevant in the light of Zide's claim that preglottalized obstruents are not phonemic.<sup>107</sup>

(4) ɔl [ɔl] 'husking pit'

(5) al [al] 'to thatch'

(6) ɔaʔ [ɔaʔ<sup>a</sup>] 'water'

(7) ɔa [ɔa] 'to do'

(8) ɔa<sup>ʔ</sup>bu [ɔa<sup>ʔ</sup>b<sup>m</sup>u] 'close-INF.TR'

(9) ɔabu [ɔabu] 'money'

Having established the contrastivity of presence vs. absence of glottal stricture in all three types, let me now come to the contrast between the three phenomena. Finding minimal pairs here is rather more complicated. The main reason for this is that the preglottalized obstruents are, by their very nature as oral obstruents, quite distinct from the purely glottal phenomena creaky voice and glottal stop. The pairs (10) and

---

<sup>107</sup> The distinctiveness can still be doubted in cases such as (8) and (9). These two can be seen as structurally different – i.e. by construing the syllable structure of (8) as /ɔa<sup>ʔ</sup>b.u/ in contrast to /ɔa.bu/ for (9). This, however, does not invalidate the fact that the distribution of both types of phonemes widely overlaps, so that no general complementary distribution can arise.



(11) as well as (12) and (13) come as close to minimal pairs as one can get. All four feature glottal constriction combined with an alveolar closure. In case of /Vʔn/ in (11) and /ʎn/ (12), this involves nasality in the form of a nasal segment, in this case the nasal stop /n/, and in (10) and (13) a nasal release of a single complex phoneme /ʔd/ [Vʔd<sup>n</sup>].

(10) ɬaʔd [ɬaʔd<sup>n</sup>] ‘for’

(11) kinɬaʔ-n [kɬaʔn] ‘at the river (river-LOC)’

(12) ɬa-nɛn [nɛn] ‘if done (do-TOP)’

(13) abgeʔd [geʔd<sup>n</sup>] ‘to ignite’

These combinations of glottal constriction, alveolar oral closure and nasality differ slightly in their phonetics. In the glottal stop plus nasal combination /Vʔn/ in (11), the vowel preceding the glottal stop as well as the nasal following it are minimally affected by the complete glottal stricture. This results in a sound event best transcribed as [aʔn]. The combination of a creaky vowel with a nasal as in (12), shows a glottal constriction that stretches over the whole articulation of the vowel and extends into the following sonorant. This sound event is best represented by the transcript [ɛn]. The preglottalized obstruents in (10) and (13) are more complicated in their articulation. The vowel starts in modal voice phonation. The glottal stricture starts parallel to the oral gesture and is clearly audible before the oral closure is complete. The stop is afterwards released in nasal plosion. This sound event may be transcribed as [aʔd<sup>n</sup>], where [ʔ] is intended to represent the temporally restricted glottal constriction at the end of the vowel.

The three types of glottal phenomena differ primarily in the relative timing of the gestures involved, i.e. glottal constriction, oral closure and velic opening, as well as in the degree of glottal stricture, i.e. complete closure in the case of /ʔ/ and partial constriction in the case of creaky voice and preglottalization.

Speaker judgments vary in their rigidity: while speakers regard a replacement of /ʔd/ by /ʔn/ as a clear audible mistake, they are less clear about the relationship between /ʔd/ and /ʎn/.

All three types of phenomena are consonantal. Interestingly, even creaky voice, which from a phonetic point of view seems to be a quality of the vowel, behaves in some respects like a consonantal phoneme. The best evidence for this comes from echo word formation. Echo words are a phonologically altered repetition of a word. One strategy used in Gorum is vowel replacement. In this formation process, the consonantal skeleton of the base word is maintained, while the vowel or vowels are replaced.

(14) ali ‘liquor’

(15) ula ‘liquor (echo word)’

(16) gagaʔ ‘cooked rice’

(17) gigiʔ ‘cooked rice (echo word)’

(18) gumar ‘winnowing’

(19) gimir ‘winnowing (echo word)’

In this process, creaky voice figures as part of the consonantal skeleton, just like the other elements with glottal constriction. Thus in the case of examples (18) and (19), the consonantal skeleton of the word is *gV.mVr*, with creaky voice a part of it, as shown by the fact that the creaky voice in the first syllable is unaffected when the original vowel pattern *u-a* is replaced in the echo word by *i-i*.

In summary, the three types of glottalized phonemes – glottal stop, preglottalized obstruents and creaky voice – have been shown to be distinct and phonemic in Gorum. They behave as a class and can be categorized by the feature [+constricted glottis]. Evidence from echo word formation suggests that all three should be grouped with consonants. This is, however, all that can be gained from a point of view centered on

the segment. All other relevant data is syllable- or even stem-related and will be discussed in the following sections.

I can now contrast my view of the phonemic status of the glottal phenomena with the two previous accounts by Aze and Zide. For Aze (1971, 1974) every occurrence of glottalization is distinctive and glottalization is a prosodic property of the syllable. However, he recognizes only one phonemic process of glottalization, so that in his analysis creaky voice phonation and the glottal stop are identical. Additionally, he regards the phenomenon here called preglottalized obstruents not as an obstruent phoneme, but as a combination of a nasal coda with prosodic glottalization. By this he can reduce all syllable structures in Gorum to four types: CV, CV?, CVN, CV?N. Zide, on the other hand, only recognizes segments and distinguishes a phonemic glottal stop from a phonemic creaky voice. Preglottalized obstruents in her analysis are obstruents. However, they are not glottalized phonemically, but allophones of the non-glottalized voiced obstruents.

Since Aze recognizes glottalization only on syllable level, the three approaches can best be compared by contrasting the possible syllable types that result from the approaches. Note how several syllable types are conflated in Aze's analysis, while Zide's analysis basically yields the same results as mine. However, she treats creaky voice as a segment /H/ and syllables with preglottalized obstruents are not phonologically glottalized in her view.

Rau	Aze	Zide
CV	CV?	CVH
CV?	CV?	CV?
CV?n	CV?n	CV?n
CVN	CV?n	CVHN
CV'O	CV?n	CVO (not glottalized)

Table 1: Comparison of the different approaches

## The Syllable

The syllable is of particular importance for the phonology of glottal constriction in Gorum, since restrictions on number and placement operate on syllable-level. Glottal constriction is restricted to the rhyme and only one occurrence is permitted per syllable.

### Syllable Structure

Gorum has a maximal CVC/CYC syllable structure. Three exceptions occur: The complex onset /dʀ/ occurs in one native word, two lexemes have a complex coda /ŋk/, and some word forms possess a coda /?n/. This last cluster can only occur as a result of the affixation of the locative marker *-n* and is very rare; an example can be found in (11) above.

The feature [+constricted glottis] is confined to the rhyme. All three types of phenomena occur with the nucleus or after it.<sup>108</sup> Additional phonetic evidence comes from the articulation of creaky voice. In words such as *q̣l* [q̣l] 'husking pit' in (20) (repeated from example 4), the glottal stricture extends into a sonorant in coda position. However, the creaky phonation never extends into a sonorant in the onset of the following syllable. Thus the /m/ in *q̣maŋ* [q̣.maŋ] 'before' is not affected by the creaky phonation of the first syllable.

(20) q̣l [q̣l] 'husking pit'

(21) q̣maŋ [q̣.maŋ] 'before'

<sup>108</sup> According to my analysis, this restriction applies to all syllables in all situations. Hence, words such as /ḍaʔb.u/ 'close-INF.TR' (from ḍaʔb 'to close') violate the maximal onset principle. Nevertheless, the phonetics and the syllabification by speakers in slow speech confirm this interpretation. An alternative analysis with the syllable structure /ḍa.ʔbu/ would be possible. Under this interpretation, the restriction on glottal constriction would only hold for the stems in the lexicon and not for actual word forms.

In addition to the positional restriction, the three glottalized elements cannot combine with each other to form a cluster. The constraints on the distribution of these sounds lead to the situation that [+constricted glottis] can occur only once in a syllable. The positional and combinatorial constraints of the three phenomena result in the syllable patterns represented in Table 2:<sup>109</sup> They are grouped here into four groups. This first group comprises open syllables together with syllables with a liquid or nasal in coda position. The second group consists of the same syllable types, but with creaky voice phonation. The third and fourth groups are syllables with a glottal stop or a preglottalized obstruent in coda position. These four groups are relevant for the following discussion of syllable weight.

(C)V	(C)Ṽ	(C)Vʔ	(C)VʔO
(C)V(j/r/l)	(C)Ṽ(j/r/l)		
(C)VN	(C)ṼN		

Table 2: Syllable structures of Gorum

### Syllable Weight

Syllable weight seems to be intimately connected to glottalization. (C)Vʔ syllables involving glottal stops and (C)VʔO syllables with preglottalized obstruents are heavy. With creaky voice, the situation is more complex. It can occur with open (C)Ṽ syllables, (C)Ṽ(j/r/l) syllables with liquids in the coda, as well as (C)ṼN syllables with nasals in coda position. Without creaky voice, open syllables and syllables with glides are light syllables, but with creaky voice they are heavy. Non-glottalized syllables with nasals are ambiguous with respect to syllable weight.

Evidence for the relation between syllable weight and creaky voice comes from nominals. Anderson and Zide (2001) propose a bimoraic constraint on nominals in Proto-Munda; a similar constraint seems to be at work in Gorum, where most nouns are disyllabic. Of the monosyllabic nouns, most have the clearly bimoraic form (C)Vʔ or (C)VʔO. There are no monosyllabic nominals with light syllables such as CV or CV(j/r/l). There is, however, a small group of nouns of the form (C)Ṽ(j/r/l). Perhaps the most telling of them is the following pair:

(22) sur ‘to hunt’

(23) sɻr ‘hunting / a hunt’

While the verb in (22) has the form CVr, the corresponding noun in (23) contains an additional creaky voice and has thus the form CṼr. This is particularly interesting given that the bimoraic constraint only applies to nouns and not to verbs. Since the glottalization seems not to be part of the root, its presence is either the source or an effect of the second mora.

