AUSTROASIATIC STUDIES
Papers from ICAAL 4

Mon-Khmer Studies Journal
Special Issue No. 2

Volume editors:
Sophana Srichampa
Paul Sidwell
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:
Title: Austroasiatic studies [electronic resource]: papers from ICAAI 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 Putthisatien,
    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 deː in Kmhmu’ ........................................................................... 112
    R. Anne OSBORNE
About ʔaɔj 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. TEMSE
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 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 Premririt (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 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 Premririt, 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) Doug 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 Suttilak 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*

\[ ba \phi \text{dokam ia kine kiei kiei baroh}. \]

*C* 2PL need OBJ 3PL=Q 3PL=Q all

‘your Father in heaven knows [(that) you need all these things]’ (6:32)

(2) *ngan leh ia kaba bha aia, [ba]*

\[ ngan \text{ioh ia ka jingim bymjukut}? \]

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,]*

\[ lymne ia ka bor U Blei]. \]

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).

---


Copyright vested in the author.
The Syntax of Khasi Adverbial Clauses

(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).

(i)

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’
The Syntax of Khasi Adverbial Clauses

(13) \[ u \; \text{hap} \; \text{jyndei} \; ha \; ka \; \text{ding}, \; \text{jyndei} \; ha \; ka \; \text{um} \; \text{ruh} \]
\[ \text{he falls much into 3SF fire much into 3SF water also} \]
\[ \text{‘he falls often into the fire and often into the water’} \]

(14) \[ U \; \text{Jonah} \; u \; \text{la} \; \text{shong} \; \text{la} \; \text{sngi} \; \text{bad} \; \text{la} \; \text{miet} \]
\[ 3\text{SM Jonah 3SM PAST sit three day and three night} \]
\[ \text{ha ka kpoh u dohkha khraw,} \]
\[ \text{in 3SF belly 3SM fish big} \]
\[ \text{‘Jonah spent three days and nights in the belly of the whale’} \]

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.
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) \( kum \ u \ briew \ uba \ u \ trai \ iling, \)

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) \( PP \)

\( na \)

\( NP \)

\( u \ briew \)
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 niut?*
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)
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 (xviii), 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.
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.

(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)

(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)  \textit{uba mynta u long},
\begin{align*}
3\text{sm}=c  & \text{ now } 3\text{sm be} \\
\text{‘which now exists’ (6:30)}
\end{align*}

(35)  \textit{mynno ba ngi la iohi ba me la thngan, ...?}
\begin{align*}
\text{when } C  & \text{ 1PL PAST see C  2SM PAST be-hungry} \\
\text{‘when did we see that you were hungry?’ (25:37)}
\end{align*}

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 \textit{lada ‘if’}.

(36)  \textit{lada un sngew ia me,}
\begin{align*}
\text{if } 3\text{SM=FUT listen OBJ youSM} \\
\text{‘if he listens to you’ (18:15)}
\end{align*}

\textbf{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)
The Syntax of Khasi Adverbial Clauses

References


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). 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-.

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
Deponent Verbs and Middle-Voice Nouns in Temiar

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.

### Table 1: Monosyllabic: gəl ‘to sit’

<table>
<thead>
<tr>
<th>VOICE</th>
<th>ASPECT</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perfective</td>
<td>Imperfective</td>
</tr>
<tr>
<td>Base</td>
<td>gəl [gəl]</td>
<td>glgəl [gəlglə]</td>
</tr>
<tr>
<td></td>
<td>‘sit’ (completed act)</td>
<td>‘sit’ (incomplete act)</td>
</tr>
<tr>
<td>Middle</td>
<td>gagəl [gəgəl]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘fall asleep’ (uncontrolledly, or all together at once)</td>
<td></td>
</tr>
<tr>
<td>Causative</td>
<td>trgəl [tərgəl]</td>
<td>trglgəl [tərgəlglə]</td>
</tr>
<tr>
<td></td>
<td>‘make (someone) sit’ (completed act)</td>
<td>‘make (someone) sit’ (incomplete act)</td>
</tr>
</tbody>
</table>

### Table 2: Sesquisyllabic: saləg ‘to sleep’ (also: ‘lie down’, ‘marry’)

<table>
<thead>
<tr>
<th>VOICE</th>
<th>ASPECT</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perfective</td>
<td>Imperfective</td>
</tr>
<tr>
<td>Base</td>
<td>sələg [sələg]</td>
<td>slgəs [sgələ]</td>
</tr>
<tr>
<td></td>
<td>‘sleep’ (completed act)</td>
<td>‘sleep’ (incomplete act)</td>
</tr>
<tr>
<td>Middle</td>
<td>sələg [sələg]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘fall asleep’ (uncontrolledly)</td>
<td></td>
</tr>
<tr>
<td>Causative</td>
<td>sərələg [sərələg]</td>
<td>sərləg [sərləg]</td>
</tr>
<tr>
<td></td>
<td>‘put (someone) to sleep’ (completed act)</td>
<td>‘put (someone) to sleep’ (incomplete act)</td>
</tr>
</tbody>
</table>

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.)

---

3 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. Their sole inflection is nominalization with -(ə)-; but they also form a progressive-cum-imperfective with the proclitic bar-~ bə-. 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). For reasons explained later, I shall refer to most of the verbs with -a- in the first syllable as ‘deponent’ verbs.

<table>
<thead>
<tr>
<th>Perfective</th>
<th>ASPECT</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>halab [halab]</td>
<td>bar-halab [bar-halab ~ bə-halab]</td>
<td>lhalab [lhalab]</td>
</tr>
<tr>
<td>‘travel downriver’</td>
<td>‘currently travelling downriver’</td>
<td>‘a journey downriver’</td>
</tr>
</tbody>
</table>

Table 3: Deponent: halab ‘to travel downriver’

<table>
<thead>
<tr>
<th>Perfective</th>
<th>ASPECT</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>golap [golap]</td>
<td>bar-golap [bar-golap ~ bə-golap]</td>
<td>gnolap [gnolap]</td>
</tr>
<tr>
<td>‘carry on shoulder’</td>
<td>‘currently carrying on shoulder’</td>
<td>‘a carrying on shoulder’</td>
</tr>
</tbody>
</table>

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 -(ə)- 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

---

4 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’.

5 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 we-bə-keykooy kɛ̃ˀ [DU-PROG-roast.PFV fish] ‘they were roasting the fish’ (instead of the expected we-bə-keykooy kɛ̃ˀ). Since these instances were rare, it is impossible to determine whether they were intentional or simply speakers’ mistakes.

6 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: sikεˀ ‘pandanus’ instead of səkεˀ (phonemically /skεˀ/). 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.

7 As noted in Section 3.2 and the Appendix below, some verbs in -(ə)- do not qualify as deponent, and should preferably be placed in the ‘disyllabic’ category.
voice’ best describes the character of these -a-inflected forms. ‘Simultative’, 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’ [hāā]-deixis), is expressed phonetically in Temiar by the low vowel a, the back consonants ˀ 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.

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. 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. I say ‘prototypically 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.

---

8 For a detailed description of the Temiar sound-system see Benjamin 1976b: 130–151. 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).

9 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.

10 An appropriate morphological pattern could nevertheless be constructed for an imperfective-middle inflection: gəl ⇒ gəglgəl ⇒ *[gεlgəl] ‘sit’; səlɔg ⇒ səgalɔg ⇒ *səgalɔg ‘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 Asian 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:11

- 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. ṣun-wawεˀ. 3PL-depart.MID.
   ‘They all departed together.’

   b. ṣun-weˀ. 3PL-depart.PFV.
   ‘They departed.’

   c. ṣun-bọ-salɔg. 3PL-PROG-sleep.MID.
   ‘They are all sleeping together.’
   [Said of a group of cats and kittens.]

   d. ṣun-bọ-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 (tataπ), 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) ṣun-bọ-tataπ słaay we-naˀ. ṣun-tap naar i'si.
   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.12 Clearly, actions

---

11 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.

12 Means (1998: 12) seems to be 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
Deponent Verbs and Middle-Voice Nouns in Temiar
carried out in or by social groups are very likely to be both determinate and simulactive, 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-~ba- ‘progressive’, as in (3a). 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 naˀ wɛU~baˀaˀog ˀəUlah.*
   After that 2DU-PROG-give.MID 3SG-lah.
   ‘After that, they were giving each other [some].’

b. Ləpas naˀ wɛU~baˀəgˀog ˀəUlah.
   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).

13 Burenhult (2005: 104) reports that Jahai, a Northern Aslian language, employs the C’a-~/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.

14 The phrase ˀəUlah 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 (mediopassively) under compulsion.

(5) Gəd ɁiUkəwɛ ̃ɛ̃s haUcɔ ̃s naˀ, gəd yəəl,
   CUT.PFV NOM-child ACC-bird that, cut.PFV finish,
   na-poˀəɔl ma-’oos, na-didih ’ɔɔk naˀ,
   3SG-cook.PFV to-fire, 3SG-boil.PFV water that,
   we-koh, we-koh, we-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 absolutive constructions exemplified in (6a), as contrasted with the agentive, non-middle, employment of the same verb in (6b):

(6) a. Kɔbɔɔ’ doh na-wawɔɔg.
    fruit this 3SG-open.MID.
    ‘This fruit split open.’

b. ᅅ- wɔɔg kɔbɔɔ’ 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 absolutive 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) ɁiUˀiid kəretɔɔh 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, dadoˀ ‘flee’ is an example; the non-middle forms doˀ and doˀ doˀ ‘run’ are also employed, but infrequently. Presumably, ‘fleeing’ is something that Temiars do under compulsion, either in terror or mock terror (as in children’s play), rather than as a fully ‘agentive’ act.15

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, absolute 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-tipuuˀ ma-lev, na-lalɔɔs, na-pɛdpoon ma-kəbəəˀ tɛɛˀ.
    3SG-deceive to:wife 3SG-return.MID, 3SG-eat.IPV to:fruit earlier.
    ‘He deceived his wife, and – believe it or not! – went back and ate the fruit.
[He had earlier (tɛɛˀ) 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-infecting, mostly disyllabic, verbs. The sole inflection they exhibit is nominalization with -n-.16

---

15 Cf. Blagden (1906: 620), entry G44, where various forms of this verb are given in different Temiar dialects: dădō [dɛdœ] and dădū’ (dăduk); dădo’ (dădōk), in the phrases ‘run off! ’ em da dok [‘em-dădu’ ‘PL.INCL<INT>~run.MID!’ or perhaps ham-dădō’ ‘2SG<INT>~run.MID!’], and ‘don’t run away’, ed da dok [probably, ‘a-dădu’ ‘VET-run.MID’].

16 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 ‘lend 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 *holab, no imperfective *hεlab, and no causative *horlab or *horblab. They do, however, take the progressive clitic bar ~ ho-, which generally serves with these verbs to express durative or imperfective meanings, as in ho-halab ‘travelling downriver’. They also form verbal-nouns by -n- infixation, as in hənabal ‘a journey downriver’. Other examples are: satah ‘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. 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 (satah), 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. 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.

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-medium positioned somewhere near their own dream-soul (rəwaay), 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 graduating slowly 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.

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

---

17 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.

18 The likely etymology of carəəh is instructive: Mon karəh ‘to fall’, Mon garəh ‘to push, shove, fall down in pieces’, carə  da ‘to drop water on the ground’, Khmer jreah (chrea) ‘tomb, 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 carə  h ‘to chop bamboo down at its base’ that could be related to carəəh.)

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

20 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 pantii, 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 -a-.

<table>
<thead>
<tr>
<th>VOICE</th>
<th>ASPECT</th>
<th>PERFECTIVE</th>
<th>IMPERFECTIVE</th>
<th>PROGRESSIVE</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>-</td>
<td>hyrezek [henrezek]</td>
<td>‘clearing up’</td>
<td>bar-hyrezek [bar-henrezek~ba-henrezek]</td>
<td>‘currently clearing up’</td>
</tr>
<tr>
<td>Middle</td>
<td>harzek [harzek]</td>
<td>‘cleared up’ (complete)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Causative</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>VOICE</th>
<th>ASPECT</th>
<th>PERFECTIVE</th>
<th>IMPERFECTIVE</th>
<th>PROGRESSIVE</th>
<th>VERBAL NOUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Middle</td>
<td>kaleek [kaleek]</td>
<td>‘to prop, be propped’</td>
<td></td>
<td>bar-kaleek [bar-kaleek~ba-kaleek]</td>
<td>‘currently propping’</td>
</tr>
<tr>
<td>Causative</td>
<td>kpleek [knkpleek]</td>
<td>‘set a prop in place’ (completed act)</td>
<td>krypleek [knyrpleek]</td>
<td>‘setting a prop in place’ (incomplete act)</td>
<td>-</td>
</tr>
</tbody>
</table>

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

Typical Temiar props (kanaleek, Table 6) are the struts set athwart a dug-out canoe, or the scaffolding (also kanaleek) 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-jahũˀˀəUnaˀ, na-kaleek
3SG-scaffold.MID yes-ACC-tree 3SG-that, 3SG-scaffold.MID
gəgevəg balas ma-jaltək ˀəh.
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-jahũˀ ‘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 *kərɛŋleek*.

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)
- *lakɔˀ* ‘to break off, come off’ (said of something falling off spontaneously: absolutive, non-agentive, intransitive), as opposed to *lerkɔˀ* ‘to break something off, remove’ (agentive, transitive).

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-lerkɔˀ basiiˀ num-gentɔk.
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–tεrgəl* ‘sit’ (Table 1) and *salɔg–səlɔg–sεrlɔg* ‘sleep’ (Table 2).
- True ‘deponents’, non-inflecting or partly-inflecting verbs with an underlying middle-voice semantic: *halab* ‘travel downriver’ (Table 3), *harek* ‘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: *lakɔˀ* ‘to break off (intrans), *lerkɔˀ* ‘to break off (trans)’, but no *ləkɔˀ* (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- ~ tert-] 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) – gol (active) – tərgəl (causative), where gol ‘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, gol, 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 gol, it is still a meaningful thing to do, given the framework of understanding that Temiar-speakers operate within.21

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)22 or one’s breathing (hemnum, the -i-infixed reduplicated form of hup), since these can be directly monitored by the individual without their needing to be controlled. The head-soul, roway, 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 roway 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 roway (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 roway 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

---

21 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.

22 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 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: reŋka’ na-ctak (door 3SG-close.MID) ‘the door closed’ as opposed to ha-ctak [ctak] reŋka’ (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 phonetic expression through iconicity.24

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.25 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.26 A small but significant category consists of words for human relations, which Temiar see as dialectical in character (cf. Benjamin 1994). The list of words for physical objects

---

23 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 the speaker.’

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

25 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.

26 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.
Deponent Verbs and Middle-Voice Nouns in Temiar

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 teɛ* ‘last night’) and a verb (as in *hɔj 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.27

- **gatũˀ** ‘snail’, cf. Shorto’s Proto Mon-Khmer reconstruction *gtooˀ*, 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.
- **kalɔɔˀ** ‘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.
- **kapiiˀ** ‘to court, flirt’, cf. Old Mon *(kaj)i ~ guñi* ‘to embellish’ (Shorto 1971:78): among Temiars this is always reciprocal, the original meaning probably being ‘to adorn each other’.28
- **kawaaˀ** ‘kinsperson’, cf. Spoken Mon *kwa* ‘companion(s)’ (Shorto 1962: 227): in Temiar, kinspersons are necessarily reciprocally so.
- **ɲanʉˀ** ‘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.
- **sakɔɔl** ‘white-haired, grey-haired’, cf. Old Mon *skɔ*, Modern Mon *həkɔˀ* ‘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 -(U)nU, 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 *coner* ‘knife’, derived from the base-form of the verb *cer* ‘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:

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

---

27 See also footnote 18, on the Mon-Khmer etymology of *carəəh* ‘to go downhill’.
• *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 *coner* ‘knife’, above, which is not *caner* because a knife acts more ‘transitively’ and consequently suffers no ‘feedback’.29 (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*.

• *lanęy* ‘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’.

• *pənɔɔh* ‘shamanic dancing space’, from *pɔɔh* ‘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 *pɔɔh*.

• *sonalɔŋ* ‘marriage, married state’, ‘undergoing marriage’, from *səlɔg* ‘to marry’. Contrast this with the verbal noun *sənlɔg* ‘marrying’.

References


29 The morphophonological changes involved in the derivation of *canuuŋ* from *cuuk* and of *lanęy* from *lek* are explained in Benjamin 1976b: 143.


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

’adees ‘to exercise (one’s body), keep moving around’: one does this to oneself
’aluur ‘to shoot/be shot backwards, of felled branch’: it undergoes its own action
’aro ‘to talk nonsense, to joke; to flirt, entice’: a socially reciprocal activity
’ayir (1) ‘to plait the interwoven bottom on a bag’: the strands hold themselves in place
babsh ‘to act undecidedly’: one undergoes one’s own indecision
bacuh ‘to imitate (an animal’s sound); to echo’: because the animal sounds through the imitator; because the sound comes back upon the sounder
bagad ‘to try something new for the first time’: both an experience and an activity
bakɔɔ ‘set a bakɔɔ trap’: such traps are self-springing
balec ‘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
babsh ‘to abuse, speak angrily to someone’: such behavior rebounds on the abuser
balae ‘to not believe’ (religion); ‘to refuse (to do something)’: such behavior rebounds on the refuser
batɔk ‘to make a loop (banatɔk) (for the straps of a basket)’: the loop holds itself in place (see the main text)
baləəc ‘to sit cross-legged’: the legs hold each other in place
bawɛk ‘to bend a springy object to shape’: the bender both pulls and is pulled by the object
cabɔl ‘to decorate one’s hair or earlobe with flowers or ornaments’: the inserted decoration holds itself in place
cacii ‘to dance (standing still while moving arms and hands)’: the dancer moves her body while
experience her centre of gravity unmoved
rabog ‘to paint (one’s hair or earlobe with flowers or ornaments)’: the inserted decoration holds itself in place
cadog ‘to sit cross-legged’: the legs hold each other in place
cadɔɔ ‘set a cadɔɔ trap’: such traps are self-springing
cakəəd ‘to be sticky, to stick’: sticky objects both ‘stick’ and ‘are stuck’
calɛ̃b ‘to have specks in one’s eye’: an undergone experience
calɔ ‘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
canay ‘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)
carɔd ‘to be interlocked’: the parts both lock and are locked
catũd ‘to be physically obstructed’ (by a crowd or object): one both obstructs and is obstructed
cayaaŋ ‘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: gerpid ‘to squeeze’, but no *gopid)
gatə to chip away (with a small knife): the action rebounds on the hand

hahasə to burst out laughing: overcome by one’s own uncontrolled laughter

halə to transport/be transported downriver: the raft is transported by the river’s flow (see the main text)

haləsə to laugh (in company): because such laughing is catching

harə to cease raining: the rain appears to stop itself

hayə to nearly do something, narrowly avoid doing something: experienced rather than carried out deliberately

jayə 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

jahə to set (sun): the sun appears to do so of its own accord

jə to stride: one gets caught up in the rhythm of the movement

kabaə to hug tightly (climbing a tree or riding piggy-back): climbers hold themselves in place

kabaə to send off tightly hugging side-shoots (plant, as on a strangling fig): the plant forms itself

kə to not know: ignorance is a state one finds oneself in, not through any actions of one’s own

kətə to boil over, flow over (water): the vessel undergoes the overflow; to give fire to (from flame or spark): the object undergoes the fire

kəə to crush (a louse etc) with one’s fingernail: one experiences the ‘crunch’ of the insect

kəeə to climb something, clasping with hands and feet: the climber both holds on and is held

kəə to eat greedily: the eater is driven by his own greed

kaəə to beat with hand or stick, without cutting: the action rebounds on the beater’s hand

kəə to burn (by beat etc): the object undergoes the fire

ləəə to assimilate oneself; ethnically mixed (person, population); Gəb ləəə person with a Temiar mother and a Malay father; Kuy ləəə type of Malay language used in some song-ritual lyrics: assimilation is something one both does and undergoes

ləəə to follow around listlessly (like a child): the child cannot help itself

məəə to suffer a kind of headache (believed to be caused by dew): an undergone experience; also a middle-voice noun

naəəah to feel sympathy or sorry for someone: one cannot help doing so

pacəə to hurry on ahead, hurry up: one gets caught up in the rhythm of the movement

pasəə to not like someone: regarded as rebounding onto the initiator

payaə to not want to court someone, reject someone: regarded as rebounding onto the initiator

ragəə to travel in a hurry: one gets caught up in the rhythm of the movement

səəə to catch (of a trap): the trap releases itself

səəə to hold food in a bulging gullet (snake, lizard): the gullet and food hold each other in place

saləə to lodge, stay with someone: regarded as a reciprocal activity

saləə to faint; to die: an undergone process

səə to wrap: because the package keeps itself wrapped

səəəə to give way (soft ground): the ground and the walker undergo the subsidence

səəə to collapse, break off, release (of fall trap): the object undergoes its own collapse (see the main text)

tabəə to hang something up for storage: the item hangs itself

taəə to hold one’s hand out: an action in expectation of receiving something; or an action in muscular dynamic equilibrium
stative verbs, adjectives

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

‘alooy ‘unripe, fresh’
babuh (1) ‘rotten (wood)’
bahul (1) ‘to be a big eater’: he is driven by his own greed
gaˀɔɔˀ~gaˀəə ‘to be shrunken with hunger (belly)’
gadal ~ kadal ‘hard pan (soil), firm (ground), caked (powder)’ (semi-deponent, ~ kɛldal, but no *kədal)
gahɛɛl ‘out of breath, weary’
galah ‘straight (hair)’
gatɔɔw ‘thin (body, cane),
gawɔɔk ‘skinny’
gayuh ‘rough or soft (stony or overgrown ground)’
harauw (1) ‘hairy, frizzled (hair, fur, spines, clothing)’
hayɔ̃ɔ̃ ‘light (weight)’
jaleeg ‘restless, promiscuous (staying in a different place each night)’
jamii ‘swollen-cheeked’
kabɔ ‘stupid, silly, ignorant, dumb’ (see the main text)
karək ‘to feel very cold; rigor mortis’
karuk ‘withered, dead (plant), weak, near death (person)’
katũũd ‘to swell (a boil, a tree-canker)’
lapood ‘soaking wet’
latah ‘bald in front’
latsb ‘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’
papo ‘crazy, mad, insane’
rahem ‘round-faced’
ratih ‘diligent (in work)’: carried away by one’s enthusiasm.
rawəəj ‘finished, used up’
sagɔɔ ‘thirsty’
sakɔɔl ‘white-haired, grey-haired’ (see the main text)
saraa ‘to depend on someone, to be up to someone to decide’: dependency is mutual
sayeep ‘dead, dried out (bamboo)’
takɛl ‘giggly, laughing a lot’: giggling is uncontrollable
talʉr ‘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.