Thus, syllable weight in Gorum is closely connected to the presence of glottalization. All syllables containing [+constricted glottis] are heavy and can be considered bimoraic, although it is not evident in all cases which is the cause and which is the effect.<sup>110</sup>

109 Additionally, there is a very small number of exceptional native lexemes – such as /lup/ [lup<sup>h</sup>] ‘big’ – which end in a voiceless obstruent. These lexemes deviate from the patterns presented here. These voiceless obstruents also do not involve glottalization, but their laryngeal component is an aspirated release.

110 The status of syllables with a nasal in coda position is ambiguous, as pointed out above. Monosyllabic nominals with this coda type tend to be (C)ṼN. However, in contrast to clearly light syllables with a liquid phoneme in coda position, there are a few exceptions. The two pronouns *miŋ* ‘I’ and *maŋ* ‘you(SG)’ as well as the singular noun *zaŋ* ‘bone’ do not contain glottalization and either violate the bimoraic constraint or are bimoraic without any interaction with glottalization.

### The Phonological Stem

The phonological stem is the next higher level of prosodic organization relevant to the phonology of glottalization in Gorum. There is a constraint on the number of glottalized syllables operating on stem-level. On the other hand, the stem is also crucial for the placement of glottalization in affectedness marking.

#### The Glottal Constraint

On the level of the phonological stem, there is a restriction on the number of syllables with glottalization that can occur. In general, only a single glottal element is allowed in a stem. Thus in reduplication of roots containing glottalization, the glottalized element is lost in the reduplicant. This is illustrated in examples (24) through (26); the reduplicant in (24) is a faithful copy of the stem, while in (25) and (26) the segment in the coda which contains the glottalization is lost in the reduplicant.

(24) zum ‘to eat’ → zumzum

(25) gaʔ ‘to eat’ → gagaʔ

(26) gaʔd ‘to cut’ → gagaʔd

This constraint on glottal elements is also at work with the causative prefix, which takes the form *ab-* with roots containing glottalization and *aʔb-*<sup>111</sup> with roots without the constricted glottis feature. In contrast to the reduplicant in (26), where the glottalized coda is lost, the coda segment – /ʔb/ or /b/ – is present in both (27) and (28), but the presence of glottalization depends on the nature of the root.

(27) ab + soʔj ‘causative + to learn’ → ab-soʔj

(28) ab + sunj ‘causative + to fall’ → aʔb-sunj

The three types of glottal phenomena form a class which can be interpreted as natural if we assume that all three involve the same feature [+constricted glottis]. Furthermore, the stem is the relevant domain here. This can be seen from example (29), which contains two glottal elements. However, the preglottalised obstruent /ʔd/ is part of the stem *dimaʔd* ‘to sleep’, while the glottal stop /ʔ/ marks affectedness on the non-past suffix *-tu*. This shows that the constraint does not extend over a stem boundary.

(29) *dimaʔd#-tuʔ* ‘sleep#-NPST:AFF’

The affectedness marker is discussed in more detail in the following section.

#### Affectedness Marking

Affectedness marking is another part of Gorum grammar in which glottalization interacts with syllable structure and the phonological stem. This morpheme marks the medium voice of a verb. It is obligatory with one class of verbs, while it has detransitivizing effects on other verb classes.

The morpheme takes two forms, depending on the syllable structure it encounters. In those cases where it combines with an open syllable, it has the form of a glottal stop, resulting in a (C)Vʔ syllable. If the coda position is occupied by a nasal, affectedness marking takes the form of creaky voice phonation, resulting in a (C)V̄N syllable.<sup>112</sup> In either case, affectedness marking takes the form of glottalization and is thus solely an instantiation of [+constricted glottis].<sup>113</sup>

<sup>111</sup> The glottalized form of the prefix could be either *aʔb-* or *qb-*. The phonetic evidence is not unequivocal, yet the preglottalized obstruent is the most likely solution.

<sup>112</sup> Other forms do not occur, as verbal suffixes only have CV or CVN syllable structures.

<sup>113</sup> This is a clear case of allomorphy. If the affectedness marking is considered in isolation, it looks very much as if the glottal stop and creaky voice were allophones conditioned by the prosodic structure. This seems to have influenced Aze’s conception of the phonological status of the three types of glottal phenomena.

The morphological behaviour of the affectedness morpheme is unique in the grammar of Gorum. Its placement is not relative to other morphemes, but solely prosodic. The vital notions determining the placement are the right stem boundary and syllable structure; it occurs in the first rhyme following the stem regardless of any other aspect of the morphological structure of the verb. Due to the general restriction on stem boundaries in Gorum to coincide with a syllable boundary, the syllable following the stem is the place of affectedness marking. This is schematically represented in (30). (31) to (34) give concrete examples of the placement relative only to the stem boundary, as well as of the allomorphy depending on the syllable structure. They also show that morpheme order and other morphological aspects are irrelevant.

(30) STEM# $\sigma^{\text{AFF}}$

(31) (ne) $\sigma$ -(ko) $\sigma$ (ko) $\sigma$ #-(tuʔ) $\sigma$  ‘I will sit’

1sA-sit-NPST:AFF

(32) (or) $\sigma$ -(giʔi) $\sigma$ #-(n-aj) $\sigma$  ‘he/she/it is not visible’

NEG:NPST-see-INF.INTR-CISL:AFF

(33) (du) $\sigma$ (ku) $\sigma$ #-(r-aj) $\sigma$  ‘they were’

be-PST-3pS:AFF

(34) (du) $\sigma$ (k#-ij) $\sigma$ -(aj) $\sigma$  ‘he/she/it is to me’

be-1sP:AFF-CISL

Example (34) is exceptional: the right boundary of the morphological stem in Gorum generally corresponds to a syllable boundary. However, in the case of the irregular verb *duku* ‘to be’, the stem boundary does not coincide with a syllable boundary, but the affectedness morpheme is nevertheless positioned directly after the boundary in the rhyme of this syllable.

### Stem Patterns

In addition to the rules discussed above, there are some strong statistical tendencies in the distribution of the three types of glottal phenomena in stems. The examples (35) to (42) give an impression of these patterns.

(35) gaʔ ‘to eat’

(36) seʔb ‘to chop’

(37) buʔl ‘to be drunk’

(38) tu.paʔd ‘to thresh’

(39) kin.daʔ ‘river’

(40) gɔ.tuŋ ‘cloth’

(41) aŋ.an.aʔd ‘door’

(42) bi.ɔ.gi ‘tomorrow’

As can be seen, there is a strong tendency for the glottal stop /ʔ/ and the preglottalized obstruents /ʔO/ to occur in the last syllable of a stem and, as such, at its right boundary, while creaky voice /ʔ̤/ prefers the penultimate syllable. In monosyllabic stems this tendency is neutralized, so that no complementary distribution arises. These patterns result in the following stem structures:

monosyllabic stems	disyllabic stems	trisyllabic stems
<i>CV?</i>	<i>CVC.CV?</i>	<i>CVC.CVC.CV?</i>
<i>CV<sup>o</sup>O</i>	<i>CVC.CV<sup>o</sup>O</i>	<i>CVC.CVC.CV<sup>o</sup>O</i>
<i>CYC</i>	<i>CYC.CVC</i>	<i>CVC.CYC.CVC</i>

Table 3: Distribution of glottalization in word stems

In a small number of stems this pattern is broken, but most of these cases seem to be frozen compounds. An example is the word *mitq̃n* ‘today’, in which creaky voice occurs in the last syllable. However, this lexeme, which speakers today view as a simplex, can be analyzed as a compound of the word *mit* ‘day’ and the synchronically unattested demonstrative *\*q̃n*, probably meaning ‘this day’. This putative demonstrative can be connected to the still used *qt* ‘that day’, parallel to the pair *q̃n* ‘hither’ and *qt* ‘thither’.

A peculiar pattern occurs with some Telugu or Oriya loan words which have a CV.CV structure in the source language. In Gorum, their first syllable became glottalized, so that the resulting prosodic structure is CṾ.CV. This pattern is not productive anymore, but at some point in the history of Gorum it must have been. The following two words are good representatives of this group.

(43) *ḍopa* ‘leaf bowl’ from Telugu *doppa*

(44) *kạḍu* ‘bangles’ from Desia Oriya *kaḍu*

The motivation for the presence of creaky voice in the first syllable of these words is unclear. None of the source languages has creaky voice as a sound phenomenon, let alone as a phoneme. Some of these lexemes, such as (43), contain a geminate in the source language, a phonological structure that does not exist in Gorum. However, since not all of these lexemes originally contained geminates, this cannot be the defining condition for the phenomenon. Whatever the motivation for the creaky voice in these words, they show that under some conditions glottalization could come into existence in lexemes that originally did not have it. This poses some problems for the historical reconstruction of this aspect of Gorum phonology. I will address this issue in the following section.

### Historical Significance

The previous discussion – especially the occurrence of creaky voice in loan words – raises some problems for the historical treatment of glottalization in Gorum. Of the three glottal phenomena, two are comparatively well understood. The glottal stop is present all over the Austroasiatic family and preglottalized obstruents are found in virtually every Munda language. Also, the history of both seems to be relatively straightforward.

Creaky voice, however, is intriguing: it occurs in several Austroasiatic languages, but its history is still nebulous. Gorum is the only Munda language in which it is phonemic. Nevertheless, it has been claimed to be reconstructable for Proto-Munda (Zide 1976 as well as Zide and Zide 1987). Diffloth (1989) argues that it goes back to Proto-Austroasiatic. Indeed the evidence from all over the language family is quite suggestive, yet the history of this phoneme is still not well understood. This may be due to a variety of reasons. In the Munda languages, it appears that several developments, besides segmental sound changes, have obscured the picture.

The most frequent occurrence of creaky voice in Gorum is probably the affectedness morpheme. This morpheme has no known direct correspondences in other Munda languages, so its history is difficult to determine. Its allomorphy with a creaky voice and a glottal stop allomorph considerably complicates the historical phonology of this morpheme: given the phonological similarity between the two allomorphs, every putative comparative evidence could pertain to either one of the two glottal phonemes.