ˀabir ‘a squirrel’
ˀabɔɔŋ ‘a land rat’
ˀacaam ‘a large squirrel’
ˀaceel ‘a large squirrel’
ˀaceel ‘a large squirrel’
ˀacoḥ ‘common grey-bellied squirrel’
ˀadeŋ ‘a small squirrel’
ˀagaɔj ~ ˀagaɔe ‘common Malay squirrel, plantain squirrel’
ˀahɔl ‘flat-headed cat’
ˀajɔh ‘rusa deer (Cervus timorensis)’
ˀajɔo, ˀajoor ‘common tree-shrew’
ˀakuub ‘forest squirrel’
ˀalaam ‘elephant’
ˀalaay ~ ˀalaaj ‘elephant’
ˀalɛɛŋ ‘a small squirrel’
ˀaləəp ‘elephant’
ˀalul ‘elephant’
ˀamaah ‘a small frog’
ˀamɔɔ ‘goat:antelope’
ˀamug ‘a white squirrel’
ˀapaam ‘a small red-tailed squirrel’
ˀapɔŋ ‘pig:tailed macaque (Macaca nemestrina)’
ˀarɔŋ ‘barking deer (Muntiacus muntjak)’
ˀataah ‘horse-tailed squirrel’
ˀataŋ ~ ˀataan ‘tiger’
ˀatuul ‘elephant’
ˀbadɔɔt ‘bearded pig’
ˀbagɛɛt ‘monitor lizard (Varanus rudicollis)’
ˀbakaan ‘small-toothed palm-civet’, ‘bear civet’
ˀbapaak ‘a white-bellied dark-backed squirrel’
ˀbarɛɛt ‘tapir’
ˀbasɛɛ ‘white-eyed gibbon’
ˀbateɛɛ ‘a gibbon’
ˀbawaj ‘pig-tailed macaque’
ˀcaɛɛ ‘a small house-mouse’
ˀcadɛɛ ‘a small squirrel’
ˀcapeɛɛ ‘solitary large male (of monkeys)’
ˀdalɔ ‘a lizard’
ˀdari ‘soft-shelled river turtle’
ˀhadaa ‘short-tailed mongoose’
ˀhagaab ‘two-horned rhinoceros’
ˀharɛŋ ‘monitor lizard (Varanus salvator)’
Deponent Verbs and Middle-Voice Nouns in Temiar

hayum ‘bamboo rat’
jajɔɔˀ ‘common palm civet; tiger-civet’
japɔŋ ‘Muller’s rat’
kapɔŋ ‘a frog sp.’
kabuc ‘monitor lizard’ (*Varanus salvator*)
kabuk ‘a green snake’
kajεεˀ ‘flat-backed land tortoise’
kahuul ‘tadpole’
karā́dy ‘a river-turtle, small box-tortoise’
kasiŋ ‘rusa deer (*Cervus timorensis*)’
kaweeb ‘Malayan sun bear’
kayii’ ‘flying lemur’
mamuuug ‘tiger’
pariˀ ‘a monitor lizard’
payā́ad ‘a monitor lizard’
sakɔl ~ sagɔɔl ‘cream-colored giant squirrel’
taŋŋ ‘wild pig’
taheɛ̃g ‘bullfrog’
taŋk ~ taŋk ‘dusky leaf-monkey’
taŋaŋ ‘a snake’
taŋɔɔr (1) ‘striped ground squirrel’
tajuw ‘snake’
talam ‘elephant’
tarok ‘a lizard’
tawɔɔh ‘white-handed gibbon’

Birds and bats

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

‘ahah ‘great slatey woodpecker’
‘akɛ̃b ‘stonechat’
‘akeɛg ‘red-headed trogon’
‘akuul ‘trogon (generic)’
‘aṭɛ̃b ‘grey-chinned minivet’
bajɔw ‘cotton teal’
baryɛt (a kind of bird?)
bayɔɔj ‘a small night-calling bird’
cabaŋ ‘black-crested yellow bulbul’
cabɛw ‘chestnut-collared kingfisher’
cacɛr ‘red-headed tailor bird’
dəŋk ‘lesser short-wing’
gagoob ‘large scimitar babbler’
kaka ‘kingfisher’
kakeh ‘hornbill (various kinds)’
kakuu ‘black hornbill’, ‘white-billed hornbill’
kasaa ‘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’
raboɔl ‘weaver finch’
sabaat ‘hawk owl’
sagur ‘a small bird’
sayah (1) ‘red-headed tree-babbler’
sayol ‘mountain nun-thrush’
sayol ‘false vampire bat’
ta’øj ‘hornbill sp.; ‘wrinkled hornbill’
tadoor ‘red jungle fowl’
tagok ‘gold-whiskered barbet’
tahâar ‘southern pied hornbill’
tahær ‘grey and buff woodpecker’
taləd ‘Scop’s owl’
tapar ‘white-eared fruit bat’
tatoh ‘maroon woodpecker’, ‘greater yellow-naped woodpecker’
tayet ‘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’
bayɔj ‘a bird’
capood ‘a very large black, or blue or striped fly; a large biting fly’
capɔɔg ‘polydesmoid millipede’
cawaas ~ cawãas ‘earwig’
galul ‘a large black mosquito’
garεεd ‘a sunset:chirping cricket’
garɔɔk ‘a night:chirping cricket’
garuuc ‘termite’
gasə ‘red stinging ant’
gatεε ‘mite’
gatũ ‘snail(shell)’ (see the main text)
jaleed ‘firefly’
jareb ‘a small noisy outdoor cricket’
jawiis ‘a seasonal cicada’
kabed ‘ant’
kaɛk ‘a black biting ant’
kasɔɔd ‘fire:ant’
lawã ‘a large red swarming tree:ant’
manay ‘scorpion’
padaaw ‘hornet’
saley ‘jewel beetle’
tabøl ‘black honey bee’
talan ‘a large beetle’
talãy ‘black millipede’
talok ‘bumble bee, carpenter bee’
tanɔy ‘dragonfly’
taro ‘house lizard’
tawãag ‘butterfly’
tawiik ‘spider’
wawah ‘a moth’
yayeed ‘an evening-sounding cicada’
Deponent Verbs and Middle-Voice Nouns in Temiar

Body parts

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

‘ayir (2) ‘itchy scalp, from louse feces’
bakoh ‘male genitals’ (sometimes ‘penis’, ‘testicles’)
balsk ‘animal’s beak or horn’
cawɔɔk ‘head, skull’
jaka’~ caka’ ‘lower jaw’ (also: ‘overhang’ of a roof)
kabooj ‘large vesicular swelling’
kadɔɔg ‘hollow of the knee-joint’
kalar ‘throat, glottis, esophagus, adam’s apple’
kalej ‘testes’
kaloɔɔr ‘snout of pig, bear, cat, etc.’
kapɔŋ ‘elbow’
kapɔɔ ‘cheek’
kakoeb ‘sternum’
karkɔ ‘knee’
katonj ‘knee-cap’
kaway ‘wing’
kaye ‘little finger’
kayood ‘fetus’: an Other within one’s Self
lagε̃ ‘caul’ ‘placenta’, ‘afterbirth’; also ‘new-born baby’: an Other within one’s Self (Benjamin 1994: 52)
panik ‘navel’
sabook ‘windpipe’
sakɔ ‘corpse’
sapal ‘upper arm’
taboo ‘thumb, big toe’
tayun ‘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’.

‘adɛg ‘a tasty wild tuber’
‘afεl ‘a cane’
‘apoos ‘wild ginger’
‘asaad ‘a large squash’
‘asɛh ‘millet (taboo name)’
‘awaat ‘bamboo’
‘awen ‘bamboo’
babuh (2) ‘a small toadstool’
badɔɔk ‘jelutong tree’
bajaaw ‘a fruiting vine’
bayas ‘a palm with edible pith’
bayoor ‘a large secondary-forest tree’
cadag ‘a long-leaved plant used for plaiting ritual crowns’
cakooj ‘bark (of tree)’
canɛh ‘pulasan fruit’
galook ‘large rotten limb of a tree, about to fall’
gareed ‘a tree (provides the barkcloth used by menstruating women)’
gayɛɛk ‘a wild fruit’
haˀog ‘a fruit-bearing forest tree’
hakɔɔr ‘a wild seasonal fruit’
hayul ‘generic term for staple food crops (non-leafy)’
haraw (2) ‘a forest flower’
kabup ‘calyx, bud’
kacuuh ‘a bitter leaf eaten with betel’
kakɔɔr ‘the bark eaten with betel chew’
karaaw ‘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
kawɔɔk ‘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’
rarɔ ‘a wild seasonal fruit’
sakɔ ‘the spiny sheath on bamboo’
sakool ‘a long leaved vegetable’
sayah (2) ‘the chaff after winnowing rice grains’
taba ‘springy branch’
tamud ‘a sweet-smelling herb’
tamɔ ‘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
kawaaw ‘kinsperson (primarily consanguineal)’ (see the main text)
lany ‘ghost’; also used of ‘terrorist’ during the Malayan Emergency
jun ‘chief’ (see the main text)
lage ‘newborn baby’ (also used to label various perinatal misfortunes)
pacɔ ‘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 pedpəəd ~ penpəəd, but no singular *papəəd)
sabat ‘a convulsive attack (associated with childhood and childbearing)’ (see Benjamin 1994: 54–55)
sape ‘pollution caused by recent death’
sarɔ ‘ghost of deceased’
sayɛɛd ‘young child’ (plural sudyɛɛd ~ senyɛɛd, but no singular *sayɛɛd)
tajar ‘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.

'abat ‘sarong’: it holds itself in place
'alεεg ‘plaited storage bag’: it forms its own shape as it fills
'apav ‘back-basket’: it is self-hanging
'apil ‘mat’: it lays itself down
'apɔk ‘tobacco pouch’: it takes its own shape
'bakɔɔ ‘large noose trap, set on tree branch’: it ‘springs’ itself
'balik ‘sky, position above’: it appears to be self-supporting
'bara ‘main beam (of house)’: because it holds itself in place
'bauw ‘fishing rod’: because it both pulls and is pulled
'cahuh ‘woven roof thatch’: it holds itself together
'cagɔɔl ‘pond’: it is self-forming
'gadaŋ ‘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
'galɛ̃l ‘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
'jala ‘thorn’: it both snags and gets snagged on people’s clothing (see the main text)
'jalb ‘tree-top’: a tree grows its own top (see the main text)
'lamun ‘bent-over sapling-spring in various kinds of trap’: it is held in dynamic equilibrium
'layeg ‘night’ (both a noun and a verb): night falls, happening spontaneously into existence of its own accord (see the main text)
'rabooŋ ‘a large backbasket’: it holds itself in place
'ran ‘back-basket’: it holds itself in place.
'sagub ‘cloud’: clouds appear to bring themselves into being (see the main text)
'salɔɔg ‘a kind of backbasket (or use when trapping)’: it holds itself in place
'sapε ‘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
'takooŋ ‘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

'abaag ‘split-bamboo internode used as eating dish’
'acag ‘to plan to do something’
'alɛh ‘to guess’
'amεs ‘small’
'apet ‘short’
'arap ‘a possession’
'adal ‘to call someone’
'habh ‘to guess’
'kamaay ~ kama ‘to store food (against future hunger)’
'kan ‘we excl.’
'la ‘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)

*lawag* ‘to mix and scoop food up with the fingers’

*lawud* ‘to cook a stew’

*marɛɛk* ‘fish-weir platform’

*palsɔ* ‘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)’

*tahɔɔr* (2) ‘hole (animal’s, in a nest or the ground)’

*takaah* ‘slope’

*tarɔg* ‘spear’

*tayerd* ‘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 ProtoPearic 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.

2. Ethnonymes in Pearic subgroup

‘Chong’, also ‘Song/Xong/Kasong’, originate from *kɔːŋ as the genuine autonym of Pearic populations. It is attested as Tchouang within “Tchouang thieves” (zhuàng zéi 撞贼) in the Tcheou Ta-kouan description of Cambodia, from the end of 13rd century (Pelliot 1902: 156; 1951: 70). In modern Khmer: ណ ជង cɔːŋ 1. “barbare”, 2. “nom de tribu à demi sauvage” (Gu esdon 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 壮 “Zhuang people”, also zhuàng/chuang 撞.

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

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

‘Pear’: from Skt. varṇa- through the literate treatment *bər > khmer bər ឈុី / ba(r)ŋ សុី pəar > 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 (Brengues 1905), they were composed of war prisoners, convicts and mountain peoples. Contrary to appearances, Pol is not an alternative of Por.

‘Somrê’, also ‘Somre/Somræ/Somray’ somrɛː/sɔmraj: formed of sreː “field, cultivate” infixed by -m-, meaning “cultivators”.

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.


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 *soːˀc. Saoch has been connected to the Khmer sœc sao c ‘scarlet fever’ (Diffloth, see in Isara 2007: 60). Guesdon (1930: 1841) gives “pimplly”. 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 thaiicized into ‘Kha Saut’ and ‘Ut’ Ut 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 Daum (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 -rl by Baradat. The final spelled *s is not consistent, it represents *-s as well as *-l. 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 ‘ʔəŋkar həŋka: “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.

<table>
<thead>
<tr>
<th>Pearic languages</th>
<th>Authors</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chong</td>
<td>Crawfurd</td>
<td>1828</td>
</tr>
<tr>
<td>Samreh d’Angkor</td>
<td>Bastian</td>
<td>1868</td>
</tr>
<tr>
<td>Khamen boran</td>
<td>Bastian</td>
<td>1868</td>
</tr>
<tr>
<td>Xong</td>
<td>Bastian</td>
<td>1868</td>
</tr>
<tr>
<td>Xong</td>
<td>Garnier</td>
<td>1873</td>
</tr>
<tr>
<td>Khamen boran</td>
<td>Garnier</td>
<td>1873</td>
</tr>
<tr>
<td>Kouys - Porth</td>
<td>Harmand</td>
<td>1878-79</td>
</tr>
<tr>
<td>Saoch</td>
<td>Pavie</td>
<td>1881-82</td>
</tr>
<tr>
<td>Sâmré (Siem Réap), Pors (Pursat)</td>
<td>Moura</td>
<td>1883</td>
</tr>
<tr>
<td>Sâauch</td>
<td>[Leclère]</td>
<td>1900</td>
</tr>
<tr>
<td>Porr (Pursat)</td>
<td>Brengues</td>
<td>1905</td>
</tr>
<tr>
<td>Saoch</td>
<td>Pannetier</td>
<td>?</td>
</tr>
</tbody>
</table>
### Table 1: Chronology of Pearic data

<table>
<thead>
<tr>
<th>agreeing</th>
<th>register</th>
<th>epithet</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chong lɔɔ (Chanthaburi)</td>
<td>Martin</td>
<td>1974</td>
<td></td>
</tr>
<tr>
<td>Chong haap (Chanthaburi)</td>
<td>Martin</td>
<td>1974</td>
<td></td>
</tr>
<tr>
<td>Somray (Kranhung, Batdambang)</td>
<td>Martin</td>
<td>1974</td>
<td></td>
</tr>
<tr>
<td>Somree (Peam Prus, Pursat)</td>
<td>Martin</td>
<td>1974</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Surekha</td>
<td>1982</td>
<td></td>
</tr>
<tr>
<td>Samree (Borai, Trat)</td>
<td>Theraphan</td>
<td>1984</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Huffman</td>
<td>1985</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Sirikarn</td>
<td>1987</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Saifon</td>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Theraphan</td>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Edmonson</td>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>Kasong (Klong Saeng, Trat)</td>
<td>Kunwadee</td>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Siri pen</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>Samre (Trat)</td>
<td>Pornsawan</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Isara</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Kasong (Klong Saeng, Trat)</td>
<td>Noppawan</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Kasong (Klong Saeng, Trat)</td>
<td>Sunee</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Chong (Klong Phlu, Chanthaburi)</td>
<td>Isara</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>Chung Yul (Kampong Som)</td>
<td>Isara</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Chung Yuy (Thung Na, Kanchanaburi)</td>
<td>Isara</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Chung (Kanchanaburi &amp; Kampong Som)</td>
<td>Isara</td>
<td>2009</td>
<td></td>
</tr>
</tbody>
</table>

#### 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.

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
</tr>
</thead>
<tbody>
<tr>
<td>[cvc]</td>
<td>[cv/c]</td>
<td>[cvc]</td>
<td>[cv/c]</td>
</tr>
<tr>
<td><strong>Clear modal</strong></td>
<td><strong>clear-creaky</strong></td>
<td><strong>breathy</strong></td>
<td><strong>breathy-creaky</strong></td>
</tr>
</tbody>
</table>

**Table 2: Pearic registers**

Examples (Siripen 2001):

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
</tr>
</thead>
</table>

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, generaly */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.

<table>
<thead>
<tr>
<th>Proto Pearic</th>
<th>Pearic dialects</th>
<th>Kg Thom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferlus</td>
<td>Headley 1985</td>
<td></td>
</tr>
<tr>
<td>*p</td>
<td>pʰ</td>
<td>p</td>
</tr>
<tr>
<td>pʰ</td>
<td>pʰ</td>
<td>p</td>
</tr>
<tr>
<td>*t</td>
<td>tʰ</td>
<td>t</td>
</tr>
<tr>
<td>tʰ</td>
<td>tʰ</td>
<td>t</td>
</tr>
<tr>
<td>*c</td>
<td>cʰ</td>
<td>c</td>
</tr>
<tr>
<td>cʰ</td>
<td>cʰ</td>
<td>c</td>
</tr>
<tr>
<td>*k</td>
<td>kʰ</td>
<td>k</td>
</tr>
<tr>
<td>kʰ</td>
<td>kʰ</td>
<td>k</td>
</tr>
</tbody>
</table>

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

In my notation, superscripts ¹ and ² are only used to differentiate the two types of correspondences.
The type */p t c k/> */pʰ tʰ cʰ kʰ/, 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 */ʔ>/ today -aya) and partially in Vietic.
Table 4: Distribution of final glottal stop

<table>
<thead>
<tr>
<th></th>
<th>Suôy (Baradat)</th>
<th>Chong (Siripen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground, soil</td>
<td>thé-é</td>
<td>tʰeʔ</td>
</tr>
<tr>
<td>Leaf</td>
<td>sla-a</td>
<td>laʔ</td>
</tr>
<tr>
<td>Stone, rock</td>
<td>thmaɪ-aʊ</td>
<td>kʰəmoʔ</td>
</tr>
<tr>
<td>Acid, sour</td>
<td>chə-ʊ</td>
<td>cʰʔ</td>
</tr>
<tr>
<td>Skin</td>
<td>trəl-o-aʊ</td>
<td>kʰəloʔ</td>
</tr>
</tbody>
</table>

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 CVʔCVʔ.

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 (crease) -ʔ- (CVʔ). As a consequence, final -ʔ must be removed from Early Proto Pearic. The system of final plosives thus had only four units *p t kʰ k.

In Chong (Siripen 2001), open syllables -O 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 -O 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 whithout final -ʔ as in Khmer (Ferlus 1992), in Katuic and in Bahnaric (Sidwell 1998, 2005).

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).
Towards Proto Pearic

Closed syllables: four registers

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[CvC]</td>
<td>[Cv'C]</td>
<td>[CvC]</td>
<td>[Cv'C]</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Stages</th>
<th>States</th>
<th>Shifts</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Proto Pearic Mon-Khmer</td>
<td>Syllables with final *ʔ.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loss of final *ʔ &gt; formation of syllables -ό</td>
<td></td>
</tr>
<tr>
<td>(2) Early Proto Pearic</td>
<td>No syllable with final -ʔ.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>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, Kautic and Khmer. In Monic, it affected only the rime *ʔ (oa -ay in modern Mon).</td>
<td></td>
</tr>
<tr>
<td>(3) Late Proto Pearic</td>
<td>No syllable with final -ʔ. Marked register /creaky/.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Devoicing of plosives initials: */b d j g/&gt;p t c k/ (in general).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reinforcement of unvoiced plosives initials: */p t c k/&gt;/pʰ tʰ cʰ kʰ/ (regular) or preservation as /p t c k/ (influence of Khmer).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formation of register contrast /clear vs breathy/.</td>
<td></td>
</tr>
<tr>
<td>(4) Pearic at registral stage</td>
<td>No final -ʔ. Four registers system: /clear/, /creaky/, /breathy/ and /breathy-creaky/.</td>
<td></td>
</tr>
</tbody>
</table>

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 Kautic 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).