Further difficulties for a comparative approach stem from the fact that in nominals such as *sur* ‘hunt’ (example 23), the creaky voice seems to be connected with the second mora and might very well be a

derivational device. In other lexemes – such as *gsuj* ‘house’ – the creaky voice does not seem to be part of the root either. In this particular case, internal and comparative evidence suggests that the root is *suj*, while *g-* seems to be a derivational prefix. Thus in some cases the creaky voice cannot be established as part of the root and its emergence at some point in history is not understood yet.<sup>114</sup>

In addition to those cases where creaky voice cannot be ascertained to be part of the etymon, there are also words where it cannot possibly be part of the original root. These loan words with a CV.CV-structure are evidence that at some point in the history of Gorum a productive pattern existed that gave rise to creaky voice in these lexemes. As long as the motivation for its emergence in these words is not understood, this poses a severe problem for the historical treatment of creaky voice.

Perhaps the most fruitful starting point for a comparative approach would focus on words in which creaky voice occurs in the root and no known pattern motivates its emergence. To my knowledge the monosyllabic verbs in (45)-(48) are the best candidates for this. The other four lexemes are also likely candidates. The only caveat is that *l̥aki* has a CVCV structure, but in contrast to the loans discussed above it is neither a loan word nor a noun. However, this list cannot resolve the fact that prosodic structure and word phonological processes seem to be the key to the understanding of this phenomenon.

(45) *b̥ul* ‘to be drunk’

(46) *d̥el* ‘to ripen’

(47) *m̥eŋ* ‘to live’

(48) *uŋ* ‘to perform a burial ritual’

(49) *̥amaŋ* ‘before’

(50) *̥gaŋ* ‘when’

(51) *l̥aki* ‘later’

(52) *bi̥gi* ‘tomorrow’

There is, I think, another option which should be considered: the phonemic creaky voice in Gorum could be the result of a split in this language, rather than a remnant of an old Proto-Munda phoneme. The tendency for a complementary distribution of creaky voice and glottal stop in stems and the allomorphy of affectedness marking could be interpreted in that direction. The current status of creaky voice in Sora and Juray would then be similar to the historical situation in Gorum.

On the other hand, Diffloth’s evidence strongly suggests that creaky voice is an old feature in the Austroasiatic family. However, the historical development, at least on the Munda side of the family, is so heavily obscured that historical reconstruction has to proceed with care.

## Conclusions

There are three phonemic types of glottalization in Gorum: the glottal stop, preglottalized obstruents, and creaky voice. These types are forms of one general phenomenon glottalization that can be represented as [+constricted glottis]. Its phonology is best understood in terms of syllable structure. Also, interactions with other elements, restrictions and regularities have to be captured on the even higher level of the phonological stem.

---

114 I am not convinced by the connection made by Zide and Zide (1987) between the loss of some instances of /s/ in Sora-Juray-Gorum and creaky voice. Since we do not understand the history of creaky voice and there are no systematic, if any, reflexes of it in other Munda languages (cf. Zide 1982, p. 367ff), it seems problematic to posit its existence in some lexemes at some earlier stage solely to explain another unexplained phenomenon.

The history of these phonemes is still poorly understood and in my opinion, a better understanding of prosodic structures in Munda languages, and especially the phonology of glottal constriction, is needed to understand the diachrony of these phenomena.

### References

- Anderson, G.D.S. and Zide, N.H. 2001. *Issues in Proto-Munda and Proto-Austroasiatic Nominal Derivation: The Bimoraic Constraint*, In: Macken, M.A. (ed.) *Papers from the 10th Annual Meeting of the Southeast Asian Linguistic Society*. Arizona State University, Tempe. 55–74
- Aze, R., 1971. *Parengi (Gorum) Phonemic Summary*. Summer Institute of Linguistics and Deccan College, Poona.
- Aze, R., 1974. *The status of glottalization in Parengi (Gorum)*. Proceedings of All India Conference of Linguistics, 1972, 84–95.
- Diffloth, G., 1989. “Proto-Austroasiatic creaky voice.” *Mon-Khmer Studies* 15, 139–154.
- Ghosh, A. 2008. “Santali”. In: Anderson, G. D. S. (ed.) *The Munda Languages*. Routledge Language Family Series. Routledge, Oxon. 11-98.
- Peterson, P. 2008. “Kharia”. In: Anderson, G. D.S. (ed.) *The Munda Languages*. Routledge Language Family Series. Routledge, Oxon. 434-507.
- Zide, A. R. K., 1963. *Gorum (Parengi) Phonology*, Unpublished M.A. thesis, University of Pennsylvania.
- Zide, A. R. K., 1966. *Some problems in the generation of Gorum obstruents and related glides*. University of Chicago (mimeo).
- Zide, N. H., 1976. “‘3’ and ‘4’ in South Munda”, *Linguistics* 174:89-98.
- Zide, A. R. K., 1982. *A reconstruction of Sora-Juray-Gorum phonology*. PhD thesis, University of Chicago.
- Zide, A. R. K., Zide, N. H., 1987. *A KM laryngeal as a conditioning factor for s-loss in Sora-Juray-Gorum*. In: Cardona, G., Zide, N. H. (Eds.), *Festschrift for Henry Hoenigswald: On the occasion of his seventieth birthday*. Narr, Tübingen, 417–431.



## Comitative PP as Head in Externally-Headed Relative Clauses in Khasi<sup>115</sup>

**Karumuri V. Subbarao**  
University of Hyderabad

**Gracious M. Temsen**  
University of Hyderabad

### Introduction

This paper attempts to discuss the case of the comitative PP as head in externally-headed relative clauses (EHRCs) (also labeled as: gap relatives) in Khasi, a Mon-Khmer language. While a sentential relative clause and a gap relative (EHRC) are both permitted with subject, direct object, indirect object, and all oblique objects (instrumental, ablative, locative PP) as head, the comitative PP is *unique* in not permitting a sentential, as well as gap relative, as head with the interpretation of *accompaniment*. A similar phenomenon is observed in all other South Asian languages and some other non-South Asian languages with regard to the comitative PP as head in gap relatives.

While there is considerable variation with regard to the positions that are ‘accessible’ in the Noun Phrase Accessibility Hierarchy (NPAH) of Keenan and Comrie (1977), in many South Asian languages of the four language families (Indo-Aryan, Dravidian, Austro-Asiatic (Mon-Khmer, Munda) and Tibeto-Burman), the modification of the comitative PP in gap relatives is not permitted in almost all South Asian languages. In the few languages in which it is permitted, specific syntactic criteria have to be fulfilled to enable the comitative PP to head the gap relative. In this paper we attempt to propose a condition which we shall label as the Thematic Eligibility Condition (TEC), and wish to demonstrate that only after fulfilling specific syntactic criteria, the requirement of TEC is satisfied to enable a comitative PP to head a relative or gap relative clause.

The processes that enable a comitative PP to head an EHRC or an IHRC include:

- (i) Comitative Adposition Incorporation in the embedded verb; as in Mizo, Hmar, Mising (Tibeto-Burman); or
- (ii) the overt occurrence of the postposition with the internal head of the comitative PP in IHRCs; as in Tenydie, Konyak and Sangtam (Tibeto-Burman); or
- (iii) the occurrence of the verbal reciprocal that functions as a group marker together with the occurrence of an adverb with the interpretation of ‘together’ in the embedded verb; as in Khasi (Mon Khmer) and Manipuri (Tibeto-Burman).

Thus, either some elements are added to the embedded predicate of the gap relative, or some syntactic process such as *Adposition Incorporation* takes place, or the internal head of an IHRC must have an overt postposition indicating accompaniment as in some Tibeto-Burman languages in the few languages that permit a comitative PP to head an EHRC or an IHRC.

---

<sup>115</sup> We are thankful to Alice Davison, Rajesh Bhatt, Hans H. Hock, George van Driem and Juanita War for their suggestions. Thanks are also due to Matt Shibatani (for Japanese and Korean), Tatiana Oranskia (for German & Russian), Ludmila Khoklova and Boris Zakharyn (for Russian) for providing either data or information concerning the occurrence of the comitative PP in relative clauses. Thanks are also due to Paul Sidwell for his continuous support in many ways.

Subbarao, Karumuri V. Gracious M. Temsen. 2011. “Comitative PP as Head in Externally-Headed Relative Clauses in Khasi.” In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.184-201.

Copyright vested in the authors.

In this paper we provide data that demonstrate that the process mentioned in (iii) above is found in Khasi in EHRCs and we provide arguments to show why such markers are included in the embedded verb.

We shall show that comitative PPs (either Prepositional or Postpositional) that are *thematically linked* to the embedded predicate freely permit the modification of the comitative PP. We argue that a thematic relation has to be established between the predicate of the embedded clause and the head of the EHRC/IHRC, in terms of either an *overt case marker with the head*, or some marker in the embedded verb to indicate accompaniment. In Khasi, we show that such a thematic relationship is established by the incorporation of the adverb *-laŋ* ‘together’, and the verbal reciprocal *-ya-* in the embedded verb. The verbal reciprocal *-ya-* in Khasi also functions as a group marker, a hortative marker, collaborative effort marker (cf. Subbarao ms.). We also show that in case the embedded predicate is inherently reciprocal, the adverb *-laŋ* ‘together’ is not needed, as a reciprocal predicate necessarily implies a combined activity by two or more than two participants.

## 2. A few facts about Khasi

Khasi, a Mon-Khmer language, is the only *non-verb-final* language in the South Asian subcontinent<sup>116</sup>. It has predominantly verb-medial structures, though there is a set of structures in which the verb obligatorily occurs in the initial position (see Tensen, 2007).

Conforming to its non-verb-final structure, Khasi has prepositions and the auxiliary verb invariably occurs to the *left* of the verb. The complementizer *ba* occurs to the left of the embedded complement clause. Khasi has subject-verb agreement and all adjectives exhibit agreement with the modifying noun. Personal pronouns such as *u* ‘he’, *ka* ‘she’, and *ki* ‘they’ are homophonous with agreement markers on the verb, the noun and the adjective as markers indicating gender. Khasi is also a null pronominal language. It has both verbal and nominal reciprocals. The verbal reciprocal performs several other functions such as a group marker, a hortative marker, and a collaborative effort marker (*cem*) (see Subbarao, ms.). Khasi has full sentential relative clauses (The English *wh*-type) and Externally-Headed Relative Clauses (EHRCs), also called as gap relatives, and conforming to its non-verb-final pattern, it does not have Internally-Headed Relative Clauses (IHRCs). The marker *ba* which occurs with adjectives also functions like an adjectivalizer in gap relatives as well as a relative pronoun.

## 3. Relative Clauses and EHRCs in Khasi

### 3.1. Relative Clauses in Khasi

The relative pronoun is *ba*, it occurs to the left of the embedded clause. We provide below two examples of Oblique Object modification. Sentence (1) is an example of a sentential relative clause. The head is a locative PP and the locative object *mē<sup>v</sup>d* ‘table’ is modified.

Locative PP as head– Relative clause

1. **ka-mē<sup>v</sup>d<sub>I</sub> ha-ka ba** u-lam u-bō? ya-ka-kōt **t<sub>I</sub>** ka-la?-kd<sup>v</sup>a?  
 f.s-table loc-3f.s which m.s-Lam 3m.s-put acc-f.s-book 3f.s-perf-break  
 ‘The table on which Lam put the book is broken.’

Sentence (2) is an example of ablative object modification in which the ablative object *jaka* ‘place’ is modified.

---

<sup>116</sup> Pnar, Mnar, Jowai and Rymbai (Mon-Khmer languages) are also verb-medial like Khasi. See Bareh (2007) for Jowai and Rymbai.

## Ablative PP as Head -Relative clause

2. **ka-jaka<sub>I</sub> na-ka ba** u-lam u-wan *t<sub>I</sub>* ka-ləŋ ka-ba- jŋāi  
 f.s-place loc-3f.s which m.s-Lam 3m.s-come 3f.s-be 3f.s-adjr- far  
 ‘The place which Lam came from is far.’

In (1) and (2) the relativized noun phrase (in bold in (1) and (2)) moves leftward to the initial position, namely, to Spec CP of the main clause. An important fact about the relative clause in all South Asian Languages (SALs) is that the relativized NP is generally *fronted*, if it is a *non-subject* (DO, IO, or OO), just as in English and French. The position from which the NP is fronted is indicated by the trace *t<sub>i</sub>*, which is coindexed with the relativized NP in (1) and (2).

## 3. 2. EHRCs in Khasi

An Externally-headed (Gap) relative that corresponds to (1) is given in (3). The embedded clause in an EHRC functions like an adjectival modifier with *ba* functioning as an adjectivalizer.

## Locative pp modified-ehrc ( gap relative)

3. ka-mē<sup>v</sup>d<sub>i</sub> [<sub>AP</sub> ba<sup>117</sup>- u-lam u-bō? ya-ka-kət  $\emptyset_{iAP}$ ] ka-la?-kd<sup>v</sup>a?  
 f.s-table adjr- s.m-Lam 3m.s-put acc-f.s-book 3f.s-perf-break  
 ‘The table Lam put the book on is broken.’

In the Gap Relative in (3) above: (i) there is a gap (a null operator) in the embedded clause (indicated by  $\emptyset$ ), as the identical noun phrase does not overtly occur; (ii) the locative preposition indicating the grammatical relation between the head of the gap relative and the embedded predicate does not occur, as EHRCs universally do not carry an adposition (preposition or postposition) with the head that is relativized; and (iii) the embedded clause occurs to the right of the head noun as modifiers in Khasi always occur to the right of the head they modify.

For each position on the Noun Phrase Accessibility Hierarchy (NPAH)<sup>118</sup> of Keenan and Comrie (1977) there is a sentential relative clause as well as a corresponding EHRC available in Khasi, except for the Comitative PP unless some elements are incorporated in the embedded verb. The positions that are relativizable in Khasi are: the Subject, the Direct Object, the Indirect Object, the Object of the Genitive, the Oblique Object, which includes the Locative, the Ablative, and the Instrumental PPs and so on.

## Formation of EHRCs with the Comitative PP as Head

This section provides a detailed description on the relativization of the Comitative PP as Head indicating *accompaniment*. In sentence (4) below, we first provide an example of the comitative PP in a simple clause.

4. u-lam u-wan bad ka-wei ka-khīnna? *ša-ka-jīŋkhawai*  
 m.s-Lam 3m.s-come com 3f.s-one f.s-child to-f.s-party  
 ‘Lam came to the party with a girl.’

<sup>117</sup> Adjectives in Khasi too carry the marker *ba*. Thus, the entire following phrase functions like an adjectival phrase (AP):

ba- u-lam u-bō? ya- ka-kət  
 ‘adjr- m,s-Lam- 3sm-put acc-3f,s-book’

<sup>118</sup> The NPAH constraints proposed in Keenan and Comrie (1977) are:

1. A language must be able to relativize subjects.
2. The strategy of relativization must apply to a continuous segment of the NPAH-scale.
3. Strategies that apply at one point of the NPAH-scale may in principle cease at any point on the scale.

In sentence (5) below, the group marker *-ya-* and the adverb *-laŋ-* ‘together’ occur with the verb. The occurrence of these two markers together imparts the specific interpretation that the act of arriving of Lam and the girl at the party happened *at the same point of time*, while sentence (4) does not necessarily have the implication that Lam and the girl came together.

5. u-lam u-ya<sup>119</sup>-wan-laŋ bad ka-wei ka-khinna? *ša-ka-jɪŋkhawai*  
 m.s-Lam 3m.s-gm-come-together com 3f.s-one f.s-child to-f.s-party  
 ‘Lam came to the party together with a girl.’

Sentences (6a & 6b) and (7) below provide evidence in support of our claim. Sentences in (6) are grammatical, while sentence (7) is not, as it is a contradiction. Thus, when two individuals perform an action *together* towards achieving a goal with a common purpose, the adverb *-laŋ-* ‘together’ and the verbal reciprocal *-ya-* which also functions as a group marker obligatorily occur with the verb.

- 6a. u-lam bad ka-wei ka-khinna? ki-ya-wan  
 m.s-Lam com 3f.s-one f.s-child pl-gm-come  
*ša-ka-jɪŋkhawai* taŋ-ba ki-khlɛm-ya-wan-laŋ  
 to-f.s-party only-that pl-neg-gm-come -together  
 ‘Lam came to the party with a girl but they did not come together.’

- 6b. u-lam u-wan *ša-ka-jɪŋkhawai* bad ka-wei  
 m.s-Lam 3m.s-come to-f.s-party com 3f.s-one  
 ka-khinna? taŋ-ba ki-khlɛm-ya-wan-laŋ  
 f.s-child only-that pl-neg-gm-come-together  
 ‘Lam came to the party with a girl but they did not come together.’

7. \*u-lam u-ya-wan-**laŋ** bad ka-wei ka-khinna? *ša-ka-jɪŋkhawai*  
 m.s-Lam 3m.s-gm-come-together com 3f.s-one f.s-child to-f.s-party  
 taŋ-ba ki-khlɛm-**ya**-wan-laŋ  
 only-that pl-neg-gm-come-together  
 ‘\*Lam came to the party with a girl but they did not come together.’

We shall demonstrate that it is the occurrence of these two verbal elements (adverb *-laŋ-* ‘together’ and the verbal reciprocal *-ya-*) that permits the comitative PP to head a relative clause and an EHRC.

We shall now discuss the case of relativization of the comitative PP in sentential relatives.

#### 4.1. Relativization of the comitative PP in Sentential Relatives

Sentence (8) below is an example of a sentential relative in which the comitative PP is relativized. In this sentence, the head of the matrix clause is *ka-khinna?* ‘f.s-child’. The relativized NP along with the preposition *bad-ka-ba* ‘with-she-who’, that is, ‘with whom’ is fronted and *t<sub>1</sub>* (trace<sub>1</sub>) in bold in (8) indicates the position in which the comitative PP originates.

---

<sup>119</sup> Note that *ya-* functions as an accusative case marker, a group marker, a hortative marker, a collaborative effort marker as well as a verbal reciprocal in Khasi.

8. ka-khinna? [<sub>S2</sub> bad-ka-ba u-lam u-ya-wan-laj t<sub>1</sub> s<sub>2</sub>]  
 f.s-child com-3f.s-adjr 3m.s-Lam 3m.s-gm-come-together  
 ka-dei ka-para (jɔŋ<sup>120</sup>)-u-kɛr  
 3f.s-be f.s-younger sibling gen-m.s-Ker  
 ‘The girl who Lam came with is Ker’s sister’

The crucial point about (8) is the obligatory occurrence of the two markers– the group marker *-ya-* and *-laj* ‘together’ which are not present in the modification of any other PP in Khasi. Recall that these two markers are not needed in a simple sentence (4) unless the act of the two participants *coming together* is focused upon as in (5). We shall argue that it is the incorporation of these two markers in the embedded verb that enables the comitative PP to be the head of an EHRC in Khasi.

#### 4.2. The Gap relative (Externally-Headed Relative Clause-EHRC)

An EHRC *cannot be formed* with the comitative PP *ka-khinna?* ‘f.s-child’ as head as (9) shows. Note that in (9) the comitative preposition *bad* ‘with’ does not overtly occur.

9. \**ka-khinna?*<sub>i</sub> [ba u-ban u-wan ø<sub>i</sub> ša-šnɔŋ<sub>AP</sub>]  
 f.s-child adjr m.s-Ban 3m.s-come all-village  
 ‘The girl with whom Ban came to the village ...’

Sentence (9) is an example of a Gap Relative (EHRC) in which neither the relativized NP occurs nor does the preposition *bad* ‘with’ occur. The reason for the non-occurrence of the comitative preposition is: externally-headed gap relatives *universally* do not carry the adposition that denotes the case relation of the head noun phrase with the embedded verb. We have indicated the gap in the embedded clause by ø<sub>i</sub> and the gap is coindexed with the subject of matrix clause *ka-khinna?* ‘f.s-child’.