<table>
<thead>
<tr>
<th>Old Chinese</th>
<th>Middle Chinese</th>
<th>transferred to Vietic, Katuic and Pearic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CvCV(C) (tenseness)</td>
<td>CV(C)/T (v. lowering)</td>
<td>T(ense)</td>
</tr>
<tr>
<td>CV(C) (laxness)</td>
<td>CV(C)/L (v. raising)</td>
<td>L(ax)</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Early Proto Vietic syll.</th>
<th>Late Proto Vietic finals</th>
</tr>
</thead>
<tbody>
<tr>
<td>*/p t c k/</td>
<td>*/s h/</td>
</tr>
<tr>
<td>*/ʔ/</td>
<td>*/m n p j r ɬ w j/</td>
</tr>
<tr>
<td>T</td>
<td>CvCVC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>CVC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹ ngang-huyễn</td>
</tr>
</tbody>
</table>

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-huyễn, sóc-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 sóc-nặng tones in Vietnamese.
8.3 Transfer and evolution of the contrast /T vs L/ in Katuic (Diffloth 1989)

<table>
<thead>
<tr>
<th>Early Proto Katuic syll.</th>
<th>Late Proto Katuic finals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*/p t c k/</td>
</tr>
<tr>
<td>T</td>
<td>*/m n p y/</td>
</tr>
<tr>
<td></td>
<td>*/r l s h w j/</td>
</tr>
<tr>
<td></td>
<td>*/Ø</td>
</tr>
<tr>
<td>L</td>
<td>p t c k</td>
</tr>
<tr>
<td></td>
<td>m n p nj</td>
</tr>
<tr>
<td></td>
<td>r l s h w j</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Early Proto Pearic syll.</th>
<th>Late Proto Pearic finals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*/p t c k/</td>
</tr>
<tr>
<td>T</td>
<td>*/m n p y/</td>
</tr>
<tr>
<td></td>
<td>*/r l s w j/</td>
</tr>
<tr>
<td></td>
<td>*/h/</td>
</tr>
<tr>
<td></td>
<td>*/Ø</td>
</tr>
<tr>
<td>L</td>
<td>p t c k</td>
</tr>
<tr>
<td></td>
<td>m n p nj</td>
</tr>
<tr>
<td></td>
<td>r l s w j</td>
</tr>
<tr>
<td></td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>v&gt;V</td>
</tr>
</tbody>
</table>

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: [pʰ tʰ cʰ kʰ] > [mʰ nʰ pʰ yʰ] while merging with nasals of the same register (Isara 2009). This phonetic change is recent because it also affects the final [sʰ>tʰ] (> [nʰ]).

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 zhí 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 (Jìa dān 賈耽), 8th century, details the land route from the Chinese possession of Kiao-tche (jiāo zhí 交趾; Sino-Vietnamese: Giao chỉ), 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-
Michel Ferlus

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 Land31 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 (婆鏤) 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

---

31 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’Étrême-Orient
BSEI: Bulletin de la Société des Études Indochinoises
ILCRD: Institute of Languages and Culture for Rural Development
JSS: The Journal of the Siam Society
MKS: Mon-Khmer Studies

Bastian, Adolf. 1868. Reise durch Kambodja nach Cochinchina, tome 3. Herman Gostenoble, Jena. [samreh 89-90; khamen boran 264-265; xong 295-298]
Ferlus, Michel. 2009. What were the four Divisions of Middle Chinese ? Diachronica 26(2): 184-213.


[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]


Towards Proto Pearic


Tcheou Ta-Kuan. 1295/1296. See Paul Pelliot 1902.


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 Gtaʔ 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 Gtaʔ 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 Gtaʔ it receives the same marking as P. The pronominal R in Gtaʔ 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. Gtaʔ 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 Gtaʔ 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 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...
Munduruku “a common function marker for dative, allative and locative.” (Blansitt 1988:180) His Function Conguity 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).” Rice & Kabata argued that direction markers tend to extend to recipient marking diachronically. 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 Purpose extending to Reason and Infinitival and with Mental to Conceptual extending to Perceiver and Experiencer.

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 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

33 See also Seppo Kittila & Silvia Luraghi’s Questionnaire for their on-going project “Differential marking of spatial relations: the case of direction with human landmarks”. (E-mail circulation).
34 For a schematic presentation of Rice and Kabata's model see Sally and Kabata, 2007: 494, Figure 21.
35 The name Lower Munda was given to three extreme south Munda languages, Remo, Gutob and Gtaʔ 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 Experience\textsuperscript{36} 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 Gtaʔ 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 Gtaʔ 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 Gtaʔ. Let us take up Remo\textsuperscript{37} data first:

Remo:

1. gitinəŋ a-niŋ mujŋ gimɛ bɛɽ:ɔʔ
   3SG ACC/DAT-1SG one goat give:PST
   ‘He gave me a goat.’

2. a-ŋɔ niŋ surɔŋ bɛʔ:ti:ŋ
   ACC/DAT-2SG 1SG medicine give:NPST:1SG
   ‘I shall give you medicine.’

3. rɑmo a-sonjə mujŋ gimɛ bɛɽ:ɔʔ
   Ramu ACC/DAT-Sonya one goat give:PST
   ‘Ramu gave Sonya a goat.’

4. aɽi a-meri mujŋ upoɑr 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\textsuperscript{38} -la, (which can also be dispensed with) as in the examples (5) and (6):

5. aɽi a-meri-la surɔŋ bɛɽ:ɔʔ
   Hadi ACC/DAT-Mary-CLIT medicine give:PST
   ‘Hadi gave medicine to Mary.’

6. rɑmo a-meri-la mujŋ upoɑr bɛɽ:ɔʔ
   Ramu ACC/DAT-Mary-CLIT one gift give:PST
   ‘Ramu gave a gift to Mary.’

\textsuperscript{36} 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).

\textsuperscript{37} Remo is the language of the Bondas distributed in and around the Bonda Hills in the Khaipur 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.

\textsuperscript{38} 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 kɔlm sunɔʔ bɛɽ-ɔ “Sabita sold her pen.” selanla niŋ-pulaj daʔ run-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 kɔlm sunɔʔ bɛɽ-ɔ and selan niŋ-pulaj daʔ run-ɔʔ.
Arun Ghosh

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. gitinɔŋ a-niŋ uʔsrɑm uɽ:ɔʔ bibɛʔ
   3SG ACC/DAT-1SG story tell-PST RED:give
   saʔ-gi-ta
   ‘He has come to tell me a story.’

8. gitinɔŋ a-niŋ ne-3 ruŋ-ɔʔ bɛɽ:ɔʔ
   3SG ACC/DAT-1SG 1-child bring:PST give:PST
   ‘He brought and gave me a child.’

9. a-joŋ-dāj
   ACC/DAT-mother-PB.KIN tell-PST
   ‘He told his mother’.

10. gitinɔŋ a-niŋ mujŋ bire tul-ɔʔ
    3SG ACC/DAT-1SG One stone throw:PST
    ‘He threw me a stone.’

11. gitinɔŋ a-niŋ mujŋ bɔl kɑnɖel-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
    ACC/DAT-house go-IMP
    ‘Go home.’    (Frank Fernandez 1967: 112)

13. niŋ (a)-ɖ io-baʔ39 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 dɑyk-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-mundlipada uj-ti-ŋ
    1SG ACC/DAT-Mudulipada go-NPST-1SG
    ‘I shall go to Mudulipada.’

39 Here R-marking morpheme a- is optional. Similarly -bɔʔ can also be dispensed with in favour of a-.
Goal Marking in Munda

17. ŋɔ ə-aŋ 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. bajaran-ɔʔ  uj-a mari gime-se run
market-D/L go-IMP and goat-meat bring
‘Go to the market and bring mutton.’

19. daktar-ɔʔ  uj-a ma a-pe suraŋ bsʔ-tɔ
doctor-D/L go-IMP who ACC/DAT-2PL medicine give-NPST
‘Go to the doctor who will give you medicine.’

20. gitinŋa  laŋa-ɔʔ  uj-ga
3SG field-D/L go-PRF
‘He went to the field.’

21. badal ari-ɔʔ  uj-suŋ-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. ninj djo-boʔ gay-t-iŋ
1SG house-D/L enter-NPST-1SG
VS
ninj a-djo gay-t-iŋ
1SG ACC/DAT-house enter-NPST-1SG
‘I enter the house.’ (Frank Fernandez 1967: 68)

23. ʃəm gus américain d̪i-ta
arrow dog-in COP-NPST
VS
ʃəm a-gusʔ d̪i-ta
arrow ACC/DAT-dog COP-NPST
‘The arrow is in the dog.’ (Frank Fernandez 1967: 68)

24. ninj korji-bagboʔ layk-t-iŋ
1SG chair-D/L sit-NPST-1SG
VS
ninj a-korji lajk-t-iŋ
1SG ACC/DAT-chair sit-NPST-1SG
‘I sit on the chair.’ (Frank Fernandez 1967: 68)

25. gulaj-ne konθa-ɔʔ onqaj-ga
boys-PL hill-D/L go-PRF
VS
gulaj-ne a-konθa onqaj-ga
boys-PL ACC/DAT-hill go-PRF
‘The boys went to the hill.’
The psychosomatic verbs\(^{40}\) encode the subject as experiencer of transfer point of bodily or mental sensations with the same R-marker, that is \(\alpha\), 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 experiencer 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. \(a\)-niŋ kuru-saʔgɑ niŋ kiaŋ sum-oʔ-niŋ
ACC/DAT-1SG hungry-PART 1SG sum eat-PST-1SG
‘As I was hungry I ate rice.’

27. \(a\)-niŋ lu-lor-du suʔ:-gu-ta
ACC/DAT-1SG RED-vomit-DESID-PRF-NPST
‘I wish to vomit.’

28. \(a\)-niŋ si susoʔ-p-dən-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 \(\alpha\)- or may be marked with \(a\)-
along with the clitic pulɑj ‘for’, as in (29)-(31):

29. sobita niŋ-pułaaj muj kɔlɔm sɔb-ɔʔ
Sabita 1SG-CLIT one pen buy-PST
‘Sabita bought a pen for me.’

29a. sobita \(a\)-niŋ muj kɔlɔm sɔb-ɔʔ
Sabita ACC/DAT-1SG one pen buy-PST
‘Sabita bought a pen for me.’

29b. sobita \(a\)-niŋ-pułaaj muj kɔlɔm sɔb-ɔʔ
Sabita ACC/DAT-CLIT one pen buy-PST
‘Sabita bought a pen for me.’

30. selɑn niŋ-pułaaj qaʔ ruŋ-səʔ
3SG 1SG-CLIT water bring-PST
‘She brought water for me.’

30a. selɑn \(a\)-niŋ qaʔ ruŋ-səʔ
3SG ACC/DAT-1SG water bring-PST
‘She brought water for me.’

31. gitinɔŋ niŋ-pułaaj semuʔ sisep-dən-ta
3SG 1SG-CLIT tree RED-cut-PROG-NPST
‘He is cutting the tree for me.’

31a. gitinɔŋ \(a\)-niŋ semuʔ sisep-dən-ta
3SG ACC/DAT-1SG tree RED-cut-PROG-NPST
‘He is cutting the tree for me.’

---

\(^{40}\) For a detailed discussion of the psychosomatic verbs in Mundari see Osada (1992: 104-109).
31b. gitinɔŋ a-niŋ-pulaj semuʔ sisep'ɖen-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. rɑmo a-sonjɑ-ul ɔ:sum:ɔʔ
Ramu ACC/DAT-Sonya-CLIT rice CAUS-eat-PST
‘Ramu caused Sonya to eat rice.’

33. a-ɔʔɔŋ-gu ɔ:lemɔ
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---\(\triangleleft\)RECIPIENT=LOCATION=EXPERIENCE =BENEFICIARY = CAUSEE

A closely related language and a close neighbour of Remo Gutob\(^{41}\) 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 \(-laj/pulaj\). The D/L-marker in respect of both human and non-human goals is \(-bo\) which optionally alternates with the R-marker \(-laj/pulaj\), 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 \(-laj/pulaj\) in case of pronominal goal and the clitic \(-laj\) 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-ɔ
3SG ACC/DAT-1SG tell-PST
‘He told me.’

37. maj o-niŋ muj-ro bol ɔtʃ-ɔ
3SG ACC/DAT-1SG one-CLAS ball throw-PST
‘He threw me a ball.’

38. sobita o-niŋ muj-ro boj ɔtʃ-ɔ
Sabita ACC/DAT-1SG one-CLAS book bring-PST
‘Sabita brought me a book.’

---

\(^{41}\) 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. ɑɽi meri-laj muj-ro upoor bɛɽ:ɔ
Hadi Mary-CLIT one-CLAS gift give-PST
‘Hadi gave a gift to Mary.’

40. badɔl sobita-nu-laj muj-ro boj rɪŋ-ɔ
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 rɪŋ-ɔ
Sabita Sonya-GEN-CLIT one-CLAS book bring-PST
‘Sabita brought a book to Sonya.’

41a. sobita sonja-nu-bo muj-ro boj rɪŋ-ɔ
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-niŋ sos lagaj-guni
ACC/DAT-1SG thirst be affected-COP
‘I am thirsty.’

43. o-niŋ besi duk-ɖeŋ-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
laj is assigned the accusative status as it is basically used for P-marking.

44. o-niŋ-laj daʔ riŋ-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 maaqru-pulaj gisĩŋ riŋ-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. niŋ oʔon-laj jon-ɖeq-nu-bo beɖ-oʔ-niŋ
1SG child-ACC/DAT mother-P.B-KIN-GEN-LOC give-PST-1SG
‘I gave the child to the mother.’
Goal Marking in Munda

47. rɑmu sonja-laj muj-ro gimeʔ  bed-oʔ
Ramu Sonya-CLIT one-CLAS goat give-PST
‘Ramu gave Sonya a goat.’ (Rajan & Rajan 2001: 32)

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

48. o-nom niŋ niŋ-nu apuŋ-lay əb-soi-tu-niŋ
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
du-tu AUX-NPST
2PL Gadaba-people British-CLIT language CAUS:learn-2PL
‘We the Gadabas are teaching the Saheb (our) language.’

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

50. niŋ bier mudlipaɖɑ i:lom-niŋ
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 Gtaʔ 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 the pronominal R is marked by the accusative-dative prefix ɑ-. Sometimes the pronominal goal receives double marking, that is, pronominal R is marked by the accusative-dative prefix ɑ- and the suffix -ke. The causative 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. ɑUniŋ muj ʈɑɑ biʔ:lɑ
ACC/DAT-1SG one rupee give:IMP
‘Give me one rupee.’

52. mɛ ɑUniŋ husrɑ ɑni:te
3SG ACC/DAT-1SG story tell:PST
‘He told me a story.’

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

---

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

44 –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 Gtaʔ in the southern part of Malkangiri district where more than 600,000 Bengalees are staying in the Dandakaranya project area.
53. ṛamu ṭonja-ke mujŋ gimi biʔ-te
   Ramu Sonya-ACC/DAT one goat give:PST
   ‘Ramu gave a goat to Sonya.’

54. nेŋ ḏomru-ke muj citi 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. ṛamu ŭjo-ke a-ʧon-te
    Ramu baby-ACC/DAT CAUS-eat-PST
    ‘Ramu fed the baby.’

56. me a-na mitŋa a-ʧon-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-neŋ besikeɖe sob-ʤon-te
    ACC/DAT-1SG very behungry-PROG-NPST
    ‘I am very hungry.’

58. tæn kiton 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-meke45 ter gisin biʔ
    ACC/DAT-3SG-ACC/DAT this fowl give
    ‘Give him this fowl.’
    (Bhattacharya 1975b:166)

60. a-me-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. nेŋ jay-ke mujŋ muŋom biʔ-te
    1SG mother-DAT one child give-PST
    ‘I gave the child to the mother.’

62. nेŋ jay-ke muŋ gisin biʔ-te
    1SG mother-DAT one fowl give-PST
    ‘I gave mother a fowl.’

45 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 Gtaʔ does not overtly mark the place-names. Compare (63) and (64) with (65):

Gtaʔ:

63. niŋ njigɖe mudulipɖa waj-e
1SG tomorrow Mudulipada go:FUT
‘I shall go to Mudulipada tomorrow.’

Gutob:

64. niŋ bijer mudlipɑ i-lom-niŋ
1SG tomorrow Mudulipada go:FUT:1SG
‘I shall go to Mudulipada tomorrow.’

Remo:

65. niŋ a-mudlipɑɖɑ uj:ti:ŋ
1SG ACC/DAT:Mudulipada go:NPST:1SG
‘I shall go to Mudulipada.’

In Gtaʔ 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 Gtaʔ may be schematised as follows:

\[ \text{ACC-DAT} ----> \text{RECIPIENT} = \text{DIRECTION} = \text{CAUSEE} = \text{EXPERIENCER} \]

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 Gtaʔ, 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. dɑkɑ em-a-d-ɪɲ-a-e
rice give-A-TM-1SG-FIN:3SG
‘He gave me rice.’

67. hɛc’-a-d-ɪɲ-a-e
come-A-TM-1SG-FIN:3SG
‘He came to me.’
(SK cf. LN. 49)

68. uni łai-a-ko-a-e
3SG tell-A-3PL-FIN:3SG
‘He will tell them.’

69. dare-ɲ benget’-a-ɮ-ɪɲ-tahɛkɑn-a
tree-1SG look-at-A-INAN-COP-COP:PST-FIN
‘I was looking at the tree.’
(Bo. Cf. LN.53)
70. ɲel-o-co-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. manɖi 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 kajj-a-t’-ko-a
    servant-PL-3SG tell-A-TM-3PL-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 inj mya-kama:y ghaliba
    you-to(IO) I one-work-obj give up
    ‘I will give you a work.’
    (Nagaraja 1999: 46)

75. inj Dic-ke ambesasa kule-c-lakken
    I he-obj.mango-bring send-per-cont
    ‘I am sending him to bring mango.’
    (Nagaraja 1999:46-47)

76. dij’ koro-ke inj-en ghaleij
    that man-ACC/DAT 1SG-D/L show
    ‘Show me that man.’
    (Bhattacharya 1975:157)

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

77. oni sim inj-ke ago-iŋ-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 beʈa-nom-te inj-te tere
    2SG 2SG-GEN son-2SG-ACC/DAT 1SG-ACC/DAT give
    ‘You give me your son.’
    (Bhattacharya 1975:158)

79. inj inj-ɑ apa-iŋ-te am-te ob-iyo-iŋ’
    1SG 1SG-GEN father-1SG-ACC 2SG-ACC CAUS-see-1SG
    ‘I will show my father to you.’
    (Bhattacharya 1975: 158)
80. in ulaga-thoŋ ɗaru-te ɗep-naŋ-g’
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 faɱo-te ara ipu baŋbaŋnda kete
Ram Shyam-ACC/DAT Self House burn about
me-gaŋ-yoŋo gam-o
NEG-say-INF say-PST
‘Ram told Shyam not to tell anybody about his burning house.’
(Pattanayak 2008:212)

82. aịŋ-te juŋ-ka gata-ro-ki ab-sop-iŋ
1SG-ACC/DAT Juang-GEN language-DEF-PL CAUS-teach-1SG
‘Teach me the language of Juang.’
(Bhattacharya1975:160)

83. am aịŋ-te goble-kon-om-te m-am-po-iŋ
2SG 1SG-ACC/DAT nephew-2SG-ACC/DAT CAUS-see-1SG
‘You will show your nephew to me.’
(Bhattacharya1975: 160)

84. ram-o-a [aịŋ-o aŋgor-te]
Ram-GEN [1SG-GEN house-ACC/DAT go-INF]
bels 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-jii a-manfra tiya
that ACC/DAT-flower-PL ACC/DAT-man give
‘Give those flowers to the man.’
(Bhattacharya 1975:162)

86. kuni a-tarbaŋ-jii a-nip-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), Birhɔɽ (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 Birhɔɽ 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.46 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

46 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 pera-ko hre'-en-a
   ‘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
   in-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 gəł-bə an-de aijn-te
   3SG If Keonjhar-DIR go-CON 1SG-ACC/DAT
   pacola m R r-ene
   shawl FUT-bring-FUT
   ‘If he goes to Keonjhar he will bring a shawl for me.’ (Pattanayak 2008: 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. sowun-ja-ŋ-ə
   smell-AM-TM-1SG-FIN
   ‘I have sensed a smell.’ (Osada 1992: 106)

Korku:

93. in-en kapara kasu-lakken
   I-dlc head ache-cont
   ‘I am having headache.’ (Nagaraja 1999: 100)

94. in-en da dadam-jen
   I-dlc water thirsty-pt
   ‘I am thirsty.’ (Nagaraja 1999: 100)
Juang:

95. *aįŋ-te emuló 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:


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 ɑ- as the permessee 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) 47

2. ACC/DAT [Korku, Birhor, Kharia, Juang and Sora]—→ RECIPIENT=DIRECTION= CAUSEE=
   BENEFICIARY=(EXPERIENCER) 48

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. Gtaʔ 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

---

47 The permissive construction is available in Santali, hence PERMISSEE is kept in paranthesis.

48 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 permissee (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**

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

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, Chaloem 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), plijh (Khmu’ Rawk) and pлях (Mlabri); ‘to wake (someone up)’: *phə́ (Proto-T’in), phə́ (Mal), phə́ (Pray), phə́ (Khmu’ Rawk) and паpуух (Mlabri). These Mon-Khmer languages belong to the Khmuic branch of the Mon-Khmer language family (Filbeck, 1972 and 1978).

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.
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.