Sentence (9) is ungrammatical because the verbal reciprocal with the interpretation of a group marker and the adverb *-laj* ‘together’ are not incorporated in the embedded verb. In contrast to (9), (10) is well-formed because the group marker *-ya-* and the adverb *-laj* ‘together’, both occur with the embedded verb, just as in sentential relatives (as in sentence (8) above).

Comitative pp as head with a verbal reciprocal *ya-* and the adverb *-laj* ‘together’

10. *ka-khinna?*<sub>i</sub> [<sub>AP</sub> ba u-ban u-ya-wan-laj ø<sub>i</sub> ša-šnɔŋ<sub>AP</sub>]  
 f.s-child adjr m.s-Ban 3m.s-gm-come-together all-village  
 ‘The girl with whom Ban came to the village ...’

The other predicates that require the group marker *-ya-* and the adverb *-laj* ‘together’ in the formation of a relative clause and a gap relative include: *mare?* ‘run’, *ya:<sup>v</sup>dkāi* ‘walk’, *jji-* ‘swim’ among others. The following examples are illustrative.

11. *ka-khinna?*<sub>i</sub> [<sub>AP</sub> ba- u-ban u-ya-mare?-laj ø<sub>i</sub><sub>AP</sub>]  
 f.s-child adjr- m.s-Ban 3m.s-gm-run-together  
 ‘The girl with whom Ban is running ...’
12. *ka-khinna?*<sub>i</sub> [<sub>AP</sub> ba u-ban u-ya-yā<sup>v</sup>dkāi -laj ø<sub>i</sub><sub>AP</sub>]  
 f.s-child adjr m.s-Ban 3m.s-gm-walk-together  
 ‘The girl Ban is walking with ...’

<sup>120</sup> The genitive marker *jɔŋ* is obligatory with personal pronouns, but may optionally be deleted if the noun that follows it is either a proper noun or a non-specific noun. E.g., *ka-kɔt-jɔŋ-ŋa* ‘my book’, *ka-kɔt-(jɔŋ-)u-bɔn* ‘Bon’s book’, *ka-tnat-(jɔŋ-)u-dēŋ* ‘A tree’s branch’.

13. *ka-khinna?*<sub>i</sub> [ <sub>AP</sub> ba u-ban u-*ya-jɔŋ* -*lay*  $\emptyset_i$  ]  
 f.s-child adjr m.s-Ban 3m.s-gm-swim-**together**  
 ‘The girl Ban is swimming with ...’

Verbs such as *ya-kren* ‘converse’, *ya-thɔ?* ‘marry’<sup>121</sup>, *ya-šɔ?* ‘fight’, *ya-kindō?* ‘meet’ which are inherently reciprocal verbs do not require the incorporation of the adverb **-lay** ‘together’, and hence, the occurrence of **-lay** with the embedded verb is prohibited in gap relatives as well as sentential relatives.

14. *ka-khinna?* [ ba u-ban u-*ya-kren*  $\emptyset_{iAP}$  ] ...  
 f.s-child adjr m.s-Ban 3m.s-*Vrec*-speak  
 ‘The girl Ban is talking with ...’

15. *ka-khinna?* [ ba u-ban u-*ya-thɔ?*  $\emptyset_{iAP}$  ] ...  
 f.s-child adjr m.s-Ban 3m.s-*Vrec*-write  
 ‘The girl Ban is married to ...’

16. *ka-khinna?* [ ba u-ban u-*ya-šɔ?*  $\emptyset_{iAP}$  ] ...  
 f.s-child adjr m.s-Ban 3m.s-*Vrec*-beat  
 ‘The girl Ban is fighting with ...’

17. *ka-khinna?* [ ba- u-ban u-*ya-kindō?*  $\emptyset_{iPP}$  ] ...  
 f.s-child adjr m.s-Ban 3m.s-*Vrec*-meet  
 ‘The girl Ban met with ...’

The other inherently reciprocal verbs that take a comitative PP are *-ya-le?kai* ‘play’, *ya-mai* ‘fight (verbally)’

We shall argue that it is the incorporation of these two markers *-ya* ‘group marker’ and *-lay* ‘together’ in the embedded predicate that enables a comitative PP to head an EHRC. Note that a comitative prepositional phrase which is neither an essential argument nor a subcategorized argument of a predicate gains the status of an argument that is thematically-linked to the embedded predicate due to the incorporation of these two markers.

The question that arises now is: What is a thematically-linked argument and what makes an NP eligible to head a relative clause and a gap relative in Khasi? First, let us enumerate the positions in which a relative clause and the gap relative occur in Khasi. These positions include the subject, the direct object, the indirect object, the object of a preposition (ablative, instrumental, locative and comitative) and the object of the genitive.

For an NP or a PP to be able to occur as an argument in a clause, it should be either a subject or an object (direct or indirect) or a PP that is thematically-linked to the predicate of the clause.

A question that arises now is: what makes an NP eligible to head an EHRC in Khasi? Can the notion of subcategorization<sup>122</sup>, for example, play any role in making an argument of the embedded predicate eligible to head an EHRC, or is there some other factor? We wish to show that subcategorization has no role to play, but it is the syntactic phenomenon of incorporation that operates on the embedded predicate that make an argument eligible to head a comitative PP.

<sup>121</sup> Note that *thɔ?* means ‘write’ and *ya-thɔ?* means ‘vrec-write’ that is ‘marry’, which literally means, ‘to write to each other’. This might have started with the tradition of writing/having to write an agreement on paper to solemnize a relationship.

<sup>122</sup> Chomsky (1965) used the term subcategorization. Subcategorization in simple terms means the number of essential arguments that a verb can take. As Lasnik and Uriagereka with Boeckx (2005: 3) put it: ‘(a) lexical entry must contain syntactic information about whether, for instance, a verb is transitive, intransitive, ditransitive, and so on (so-called *subcategorization*).’ See Chomsky (1965) for further details.

Before we proceed further to explicate the syntactic phenomenon involved, we shall first mention some facts related to the issue. There are two specific sets of verbs that can take a comitative PP in a clause. Set 1 includes embedded verbs such as *wan* ‘come’, *leyt* ‘go’, *mare?* ‘run’ etc. which take a comitative PP with the interpretation of *accompaniment*. This set of verbs is in *no way linked* to the embedded verb either thematically or in terms of subcategorization. Set 2 includes embedded verbs such as *ya-kren* ‘talk’, *ya-tho?* ‘marry’, *ya-šo?* ‘fight’ which are inherently reciprocal verbs and they are not verbs indicating *accompaniment*. In such sentences, the subject of the embedded clause and the head of the comitative PP are in a reciprocal relationship. The object that occurs with such a set of verbs is a *subcategorized argument*. Furthermore, in an EHRC, the argument of the embedded clause that heads an EHRC must be in a *specific grammatical relation* with the embedded predicate. The arguments that qualify universally to head an EHRC include subject, direct object, indirect object, object of the genitive and the oblique object that includes locative, ablative and instrumental PPs.

A comitative PP can head an EHRC in South Asian languages if and only if either:

- (i) postposition Incorporation in the embedded verb takes place in a verb-final language, or
- (ii) the verb carries a marker that indicates some group activity together with the occurrence of an adverb with the interpretation of ‘together’, or
- (iii) the head of an Internally-Headed Relative Clause (IHRC) carries overtly the comitative PP (Subbarao 2010 a).

Let us examine if the notion of subcategorization mentioned above is of any relevance in accounting for the modification of the comitative PP. As mentioned above, the NPs that can head an EHRC are: Subject, Direct Object, Indirect Object, Oblique Object (Locative, Ablative, Instrumental, etc.). Note that Subject is an external argument of the embedded verb and hence, it is not a subcategorized argument. Direct and indirect objects are subcategorized arguments of the embedded verb.

We now look at the Oblique Objects. Locative PP is a subcategorized argument in case of predicates such as *keep*, *place*, *live* etc. Predicates such as *sleep*, *sit*, *walk*, *run* etc. that take a locative PP too are thematically linked to the predicate of the embedded clause though they are not subcategorized arguments of the predicate.

The instrumental PP and the ablative PP are *not* subcategorized arguments of the embedded verb, but they too are *thematically linked* to the (embedded) verb as the occurrence of the instrumental and ablative PP is crucially dependent on the nature of the predicate. A predicate such as *cut*<sup>123</sup> or *slice* for example invariably takes an instrumental PP and predicates such as *get down*, *alight*, *bring* for example require a PP which is source or an entity that potentially moves.

We now consider the comitative PP that denotes *accompaniment*. The comitative PP is in no way connected/ related to the embedded predicate either in terms of subcategorization or thematic relation. Predicates such as *come*, *go*, *walk*, *swim* can take a comitative PP. In fact, that is the reason why the comitative PP does not form the head of an EHRC in any South Asian language and some genetically unrelated languages such as Japanese and Korean unless: (i) the embedded predicate carries a special marker or markers that indicate accompaniment and some group activity; or (ii) the comitative PP as head in an Internally-Headed Relative Clause (IHRC) carries an overt comitative postposition; or (iii) the embedded

<sup>123</sup> Let us consider (i) from Khasi.

(i) *ka-tari*<sub>i</sub> [ba u-ban u-ɔt ya-u-sə? ø<sub>i</sub>] ka-ləŋ ka-ba-ləŋ  
 f.s-knife adjr m.s-Ban 3m.s-cut acc-m.s-fruit 3f.s-be 3f.s-adjr-blunt  
 ‘The knife Ban cut the fruit with is blunt.’

Note that in (i) there is no special marker that occurs with the embedded verb that shows the relationship between *ɔt* ‘cut’ and *tari* ‘knife’. However, the predicate *ɔt* ‘cut’ and the NP *tari* ‘knife’ are thematically linked as the verb *cut* requires an instrument such as *knife* for cutting.

predicate in an EHRC carries an incorporated comitative postposition (Subbarao 2010 a). Thus, such special markers include either an incorporated postposition in the embedded predicate as in Mizo, Hmar or Thadou (Tibeto-Burman) or the comitative postposition that overtly occurs with the head of a comitative PP in an Internally-Headed Relative Clause (IHRC) as in Sangtam or Tenyidie, or Konyak (Tibeto-Burman) or the occurrence of a group marker and an adverb that indicates *togetherness* in the embedded verb as in Khasi (Austro-Asiatic) and Manipuri (Tibeto-Burman) (Subbarao 2010 a, c).