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Labial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>Vl. unaspirated</td>
<td>p</td>
<td>t</td>
<td>c</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>Vl. Aspirated</td>
<td>ph</td>
<td>th</td>
<td>kh</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vd.</td>
<td>b</td>
<td>d</td>
<td>j</td>
<td>g</td>
</tr>
<tr>
<td>Nasal</td>
<td>Vl.</td>
<td>hm</td>
<td>hn</td>
<td>ʰn</td>
<td>ʰŋ</td>
</tr>
<tr>
<td></td>
<td>Vd.</td>
<td>m</td>
<td>n</td>
<td>ɲ</td>
<td>ŋ</td>
</tr>
<tr>
<td>Fricative</td>
<td>Vl.</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aproximant</td>
<td>Vl.</td>
<td>hw</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vd.</td>
<td>w</td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pregottalised</td>
<td>ʔw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Consonant clusters

\[
\begin{align*}
\text{Cw} & \quad \text{phw thw kw khw lw} \\
\text{Cl} & \quad \text{pl phl bl kl khl} \\
\text{Cj} & \quad \text{pj phj bj kj qj mj sj} \\
\text{NC (l/j)} & \quad \text{mph mpl mphl nth ns ŋk ŋkh ŋɡ ŋkl ŋkj}
\end{align*}
\]

### Vowels

#### Monophthongs

\[
\begin{array}{ccccccc}
\text{Front} & \text{Central} & \text{Back} \\
\text{High} & \text{Short} & \text{Long} & \text{Short} & \text{Long} & \text{Short} & \text{Long} \\
\text{Mid} & \text{Short} & \text{Long} & \text{Short} & \text{Long} & \text{Short} & \text{Long} \\
\text{Low} & \text{Short} & \text{Long} & \text{Short} & \text{Long} & \text{Short} & \text{Long} \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Diphthongs} & \text{io} & (iɔ) & uɔ \\
\text{uɔ} & 50
\end{array}
\]

### Tones

\[
\begin{array}{ll}
\text{́ (high)} & \text{̀ (low)}
\end{array}
\]

\*\*Proto-T’in (Proto-Lua’)* -r has become *-ŋ in Yotdoiwatthana Mal, for example, *phar > phāı́ ‘to fly’; *thar > thā́ ‘rope, cord, string’; *khe:r > khē:ŋ ‘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 reflect tonal evolution in Mal, three more Mal varieties spoken in Ban Kwet, Ban Phu Kok, Ban Yotdoiwatthana 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: iː; eː; ɛː; ɨː; əː; aː; uː; oː; ɔː. 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:iː; e:eː; ɛ:ɛː; ɨ:ɨː; ə:əː; a:aː; u:uː; o:oː; ɔ:ɔː, 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 Patthawi 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 Yotdoiwatthana 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/ and velars /k kh ɡ/ and nine long vowels, i.e. /iː eː ɛː ɨː əː aː uː oː ɔː/ 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.
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).

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.
Generally, the short front vowels /i e ɛ/ have a higher F1 but a lower F2 than those of their long counterparts /iː eː ɛː/. This means that the short front vowels are lower and more centralised. For central vowels, /ɨ/ is lower than /ɨː/ since /ɨ/ has a higher F1, whereas /ɔ/ and /a/ are higher and more centralised than /ɔː/ and /aː/ due to the fact that /ɔ/ and /a/ have a lower F1 but a higher F2. With regard to back vowels, the short vowels /ʊ o ɔ/ have a higher F1 than their long counterparts /ʊː oː ɔː/. This indicates that the short vowels are lower than the long vowels and that /ɔ/ is more peripheral than /ɔː/, 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-əː/ and /a-aː/, and the F2 of the pairs /e-ɛː/, /i-iː/ and /a-aː/.

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.

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, /kǎan/ ‘work’ (loanword) : /kaan/ ‘defeated’ (loanword); /cǎaŋ/ ‘be able’ (loanword) - /caaŋ/ ‘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 Jirananthanaporn (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.

<table>
<thead>
<tr>
<th>Syllable structure</th>
<th>Phonetic characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open and closed smooth syllable</td>
<td>High-falling, Low rising</td>
</tr>
<tr>
<td>Checked syllable</td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>Level</td>
</tr>
<tr>
<td>Long</td>
<td>Level, Low-rising</td>
</tr>
<tr>
<td>Syllable with short vowel and /-h/</td>
<td>Level</td>
</tr>
</tbody>
</table>

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).
A Synopsis of Mal Phonetics

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

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)


———. 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)


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)


Appendix

Distribution of Thin (Mal) Ethnic Group in Nan Province, Thailand
Mon Converging Towards Thai Models: Evidence from a Mon Historical Text

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 19th-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 Rājāvaṁsa Kathā 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. 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 13th to the latter part of the 14th centuries AD.

51 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.

52 The Mon title is "Gaḷāṇ Khāw Bduh Dthw Smiṅ Dūh Hāṃsā, meaning, “On the Origin of the Haṃsāvatī 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.

The language of the Rājāvamsa Kathā, particularly the narrative of Rājādhirāj, appears highly “Siamified.” 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 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 20th century. Mons settled scattered throughout what is now Central Thailand. The usual understanding is that these Mons arrived from Burma in waves starting in the 17th century, although there is mention of Mon-speaking communities at Ayutthaya in earlier centuries. 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. 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

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.

---

55 The printed text has been available in Siam and Burma for close to one hundred years and has been the focus of scholarly attention.
56 I use the term “Siamified” because this process started before there was the modern state of Thailand.
57 That is, Mons from inside Burma, in opposition to the Siamese or Thai Mons, the ไทยรามัญ Thai Rāmañ.
60 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.
61 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 19th 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.

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. 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 kʰɔ̌ːŋ or “thing,” which has lost its original meaning, but indicates possession, as in:

<table>
<thead>
<tr>
<th>Thai</th>
<th>Burma Mon</th>
</tr>
</thead>
<tbody>
<tr>
<td>kʰonöm kʰɔ̌ːŋ dëk</td>
<td>kwaiŋ kon</td>
</tr>
</tbody>
</table>

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

In the Rājādhirāj narrative, we find a seemingly incongruous use of the word krɔ̤p, 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 kʰɔ̌ːŋ. Because of the possible reading of the Mon term

---

62 This is a pseudonym.
63 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.
64 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.
65 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 Matichon dictionary of Siamese Mon does in fact list krɔ̤p as having the Thai grammaticalized usage. See คณะกรรมการจัดทําพจนานุกรมมอญ:ไทย พจนานุกรมมอญ-ไทย อุปสมบัติ (กรุงเทพฯ: นสพ.น, 2548), p. 143. [Committee for the Preparation of the Mon-Thai Dictionary, *Mon-Thai Dictionary, for Thai Mon*, (Bangkok: Matichon, 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 krɔ̀p.

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 krɔ̀p. 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 Harināvatī, 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 rājaśastra, here rɛ̀rcəst, following the Thai pronunciation ราชศาสตร์. In Thai usage, this is a law promulgated by the king in accordance with the principles of the Dharmaśastra. Following the interpretational lens of Thai, at first glance the phrase ɲɛ̀h mə tɔ̤h cəno̤ k implies, “the person who becomes/will become great.” A more likely interpretation would be the common Thai expression, ผู้เป็นใหญ่, a common description of someone who is a “superior.”

66 Unless otherwise indicated, page numbers are all from the Rājādhirāj section of Phra Candakantā’s Rājāvaṁsa Kathā.

67 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.

68 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.
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 kṛṣp 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 စွပ်ရှားစူးဒီယို Slapat Rājāvaṅ Datow Smiṅ or “(Treatise) on the History of the Kingly Lineage,” sometimes referred to in English as “The History of Pegu.”

p 17 (Slapat Rājāvaṅ Datow Smiṅ)

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 ว่า wâː, a grammaticalized verb meaning “say.” The fact that wâː 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 Rājādhirāj repeated examples of ကြာ kgh, a Literary Mon verb meaning “say, tell,” being used exactly following the Thai pattern of wâː. Consultation both with Burma Mon speakers and

---

69 Title of ambiguous meaning, from “king” through “local leader” or “headman.”
70 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 Smiṅ Roṅ, but the final word is an assertive focus particle. Following tradition, I have given titles in Indic transliteration, not transcription.
71 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.”
72 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 kɛ̤h is the second element in the Literary Mon phrase hɒm kɛ̤h, an elaborate expression meaning “to say” used in formal language.\footnote{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.} Instances of uses outside of this literary collocation may provide stronger evidence for the Mon replication of the Thai model, as in:

Lord Rājādhīrāj 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:

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

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.
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. Burmese là and Mon ha. Questions involving relative words, however, are marked with different particles, usually lè in Burmese and rao in Mon. In contrast, Thai, uses a variety of expressions to ask absolute questions, including ไหม mǎi, often with an expectation of confirmation. Another strategy is หรือ rɯ̌, often followed by ปล่อย plɔ̀ːy “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:

(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:

The Thai word rɯ̌ 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

---

75 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.
76 Burmese has alternate forms for both absolute and relative questions, though these two are the most common.
77 This word is the result of a typical sound change in Thai Mon – the Burma Mon equivalent is pʰyɒn, made up of causative pə + cɜ ̤n “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.
78 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:

*lamyñ miʔ ket lamyym ʔoə ha, palaʔ lamyym ʔoə ha*

now mother take  life 1S INT, free  life 1S INT

The last example sentence begins Thai and ends Burmese. The Thai honorific prefix พระ phráʔ appearing in the title of Min Gaung’s name, while at the end is the phrase หู seŋ pəwaʔ, highly reminiscent of the Burmese phrase အိုးလေးဗြဲ ma hoo သံ အိုးဗြဲ, 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:

*ha təlaʔ.ɲɛ̤h  pʰraʔ mɛ ̤aŋ  kɔ̤ŋ    kɒ ̤  lik   ne ̤əŋ nɔʔ  kɔ̤h*

fact lord Phra Maṅ Kaṁṅ GIVE exist letter bring this TOP

*kəsɔp    krɜ ̤h  nɛ ̤ək lo ̤ n.   cʰəkɛ ̤h   tɔ̤h  ekarat   nʊm  kɔ*

thought harm deep very. but be king exist with

*ʔat  kɒ pɜ ̤  ləpɛ ̤h   səray  coiŋ cʰeh*

request give officer soldier hero elephant horse

_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 kʰɔ̌ː, 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 kʰɔ̌ː 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.

*ʔat ko pɔ₃ lapgh soray coiŋ cʰeh*

request give officer soldier hero elephant horse
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 ʔat kn suggests Thai ขอให้ kʰɔ̌ː 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.\textsuperscript{79}

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 ʔat kn suggests Thai ขอให้ kʰɔ̌ː 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.\textsuperscript{79}

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 Don Keh Htaw, “The Story of Golden-Nib Mi Dong” from 19\textsuperscript{th}-century Burmā.\textsuperscript{80} The following features paratactic repetition of the verb “request.”

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 cʰûay “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 cʰûay into Mon: translating it into its Mon equivalent rûm bûng, or using cʰoə, the closest phonetic rendering

\textsuperscript{79} A direct translation such as ขอให้ข้าจะได้รับการตอบแทนบุญคุณเจ้าจนหมดชีวิตหน่อย does not appear acceptable, while ข้าขอตอบแทนบุญคุณเจ้าจนหมดลมหายใจ may be more so.

\textsuperscript{80} Nai Lokasīri, ed., Wotthu Mi Doṅ 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 ….”

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

```
Mjt? pʰaɨŋ cʰə hom ko ɲiʔ kóm
mother HELP HELP say GIVE LITTLE also
```

Mother, please tell him for me.

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

```
təlaʔkṳn pʰyeh mettā cʰəʔā hɒm kóm ɲiʔ ko ngə nɔʔ kóm
monk, throw mettā HELP go say GIVE Baññā Noy LITTLE.
```

Your reverence, please81 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 ʈʰɯ̌ ŋ 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 ʈʰɯ̌ ŋ 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 a few direct Thai loanwords: the term kʰít means think, plan, consider, wonder, be of the opinion. In Burma Mon, the various meanings of kʰít tend to be broken up into disparate expressions, some borrowed from Burmese. Because kʰít 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:

```
ʔat pʰyeh mettā klon kəlon ɲɔʔ kóm ɲiʔ oə.ɗoik
request throw mettā do work this GIVE 1S.vassal LITTLE.
```

"Please do this work for me."

81 The collocation throw mettā is used in both Thai (แผ่เมตตา pʰæ̀ː mêːttaː) 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:

```
ʔat pʰyeh mettā klon kalon nɔʔ ko ʔoə.ɗoik ɲiʔ
```

"Please do this work for me."

In Thai Mon, the pronoun ʔoə.ɗoik is not used only when talking to monks but as a general polite 1S pronoun. See ละออ แป้นเจริญ, ระบบไวยากรณ์มอญ (มหาวิทยาลัยศิลปากร, 2526), p. 138. [La-or Paencaroen. Mon Grammatical Systems. MA Thesis, Silapakorn University, 1983].
Let Banñā Noy think of Lady Me Thao; let Banñā Noy miss Lady Me Thao.

To the Burma Mon, the above sentence may sound something like, “Let Banñā 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:

Let Banñā Noy miss Lady Me Thao.

Search

In Thai, another word used in a similar way is the หา hǎː 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. 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 kláy, ‘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:

Let Banñā Noy come to me, Have Banñā Noy come to me.

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

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 nɔ́ːy, “small” or “junior.”

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.
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 wâː* “be of the opinion; think” and *hěn dûay* “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 ɲat, which can only mean “see” and does not match the semantic range of the Thai. The expression ɲat *kɛ̤h* 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 ɲat *kɒm*, rendering *hěn duây*, 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.

In Burma Mon, this might be rendered as:

\[
\text{tɛ̅h \?ɔiŋ.tɜ̅ŋ lɔ̅ kла raʔ} \quad \text{HIT endure set-down before FOC, 1S think}
\]

We may note that the above Burma Mon sentence not only employs a Mon rendering of the Burmese ɔɔ̃ ʔtʰ, 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:

\[
\text{ʔɒt cɒt} \quad \text{use up mind}
\]

*be fed up, be uninterested*
This is rendered in Burma Mon as *cotent.* 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:

```
အောင် မြင်နေပါသည်။ အောင် မြင်ပြီး ခြင်းများ ရှုပ်ပြောင်း မရှိပါ။
```

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* 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:

```
ပြါးရွှင် ပုံ ကြည့်ပါ။ အောင် မြင်ပြီး မှန်ကန် မရှုပ်ပြောင်း
```

in forces 3P TOP, do not mind small FOC

Calque of the Thai Collocation *ขอโทษ* *kʰɔtʰoːt*

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 kʰɔtʰ*, *request* as discussed above, and the Sanskrit *โทษ tʰɔtʰ*, meaning *crime, offense; sin, fault*. But 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:

```
ကြောင့် သီးခြား သူများ သုံးပြီး လှည့်ပြောင်း
```

....(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. With few exceptions, these words are rendered in a Pāli or Pali-cized 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.

---

84 This particular expression has no direct Thai equivalent. It is, however, exactly parallel to the Burmese expression of the same meaning, ပေါင်းစံုစာရင်း.

85 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.
Table 1: Mon Words Following the Thai Use of Pāli-Sanskrit

| Mon |

<table>
<thead>
<tr>
<th>Pāli/Sanskrit</th>
<th>Thai Usage</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>tejānubhāva</td>
<td>เตเจนุภาพ</td>
<td>king’s power, influence</td>
</tr>
<tr>
<td>pāḷi?</td>
<td>pradeśa</td>
<td>ประเทศ</td>
</tr>
<tr>
<td></td>
<td>pradeśa</td>
<td>country</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Pāli word means variety in Burma. In Thai Mon, this term replaces  kwan “city” in the collocation kwan dny “country,” creating kwan pāḷi?</td>
</tr>
<tr>
<td></td>
<td>prayojana</td>
<td>ประโยชน์</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Pāli kāṇa from Pāli guṇa is used with similar meaning.</td>
</tr>
<tr>
<td></td>
<td>paṁña</td>
<td>ปัญหา</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mon can use a Sanskritic form, prayojana</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Thai usage covers two meanings not connected in Mon. For method, lọkākāna can be used.</td>
</tr>
<tr>
<td></td>
<td>samaya</td>
<td>สมัย</td>
</tr>
<tr>
<td></td>
<td></td>
<td>khet derived from Pāli khetta is used.</td>
</tr>
<tr>
<td></td>
<td>saonāyana</td>
<td>สมัย</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No such specialized usage appears to exist.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instead of “Pālifying” the spelling to something like *saggagata, the Thai pronunciation has been rendered in Mon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In Burma, this rare word means nonentity. In Thai, it is used as a verb, for which Mon uses kākāna.</td>
</tr>
<tr>
<td></td>
<td>saowana</td>
<td>อาศัย</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A variety of native expressions, mọ̤ n̥ ọ̤ ṭ o or paʔ hətao, can all be used.</td>
</tr>
</tbody>
</table>

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  eːŋ meaning (one)self, which is borrowed into Mon as ʔoiŋ. We find a curious usage of the word ɓoə, which in Mon means bean, pea, but from the context appears to be the Thai word  bùə, rice-gall midge.

86 Some of the Thai Mon pronunciations are conjectural.
87 The second element is derived from an earlier pronunciation of Burmese, now  lọkākāna but once something like  lọkākāna. |
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 oundingBox, which in Mon means *corrupted, rotten*, rendering the sentence unintelligible:

```
p 382
mæŋ ray kawɔl tɔ̤h 3P 3cɛl pɔ hɔ̤n mæŋ ha.
```
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 tæ̀ː, 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
ัสต cʰaʔ təlaʔ ʔoə raʔ
usat
only lord 1S FOC
```

The Mon sentence, which appears in any case incomplete, is wholly unintelligible without reference to the Thai. The native Mon sут 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 Hon⁹⁹
ัสตต เหรจ้า ที่ คิด ยก
sut täː pʰráʔ.câw pʰîː  kʰít tʰɜ̄t
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
คิด ก็ ไป อามาย เขา นั้น ไม่ หมด แล้ว
kʰít cà pay ?ùbaːy kʰâw nán mây mòt lǽːw
think IRR go stratagem who even use.up PERF
```

Parsing the sentence to render ɲɛ̄h.kɔ̄h 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 ɲɛ̄h kɔ̄h, which could then mean *that person*, another meaning is possible:

```
คิด จะ ไป อามาย เขา นั้น ไม่ หมด แล้ว
kʰít cà pay ?ùbaːy kʰâw nán mây mòt lǽːw
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.

---

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 17th 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

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 19th 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 90 that are based on an ‘Eastern’ Khmu dialect. 91 (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:

90 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).
91 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.


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; Aguetatt 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 Uthai Thani 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).

<table>
<thead>
<tr>
<th>Village name</th>
<th>Thai name</th>
<th>Local Khmu name (IPA)</th>
<th>No. of households</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nam Paan</td>
<td>น้ําปาก</td>
<td>/ʔom paːn/</td>
<td>65</td>
<td>450</td>
</tr>
<tr>
<td>Huay Klaep</td>
<td>ห้วยแกลบ</td>
<td>/ʔom klɛːp/</td>
<td>71</td>
<td>378</td>
</tr>
<tr>
<td>Nam Lu</td>
<td>น้ําหลุ</td>
<td>/ʔom lúʔ/</td>
<td>74</td>
<td>453</td>
</tr>
<tr>
<td>Huay Moi</td>
<td>ห้วยมอย</td>
<td>/ʔom mɔːj/</td>
<td>24</td>
<td>244</td>
</tr>
<tr>
<td>Ban Mai Chaidaen</td>
<td>บ้านชายแดน</td>
<td>/ʔom sêːn/</td>
<td>39</td>
<td>209</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>273</strong></td>
<td><strong>1734</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: Name and population of Khmu villages in Chon Daen sub-district, 2008 (Songkhwae 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 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. 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

---

92 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 Songkhwae 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:

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

<table>
<thead>
<tr>
<th>Ethnolinguistic vitality factor</th>
<th>Significant changes or observations reported</th>
<th>Overall future language maintenance impact (positive, negative, or neutral)</th>
<th>Landweer vitality scale rating where applicable</th>
</tr>
</thead>
</table>
| 1. Intergenerational language transmission | ▪ Children speaking K in the home  
▪ K the language of communication between 3 generations within households | positive | Not applicable |
| 2. Population and group dynamics | ▪ Dramatic increase in young women working outside village  
▪ Few K with K marriages  
▪ Attitude toward mixed marriages increasingly neutral  
▪ Few non-Khmu immigrants in the villages | negative/neutral | 2 |
| 3. Domains of language use | ▪ K the primary language of homes in village (children’s first language)  
▪ K the primary language used in daily village life  
▪ K the ‘most fluent’ language for nearly all respondents  
▪ ST used exclusively in education; urban workplace  
▪ Earlier exposure to ST at day care for 2-3 year olds  
▪ M used for local commerce | positive | 3 |
| 4. Access to a stable and acceptable income base within the locale | ▪ Lack of land rights  
▪ Market for cash crops unstable  
▪ Subsistence rice farming; large debt  
▪ Increased demand for consumer goods through broadcast media exposure | negative | 1.5 |
| 5. Social outlook regarding and within the speech community | ▪ Positive attitude toward culture/language and local language development  
▪ Disappearing visible cultural distinctives; ‘flexibility’  
▪ Youth desire and pursue modern lifestyle | positive/neutral | 2 |

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.

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

Table 3: Skill level choices for reading skills

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

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 an 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


Michelle M. Miller, Timothy M. Miller

Multiple uses of the pronoun *deː* 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 *deː*. Unlike the other personal pronouns, *deː* 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 *deː* over one of the specific pronouns? How do the hearers know who *deː* is referring to? Are the contextual factors involved in interpreting the meaning of *deː* purely pragmatic, or can syntax help? This paper explores the syntactic and pragmatic constraints that govern the use and interpretation of *deː*.

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 *deː* 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 *deː* can ‘replace all other pronouns’, and can also mean ‘alone’. Building on these findings, this paper presents four structurally defined uses of *deː*, one of which has multiple pragmatic functions.

**Syntactically Defined Uses of *deː***

These four uses of *deː* occur in distinctive syntactic contexts. Where there is an antecedent reference in a clause, *deː* has a co-referential function. Where there is no antecedent, *deː* is interpreted as a first person meaning that is unspecified for number. As the final element in a clause, *deː* means ‘alone’. When it occurs in the verb complex, *deː* combines with a verbal element to add aspectual or adverbial information to the main verb.

**Co-referential function**

In a well-formed sentence *deː* is used anaphorically for any pronominal co-reference within a clause and across some clause boundaries. This syntactic context requires an antecedent reference for *deː*. In (1), *deː* is co-referent with the antecedent first person singular pronoun *ʔoʔ*.

(1)  
leʔ ʔoʔ jɛt jɔʔ mɑʔˈkɨn deː i/#ʔoʔ i  
and 1sg stay with aunt CO-REF/ 1sg  
‘And I stayed with my aunt’

---

93 In the Kmhmu’ possessive, possessors follow the possessed entity.
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 *hɨrˈ ňɨɑm deː ‘his heart/mind’.

(2)  
\[ gaː; \_ gəʔ \_ taŋ hɨrˈ ňɨɑm deː; \_ lə? \_ lə? \]
3sgm so_then set_up heart CO-REF good good

‘So he made up his mind firmly.’

In alienable possession also, *deː* may refer to the possessor, as shown in (3), where the possessor in *ʔom deː ‘her water’ is co-referent with the subject *naː ‘3sgf ‘.

(3)  
\[ naː; \_ gəʔ \_ gə:k \_ ʔom deː; \]
3sgf so_then carry_on_shoulder water CO-REF

‘Then she carried her water on (her) shoulder…’

When *deː* has a co-referent function, the antecedent is usually the subject of the clause, as shown in (4) where *deː* is co-referent with *jaʔ gi naːj ‘that woman’.

(4)  
\[ jaʔ \_ gi naːj, \_ an kɔːn ňim \_ deʔ məak deː; \]
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 *deː* is co-referent with *kɔːn ňim ‘young person’, the non-subject argument in the clause.

(5)  
\[ jaʔ \_ gi naːj, \_ an kɔːn ňim \_ hak deʔ məak deː; \]
woman that give young_person nevertheless get hat CO-REF

‘That woman gave the young person his/her hat.’

In order for *deː* 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 *deː* 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 *deː* is the subject of the complement clause *deː ci haːn kʰɨˈˀniʔ ‘I will die now’, and is co-referent with the subject of the matrix clause *ʔoʔ ‘1sg’.

(6)  
\[ ʔoʔ \_ iŋɔʔ \_ deː; \_ ci haːn kʰɨˈˀniʔ \]
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 *gaː ‘3sgm’ is co-referent with *kɔːn ‘child’ the object in the matrix clause.

(7)  
\[ ʔoʔ \_ mə:k kɔːn \_ an gaː; \_ jɔh \]
1sg tell child IMP 3sgm go

‘I told the child to go.’
If *deː* is used in this context, as shown in (8), the sentence is ungrammatical because it has no meaningful interpretation.

(8)

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 *deː* is the subject of the clause *pʰɨa deː*: *ʔoʔ iː ʔɑːn deː ʲi/ ʲi* *jɔh* ‘so that he would become richer than before’ and co-referent with the subject in the main clause *tɑːj goː*: ‘his older brother’.

(9)

*taːj* *goː*: *deː* *suan kap tuːt ʰɾuːl* *pʰia*

older_sibling 3sgm get garden with fig_tree PURP

*deː*: ʰiː *ʔah ʰiː: raˈmaːŋ kʰian liːn əŋ* *ŋan* ‘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 *deː* is co-referent with it.⁹⁴

(10)

₁sg stay with aunt CO-REF

And I stayed with my aunt.’

₀*  jɛt jɔʔ *maʔˈkɨn* *deː*;

but 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 *deː* is used as a co-reference in the first clause, but in the 2nd clause the specific first person singular pronoun *ʔoʔ* must be used to identify the same referent.

(11)

*ʔoʔ* *jɛt jɔʔ* *maʔˈkɨn* *deː*;

₁sg stay with aunt CO-REF

‘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 *deː* has implications for identifying clause boundaries and the nature of the relationship between clauses. Where *deː* 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.

—

⁹⁴ If this construction is analysed without a null subject, then a tail-head linkage is a case where the antecedent of *deː* 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)

\[
gɔʔ ɗɑʔ ʔoː mok niʔ
\]

so then far oh! mountain this

‘Oh! (it) is far away, this mountain.’

\[
mia de: ɡa: kɪŋ’dɹɪ.m mok da? ʾnɪŋ
\]

when unspecified climb underneath mountain at up_th ere

\[
pa’mɑ:n sa;m si: ɬak niʔ ɬe?
\]

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)

\[
ʔɑn rɑ’wɑːj niʔ tu? pok meʔ ɡəː ɡɔʔ pok mɑh gοn deː tɛːŋ m ̥ əh jɔʔ ɡəː ʔɑm bɨɑn
\]

COND tiger this want bite INDEF 3sgn so_then bite 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.95

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 de?, 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)

\[
gon de? deː tɛːŋ ɱɑh jɔʔ ɡəː ʔam bian
\]

person POSS unspecified do INDEF with 3sgn NEG can

‘…our people could not do anything to it.’

\[
*ɡon n̥ɑm deː tɛːŋ ɱɑh jɔʔ ɡəː ʔam bian
\]

person big unspecified do INDEF with 3sgn NEG can

‘…we big people could not do anything to it.’

---

95 gon de: is less restrictive than gon ?iʔ ‘we (specific group) people’, which includes the speaker and their family or a specific group identified in the discourse context.
Thus it seems that gon deː may be a compound. Kmhm’ speakers think of the construction as two words and it is difficult to come to a definitive answer on this.

In other speech situations, deː 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 deː in this sentence. It is interpreted using cultural knowledge as a first person plural meaning, the family of the speaker, i.e. kʊŋ deː: ‘our village’.

(15) ŋɔːr ruːŋ tɑŋ dɑʔ wiɑŋ rɔːt dɑʔ kuŋ
deː ɡɨˈniʔ
road rough starting at Vientiane arriving at village
deː 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 deː: in this sentence. It is interpreted using the context of the discourse as a first person singular meaning.

(16) haːt ʃoŋ ʔoʔ lɑw sɑh ɨː ʔam laʔ ɨː
and_then father 1sg say COMP IRR NEG good eat
ʔomˈsɑː jɔː r sɑh deː sɨrˈmɑʔ
tea because unspecified have_fever
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 deː. When deː occurs as the final element in a clause, it can mean ‘alone’ or ‘by oneself’⁹⁶. This is seen in (17).

(17) ɡɑt giː ɡəː ɡɔʔ jɔh
turn this_one 3sgm so_then go oneself
deː ‘Then (at) this time he went alone.’

This meaning is made clearer to the audience by the optional insertion of the counter expectation marker hɑk ‘nevertheless’, as shown in (18).

---

⁹⁶ 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”.

---
Then (at) this time he went alone."

'Where the subject is plural, this sense of *de:* can also be used, see (19).

'(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 *go?* ‘so then’, which, when it occurs, always follows the subject and precedes the verb complex. When *de:* appears in this context with *go?*, it always follows *go?* 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.

'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).

'Whatever (he) did (I) didn't get to drink water'.

Negation of *bian de:* indicates that an action was not achieved, as shown in (22), where *ʔam bian de:* *ʔak ʔom* means ‘didn’t get to drink water’.

‘Whatever (he) did (I) didn’t get to drink water.’
Multiple Uses of the Pronoun de: in Kmhmu

$t\text{aŋ de: RESULT}$

When used as a main verb $t\text{aŋ}$ 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 $t\text{aŋ de:}$ in the result clause $g\text{a}: \ t\text{aŋ de: } p\text{lɔ: j} ‘$then it released (him)$’.

(23)
\[
g\text{a}: \ g\text{a}j \ t\text{ɔ:t} \ mi:t \ da? \ guaŋ \ da? \ gi:\
\]
3sgm but then draw out knife at waist at this one
‘But then he drew out a knife at (his) waist here.’

$j\text{ɔh} \ t\text{ɔ:t}: \ g\text{a}: \ t\text{aŋ de: } p\text{lɔ: j}
\]
go approach 3sgn then release
‘(As he) went (and) approached, then (as a result) it released (him).’

The conjunction $g\text{ɔʔ}$ can be inserted after the subject and before $t\text{aŋ de:}$ as shown in (24).

(24)
\[
g\text{a}: \ g\text{ɔʔ} \ t\text{aŋ de: } p\text{lɔ: j}
\]
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)
\[
g\text{a}: \ t\text{aŋ } p\text{lɔ: j}
\]
3sgn then release
‘Then it released (him).’

When $de:$ is omitted, however, the inclusion of $g\text{ɔʔ}$ produces an ungrammatical sentence, as shown in (26).

(26)
\[
* \ ga\text{: } g\text{ɔʔ} \ t\text{aŋ } p\text{lɔ: j}
\]
3sgn so then then release
‘So then it released (him).’

Negation of $t\text{aŋ de:}$ is not possible, probably for semantic reasons.

$p\text{ʰɔː de: ‘almost, just about’}$

When used as a main verb $p\text{ʰɔː:}$ 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) $p\text{ʰɔː: de:}$ is used to add aspectual meaning to the adjectival state verb $b\text{aḥ}$ ‘to be light’.

(27)
\[
\ddot{\text{ʔiʔ}} \ r\text{əh} \ j\text{ɔh} \ da? \ pe? \ η\text{iən} \ b\text{riʔ} \ p\text{ʰɔː: de: } b\text{aḥ}
\]
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)
\[
* \ b\text{riʔ} \ p\text{ʰɔː: } b\text{aḥ}
\]
environment almost light
‘It was almost light.’
Insertion of the conjunction ɡəʔ is grammatical, as shown in (29).

(29)  

\begin{tabular}{llll}
\textit{briʔ} & ɡəʔ & \textit{pʰɔː \text{ de:}} & \textit{bah} \\
\end{tabular} 

environment so then almost light

‘It was almost light.’

Negation of \textit{pʰɔː \text{ de:}} is not possible, perhaps due to semantic constraints.

\textit{mian \text{ de:}}: ‘as if’

When used as a main verb \textit{mian} means ‘to be like, to be the same as’. In combination with \textit{de:} in the verb complex, it signifies an event or state which appears to be true, as shown in (30).

(30)  

\begin{tabular}{llll}
\textit{ur} & \textit{gi:} & \textit{mian \text{ de:}} & \textit{t̥i: \text{ lam}} \\
\end{tabular} 

stew this one as if IRR delicious

‘This stew looks as if (it) will be delicious.’

Omission of \textit{de:} produces a grammatical sentence with no change of meaning, as seen in (31).

(31)  

\begin{tabular}{llll}
\textit{ur} & \textit{gi:} & \textit{mian} & \textit{t̥i: \text{ lam}} \\
\end{tabular} 

stew this one as if IRR delicious

‘This stew looks as if (it) will be delicious.’

The indefinite pronoun \textit{məʔ} ‘someone, something’ can be substituted for \textit{de:} with no change in meaning as shown in (32).

(32)  

\begin{tabular}{llll}
\textit{ur} & \textit{gi:} & \textit{mian \text{ meʔ}} & \textit{t̥i: \text{ lam}} \\
\end{tabular} 

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 \textit{de:} gives some support to the idea that \textit{de:} retains some of its referential functions in this context.

\textit{ʔɔːr jɔʔ \text{ de:}}: ‘together, with one another’

When used as a main verb \textit{ʔɔːr} means ‘to lead, to accompany’ and \textit{jɔʔ} means ‘together’. In combination with \textit{de:} in the verb complex, \textit{ʔɔːr jɔʔ \text{ 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) \textit{ʔɔːr jɔʔ \text{ de:}} precedes the main verb \textit{həj kʰrɑ̃} ‘to laugh at’ and indicates that the other people all joined together to laugh at them.

(33)  

\begin{tabular}{lllll}
\textit{leʔ} & \textit{briaj} & \textit{gɔʔ} & \textit{ʔɔːr jɔʔ \text{ de:}} & \textit{həj kʰraʃ} \text{ si’na:} \\
and other people so then together laugh at 3du \\
\end{tabular} 

‘And other people together laughed at them.’

Omission of \textit{de:} is grammatical and brings no change of meaning, as seen in (34).

(34)  

\begin{tabular}{lllll}
\textit{leʔ} & \textit{briaj} & \textit{gɔʔ} & \textit{ʔɔːr jɔʔ} & \textit{həj kʰraʃ} \text{ si’na:} \\
and other people so then together laugh at 3du \\
\end{tabular} 

‘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’; hoː ʔoʔ lɑw bɔ? ‘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 goː j bɔ? ‘I had never eaten (it)’, de: law bɔ? ‘I said (I) did (want to eat it)’, de: diː m ‘I believed’. When she resumes the events of the story, she reverts to the specific pronoun ʔoʔ ‘1sg’; ʔoʔ goː gleː t gleː t jɔh ‘I licked (the ice-cream), licked (and) went along’.

(35)  
hoː ʔoʔ lɑw bɔ? ˀnɑj de: ʔam goː j bɔ? ˀmoːj  
and_then 1sg say eat but unspecified NEG ever eat not_one  
bat ʔam goː j bɔ? ˀmoːj 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 bɔ? de: diː m sah goː tiː 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 goː gleː t gleː t jɔh hoː ˀnɑj ʔoʔ ˀnaːm jɔh kaːl  
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.’

(36)  
nɑːjmɔː ɡɔʔ ləŋ joŋ ʔoʔ  
doctor so_then scold father 1sg get COMP father 1sg stupid  
‘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’mɑʔ ‘I had a fever…’, de: goː pɔːŋ jɔh laː ‘I still went out to play’, de: sir’mɑʔ ‘I had a fever’, de: məh kɔːn ȵɛʔ ‘I was a small child’. When she returns to the storyline events, she resumes using the specific pronoun ʔoʔ; seh ʔoʔ de? dom ʔiːː ‘(They) put saline into me’.

(36)  
jɔːr sah ȵaːm ɲɔːŋ jɛt daʔ kʊn de: sir’mɑʔ siː giː  
because when yet stay at village unspecified have_fever today  
pim’gi: ʔam sir’mɑʔ sir’mɑʔ beː p ʨiː waːŋ miː  
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.’
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.

“So (because it) was like that, they said (it) was malaria.”

“(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 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.

“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).
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 *briang,* 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 *briang.*

(39)

*de:* unspecified

*goʔ* unspecified

*sirʔeŋ* unspecified

*briang* unspecified

*ʔoh* unspecified

*briang* unspecified

*sirʔeŋ* unspecified

*de:* 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 *ɗeʔ* ‘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


About ʔaɔj in Contemporary Khmer.

Denis Paillard
Laboratoire de linguistique formelle
Université Paris Diderot

“GIVE-verbs” 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.

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) Thaï (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’: $N_1$ ʔaɔj $N_2$ / $N_1$ ʔaɔj $N_2$ $V N_j$

1a) koat ʔaɔj baːj cʰkaɛ maɔŋ pɔnman
3sg ʔaɔj rice dog hour how many
“At what time does he give the dogs their meal?”

1b) koat ʔaɔj baːj cʰkaɛ siː maɔŋ pɔnman
3sg ʔaɔj rice dog eat hour how many
“At what time he will give the dogs their meal?” (they are starving)

---

97 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.
98 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).
As regards work construction, he gives the workers 5000 riels a day” (workers salary)

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 \( \ddot{a}a \ddot{j} \) either alone or combined with another verb (namely \( t\ddot{a}i \) ‘go’ in most cases). With \( \ddot{a}a \ddot{j} \) alone, the Ns stand for arguments (‘agent’, ‘object’, ‘beneficiary’ of \( \ddot{a}a \ddot{j} \) and the interpretation is that of transfer. With \( \ddot{a}a \ddot{j} \) combined with another verb, the Ns coming after \( \ddot{a}a \ddot{j} \) have a different status; those Ns are arguments of the second verb, whereas the preceding N only is an argument of \( \ddot{a}a \ddot{j} \). 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 \( \ddot{a}a \ddot{j} \) expressing the transfer and \( \ddot{a}a \ddot{j} \) as a ‘causative’ is quite blurred. With the transfer value, \( \ddot{a}a \ddot{j} \) 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 \( \ddot{a}a \ddot{j} \). Contrary to \( \ddot{a}a \ddot{j} \), \( cuun \) can combine with the “causative” prefix \( b\ddot{a}n \). This difference can already be found in old Khmer

\textbf{Benefactive}\textsuperscript{99}: \( N_1 \ V (N_2) \ \ddot{a}a \ddot{j} N_3 / \ N_1 \ V (N_2) \ \ddot{a}a \ddot{j} t\ddot{a}i \ N_3 \)

This second use of \( \ddot{a}a \ddot{j} \) 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 \( \ddot{a}a \ddot{j} \) is to point out who benefits by this process.

3) \( s\ddot{a}m \ c\ddot{a}h \ t\ddot{a}laj \ \ddot{a}a \ddot{j} k\ddot{t}nom \)  
ask descend cost \( \ddot{a}a \ddot{j} \text{ s}g \)  
“Reduce the price for me”

4) \( j\ddot{a}k \ c\ddot{a}naj \ t\ddot{a}m \ nuh \ m\ddot{a}k \ \ddot{a}a \ddot{j} \ ma? \ p\ddot{a}\ddot{a} \)  
take pot big dem come \( \ddot{a}a \ddot{j} \text{ mother part.} \)  
“Bring me this big pot” (the mother to her daughter)

5a) \( koat \ t\ddot{a}g \ s\ddot{a}p\ddot{a} \ddot{j} \ (\ddot{a}a \ddot{j} \text{ cuun}) \ m\ddot{a} \ddot{a}j \ koat \)  
3sg buy book \( \ddot{a}a \ddot{j} \text{ cuun mother} \text{ 3sg} \)  
“He bought a book for his mother”

\textsuperscript{99} We take over this terminology from the literature dealing with GIVE-verbs.
5b) koat tɔŋ sıipʰɔi ʔaɔj (cuun) tɔi mdaːj koat
   3sg buy book ʔaɔj cuun go mother 3sg
   “He bought a book, intended for his mother”

The benefactive is the second case where ʔaɔj can commute with cuun. As shown by (5b), ʔaɔj 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 ʔaɔj: buying the book, on the one hand; allocating the book to the mother on the other hand.

**Delegative: N₁ V (N₉) ʔaɔj N₃**

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) kʰɲɔm bık sk laːn ʔaɔj nizjuʔ juːnnesko
   1sg drive car ʔaɔj manager UNESCO
   “I drive the UNESCO manager’s car as his driver”

7a) bαkpraɛ ʔatʰabɔt nih ʔaɔj kʰɲɔm pʰαŋ
   translate text dem. ʔaɔj 1sg part.
   “Translate this text for me!”

7b) kʰɲɔm bαkpraɛ ʔatʰabɔt nih ʔaɔj koat
   1sg translate text dem. ʔaɔj 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 ʔaɔj 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. N₁ kuɜ ʔaɔj V

kuɜ specifies the preceding term as the very best item liable to fill such or such place of argument of the verb coming after kuɜ. Between kuɜ and the V it is possible to insert ʔaɔj.

8) bαhcʰnaɔt lɘːʔ nih kʰmiz naʔnàʔpaʔ naː kuɜ ʔaɔj ca p ʔaːrɑm sɑh
   election time dem. neg. exist party indef. kuɜ ʔaɔj cap ʔaːrɑm sah
   “Regarding these elections, there isn’t one single party worthy of attention”

9a) bαŋ mizmɔp ciːwɛt bαŋ kamsat mɛːn
   older sibling name life older sibling miserable true
   kuɜ ʔaɔj ʔaːnɜt nah
   kuɜ ʔaɔj take pity very
   “Meanup, you have an unhappy life inspiring me compassion”

9b) bαŋ mizmɔp ciːwɛt bαŋ kamsat mɛːn kuɜ ʔaːnɜt nah
   older sibling name life older sibling miserable true kuɜ take pity very
   “Meanup, it is true that you have an unhappy life, one feels compelled to compassion for you (one =

---

100 kuɜ originates in a verb (‘suit’); it keeps this predicative value in some expressions such as kuɜ pom kuɜ (kuɜ neg kuɜ) ‘is it all right or not?’ and kuɜ ʔaɔj ri kuɜ pom ʔaɔj ‘should we give it to him or not?’
The difference between (9a) (presence of ʔaɔj) and (9b) lies in the fact that the relation between the N standing as kua scope and the verb is actualized, whereas with kua 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 kua 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 kʰəŋ mok koat lasŋ kraham kua ʔaɔj kʰlaːc
be angry face 3sg ascend red kua 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 Ø / 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) nsi o sjism (sdap ke: nijisj xsn)
remain o still (listen people speak first)
“Keep still! (listen to the others first)”

11b) nsi ʔaɔj sjism (lec tu:k ʔɛjlej hasu)
remain ʔaɔj still (sink boat at this time PART.)
“Keep still! Or you’ll make the boat sink!”

12a) hæct ʔɛj kaː coːlcit nijisj kʰlaŋ ʔaŋcʃy
reason what PART. like speak loud so
“But why do you enjoy speaking so loudly!”

12b) sɔm nijisj ʔaɔj kʰlaŋ ciʃŋ nsŋ baːn te:
part.politeness speak ʔaɔj loud more deict. can PART.
“Could you speak louder?”

13a) nijisj muʃj muʃj baːn te:
speak one one can PART.
“Could you speak slowly?”

13b) * nijisj ʔaɔj muʃj muʃj baːn te:
speak ʔaɔj one one can PART.

In (13b), the term muʃj muʃj specifying ‘speak is not a predicate (‘slowly’ is the reduplication of the numeral ‘one’), which accounts for the impossibility to have ʔaɔj:

14a) nijisj kʰlej kʰlej tʃi ʔah pe:l hasu
speak short short go finish time PART.
“Make it short we have no time left!”

14b) nijisj ʔaɔj kʰlej tʃi...
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) rɔt ʔaɔj lɨɘn
   run ʔaɔj fast
   “Run fast”

15b) koat rɔt ø (*ʔaɔj) lɨɘn
   3sg run ø *ʔaɔj fast
   “He runs fast” (ʔaɔj is impossible in an assertion)

15c) koat rɔt baːn lɨɘn
   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) jɔk kambɘt nuh tɘɨ ʔaɔj paː samliɛŋ ʔaɔj mʊt tɘɨ
   take knife dem. go ʔaɔj dad sharpen ʔaɔj sharp par t.
   “Bring this knife to your father, for him to sharpen it”

In this use, it is not possible to have the modal negation kɔm unlike what is to be found with the permissive (1.6) and the causative (1.7).

**Permissive: ʔaɔj N₂ V / N₁ʔaɔj N₂ V (N₃)**

Permissive refers to constructions where ʔaɔj is followed by a sequence N₂ V (N₃) where N₂ refers to a human being and V to a process aimed at by N₂, 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 kʰɲɔm tsi pʰaːŋ
    ʔaɔj 1sg go also
    “Let me come with you!”

17b) koat ʔaɔj Daraː tɘɨ
    3sg ʔaɔj Dara go
    “He let Dara come with him”

**Causative: ʔaɔj N₂ V / N₁ʔaɔj N₂ 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 N₁ and the process expressed by the V in the sequence coming after ʔaɔj.

18a) pe:l koat mɔk dal ʔaɔj koat cam kʰɲɔm 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 wɪs cam kʰɲɔm bantec
    3sg ʔaɔj 3sg wait 1sg a little
    “He told him to wait for me for a while”

19) nijiɜj ʔaɔj kʰɲɔm sdap pʰaːŋ
    tell ʔaɔj 1sg listen PART.
    “Tell me about it for me!”
1.8. P1 ʔaɔj P2:  \( N_1 V \ ʔaɔj \ N_2 V \  (Z) / N_1 V N_2  \ ʔaɔj \ V \  (Z) \)

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)  
\[
\text{kʰɲɔm noam koat mɔk } O \text{mɘːl ptʰɛah} \\
\text{1sg lead 3sg come } O \text{ look house} \\
\text{“I bring him along to see the house”: either 1) “for him to see my house and know how it looks like”; or} \\
\text{2) “for him to know where I’m living”.
\]

20b)  
\[
\text{kʰɲɔm noam koat mɔk ʔaɔj mɘːl ptʰɛah} \\
\text{1sg lead 3sg come ʔaɔj look house} \\
\text{“I bring him along to see the house”: either 1) the house is for sale; or 2) someone is needed as a caretaker for the house.
\]

In (20a) (with 0) 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 at a target involving the \( N_2 \) subject of V2, which is confirmed by the impossibility to have ʔaɔj in (21) and, conversely that of 0 in (22a):

21)  
\[
\text{sʔaɛk peːl kʰɲɔm noam kojn } O \text{ (*ʔaɔj)} \\
\text{to morrow time 1sg lead child 0} \\
\text{tɘɨ saːlaː cam kʰɲɔm chist cɔːl len} \\
\text{go school wait 1sg. existence enter play} \\
\text{“To morrow, I’ll avail on the occasion of taking my son to school to come and see you”}
\]

22a)  
\[
\text{kʰɲɔm cis nɛak noam wiʃ ʔaɔj (*0) skoal caowaːj kʰaɛt} \\
\text{1sg. be person lead 3sg. ʔaɔj know chief province} \\
\text{“I’m the one who did introduce him to the province governor”}
\]

22b)  
\[
\text{kʰɲɔm cis nɛak noam ʔaɔj wiʃ skoal caowaːj kʰaɛt} \\
\text{1sg. be person lead ʔaɔj 3sg. know chief province} \\
\text{“(he wanted to be introduced to the province governor and) I am the one who managed a meeting”}
\]

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)  
\[
\text{nɛak naː noam wiʃ ʔaɔj skoal kɑnlaɛŋ nih} \\
\text{person indef. lead 3sg. ʔaɔj know place dem.} \\
\text{“Who made him know this place?” (he wasn’t supposed to know this place)
\]

23b)  
\[
\text{nɛak naː noam ʔaɔj wiʃ skoal kɑnlaɛŋ nih} \\
\text{person indef. lead ʔaɔj 3sg. know place dem.} \\
\text{“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 ʔaɔj 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 ʔaɔj:

24) kʰɲɔm soːmtoh kʰɲɔm noam ʔaɔj baŋ piʔbaʔ daɔjsaʔ kʰɲɔm
   1sg. beg forgiveness 1sg lead ʔaɔj 2sg difficult because 1sg
   “I beg your pardon for getting you into a difficult situation”

25) ʔaɛŋ tʰweː ʔaɲcɜŋ ʔaɛŋ noam ʔaɔj keː mɘːl ŋiɜj dɑl kruɜsaː
   2sg. do so 2sg. get ʔaɔj 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 in the sequence:

26) A mother is complaining about his son who doesn’t want to study:

26a) ʔaː muɛj nɜŋ baɛ kʰɲɔm mɜn bɑŋkʰɑm wiɛ
   anaph. one deict. if 1sg. neg. force 3sg.
   ʔaɔj riʃn teː wiɛ ʔat riʃn teː
   ʔaɔj 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) ʔaː muɛj nɜŋ baɛ kʰɲɔm mɜn bɑŋkʰɑm ʔaɔj
   anaph. one deict. if 1sg. neg. force ʔaɔj
   wiɛ riʃn teː wiɛ ʔat riʃn teː
   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 ʔaɔj emphasizes the interdependence between mother and son as regards studying, whereas in (26b) (subject of P2 coming after ʔaɔj) ‘studying’ is presented as an aim the mother endeavours to reach. It can also be observed that ʔaɔj can hardly be removed.

27a) caowːaʃ sɾɛj ʔaɛŋ mɜn baːn bɑŋkʰoap ʔaɔj ʔaɛŋ
   boss woman 2sg neg. get order ʔaɔj 2sg
   tʰweː teː ʔaɛŋ tʰweː kʰluɜn ʔaɛŋ taɘ
   do part. 2sg do body 2sg restr.
   “Your boss did not tell you to do it, you did it on your own ?!”

27b) caowːaʃ sɾɛj ʔaɛŋ mɜn baːn bɑŋkʰoap ʔaɛŋ ʔaɔj
   boss woman 2sg neg. get order 2sg ʔaɔj
   tʰweː teː koat ʔaɔj neah pʰseːŋ tʰweː taɘ
   do part. 3sg ʔaɔj person else do part.
   “Your boss did not ask you to do it, she asked somebody else to do it”

27c) caowːaʃ sɾɛj kʰɲɔm bɑŋkʰoap mɘdaːj kʰɲɔm ʔaɔj
   boss woman 1sg order mother 1sg ʔaɔj
   kʰɲɔm rʃːk lʊj ʔaɔj koat
   1sg seek money ʔaɔj 3sg
   “My boss ordered my mother to ask me to go and fetch money for her (the boss)”
In (27a) and (27b), the verb *bɑŋkoap* ‘order’ (P1) comes under a negation: with the ante position of N₂ (ex. 27a), the independence of P2 is reinforced: P1 is not the event which brings P2 about. With the postposition of N₂ (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 *bɑŋkoap* ‘order’ and a postposed N, the N object takes over from the agent of *bɑŋkoap* as regards the validation of the process referred to by P2.

When the V of P1 is a strict causative verb, such as *bɑndɑl* ‘cause’, N₂ postposition is only possible: owing to its semantic value, the verb *bɑndɑl* cannot express a relation between the two subjects:

(28)