Regarding the case of the comitative PP in Khasi, as we have seen, a comitative PP cannot head an EHRC. The ungrammaticality of (9) above is illustrative. However, when the verbal reciprocal functioning as a group marker and the adverb indicating *together* occur with the embedded verb, the sentence is grammatical as in (10)-(13).

The question that arises is: What makes the head of the comitative PP eligible to head an EHRC in Khasi? We wish to claim that just as an instrument such as a *knife*, or *saw* is *thematically linked* to verbs such as *cut*, *slice*, *saw*, the occurrence of the group marker and the adverb indicating *together* with the embedded verb makes the object of the comitative PP eligible to get thematically-linked. Thus, the Thematic Eligibility Condition (TEC) is met and thus, it is this thematic linking that makes the comitative PP qualified to head an EHRC in Khasi.

The object of the genitive is not an argument of the embedded verb and hence, we ignore it for the present discussion.

### 5. Evidence in support of Incorporation in the embedded predicate in a comitative PP in an EHRC

We present below three pieces of evidence in support of our hypothesis concerning thematic linkage.

(i) The first piece of evidence comes from the **grammaticalization of the** verbal reciprocal *-ya-* and the adverb *-laŋ* ‘together’ to form a new predicate. These two elements together constitute a predicate in Khasi that imparts the meaning of ‘meet’ / ‘get together’ / ‘gather in one place’ / ‘date someone’ as in (18).

18.   ya-laŋ  
      vrec -together  
      ‘meet’ / ‘get together’ / ‘gather in one place’ / ‘date someone’

When the nominalizer *jiŋ* occurs with the predicate *ya-laŋ* ‘v rec -together’, it imparts the interpretation of ‘meeting’.

19.   jiŋ-   ya-   laŋ  
      nozr- vrec - together  
      ‘meeting’

The fact that the predicate *ya-laŋ* has the interpretation of ‘get together/gather in one place’ / ‘date someone’ shows that the interpretation of ‘togetherness’ is imparted by the combination of *-ya-* and *-laŋ*, and that is precisely what helps the comitative PP to head the EHRC in Khasi.

(ii) The second piece of evidence comes from inherently reciprocal predicates such as *ya-šɔʔ* ‘fight’, *ya-kɾɛn* ‘talk’, *ya-thɔʔ* ‘marry’, *ya-kindɔʔ* ‘meet’, *ya-kdup* ‘hug’/‘embrace’, *ya-dɔʔ* ‘kiss’ which subcategorize for a direct/indirect object which may take a comitative postposition. Recall that these verbs do not indicate accompaniment. In this set of reciprocal predicates, there are a few predicates that *also* take an accusative object as well in which case they are examples reflecting group activity and not a reciprocal activity in Khasi. We found only two predicates so far which take either an accusative or a comitative postposition. These are *ya-thɔʔ* ‘marry’ and *ya-kindɔʔ* ‘meet’.

20.   ka-šɪda   ka-ya-thɔʔ   ya-/bad-   u-ban  
      f.s-Shida   3f.s-marry   acc/com   m.s-Ban  
      ‘Shida married Ban.’



21. ka-šida ka-ya-kindō? <sup>124</sup> ya-/bad- u-ban  
 f.s-Shida 3f.s-meet acc/com m.s-Ban  
 ‘Shida met Ban.’

With such predicates when a comitative PP heads either a sentential relative clause or an EHRC, only the verbal reciprocal *-ya* which is inherently present with the predicate occurs and the adverb *-laŋ* ‘together’ cannot occur. Sentences (22)-(25) are illustrative<sup>125</sup>.

22. ka-khinna? [ba u-ban u-ya-šō? ] ka-dei  
 f.s-child adjr m.s-Ban 3m.s-vrec-beat 3f.s-be  
 ka-paralok jōŋ-ŋa  
 f.s-friend gen-1s  
 ‘The girl Ban is fighting with is my friend’

23. ka-khinna? [ba u-ban u-ya-kreŋ ] ka-dei  
 f.s-child adjr m.s-Ban 3m.s-vrec-speak 3f.s-be  
 ka-paralok jōŋ-ŋa  
 f.s-friend gen-1s  
 ‘The girl Ban is talking with is my friend.’

24. ka-khinna? [ba u-ban u-ya-thō? ] ka-dei  
 f.s-child adjr m.s-Ban 3m.s-vrec-write 3f.s-be  
 ka-paralok jōŋ-ŋa  
 f.s-friend gen-1s  
 ‘The girl Ban is married to is my friend.’

25. ka-khinna? [ba- u-ban u-ya-kindō? ] ka-dei  
 f.s-child adjr m.s-Ban 3m.s-vrec-meet 3f.s-be  
 ka-paralok jōŋ-ŋa  
 f.s-friend gen-1s  
 ‘The girl Ban met with is my friend’

In case the predicate is inherently reciprocal, the specific occurrence of the adverb *-laŋ* ‘together’ with the embedded predicate imparts the interpretation that the subject of the embedded clause together with someone else is performing the action *with/for/against* the subject of the matrix clause as in sentences (26)-(28).

26. ka-khinna? [ba- u-ban u-ya-šō?-laŋ] ka-dei  
 f.s-child adjr- m.s-Ban 3m.s-gm-beat-together 3f.s-be  
 ka-paralok jōŋ-ŋa  
 f.s-friend gen-1s  
 ‘The girl who(m) Ban along with someone else is beating (someone) is my friend.’

<sup>124</sup>The verbal reciprocal *-ya* is grammaticalized and hence, *ya-kindō?* is lexicalized as one word.

<sup>125</sup>Note that the marker *-ya* in these sentences functions as a verbal reciprocal marker, and not as a group marker.

27. ka-khinna? [ba u-ban u-ya-kren-laŋ ] ka-dei  
 f.s-child adjr m.s-Ban 3m.s-gm-speak-together 3f.s-be

ka-paralok jəŋ-ŋa  
 f.s-friend gen-1s

‘The girl who(m) Ban together with someone else is talking to is my friend.’

28. ka-khinna? [ba u-ban u-ya-thəʔ-laŋ ]  
 f.s-child adjr m.s-Ban 3m.s-gm-write-together

ka-dei ka-paralok jəŋ-ŋa  
 3f.s-be f.s-friend gen-1s

‘The girl Ban together with someone else is marrying is my friend.’

[That is, Ban and some other person are getting married to the same girl on the same occasion.<sup>126</sup>]

(iii) The third piece of evidence comes from a set of languages that permit the comitative PP as head (Let us call it Set 1) as against the other set (Let us call it Set 2) that do not.

As we mentioned above, the comitative PP is not linked either syntactically or thematically to the embedded predicate. Hence, unless the embedded predicate carries a special marker that establishes a relationship with the head of the comitative PP, a comitative PP cannot head an EHRC.

Set 1: Dravidian and Indo-Aryan languages belong to Set 1. Dravidian languages do not permit the embedded predicate to carry a specific marker such as ‘together’ and hence, none of the Dravidian languages permit an EHRC with a comitative PP as the head with the interpretation of *accompaniment*, though an oblique PP such as locative, ablative and instrumental can form an EHRC in Dravidian (Subbarao ms. & Subbarao 2010 b). The following examples with the comitative PP as head in Dravidian are illustrative.

EHRC with the comitative PP as head – not permitted

Malayalam (Dravidian)

29. \*bābu vanna ezuttə  
 Babu come.adjr letter  
 ‘The letter with which Babu came.’ (Asher and Kumari 1997: 61)

Tamil (Dravidian)

30. \*kumār veļiye pōna umā  
 Kumar out go.pst.adjr Uma  
 ‘Uma, with whom Kumar went out.’ (Annamalai 1997: 78)

Telugu (Dravidian)

31. \*nēnu veļl-in-a saraļa  
 I go-pst-adjr Sarala  
 ‘Sarala with whom I went.’ (Ramarao 2003: 79)

---

<sup>126</sup> In the existing social set up, it is not possible for two persons to get married to the same girl but that is what the interpretation of the sentence is.

## Kannada (Dravidian)

32. \*huḍuga (ø-jote) banda suṣīla  
 boy (ø-with) came.pst.adj Susheela  
 ‘Susheela, who the boy came with’ (Sreedhar 1990:58)

Most of the Indo-Aryan languages do not permit the modification of an oblique PP as head. Though a few of them like Marathi, Konkani, Nepali and Oriya (IA) do permit the modification of oblique objects, they do not permit the modification of the comitative PP as head (Subbarao, in press). Cambridge: Cambridge University Press.)

## Marathi (IA)

33. \*karuṇa ge- le-l-i mulgi  
 Karuna go- Perf Pple.f girl  
 ‘The girl with whom Karuna went.’ (Lalita Dhareshwar p.c.)

In most of the Tibeto-Burman languages too a comitative PP cannot head an EHRC (Subbarao, in press). In Manipuri (Tibeto-Burman) the comitative PP can head an EHRC provided:

- (i) the verbal reciprocal *-na* occurs with the embedded verb, and
- (ii) the adverb *min* ‘together’ occurs with the embedded verb.

## EHRC with the comitative PP as head

## Manipuri (Tibeto-Burman)

34. tomba-nə lak-min-na-bə nupi-du pha-i  
 Tomba-nom come-together-vrec-nozr girl-def good-[-fut]  
 ‘The girl with whom Tomba came is good.’ (Subbarao, in press)
35. tomba-nə pha-min-na-bə nupimaca-du magi macanupi- ni  
 Tomba-nom sit-together-vrec-nozr girl-def his- daughter- is  
 ‘The girl with whom Tomba sat is his daughter.’ (N. Pramodini, p.c.)

In Manipuri the reciprocal marker *-na* together with the incorporated adverb *min* ‘together’ imparts the interpretation of *doing an act together*. The other verbs that require both *min* and *na* are verbs such as: *stand, swim, crawl, walk, run, cross the river, jump* etc. just as in Khasi.