```
koat kʰmizn tʰwɘː ʔaj dael mizn praːjaʃj dal kʰmḥon teː.
```

3SG. NEG. do INDEF. REL. have interest reach Firm PART.

```
koat bɑndɑl ʔaʃj taj kʰmḥon kʰat laʃj nɘj
```

3SG. cause ʔaʃj 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 *tʰwɘː* (‘do’) as a causative verb and with an inanimate N as the subject of the V of P2.

(29a)

```
tʰwɘː laːn ʔaʃj (*Ø) cʰɛh sen cam tsí
```

make car ʔaʃj *Ø work first wait go

“Fix the car before you leave!” (manage to make it work before leaving)

(29b)

```
neak naː tʰwɘː laːn ʔaʃj (*Ø) cʰɛh
```

person indéf. make car ʔaʃj *Ø work

“Oh who fixed the car?” (I thought there was no one to fix it)

(29c)

```
neak naː tʰwɘː ʔaʃj (*Ø) laːn cʰɛh
```

person indéf. make ʔaʃj *Ø car work

“Oh who made the car work?” (I don’t want it to be put in working order)

(29d)

```
peːl kʰɲɔm tʰwɘː Ø (*ʔaʃj) laːn cʰɛh hajj
```

time 1sg. make Ø *ʔaʃj car work part.

```
kʰɲɔm kːɾˈpɛcm dam baj
```

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 (laːn ‘car) before or after ʔaʃj; b. the presence or not of ʔaʃj between P1 and P2. In (29a) and (29b), laːn must necessarily stand before ʔaʃj: P1 is first. In (29c) laːn must necessarily stand after ʔaʃj: 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 ʔaʃj 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 ʔaʃj 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 ʔaʃj, the event expressed by P1 is prominent (the autonomy of P2 is quite relative). When N2 comes after ʔaʃj, 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: ʔaʃj+tæc p and kʰm+ʔaʃj+tæc p

In this case ʔaʃj is at the initial position\(^{101}\) and is followed by tæc (‘only). ʔaʃj combined with tæc, 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,

\(^{101}\) It can be preceded by the modal negation kʰm; 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) wiɛ sɔkcɛt twɛ: ta:m ʔaŋŋ twajʔah ə ʔaɔj tae ʔaŋŋ niʃiʃ ʃəd: nʃŋ wiz
3sg. agree do follow 2sg. all ə ʔaɔj tae 2sg. speak well with 3sg.
“He agrees to do anything you want, on the only condition that you speak kindly to him”

31) ʔaɔj tae ba:n twi ʔaɔj kʰɔm twɛ: ʔɛj kə: kʰɔm twɛ: dæʔ ʔaɔj tae 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 fulfill it”

32) kʰmeːŋ nʃŋ sralaŋ ʔow wiɛ ʔaɔj tae wiɛ
child Deict. love father 3sg. ʔaɔj tae 3sg.
ʔaɔj taɛ baːn tɘɨ ʔaɔj kʰɲɔm twɘː ʔɛj kɑː kʰɲɔm twɘː dæʔ
“On the only condition that I’m free to go away, I can be asked anything, I will not refuse to fulfill it”

33) wiɛ sɔkcɛt twɛ: ta:m ʔaŋŋ twajʔah kɔm
3sg. agree do following 2sg. everything kɔm
ʔaɔj tae ʔaŋŋ niʃiʃ kə:kraʔ nʃŋ wiz
“He agrees to do anything you want on the only condition that you do not speak to him nastily”

34) wiɛ krɔantɛ jɔːk ʔaŋŋ twi kʰɔm mən ʔɛj tɛː kɔm ʔaɔj
3sg. only take 2sg. go put in jail neg. indéf part. kɔm ʔaɔj
taɛ wiɛ waj təɛmtiɜt nʊh tok ciʃ səmnaːŋ tɛː
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 tae or kɔm ʔaɔj tae can be found:

34a) kʰmeːŋ nʃŋ mən dəŋ ciʃ jaːŋmeːc wiɛ ʔaɔj
child deict neg know be how 3sg ʔaɔj
taɛ wiɛ kʰɔŋ ʔow wiɛ wiɛ jɔm pʰliːm
“I don’t know what is the matter with this child: as soon as he sees his father, he starts crying”

34b) kʰmeːŋ nʃŋ mən dəŋ ciʃ jaːŋmeːc wiɛ kɔm ʔaɔj
child deict neg know be how 3sg kɔm ʔaɔj
taɛ wiɛ kʰɔŋ ʔow wiɛ wiɛ jɔm pʰliːm
“I don’t know what is the matter with this child: seeing his father is enough to start crying right away”

With ʔaɔj tae, 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, kəm 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.

\[
\begin{align*}
35) \quad & \text{mun nɜŋ tʰa: (*sdej) ʔaɔj ke: məl kʰluam ʔaɔj səm} \\
              & \text{before part. tʰa: *sdej ʔaɔj people look oneself 2sg first}
              \\
\text{“Before criticizing the others, just look at yourself!” (Someone criticizes the people around him about what they wear)}
\end{align*}
\]

\[
\begin{align*}
36a) \quad & \text{baʃ wiz mɔ:k jɨt kəm tʰa: ʔaɔj wiz} \\
              & \text{if 3sg come late neg.mod. tʰa: ʔaɔj 3sg}
              \\
\text{“If he’s late, don’t be critical!”}
\end{align*}
\]

\[
\begin{align*}
36b) \quad & \text{baʃ wiz mɔ:k jɨt kəm sdej ʔaɔj wiz} \\
              & \text{if 3sg come late neg.mod. sdej ʔaɔj 3sg}
              \\
\text{“If he’s late, don’t rebuke him!”}
\end{align*}
\]

37) mother asks her husband not to rebuke their son for being out every night

\[
\begin{align*}
37a) \quad & \text{peːl wiz mək kəm sdej ʔaɔj wiz cam} \\
              & \text{time 3sg. come neg.mod. sdej ʔaɔj 3sg. wait}
              \\
\text{ʔaɛk nijɛj pənjɔl wiz təm samruɛj tʃi tomorrow speak explain 3sg. following quiet go}
              \\
\text{“When he comes back, don’t shout at him! Wait until tomorrow to have a quiet explanation with him!”}
\end{align*}
\]

\[
\begin{align*}
37b) \quad & \text{peːl wiz mək kəm tʰa: ʔaɔj wiz cam} \\
              & \text{time 3sg. come neg.mod. tʰa: ʔaɔj 3sg. wait}
              \\
\text{ʔaɛk nijɛj pənjɔl wiz təm samruɛj tʃi tomorrow speak explain 3sg. following quiet go}
              \\
\text{“When he comes back, don’t criticize him! Wait until tomorrow to have a quiet explanation with him!”}
\end{align*}
\]

38) baʃ wiz mɔ:k jɨt kəm tʰa: (*sdej) ʔɛj ʔaɔj wiz

\[
\begin{align*}
\text{if 3sg come late neg.mod. tʰa: *sdej indef ʔaɔj 3sg}
\text{“If he’s late, don’t tell him anything!”}
\end{align*}
\]

39) ʔaɛŋ sdej (*tʰa:) ʔaɔj koat rɨɜŋ ʔɛj pə tʃi

\[
\begin{align*}
\text{2sg. sdej tʰa: ʔaɔj 3sg. matter indef. be true go}
\text{“What are you rebuking him about exactly?”}
\end{align*}
\]

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 ʔɛj refers to the class of possible words, and tʰa: only is possible; whereas in (39) rɨɜŋ ‘story’ refers to an event (which has to be identified) and sdej only is possible\(^{102}\).

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

\[^{102}\text{This difference has to be backed out by a systematic study of these two verbs} \]
night") gets N₁ to say something about it: a reproachable behaviour leads to N₁ seizing verbal power on N₂; 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 ʔaɔj 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 ʔaɔj. This presentation has also made it possible to bring out a characteristic feature at work in all the uses of ʔaɔj: the partial autonomy of the sequence coming after ʔaɔj. As a rule, this autonomy of the sequence following ʔaɔj comes together with the - possible or necessary - presence in this sequence of a V different from ʔaɔj. 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 ʔaɔj, it should be noted that this is the case with the following uses: benefactive, delegative, jussive, P₁ ʔaɔj P₂ and criticize/rebuke.

We hereafter resume the ten types of uses in reference to the autonomy of the sequence coming after ʔaɔj.

**Transfer** is characterized by the possibility to introduce a verb after ʔaɔj. The presence of a V (normally tɘɨ ‘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 ʔaɔj. 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 ʔaɔj 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 ʔaɔj. The presence of tɘɨ on the right of ʔaɔj reinforces the subordination of the process achieved by S₁ to the interest of S₂.

**Permissive** and **causative**: the compulsory presence of a verb after ʔaɔj means that the subject-agent of ʔaɔj is taken into account as it allows the event referred to by the V to be possible or achieved.

**P₁ ʔaɔj P₂** is characterized by the compulsory presence of a V both on the left (P₁) and on the right (P₂) of ʔaɔj. 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 ʔaɔj, the event referred to by P₁ gets its own autonomy as an event.

**Jussive and kuɛ**: alternative Ø / ʔaɔj. The mere possibility of a construction without ʔaɔj means that the event is referred to mainly by the verb other than ʔaɔj; This V is on the left in the jussive case, on the right in the kuɛ case. In the jussive case, the predicate on the right of ʔaɔj 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 ʔaɔj 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 ʔaɔj. 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 ʔaɔj: the only thing which matters is the (actual or aimed at) achievement of the only event referred to by the sequence coming after ʔaɔj.

**Criticize – rebuke**: the impossibility to have a V coming after ʔaɔj lies on a mechanism comparable to that described in the delegative case, but for one difference: the V which is not explicit on the right of ʔaɔj is present in the left context and not directly in the sequence before ʔaɔj. 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 ʔaɔj preceded by the verbs ʔa: and sdesk, 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 ʔaɔj as a subject of whom is said that he is responsible for a process.


As indicated before, it is not possible to account for these ten uses of ʔaɔj basing on a central value called transfer. It appears that ʔaɔj 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 ʔaɔj E₂

The relation thus established between E₁ and E₂ gives ʔaɔj a causative dimension, but this relation should not be reduced to a mere causal relation. In 2. the autonomy of the sequence coming after ʔaɔj 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 ʔaɔj can therefore be schematized as follows:

\[ S_1 \ p_1 \ (E_1) \ ʔaɔj \ S_2 \ p_2 \ (E_2). \]

This semantic function is at work in all the uses of ʔaɔj. All these uses correspond to various modes of realization coming from the specific units embodying the sequences S₁, p₁ and S₂, p₂. The sequence corresponding to the use P₁ ʔaɔj  P₂ gives the largest extension: S₁, p₁ and S₂, p₂ correspond each to a clause formed on the pattern N V (XY). The transfer value corresponds to the minimal extension, ʔaɔj being the only verb, S₁, p₁ and S₂, p₂ reducing to Ns.

From this viewpoint, describing such or such use of ʔaɔj consists in interpreting the different constituents of the utterance as framed by the S₁, p₁ ʔaɔj S₂, p₂ pattern. We hereafter present a table showing the organization of the ten uses of ʔaɔj with an indication of the element of the abstract pattern realized by each constituent. When an element of the abstract pattern is not realized, we use the Ø 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.

<table>
<thead>
<tr>
<th>Uses</th>
<th>S₁</th>
<th>p₁</th>
<th>ʔaɔj</th>
<th>S₂</th>
<th>p₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer</td>
<td>N</td>
<td>Ø</td>
<td>ʔaɔj</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>Ø</td>
<td>ʔaɔj</td>
<td>N</td>
<td>N</td>
<td>V + N</td>
</tr>
<tr>
<td>Benefactive</td>
<td>N</td>
<td>V + N</td>
<td>ʔaɔj</td>
<td>Ø</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>V + N</td>
<td>ʔaɔj</td>
<td>Ø</td>
<td>V + N</td>
<td></td>
</tr>
<tr>
<td>Delegative</td>
<td>N</td>
<td>V + (N)</td>
<td>ʔaɔj</td>
<td>N</td>
<td>Ø</td>
</tr>
<tr>
<td>kuɜ</td>
<td>N</td>
<td>kuɜ</td>
<td>ʔaɔj</td>
<td>(N)</td>
<td>V</td>
</tr>
<tr>
<td>Jussive</td>
<td>the addressee</td>
<td>V</td>
<td>ʔaɔj</td>
<td>Ø</td>
<td>PRED</td>
</tr>
<tr>
<td>Permissive</td>
<td>(N)</td>
<td>Ø</td>
<td>ʔaɔj</td>
<td>N</td>
<td>V</td>
</tr>
<tr>
<td>Causative</td>
<td>(N)</td>
<td>Ø</td>
<td>ʔaɔj</td>
<td>N</td>
<td>V</td>
</tr>
<tr>
<td>P₁ ʔaɔj P₂</td>
<td>N</td>
<td>V</td>
<td>ʔaɔj ʔaɔj</td>
<td>N</td>
<td>V</td>
</tr>
<tr>
<td>N</td>
<td>V + N (= S₂)</td>
<td>ʔaɔj</td>
<td>Ø</td>
<td>V</td>
<td></td>
</tr>
</tbody>
</table>

103 Keeping ʔaɔj 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 S₁, p₁ (E₁) and S₂, p₂ (E₂).

104 In this table, we only mention the constituents in direct relation with the abstract pattern S₁, p₁ (E₁) ʔaɔj S₂, p₂ (E₂).
Conclusion

This approach of ʔaɔj 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 ʔaɔj in its various uses.

The semantic identity of ʔaɔj 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 ʔaɔj 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

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEICT</td>
<td>deictic or demonstrative</td>
</tr>
<tr>
<td>RELAT</td>
<td>relative</td>
</tr>
<tr>
<td>INDEF</td>
<td>indefinite</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>1SG</td>
<td>1st person of singular personal pronoun</td>
</tr>
<tr>
<td>2SG</td>
<td>2nd person of singular personal pronoun</td>
</tr>
<tr>
<td>3SG</td>
<td>3rd person of singular personal pronoun</td>
</tr>
</tbody>
</table>

References

Gorgoniev, Iou (1966) *Grammatika khmerskogo jazyka*, Moskva
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)

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.

Table 1: Eight stages of language endangerment according to Fishman’s GIDS
(Adapted from Suwilai Premsrirat, 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 “Auut”. Fewer than 50 speakers are found in Srissawat 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.
### 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).

---

<table>
<thead>
<tr>
<th>Number</th>
<th>CHONG</th>
<th>CHUNG</th>
<th>KASONG</th>
<th>SAMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>mq̃j</td>
<td>mj̃j</td>
<td>môj</td>
<td>môj</td>
</tr>
<tr>
<td>Two</td>
<td>phâj</td>
<td>prâ:</td>
<td>pâ:</td>
<td>pâ:</td>
</tr>
<tr>
<td>Three</td>
<td>phèw</td>
<td>phèw</td>
<td>phè:</td>
<td>phè:</td>
</tr>
<tr>
<td>Four</td>
<td>phôn</td>
<td>phôn</td>
<td>phôn</td>
<td>pʰe:n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kinship Terms</th>
<th>CHONG</th>
<th>CHUNG</th>
<th>KASONG</th>
<th>SAMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>mq:</td>
<td>mq:</td>
<td>mîp</td>
<td>mîp?mê:?</td>
</tr>
<tr>
<td>Father</td>
<td>?âp</td>
<td>?âp</td>
<td>kʰâp</td>
<td>kʰâp</td>
</tr>
<tr>
<td>Husband</td>
<td>kâł̃ŋ</td>
<td>kʰsłoŋ</td>
<td>kł̃ŋ</td>
<td>kł̃ŋ</td>
</tr>
<tr>
<td>Offspring (children)</td>
<td>kʰe:n</td>
<td>kʰe:n</td>
<td>kʰe:n</td>
<td>kʰe:n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natural phenomena</th>
<th>CHONG</th>
<th>CHUNG</th>
<th>KASONG</th>
<th>SAMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>tʰək</td>
<td>tʰək</td>
<td>tək</td>
<td>tək</td>
</tr>
<tr>
<td>Fire</td>
<td>pʰəw</td>
<td>pləw</td>
<td>pləw</td>
<td>pləw</td>
</tr>
<tr>
<td>Mountain</td>
<td>kʰnəŋ</td>
<td>nəŋ</td>
<td>nəŋ</td>
<td>nəŋ</td>
</tr>
<tr>
<td>Rain</td>
<td>kəmâ:</td>
<td>kəmâ:</td>
<td>kamâ:</td>
<td>kamâ:</td>
</tr>
<tr>
<td>Sun</td>
<td>tʰəŋiʔ</td>
<td>tʰəŋiʔ</td>
<td>tʰəŋiʔ</td>
<td>sənə</td>
</tr>
</tbody>
</table>

---

`baːn luŋ haŋ tɔː pʰaːj se:R1`

house Uncle: Hang sells twenty)

‘LungHang’s shop sells (it) for twenty baht.’

`kʰɔː R3 naj meːnəm ʔiːn R1 raːj R3 tuː`

crocodile in river have ten clas.

‘There are ten crocodiles in the river.’

`pʰəj R3 cʰalaː:t kwaː dac:R1 əna`

it clever than they fp.

‘It is smarter than the others.’

`muː pʰəj R1 cʰə:p wiwuːt sa:R1`

they like argue each other

‘They like to argue with each other.’

`pʰən kʰaːkʰaːj məŋ R1 sa:R1`

friend trade together

‘They do business together.’
muː pʰǝj R1 si: haː kʰon
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.

pʰǝj R1 ꜫiŋ R1 hoːc R1 ꜫiŋ R1
it not die die not
‘It does not die.’

Examples above illustrate negative constructions in Chong and Chung that are different from Thai but similar to Khmer. This confirms the settlement of Pearsic 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.

<table>
<thead>
<tr>
<th>English Gloss</th>
<th>KASONG</th>
<th>Alternative</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger and elder sibling</td>
<td>ꜫoːt ꜫхиɲ ꜫхиɲ ꜫoːt ꜫ phéː nɔŋ</td>
<td>ꜫхиɲ ꜫoːt ꜫхиɲ ꜫoːt ꜫ phéː nɔŋ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>younger elder (sibling) elder younger (sibling) elder younger (sibling)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Husband and wife</td>
<td>ꜫaŋɪ ꜫkʰɛŋ ꜫkʰɛŋ ꜫaŋɪ ꜫ phuː miː</td>
<td>ꜫkʰɛŋ ꜫaŋɪ ꜫphuː miː</td>
<td></td>
</tr>
<tr>
<td></td>
<td>wife husband husband wife husband wife</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wife and children</td>
<td>ꜫaŋɪ ꜫkʰɛn ꜫkʰɛn ꜫaŋɪ ꜫ lu ꜫɛː ꜫmɛː ꜫnɔː ꜫŋ</td>
<td>ꜫkʰɛn ꜫaŋɪ ꜫlu ꜫɛː ꜫmɛː ꜫnɔː ꜫŋ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>wife child child wife child wife</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English Gloss</th>
<th>CHONG</th>
<th>Alternative</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face and eyes</td>
<td>ꜫoːr ꜫŋaːj ꜫŋaːj ꜫoːr ꜫ nː tː ꜫ ꜫnː tː</td>
<td>ꜫŋaːj ꜫoːr ꜫŋaːj ꜫoːr ꜫ nː tː</td>
<td></td>
</tr>
<tr>
<td></td>
<td>eye face</td>
<td>face eye</td>
<td></td>
</tr>
<tr>
<td>Younger and elder sibling</td>
<td>ꜫoːr ꜫlɪŋ ꜫlɪŋ ꜫoːr ꜫ ꜫpʰɔ ꜫmɛː ꜫnɔː ꜫŋ</td>
<td>ꜫlɪŋ ꜫoːr ꜫlɪŋ ꜫoːr ꜫ ꜫpʰɔ ꜫmɛː ꜫnɔː ꜫŋ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>younger elder (sibling) elder younger (sibling) elder younger (sibling)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother and father</td>
<td>ꜫmɛː ꜫŋaːj ꜫŋaːj ꜫmɛː ꜫ ꜫpʰɔ ꜫmɛː ꜫnɔː ꜫŋ</td>
<td>ꜫŋaːj ꜫmɛː ꜫŋaːj ꜫmɛː ꜫpʰɔ ꜫmɛː ꜫnɔː ꜫŋ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mother father</td>
<td>father mother</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Affixes</th>
<th>Function</th>
<th>Example</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pa-/</td>
<td>Locative</td>
<td><code>dɨŋ</code> R1 ‘on’</td>
<td><code>pədɨŋ</code> R1 ‘above’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>muːn</code> R1 ‘behind’</td>
<td><code>pəmuːn</code> R1 ‘behind’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>reʔ</code> R1 ‘in’</td>
<td><code>pəreʔ</code> R1 ‘inside’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>tʰɛːw</code> R2 ‘other’</td>
<td><code>pətʰɛːw</code> R2 ‘elsewhere’</td>
</tr>
<tr>
<td>/ma-/</td>
<td>Causative</td>
<td><code>həc</code> R1 ‘die’</td>
<td><code>məhəc</code> R1 ‘to kill’</td>
</tr>
</tbody>
</table>

#### KASONG

<table>
<thead>
<tr>
<th>Affixes</th>
<th>Function</th>
<th>Example</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>/-n/-</td>
<td>Instrumental</td>
<td><code>khêt</code> ‘to comb’</td>
<td><code>kənɛt</code> ‘comb’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>kɛw</code> ‘to harvest’</td>
<td><code>kənɛw</code> ‘sickle’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>pək</code> ‘to wrap’</td>
<td><code>pənək</code> ‘package’</td>
</tr>
<tr>
<td></td>
<td></td>
<td><code>kək</code> ‘to carry on the shoulder’</td>
<td><code>kənək</code> ‘shoulder pole’</td>
</tr>
</tbody>
</table>

*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.

<table>
<thead>
<tr>
<th></th>
<th>R1 (Mid clear)</th>
<th>R2 (High creaky / Glottal constriction)</th>
<th>R3 (Low breathy)</th>
<th>R4 (Low breathy followed by high glottal constriction)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chong</strong></td>
<td>kətak ‘peanut’</td>
<td>kətak ‘tongue’</td>
<td>tʰak ‘water’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kləŋ ‘bone’</td>
<td>kwələŋ ‘husband, male’</td>
<td></td>
<td>kələŋ ‘temporary wood bridge’</td>
</tr>
<tr>
<td><strong>Chung</strong></td>
<td>tak ‘bean, peanut’</td>
<td>tək ‘tongue’</td>
<td>tʰak ‘water’</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mələŋ ‘eel’</td>
</tr>
<tr>
<td><strong>Kasong</strong></td>
<td>R1 (mid)</td>
<td>R2 (high-falling)</td>
<td>R3 (operated low)</td>
<td>R4 (operated, mid, high falling)</td>
</tr>
<tr>
<td></td>
<td>tak ‘big’</td>
<td>katək ‘tongue’</td>
<td>tək ‘water’</td>
<td>tək ‘wet’</td>
</tr>
<tr>
<td></td>
<td>kləŋ ‘bone’</td>
<td>kləŋ ‘husband’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kənə ‘green frog’</td>
<td>kənə ‘long’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lələ ‘rotten worm’</td>
<td>pək ‘three’</td>
<td>pək ‘water’</td>
<td>pək ‘watch’</td>
</tr>
<tr>
<td></td>
<td>ɛkə ‘dog’</td>
<td>ɛkə ‘sour’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Suwilai Premsrirat

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.

<table>
<thead>
<tr>
<th>Survey Pearic languages and Stage of endangerment</th>
<th>CHONG</th>
<th>CHUNG (Sa-oc)</th>
<th>KASONG</th>
<th>SAMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stage 8</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

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 Chung 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
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

<table>
<thead>
<tr>
<th>Language Activities</th>
<th>Revitalization</th>
<th>Pearic Languages and Stage of Endangerment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CHONG (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHUNG (Sa-oc) (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KASONG (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAMRE (8)</td>
</tr>
<tr>
<td>1. Community study / Language Situation survey / language vitality / Attitude Survey</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>2. Orthography Development</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>3. Literature/book in Vernacular Language</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>5. Learning-Teaching-Reading Materials</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td>- Primer</td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>- TPR lessons</td>
<td></td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>6. Community Learning Center (outside school)</td>
<td>✓</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>--</td>
</tr>
</tbody>
</table>

### 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


Appendix

Photograph 1-2: Pa Cin and Nai Chian, fluent speakers of Northern Chong dialect

Photograph 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).

Photograph 5-6: The Chong speakers working with professional linguists (Left). The last three speakers of the Eastern Chong Dialect (Right)

Photograph 7-8: The well-known Kamnan Chern, the first Chong project leader (Left) The middle age key persons of Chong revitalization program (Right).
Photograph 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).

Photograph 11-12: Chung Speakers in Cambodia (Left). Chung Speakers in Thailand (Right).

Photograph 13-14: Thai-based Chong language development

Photograph 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 Singhbum 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.
2 Consonants

The phonemic consonant inventory of Ho is shown in table 1. 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.

<table>
<thead>
<tr>
<th>Stops</th>
<th>Labial</th>
<th>Dental</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fricatives</td>
<td>p, b</td>
<td>t, d</td>
<td>ʈ, ɖ</td>
<td>c, j</td>
<td>k, g</td>
<td>?</td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
<td>n</td>
<td>ɲ</td>
<td>&lt;ŋ</td>
<td>η</td>
<td>h</td>
</tr>
<tr>
<td>Flaps</td>
<td>r</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>l</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glides</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 /ʈ/ and /ɖ/ to measure (retroflex stops only appear in borrowed words). Additionally, I omitted tokens of /ɕ/ and /j/ because they are phonetically too fricative-like to measure VOT.

<table>
<thead>
<tr>
<th>stop (no. of tokens)</th>
<th>average VOT time (ms)</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>k (10)</td>
<td>49</td>
<td>38 – 64</td>
</tr>
<tr>
<td>p (6)</td>
<td>30</td>
<td>17 – 43</td>
</tr>
<tr>
<td>t (8)</td>
<td>25</td>
<td>14 – 23</td>
</tr>
<tr>
<td>g (10)</td>
<td>-68</td>
<td>-46 – -48</td>
</tr>
<tr>
<td>b (10)</td>
<td>-89</td>
<td>-54 – -119</td>
</tr>
<tr>
<td>d (10)</td>
<td>-78</td>
<td>-47 – -112</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>p : b</th>
<th>/puraʔ/</th>
<th>‘much, many’</th>
<th>/buraʔ/</th>
<th>‘to draw or ladle out’</th>
</tr>
</thead>
<tbody>
<tr>
<td>/capa/</td>
<td>‘draw picture’</td>
<td>/caba/</td>
<td>‘finish’</td>
<td></td>
</tr>
</tbody>
</table>
We can also see contrast between the palatal and velar stops, as the following words show:

\[
\begin{align*}
\text{c : k} & \quad /\text{cipin}/ \quad \text{‘how much, many’} \quad /\text{kimin}/ \quad \text{‘the wife of one’s son’} \\
& \quad /\text{-ici}/ \quad \text{‘caus suffix’} \quad /\text{ikir}/ \quad \text{‘deep’} \\
\text{j : g} & \quad /\text{jom}/ \quad \text{‘eat’} \quad /\text{gom}/ \quad \text{‘wheat’} \\
& \quad /\text{sajaw}/ \quad \text{‘decorate’} \quad /\text{saga}/ \quad \text{‘type of grass seed’}
\end{align*}
\]

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 \([t]\), and \([t]\) and \([c]\), so that /cimip/ ‘how many’ may also be pronounced /timmip/. Similarly, ‘how do you say’ may be said as /tilekepe kadije/ or /cilekepe kajje/ (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.

\[
\begin{align*}
\text{t : \text{ʈʈ \ʈʈ}} & \quad /\text{mata}/ \quad \text{‘ripe’} \quad /\text{matα}/ \quad \text{‘whey’} \\
& \quad /\text{tuʔ}/ \quad \text{‘mulberry’} \quad /\text{tu}/ \quad \text{‘squirrel’} \\
\text{d : \text{ɖɖ \ɖɖ}} & \quad /\text{duku}/ \quad \text{‘sorrow’} \quad /\text{duki}/ \quad \text{‘urine’} \\
& \quad /\text{didi}/ \quad \text{‘a vulture’} \quad /\text{dzi}/ \quad \text{‘to stand on tiptoes’}
\end{align*}
\]

There is however, free variation between /t/ and /ʈʈ/ and /d/ and /ɖɖ/ in some words, as in the following examples.
Aspects of Ho Phonetics and Phonology

157

(1) \[\text{[tai]} \sim \text{[ʈai]} \ 'stay, remain'\]
\[\text{[data]} \sim \text{[ɖata]} \ 'tooth'\]
\[\text{[dandɪ]} \sim \text{[ɖandɪ]} \ 'a small handle, connecting device'\]

There are minimal pairs establishing the phonemic status of /ʈ/ and /ɽ/.

\[\text{ʈʈ ʈʈ : ɽɽ ɽɽ}\]

/buʈa/ ‘tree trunk’
/jutɪ/ ‘defile’
/goʈa/ ‘whole’

/burʈa/ ‘old man’
/juri/ ‘friend, companion’
/goɾa/ ‘ground, land’

Again, Anderson et al. (2008:202) note that it is also possible to get some free variation between /ʈ/ and /ɽ/, and give the following examples from Deeney (2005).

(2) \[\text{[peɽe]} \sim \text{[peʈe]} \ 'pluck twig or small branch with one or both hands'\]
\[\text{[kaːʈɔb]} \sim \text{[kaːʈɔb]} \ 'crab'\]
\[\text{[pora]} \sim \text{[pora]} \ 'intestines' \ (Deeney 2005)\]

The fact that /ʈ/ can vary freely with both /t/ and /ɽ/ in some instances leads Anderson et al. (2008) to question the phonemic status of /ʈ/. 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 intervocally than initially. In fact, there is only one instance of variation initially ([tai] ~ [ʈai] ‘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, ḏ, ɖ\} to be prenasalized, e.g., /bandor/ \(\rightarrow\) [ʰbandor] ‘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 [ʰmanda] ‘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 [ʰmanda] ‘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.

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>j</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>prenasalized</td>
<td>91% (247)</td>
<td>78% (134)</td>
<td>90% (165)</td>
<td>58% (107)</td>
<td>77% (20)</td>
</tr>
<tr>
<td>not prenasalized</td>
<td>9% (23)</td>
<td>22% (38)</td>
<td>10% (19)</td>
<td>42% (78)</td>
<td>23% (6)</td>
</tr>
<tr>
<td>total no. of tokens</td>
<td>279</td>
<td>172</td>
<td>184</td>
<td>185</td>
<td>26</td>
</tr>
</tbody>
</table>

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 /ɟ/ 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 prenasalized 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)  \[ jibon \] \sim \[ jiβon \] ‘life’
\[ babata \] \sim \[ baβata \] ‘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’, /suit/ ‘suit, salwar kameez’ and /biskut/ ‘biscuit’, /kop/ ‘cup’ (from English) and /camac/ ‘spoon’ (from Hindi). I have no examples of word final /ʈ/.

Voiced stops /b/ and /ɖ/ 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 [la’d̚] ‘bread’