In Japanese and Korean a comitative PP may head an EHRC if the adverb with the interpretation of ‘together’ is incorporated, just as in Khasi and Manipuri.

## Japanese

36. [matt- ga \*(issyoni) kita] onna- no- ko  
 Matt- nom together came woman- of- child  
 ‘The girl Matt came with’ (Masayoshi Shibatani, p.c.)

In Korean too the occurrence of the adverb *hamkkey* ‘together’ with the embedded predicate is obligatory in an EHRC with a comitative PP as head.

Korean

37. yeongsig- i \*(hamkkey)- on yeca  
 Yeongsig- *nom together- come girl*  
 ‘The girl Yeongsig came with’ (Masayoshi Shibatani, p.c.)

Other intransitive verbs such as *swim, walk, run, sit, crawl* etc. too permit the comitative PP to head an EHRC if the adverb ‘together’ occurs with the embedded predicate in Japanese and Korean (Matt Shibatani, p.c.). The occurrence of the adverb ‘together’ in Japanese and Korean clearly demonstrates that a comitative PP can head an EHRC if and only if the predicate can get ‘*thematically related*’ to the embedded predicate.

In German (Tatiana Oranskia, p.c.) and Russian (Ludmila Khoklova, Boris Zakharyn & Tatiana Oranskia, p.c.) a comitative PP cannot head an EHRC as these languages do not permit adverb incorporation just as Dravidian languages cannot.

Thus, both Khasi (Mon-Khmer) and Manipuri (Tibeto-Burman) employ an identical process in the formation of the Gap Relative with the comitative PP as head, namely, the addition of the verbal reciprocal/group marker and the adverb denoting ‘together’ to the embedded verb to enable a comitative PP to head an EHRC.

Consequently we may reconsider the case of the comitative PP in Khasi. The predicates *wan* ‘come’ in (10), *marɛ?* ‘run’ in (11), *ya.<sup>y</sup>dkāi* ‘walk’ in (12), and *jji* ‘swim’ in (13) are not thematically linked to the head of the comitative PP. That is the reason why a sentence such as (9) repeated here as (38) is ungrammatical in Khasi.

38. \**ka-khinna?*<sub>i</sub> [ba u-ban u-wan ø<sub>i</sub> ša-šnɔŋ<sub>pp</sub>]  
 f.s-child adjr m.s-Ban 3m.s-come all-village  
 ‘The girl with whom Ban came to the village ...’

When the group marker *-ya-* and the adverb *-laŋ-* ‘together’ occur with the embedded predicate, a predicate such as *wan* ‘come’ gets thematically linked to the head of the comitative PP. Hence, a comitative PP qualifies itself to become the head of an EHRC. If a predicate is inherently reciprocal, only the group marker which also functions like a verbal reciprocal is required and the occurrence of the adverb *-laŋ-* ‘together’ alone is sufficient to establish a thematic link.

Therefore, it is the thematic link that permits either an instrumental PP or a comitative PP to head an EHRC in Khasi. Our generalization also has predictive power; a predicate such as *jji* ‘swim’ or *pār* ‘crawl’ which does not get thematically linked to the head of the comitative PP obligatorily requires *-laŋ* ‘together’. Our predication is rightly borne out in (39) and (40) in Khasi.

39. *u-khinna?* [ba u-ban u-*ya-jji-laŋ*] u-dei u-paralok jɔŋ-ŋa  
 m.s-child adjr m.s-Ban 3m.s-**gm-swim-together** 3m.s-be m.s-friend gen-1s  
 ‘The boy Ban is swimming with is my friend.’
40. *u-khinna?* [ba u-ban u-*ya-pār-laŋ*] u-dei u-paralok jɔŋ-ŋa  
 m.s-child adjr m.s-Ban 3m.s-**gm-crawl-together** 3m.s-be m.s-friend gen-1s  
 ‘The boy Ban is crawling with is my friend.’

Our analysis has general predictive power too. A comitative PP cannot head an EHRC or an IHRC in any language unless the embedded verb carries a marker that indicates accompaniment in an EHRC or the internal head in an IHRC carries an overt adposition. Further, if it is a predicate which is [+inherently

reciprocal] in a language, the adverb ‘together’ may not occur. Manipuri<sup>127</sup> and Khasi provide evidence in support of this claim.

An inherently reciprocal predicate such as *khat* ‘fight’ or *luhong* ‘marry’ does not require the adverb *min* ‘together’ to occur with the embedded predicate. Only the verbal reciprocal is required.

Manipuri (Tibeto-Burman)

41. tomba- na khat- na- ba nupimaca- du eigi imanabi- ni  
 Tomba- nom fight- vrec- nozr girl- def my female friend- is  
 ‘The girl who Tomba fought with is my friend. (N. Pramodini, p.c.)’

The fourth piece of evidence comes from clefts in Khasi. In clefts the clefted noun phrase is fronted and a new CP node is created to accommodate the clefted noun phrase and the verb *dɛi* ‘be’ in Khasi. The clefted noun phrase carries the preposition that it carries in a simple clause.

An example of a simple clause with a locative PP is given in (42) and the corresponding clefted sentence is given in (43).

Khasi (Mon-Khmer)

42. u-wan u-bōʔ ya-ki-kət ha-lər-ka-mēc  
 m.s-Wan 3m.s-put acc-pl-book loc-surface-f.s-table  
 ‘Wan puts the books on the table.’
43. dɛi ha-lər-ka- mēc ba u-wan u-bōʔ ya-ki-kət  
 be loc-surface-f.s-table comp m.s-Wan 3m.s-put acc-pl-book  
 ‘It is on the table that Wan puts the books.’

When a comitative PP is clefted, as expected, the occurrence of the verbal reciprocal/group marker *-ya-* and of the adverb *-laŋ-* ‘together’ is obligatory as in (44).

44. dɛi bad-u-paralək jəŋ-ŋa ba u-ban u-ya-ŋji-laŋ  
 be com-m.s-friend gen-1s comp m.s-Ban 3m.s-gm-swim-together  
 ‘It is with my friend that Ban is swimming.’

However, if the predicate is inherently marked reciprocal, only the verbal reciprocal *-ya-* occurs and the adverb *-laŋ* ‘together’ does not as in (45).

45. dɛi bad-ka-paralək jəŋ-ŋa ba u-ban u-ya-māi  
 be com-f.s-friend gen-1s comp m.s-Ban 3m.s-vrec-fight (verbally)  
 ‘It is with my friend that Ban is fighting (verbally).’

Thus, the occurrence of the markers *-ya-* and *-laŋ* in clefts parallels the one in EHRCs with the comitative PP as head. This further substantiates our claim that the comitative PP as head requires the verbal reciprocal *-ya-* and the adverb *-laŋ* with predicates that are [-reciprocal] and only the verbal reciprocal *-ya-* with predicates that are [+reciprocal].

To account for the facts, we propose the Thematic Eligibility Condition (TEC) which states that:

Each predicate in the embedded clause of the EHRC has to meet ‘Thematic Eligibility Condition’ (TEC) for it to be able to accept a comitative PP as head. ‘Such eligibility is the result of language-specific syntactic processes such as Postposition Incorporation, the incorporation of the verbal reciprocal as a group

<sup>127</sup> Thanks to N. Pramodoni for a helpful discussion on this issue in Manipuri.

marker and the adverb *together* or the incorporation of the adverb *together* alone in the embedded predicate or the overt presence of the comitative postposition with the head in an internally-headed relative clause.” (Subbarao, in press)

## 6. Conclusion

In this paper, we have shown that the comitative PP as head in relative clauses and EHRCs in Khasi requires the verbal reciprocal *-ya-* and the adverb *-lay-* ‘together’ with the embedded predicate, when the comitative PP is *not* thematically related to the subject of the embedded predicate. If it is an inherently reciprocal predicate, then the occurrence of *-lay-* ‘together’ is prohibited, as it imparts the interpretation that the subject of the matrix clause and the head of the comitative PP are together performing the action for/against someone else.

Thus, for a predicate to be able to license a comitative PP as head of an EHRC/IHRC, the embedded predicate must be thematically eligible/ qualified. Such eligibility is achieved by any one or two of the following processes that effect the embedded predicate:

- (i) The presence of an adposition (due to incorporation in the verb or occurrence with the head in an IHRC), as in Tibeto-Burman languages,
- (ii) The addition of an adverb such as ‘together’ to the predicate as in Japanese and Korean and Manipuri (Tibeto-Burman).

In Appendix 1 and 2 below we provide data from Khasi which illustrates the various positions that can be relativized in Khasi.

## Appendix I

### Relative Clauses in Khasi

#### Subject Modification

- (1) u-brēw      [u    ba    la-āy    ya-ka-pisa    ha-u-kh̄inna?]    u-dei    u-paralok    jɔŋ-ŋa  
 m-human    3ms    adjr    pst-give    acc-f-money    dat-m-child    3ms-be    m-friend    gen-1s  
 ‘The man who gave the money to the boy is my friend.’

#### Direct and Indirect objects Modification

##### Direct object

- (2) ka-pisa                    [(ya-)ka    ba    u-brēw    u-la-āy  
 f-money                    acc-3fs    adjr    m-human    3ms-pst-give  
 ha-u-kh̄inna?]    ka-dei    ka    ba                    la?-jɔt  
 dat-m-child    3ms-be    3fs    adjr                    perf-torn  
 ‘The money which the man gave to the boy is torn.’

##### Indirect object

- (3) u-kh̄inna?    [(ha-)u    ba    u-brēw    u-la-āy    ya-ka-pisa]  
 m-child    dat-3ms    adjr    m-human    3ms-pst-give    acc-f-money  
 u-lɔŋ        u            ba    jrɔŋ  
 3ms-be    3ms        adjm    tall  
 ‘The boy who the man gave the money to is tall.’

**Oblique object Modification**

## Locative

- (4) *ka-mē'd [ha -ka ba ηa-bo? ya-ka-kət] ka-la?-kd<sup>y</sup>a?*  
 f-table loc-3fs adjr 1s-put acc-f-book 3fs-perf-break  
 'The table on which I put the book is broken.'

## Ablative

- (5) *ka-jaka [na-ka ba u-(la)-wan] ka-ləŋ ka-ba jŋāi*  
 f-place abl-3fs adjr 3ms-pst-come 3fs-be 3fs-adjm far  
 'The place which he came from is far.'

## Instrumental

- (6) *ka-tari [da-ka ba u-(la)-ət ya-u-sə?] ka-ləŋ ka-ba ləŋ*  
 f-knife instl-3fs adjr 3ms-pst-cut acc-m-fruit 3fs-be 3fs-adjr blunt  
 'The knife with which he cut the fruit is blunt.'

## Comitative

- (7) *ka-khīnna? [bad-ka ba ηa-ya-wan-laŋ]*  
 f-child com-3fs adjr 1s-vr-come-together  
*ka -dei ka-para (jəŋ)-u-meban*  
 3fs-be f-younger sibling gen-m-Meban  
 'The girl who came with me is Meban's sister.'

## Object of the Genitive Modified

- (8) *u-brēw [u ba ya-ka-yēŋ jəŋ-u la-pinjət da-ka-ēr yəŋ] u-daŋ-yām*  
 m-human 3ms adjr acc-f-house gen-3ms pst-destroy instl-f-storm 3ms-prog-cry  
 'The man whose house was destroyed by the storm is crying.' Lit: 'The man he that his house was destroyed by the storm is crying.'

- (9) *u-brēw [u ba ka-yēŋ jəŋ-u ka-(la)-sha?-pinjət ha-ka-ēr yəŋ] u-daŋ-yām*  
 m-human 3ms adjr f-house gen-3ms 3fs-pst-SA-destroy loc-f-storm 3ms-prog-cry  
 'The man whose house got destroyed in the storm is crying.' Lit: 'The man he that his house got destroyed in the storm is crying.'

## Object of comparison

- (10) *u-brēw [u ba u-lan u-kham-jrəŋ (ban ya-u)] u-dei u-paralək jəŋ-ηa*  
 m-human 3ms adjr m-Lan 3ms-more-tall than acc-3ms 3ms-be m-friend gen-1s

‘The man who Lan is taller than is my friend.’

- (11) *u-brēw*      [***ya-u-ba***      *u-lan*      *u-kham-jrɔŋ*]      *u-dɛi*      *u-paralɔk*      *jɔŋ-ŋa*  
 m-human      acc-3ms-adjr      m-Lan      3ms-more-tall      3ms-be      m-friend      gen-1s  
 ‘The man who Lan is taller than is my friend.’

## Appendix II

### Participles in Khasi

#### Subject Modification

- (1) *u-brēw*      [***ba***      *āy*      *ya-ka-pisa*      *ha-u-khɪnna?*]      *u-dɛi*      *u-paralɔk*      *jɔŋ-ŋa*  
 m-human      adjr      give      acc-f-money      dat-m-child      3ms-be      m-friend      gen-1s  
 ‘The man who gave the money to the boy is my friend.’

#### Direct and Indirect objects Modification

##### Direct object

- (2) *ka-pisa*      [***ba***      *u-brēw*      *u-āy*      *ha-u-khɪnna?*]  
 f-money      adjr      m-human      3ms-give      dat-m-child  
  
*ka-dɛi*      *ka*      *ba*      *la?-jɔt*  
 3ms-be      3fs      adjr      perf-be torn  
 ‘The money which the man gave to the boy is torn.’

##### Indirect object

- (3) *u-khɪnna?*      [***ba***      *u-brēw*      *u-āy*      *ya-ka-pisa* ]  
 m-child      adjr      m-human      3ms-give      acc-f-money  
  
*u-dɛi*      *u*      *ba*      *jrɔŋ*  
 3ms-be      3ms      adjm      tall  
 ‘The boy who the man gave the money to is tall.’

#### Oblique object Modification

##### Locative

- (4) *ka-mē<sup>y</sup>d*      [***ba***      *ŋa-bo?*      *ya-ka-kɔt*]      *ka-la?-kd<sup>y</sup>a?*  
 f-table      adjr      1s-put      acc-f-book      3fs-perf-break  
 ‘The table on which I put the book is broken.’

##### Ablative

- (5) *ka-jaka*      [***ba***      *u-wan*]      *ka-lɔŋ*      *ka-ba*      *jŋāi*  
 f-place      adjr      3ms-come      3fs-be      3fs-adjm      far  
 ‘The place which he came from is far.’



## Instrumental

- (6) *ka-tari* [*ba u-ɔt ya-u-sO?*] *ka-lɔŋ ka-ba looñ*  
 f-knife adjr 3ms-cut acc-m-fruit 3fs-be 3fs-adjm blunt  
 ‘The knife with which he cut the fruit is blunt.’

## Comitative

- (7a) *ka-khinna?* [*ba ŋa-ya-wan-laŋ*] *ka -dei*  
 f-child adjr 1s-vr-come-together 3fs-be

*ka-para* (*jɔŋ*)-*u-meban*  
 f-younger sibling gen-m-Meban  
 ‘The girl who came with me is Meban’s sister.’

- (7b) *ka-khinna?* [*ba ya-wan-laŋ bad-ŋa*]  
 f-child adjr vr-come-together com-1s

*ka -dei ka-para* (*jɔŋ*)-*u-meban*  
 3fs-be f-younger sibling gen-m-Meban  
 ‘The girl who came with me is Meban’s sister.’

## Object of the Genitive Modified

- (8) \**u-brēw* [*ba ya-ka-yēŋ jɔŋ-u*]  
 m-human adjr acc-f-house gen-3ms

*la-pɪŋjɔt da-ka-ēr yɔŋ*] *u-daŋ-yām*  
 pst-destroy instl-f-storm 3ms-prog-cry

‘The man whose house was destroyed by the storm is crying.’ Lit: ‘The man he that his house was destroyed by the storm is crying.’

- (9) \**u-brēw* [*ba ka-yēŋ jɔŋ-u ka-(la)-sha?-pɪŋjɔt*]  
 m-human adjr f-house gen-3ms 3fs-pst-SA-destroy

*ha-ka-ēr yɔŋ*] *u-daŋ-yām*  
 loc-f-storm 3ms-prog-cry

‘The man whose house got destroyed in the storm is crying.’ Lit: ‘The man he that his house got destroyed in the storm is crying.’

## Object of comparison

- (10) \**u-brēw* [*ba u-lan u-kham-jrɔŋ (ban ya-u)*]  
 m-human adjr m-Lan 3ms-more-tall than acc-3ms

*u-dei u-paralɔk jɔŋ-ŋa*  
 3ms-be m-friend gen-1s

‘The man who Lan is taller than is my friend.’

- (11) \**u-brēw* [*ba u-lan u-kham-jrɔŋ*] *u-dei u-paralɔk jɔŋ-ŋa*  
 m-human adjr m-Lan 3ms-more-tall 3ms-be m-friend gen-1s

‘The man who Lan is taller than is my friend.’

## Abbreviations

1	first person;
3	third person;
acc	accusative;
adjr	adjectivalizer;
all	allative;
com	comitative;
def	definite
DO	direct object;
f	feminine;
fut	future tense;
gen	genitive;
gm	group marker;
inf	infinitival marker;
IO	indirect object;
loc	locative;
m	masculine;
nom	nominative;
nozr	nominalizer;
OO	oblique object;
perf	perfective aspect;
pst	past tense;
s	singular;
SA	self-affective
VRec	verbal reciprocal;

## References

- Annamalai, E., 1997. *Adjectival Clauses in Tamil*. ILCAA, Tokyo University of Foreign Studies, Tokyo, Japan.
- Asher, R. E. & T. C. Kumari, 1997. *'Malayalam'* London and New York: Routledge.
- Bareh, C., 2007. '(A) Descriptive analysis of the Jowai and Rymbai dialects of Khasi', Ph.D. thesis, North Eastern Hill University, Shillong, India.
- Chomsky, N., 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Lasnik, H. and J. Uriagereka with C. Boeckx. 2005. *Minimalist syntax*. Malden: Blackwell.
- Ramarao, C., 2003 *Telugu velugulu* ('Telugu Brightness' in Telugu). Hyderabad: Navodaya Book House.
- Sridhar, S. N., 1990 *Kannada*. London: Routledge.
- Subbarao, K.V. *South Asian Languages: A Syntactic Typology*. ms. (in press).
- Subbarao, K.V. 2010 a. Comitative PP as Head in Internally-Headed Relatives Clauses in South Asian Languages. Keynote address presented at International Conference on South Asian Languages (ICOSAL) at Punjabi University, Patiala, India.
- Subbarao, K.V. 2010 b. Dravidian Languages: A Syntactic Typology. Keynote address delivered at the seminar on *Language and Genetic Architecture of Dravidian* at Dravidian University, Kuppam, India.
- Subbarao, K.V. 2010 c. Tibeto-Burman: A Syntactic Typology. Keynote address delivered at the seminar on *The Languages of Nagaland* at Nagaland University, Kohima, India.
- Temsen, G. M., 2006. *Embedded Clauses in Khasi*. PhD thesis, Delhi University, Delhi.
- Temsen, G.M. 2007. *What is the Order?* Paper presented at the National Seminar on *Emerging Linguistic Scene in North East India*. North Eastern Hill University, Shillong, India.

This is the first of two volumes of papers from the fourth International Conference on Austroasiatic Linguistics (ICAAL4), which was held at the Research Institute for Language and Culture of Asia, Salaya campus of Mahidol University (Thailand) 29-30 October 2009. Participants were invited to present talks related to the meeting theme of 'An Austroasiatic Family Reunion', and some 70 papers were read over the two days of the meeting. Participants came from a wide range of Asian countries including Thailand, Malaysia, Vietnam, Laos, Myanmar, India, Bangladesh, Nepal, Singapore and China, as well as western nations.

Published by:  
SIL International, Dallas, USA  
Mahidol University at Salaya, Thailand  
Pacific Linguistics, Canberra, Australia

ISBN 9780858836419