/porob/ \→ [por̚b̚] or [po’ro’b̚n] ‘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/, /ɡ/ and /ɟ/ 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.106 The glottal stop is always followed by an echo vowel which has the same quality as the preceding vowel.

(5) \(/da\tilde{\text{i}}/ \rightarrow [da\text{\text`a}]\) ‘water’
\(/\text{se}ta\tilde{\text{i}}/ \rightarrow [se\text{\text`a}]\) ‘morning’

I represent these words phonemically with the glottal stop, e.g., /da\tilde{\text{i}}/ ‘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) /\text{ij}u/ ‘shout, call out’ \(/\text{ij}u\tilde{\text{i}}/\) ‘fall from a height’
/\text{j}o:/ ‘fruit’ \(/\text{j}o\tilde{\text{i}}/\) ‘sweep’

As in the examples with the nasal release, the echo vowel can be heard in the citation form of polysyllabic words, such as /se\text{\text`a}/ ‘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 (/\text{j}inis/ ‘thing’ being an exception).

(7) /sarkam/ ‘leaf’
/\text{ra}s/ ‘juice’

\(/\text{h}/\) can only appear in onset position.

(8) /\text{hi}si/ ‘twenty’
/\text{he}nde/ ‘black’

In some words, it seems that [s] is in free variation with [ʃ], e.g., [\text{san}ɖi] or [ʃanɖi] ‘rooster’, [hõʔoso] or [hõʔoʃo] ‘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.

<table>
<thead>
<tr>
<th>m</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>/\text{mana}/</td>
<td>‘forbid’</td>
</tr>
<tr>
<td>/\text{hambal}/</td>
<td>‘heavy’</td>
</tr>
<tr>
<td>/\text{nimin}/</td>
<td>‘contents’</td>
</tr>
<tr>
<td>/\text{ikum}/</td>
<td>‘kneel’</td>
</tr>
</tbody>
</table>

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 intervocally 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) [\text{\text`ma}ŋdi] ‘food’
[\text{\text`gaŋti}] ‘small bell’

106In 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 intervocally. 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/.

\[
\begin{align*}
\text{\['siŋgi\] } & \quad \text{‘sun’} \\
\text{\['taŋku\] } & \quad \text{‘pit (of fruit)’}
\end{align*}
\]

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].

\[
\begin{array}{|l|l|l|}
\hline
| n & \text{ɲ} & \text{ŋ} \\
|---|---|---|
| [isin] & ‘cook’ & [tisip] & ‘today’ \\
| [heben] & ‘bitter’ & [setep] & ‘spring, fountain’ \\
| [susun] & ‘dance’ & [holoɲ] & ‘tumeric’ \\
\hline
\end{array}
\]

\textbf{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\}:

\[
\begin{array}{|l|l|l|}
\hline
| \text{ɲ} & \text{ŋ} & \\
|---|---|---|
| [beteɲ] & ‘wait’ & [unuɲ] & ‘play’ \\
| [seteɲ] & ‘spring, fountain’ & [roɲ] & ‘color’ \\
\hline
\end{array}
\]

\textbf{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].

\[
\begin{align*}
\text{[apuŋ] } & \quad \text{‘father - vocative form’} \\
\text{[apuɲ] } & \quad \text{‘my father’} \\
\text{[run] } & \quad \text{‘to husk’} \\
\text{[ruɲ] } & \quad \text{‘sensation of having a limb asleep’}
\end{align*}
\]

They claim that this is evidence of phonologically “front vs. back [u]” (Anderson et al. 2008:202). I measured the formants of [u] for [run] ‘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 [run] 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]:

\[
\begin{align*}
\text{[an] } & \quad \text{‘dawn’} \\
\text{[aɲ] } & \quad \text{‘I’}
\end{align*}
\]
Anderson et al. record [germoɲ] ‘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., [tisɨŋ] ‘today’ rather than [tisŋ] (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 [n] 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, [iminan] ‘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], [ɳ], 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 /ŋ/ phoneme is a flap. Both /l/ and /r/ can appear in all positions and they are contrastive.

(13) /kamal/ ‘lotus’ /kamar/ ‘blacksmith’
    /lo:/ ‘burn’ /ro:/ ‘dry’
    /jalom/ ‘net’ /jarom/ ‘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 /ʃ/, 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 /ʃ/. It only seems to occur after the back vowels {u, o} and before /a/, as in (2.5):

(15) /towə/ ‘milk’
    /guwa/ ‘betel nut’

The palatal glide /ʃ/ 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/), /hōjaʃ/ ‘father-in-law’.

(16) /tuju/ ‘jackal’
    /tajom/ ‘after’
    /hojo/ ‘air’
    /dijan/ ‘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’ /palți/ ‘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 /j/ 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/j variation in /jaron/ ‘egg’ are spoken within phrases of each other, by the same person.

(19) /..mendo ażu-qe-a binj jaronom.
...but bring-T/A-TR-FIN snake egg.

Endo binj jaronom-aʔ agu-le-q-redo.../
So snack egg-POS 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) [rulbın] ~ [nurbın] ‘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) [lapan-lapan] ~ [langab-langab] ‘out of breath’
[siriţi] ~ [simiɾi] ‘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 [bepeta] ~ [tepeba] ‘meet’ from /beɾa/ ‘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.

<table>
<thead>
<tr>
<th>/mi Çalış/</th>
<th>‘sweetmeats’</th>
<th>/ci Cic/</th>
<th>‘a letter’</th>
<th>/bita/</th>
<th>‘length between the tip of the thumb and the tip of one finger’</th>
</tr>
</thead>
<tbody>
<tr>
<td>/meta/</td>
<td>‘to say to’</td>
<td>/cetan/</td>
<td>‘above’</td>
<td>/beğa/</td>
<td>‘to arrive, reach’</td>
</tr>
<tr>
<td>/mata/</td>
<td>‘to ripen’</td>
<td>/caštă/</td>
<td>‘to split, crack open’</td>
<td>/bake/</td>
<td>‘leave, abandon’</td>
</tr>
<tr>
<td>/moča/</td>
<td>‘thick, fat’</td>
<td>/coka/</td>
<td>‘frog’</td>
<td>/boka/</td>
<td>‘a stupid person’</td>
</tr>
<tr>
<td>/muča/</td>
<td>‘nose’</td>
<td>/cuču/</td>
<td>‘mouse’</td>
<td>/buğa/</td>
<td>‘tree trunk’</td>
</tr>
</tbody>
</table>

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.
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.

<table>
<thead>
<tr>
<th></th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>311.3</td>
<td>2042.92</td>
</tr>
<tr>
<td>e</td>
<td>426.79</td>
<td>1856.29</td>
</tr>
<tr>
<td>a</td>
<td>592.29</td>
<td>1286.73</td>
</tr>
<tr>
<td>o</td>
<td>455.2</td>
<td>1021.58</td>
</tr>
<tr>
<td>u</td>
<td>325.33</td>
<td>1034.22</td>
</tr>
</tbody>
</table>

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. noticed this for Mayurbhanj Ho and it is also a feature of Chaibasa Ho.

(25) /kula/ → [kule] ‘tiger’
/luga/ → [luge] ‘nest’
/misa/ → [mise] ‘once’

but,

/gaɾa/ → [gaɾa] ‘river’
/boja/ → [bojo] ‘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.
Aspects of Ho Phonetics and Phonology

(26) /enga mĩɖi/ → [ẽŋga mĩɳɖi] ‘ewe’
(27) /hilaŋ/ → [hilẽŋ] ‘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/ → [‘kakəla] ‘to shout’
/tisɪɲ/ → [tə’siɲ] ‘today’
/dudulum/ → [duda’lum] ‘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 [duda’lum] ‘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](image)

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.

<table>
<thead>
<tr>
<th>[kani]</th>
<th>‘a pointed edge’</th>
<th>[kaːni]</th>
<th>‘story’</th>
</tr>
</thead>
<tbody>
<tr>
<td>[med]</td>
<td>‘eye’</td>
<td>[meːd]</td>
<td>‘iron’</td>
</tr>
<tr>
<td>[agu]</td>
<td>‘bring’</td>
<td>[aːgu]</td>
<td>‘lower’</td>
</tr>
<tr>
<td>[cera]</td>
<td>‘diarrhea’</td>
<td>[ceːra]</td>
<td>‘beautiful’</td>
</tr>
<tr>
<td>[gom]</td>
<td>‘wheat’</td>
<td>[goːm]</td>
<td>‘to accompany someone’</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Short range</th>
<th>Short mean</th>
<th>Long range</th>
<th>Long mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>51-101 ms</td>
<td>71 ms</td>
<td>102-316 ms</td>
<td>203 ms</td>
</tr>
<tr>
<td>e</td>
<td>55-102 ms</td>
<td>71 ms</td>
<td>152-323 ms</td>
<td>212 ms</td>
</tr>
<tr>
<td>a</td>
<td>54-106 ms</td>
<td>80 ms</td>
<td>140-280 ms</td>
<td>202 ms</td>
</tr>
<tr>
<td>o</td>
<td>64-98 ms</td>
<td>83 ms</td>
<td>145-274 ms</td>
<td>217 ms</td>
</tr>
<tr>
<td>u</td>
<td>44-91 ms</td>
<td>75 ms</td>
<td>131-352 ms</td>
<td>203 ms</td>
</tr>
</tbody>
</table>

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 \( \tilde{r} \). If we compare Mundari and Ho vocabulary (table 11), we can see that in some words where Mundari has an /\( \tilde{r} \)/, Ho has lost the consonant and is left with a long vowel.

<table>
<thead>
<tr>
<th>Mundari</th>
<th>Ho</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ho(\tilde{r})o]</td>
<td>[ho:]</td>
<td>‘man’</td>
</tr>
<tr>
<td>[du(\tilde{r})um]</td>
<td>[du:m]</td>
<td>‘sleep’</td>
</tr>
<tr>
<td>[sepe(\tilde{r})ed]</td>
<td>[sepe:d]</td>
<td>‘young man’</td>
</tr>
<tr>
<td>[r(\tilde{r})e(\tilde{r})]</td>
<td>[r(\tilde{r})]</td>
<td>‘joy in the company of others’</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Nasalized Vowel</th>
<th>Oral Vowel</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ija/</td>
<td>/ija:/</td>
<td>‘grandmother’</td>
</tr>
<tr>
<td>/r(\tilde{e})jo/</td>
<td>/r(\tilde{e})jo/</td>
<td>‘even if, although’ (verbal suffix)</td>
</tr>
<tr>
<td>/b(\tilde{a})l/</td>
<td>/b(\tilde{a})/</td>
<td>‘to burn a hole into’</td>
</tr>
<tr>
<td>/s(\tilde{a})i/</td>
<td>/s(\tilde{a})/</td>
<td>‘foul smell, stench’</td>
</tr>
<tr>
<td>/e(\tilde{a})/</td>
<td>/e(\tilde{a})/</td>
<td>‘to erase, extinguish’</td>
</tr>
<tr>
<td>/r(\tilde{a})sa/</td>
<td>/r(\tilde{a})sa/</td>
<td>‘joy, delight’</td>
</tr>
<tr>
<td>/d(\tilde{a})si/</td>
<td>/d(\tilde{a})si/</td>
<td>‘thirty’</td>
</tr>
<tr>
<td>/(\tilde{a})r/</td>
<td>/(\tilde{a})r/</td>
<td>‘leather, hide’</td>
</tr>
</tbody>
</table>

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) /s\(\tilde{u}\)/ ‘foul smell, stench’
| /e\(\tilde{u}\)/ | ‘to erase, extinguish’ |
| /r\(\tilde{a}\)sa/ | ‘joy, delight’ |
| /d\(\tilde{a}\)si/ | ‘thirty’ |
| /\(\tilde{a}\)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.

\[(30) \quad /\text{bai}/ \text{ ‘work, build’} \]
\[/\text{hau}/ \text{ ‘red ant’} \]
\[\ɟ/\text{umbui}/ \text{ ‘glutton’} \]

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.

<table>
<thead>
<tr>
<th>Diphthong</th>
<th>Mean length (ms)</th>
<th>Range</th>
<th>No. of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ai/</td>
<td>223ms</td>
<td>126-292</td>
<td>11</td>
</tr>
<tr>
<td>/au/</td>
<td>216ms</td>
<td>150-330</td>
<td>6</td>
</tr>
<tr>
<td>/ui/</td>
<td>206ms</td>
<td>154-305</td>
<td>11</td>
</tr>
</tbody>
</table>

*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 [baiː], 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:

\[(31) \quad /\text{toroe}/ \text{ ‘ashes’} \rightarrow [\text{toroʔe}] \]
\[/\text{moroe}/ \text{ ‘acid, sour’} \rightarrow [\text{moroʔe}] \]
\[/\text{aeʔ}/ \text{ ‘3.sg’} \rightarrow [\text{aʔeʔ}] \]

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 | VCV | /ũr/ | ‘hide, leather, animal skin’ |
| VVC | /aʔ/ | ‘herb’ |
| CV | /na/ | ‘drink’ |
| CVV | /tu:/ | ‘squirrel’ |
| /bai/ | ‘work’ |
| CVC | /cur/ | ‘surround, gush’ |
| CVVC | /jux/ | ‘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.rap/ | ‘maybe’ |
| VC.CV | /en.kai/ | ‘like that’ |
| VC.CVC | /aŋ.kar/ | ‘sense’ |
| CVC.CV | /kun.tu/ | ‘wooden post’ |
| CVC.CVC | /ban.dor/ | ‘monkey’ |

| Trisyllabic | V.CV.CV | /a.do.wa/ | ‘husked without boiling’ |
| V.CV.CVC | /e.pa.rap/ | ‘quarrel’ |
| CV.CV.CV | /sa.sa.ti/ | ‘torment’ |
| CV.CV.V | /ku.la.e/ | ‘rabbit, hare’ |
| CV.CV.CVC | /ko.lo.wad/ | ‘to give a loan’ |
| CV.CVC.CV | /go.pon.de/ | ‘quarrel, recip.’ |
| CV.CVC.CVC | /ka.ram.caʔ/ | ‘fox’ |
| VC.CV.CV | /an.gu.ʔ/ | ‘finger’ |
| CVC.CV.CV | /sin.du.ɾi/ | ‘vermilion’ |

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’
[c:a:s] ‘cultivate, farm’
[kon.’qaid] ‘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.

<table>
<thead>
<tr>
<th>Onset</th>
<th>Coda</th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>3</td>
<td>only borrowings</td>
<td>s</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>only borrowings</td>
<td>h</td>
</tr>
<tr>
<td>t</td>
<td>3</td>
<td>only borrowings</td>
<td>m</td>
</tr>
<tr>
<td>d</td>
<td>3</td>
<td>only borrowings</td>
<td>n</td>
</tr>
<tr>
<td>l</td>
<td>3</td>
<td>only borrowings</td>
<td>η</td>
</tr>
<tr>
<td>q</td>
<td>3</td>
<td>only borrowings</td>
<td>p</td>
</tr>
<tr>
<td>c</td>
<td>3</td>
<td>only borrowings</td>
<td>η</td>
</tr>
<tr>
<td>j</td>
<td>3</td>
<td>only borrowings</td>
<td>r</td>
</tr>
<tr>
<td>k</td>
<td>3</td>
<td>only borrowings</td>
<td>η</td>
</tr>
<tr>
<td>g</td>
<td>3</td>
<td>only borrowings</td>
<td>l</td>
</tr>
<tr>
<td>?</td>
<td>7</td>
<td>3</td>
<td>w</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>j</td>
</tr>
</tbody>
</table>

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 /m/, 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’
     /ʈuʈka/ ‘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 [ɔ] 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.

![Vowel plot](image)

*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 [ɔ].

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’ +/ʔ/ ‘passive’ → [hasu:ʔ]

(37) /abu/ ‘1.PL.INCL’ +/leka/ ‘like, as’ → [abu:like]

(38) /en/ ‘that’ +/kiɲ/ ‘dual’ → [inkiɲ]

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-d-a] ‘eat-T/A-TR-FIN’
    [men-ke-d-a] ‘say-T/A-TR-FIN’
    [cike-ke-d-a] ‘do-T/A-TR-FIN’ (/cika/ ‘do’)

(40) [tai-ki-n-a] ‘stay-T/A-ITR-FIN’

(41) [atom-idi-ki-d-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 [atu:m] and *-redo* does not become [rito]. 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).

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.

There are also some examples of disharmonicity. For example, the plural morpheme */-ko/* sometimes harmonizes to */-ku/* when that word contains a high vowel (48).

However, this does not seem to be consistent. In the following examples, */-ko/* follows a high vowel but does not harmonize.

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.
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] ‘vermilion’

['kun.tr] ‘wooden post’

['ce.ra] ‘pretty’

[ra.'cam] ‘cut with scissors’

Ho also has superheavy syllables and these always attract emphasis:

(55) [kon.'daid] ‘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.k[a].rã] ‘snacks served with rice beer’

[ge.je.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’

['pa.m.pal] ‘butterfly, moth’

However, in some instances of a homorganic nasal-stop sequence, the final syllable can also be prominent:
(58) [bap."caw] ‘save, salvation’
   [ban."dor] ‘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

Online version
<http://www.ethnologue.com>
Haumann, O. Reichmann, H. E. Wiegard L. Zgusta (eds), An International Encyclopedia of
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 /ˀb, ˀd, ˀɟ, (ˀg)/, such as /ˀd/ in (2), and vowels with creaky voice articulation, i.e. /a̰ ḛ ḭ o̰ ṵ/, as /a̰/ in (3).

(1) ɖaʔ [ɖaʔ] ‘water’
(2) ɖaˀd [ɖaˀdⁿ] ‘for’
(3) a̰l [a̰l̰] ‘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 perceptionally 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. Perceptionally, 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.
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)Vʔ(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 seems 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.

(4) al [a̰l] ‘husking pit’
(5) al [al] ‘to thatch’
(6) ɖaʔ [ɖaʔᵃ] ‘water’
(7) ɖa [ɖa] ‘to do’
(8) ɖaˀbu [ɖaˀbᵐu] ‘close-INF.TR’
(9) ɖabu [ɖabu] ‘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

---

107 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 /daˀbˌu/ in contrast to /daˌbu/ 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 /Yn/ (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/ [Vʔdⁿ].

(10) ɖaˀd [ɖaˀdⁿ] ‘for’
(11) kinɖaʔ:n [ɖaʔn] ‘at the river (river:LOC)’
(12) ɖa:nḛn [nḛn̰] ‘if done (do:TOP)’
(13) abgeˀd [geˀdⁿ] ‘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 [eq]. 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 [a/dⁿ], 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 /d/ by /ʔn/ as a clear audible mistake, they are less clear about the relationship between /d/ and /Yn/.

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) gṵmar ‘winnowing’
(19) gḭmir ‘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 gF.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 phonicemic 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, CVʔ, CVN, CVʔN. 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.

<table>
<thead>
<tr>
<th>Rau</th>
<th>Aze</th>
<th>Zide</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVʔ</td>
<td>CVʔ</td>
<td>CVH</td>
</tr>
<tr>
<td>CVʔ</td>
<td>CVʔ</td>
<td>CVʔ</td>
</tr>
<tr>
<td>CVʔn</td>
<td>CVʔn</td>
<td>CVʔn</td>
</tr>
<tr>
<td>CVN</td>
<td>CVʔn</td>
<td>CVHN</td>
</tr>
<tr>
<td>CVʔO</td>
<td>CVʔn</td>
<td>CVO (not glottalized)</td>
</tr>
</tbody>
</table>

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/CVC syllable structure. Three exceptions occur: The complex onset /ɖʔ/ occurs in one native word, two lexemes have a complex coda /ŋk/, and some word forms possess a coda /ʔn/. 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. Additional phonetic evidence comes from the articulation of creaky voice. In words such as al [a̰l̰] ‘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 ãman [ã.man] ‘before’ is not affected by the creaky phonation of the first syllable.

(20) al [a̰l̰] ‘husking pit’
(21) ãman [ã.man] ‘before’

108 According to my analysis, this restriction applies to all syllables in all situations. Hence, words such as /ɖaˀb.u/ ‘close-INF.TR’ (from ɖaˀb ‘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 /ɖa.ˀbu/ 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 combinatorical constraints of the three phenomena result in the syllable patterns represented in Table 2.\(^{109}\) 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.

<table>
<thead>
<tr>
<th>Syllable Structure</th>
<th>Syllable Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C)V(J/R/L)</td>
<td>Light</td>
</tr>
<tr>
<td>(C)Vʔ</td>
<td>Heavy</td>
</tr>
<tr>
<td>(C)VˀO</td>
<td>Heavy</td>
</tr>
<tr>
<td>(C)V(N)</td>
<td>Bimoraic</td>
</tr>
</tbody>
</table>

*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)VˀO syllables with preglottalized obstruents are heavy. With creaky voice, the situation is more complex. It can occur with open (C)V syllables, (C)V(j/r/l) syllables with liquids in the coda, as well as (C)V(N) 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-glossitized 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)VŐ. 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)V(j/r/l). Perhaps the most telling of them is the following pair:

(22) sur ‘to hunt’
(23) sṵr ‘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 CVr. 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.\(^{110}\)

\(^{109}\) Additionally, there is a very small number of exceptional native lexemes – such as /lup/ [lupʰ] ‘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)V(N). 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 zan ‘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) gaˀd ‘to cut’ → gaga’d

This constraint on glottal elements is also at work with the causative prefix, which takes the form ab- with roots containing glottalization and aˀbU with roots without the constricted glottis feature. In contrast to the reduplicant in (26), where the glottalized coda is lost, the coda segment – /ʰ/ or /b/ – is present in both (27) and (28), but the presence of glottalization depends on the nature of the root.

(27) ab + soˀɟ ‘causative + to learn’ → ab:soˀɟ
(28) ab + suŋ ‘causative + to fall’ → aˀb:suŋ

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 /ˀd/ is part of the stem ɖimaˀd ‘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) ɖimaˀd#: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)V̰ N syllable. In either case, affectedness marking takes the form of glottalization and is thus solely an instantiation of [+constricted glottis].

---

111 The glottalized form of the prefix could be either aˀbU or a̰ bU. The phonetic evidence is not unequivocal, yet the preglottalized obstruent is the most likely solution.

112 Other forms do not occur, as verbal suffixes only have CV or CVN syllable structures.

113 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#σAFF
(31) (ne)_(ko)(ko)_(#-(tuʔ))_σ ‘I will sit’
1sA:sit-NPST:AFF
(32) (or)_(g̊i̊)_(#-(n-q))_σ ‘he/she/it is not visible’
NEG:NPST-see-INF.INTR-CISL:AFF
(33) (ɖu)_(ku)_(#-(r-imitive))_σ ‘they were’
be-PST-3pS:AFF
(34) (ɖu)_(k#-i)_(a)_(aj)_σ ‘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 ɖuku ‘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) bu’il ‘to be drunk’
(38) tu.pa’d ‘to thresh’
(39) kin.qaʔ ‘river’
(40) go̰.tuŋ ‘cloth’
(41) aŋ.an.a’d ‘door’
(42) bi.o̰.gi ‘tomorrow’

As can be seen, there is a strong tendency for the glottal stop /ʔ/ and the preglottalized obstruents /ˀO/ to occur in the last syllable of a stem and, as such, at its right boundary, while creaky voice /V̰/ 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

\begin{tabular}{l l l}
\textit{CVʔ} & \textit{CVC.CVʔ} & \textit{CVC.CVC.CVʔ} \\
\textit{CV’O} & \textit{CVC.CV’O} & \textit{CVC.CVC.CV’O} \\
\textit{CYC} & \textit{CYC.CVC} & \textit{CVC.CYC.CVC} \\
\end{tabular}

\textbf{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 \textit{mitan} ‘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 \textit{mit} ‘day’ and the synchronically unattested demonstrative \textit{*gn}, probably meaning ‘this day’. This putative demonstrative can be connected to the still used \textit{gt} ‘that day’, parallel to the pair \textit{gn} ‘hither’ and \textit{gt} ‘thither’.

A peculiar pattern occurs with some Telugu or Oriya loan words which have a \textit{CV.CV} structure in the source language. In Gorum, their first syllable became glottalized, so that the resulting prosodic structure is \textit{CV̰ 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) ɖo̰pa ‘leaf bowl’ from Telugu \textit{doppa}

(44) ka̰ɖu ‘bangles’ from Desia Oriya \textit{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.

\textbf{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 \textit{sṵr} ‘hunt’ (example 23), the creaky voice seems to be connected with the second mora and might very well be a
Notes on Glottal Constriction in Gorum

derivational device. In other lexemes – such as *gsug* ‘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 *sug*, 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.\(^{114}\)

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 *la̰ki* 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) bṵl ‘to be drunk’
(46) dḛl ‘to ripen’
(47) mḛŋ ‘to live’
(48) ūŋ ‘to perform a burial ritual’
(49) ḗmaŋ ‘before’
(50) аqaj ‘when’
(51) лa̰ki ‘later’
(52) biggi ‘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


Comitative PP as Head in Externally-Headed Relative Clauses in Khasi

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 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.

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.


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 -lay- ‘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 -lay- ‘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. It has predominantly verb-medial structures, though there is a set of structures in which the verb obligatorily occurs in the initial position (see Temsen, 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 -he’, -she’, and -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’d ‘table’ is modified.

Locative PP as head– Relative clause

1. ka-mē’d, ha-ka ba u-lam u-bōʔ ya-ka-kət t₁ ka-laʔ?-kd’aʔ
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.

---

116 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, na-ka ba** u-lam u-wan t, ka-lông ka-ba- jįjaì
   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, 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)**

   f.s:table adjr- s.m-Lam 3m.s-put acc-f.s:book 3.f.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 o), 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)118 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-khìnna? ša-ka-jįgkhawai**
   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.’

117 Adjectives in Khasi too carry the marker ba. Thus, the entire following phrase functions like an adjectival phrase (AP):

   ba- u-lam u-böʔ ya- ka-kot
   ‘adjr- m.s-Lam- 3sm-put acc-3f.s:book’

118 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\textsuperscript{119} wan-laŋ bad ka-wei ka-khînaʔ ša-ka-įŋkhawai
   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-khînaʔ ki-ya wan
   m.s:Lam com 3f.s:one f.s:child pl:gm:come
   ša-ka-įŋkhawai taŋ:ba ki-khîm-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-įŋkhawai bad ka-wei
   m.s:Lam 3m.s:come to:f.s:party com 3f.s:one
   ka-khînaʔ taŋ:ba ki-khîm-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.’

   m.s:Lam 3m.s:gm:come-together com 3f.s:one f.s:child to:f.s:party
   taŋ:ba ki-khîm-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-khînaʔ ‘f.s:child’. The relativized NP along with the preposition bad-ka-ba ‘with:she:who’, that is, ‘with whom’ is fronted and t\textsubscript{1} (trace\textsubscript{1}) in bold in (8) indicates the position in which the comitative PP originates.

\textsuperscript{119}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? [s₂ bad-ka-ba u-lam u-ya-wan-laŋ t₁ s₂]
  f.s-child com-3f.s-adjr 3m.s-Lam 3m.s-gm-come-together
  ka-dεi  ka-para (jɔŋ₁²⁰)-u-ker
  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 -laŋ ‘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-khinnaʔ [ba u-ban u-wan o, ša-šnəŋ AP]
   f.s-child adjr m.s-Ban 3m.s-gm-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 o, and the gap is coindexed with the subject of matrix clause ka-khinnaʔ? ‘f.s-child’.

Sentence (9) is ungrammatical because the verbal reciprocal with the interpretation of a group marker and the adverb -laŋ ‘together’ are not incorporated in the embedded verb. In contrast to (9), (10) is well-formed because the group marker -ya-and the adverb -laŋ ‘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 -laŋ-‘together’

10. ka-khinnaʔ [AP ba u-ban u-ya-wan-laŋ o, ša-šnəŋ AP]
    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 -laŋ ‘together’ in the formation of a relative clause and a gap relative include: marɛʔ ‘run’, ya-ʔdkāi ‘walk’, jŋi-‘swim’ among others. The following examples are illustrative.

11. ka-khinnaʔ [AP ba- u-ban u-ya-marsʔ-laŋ o, AP]
    f.s-child adjr m.s-Ban 3m.s-gm-run- together
    ‘The girl with whom Ban is running …’

12. ka-khinnaʔ [AP ba u-ban u-ya-ʔdēŋ- laŋ o, AP]
    f.s-child adjr m.s-Ban 3m.s-gm-walk- together
    ‘The girl Ban is walking with …’

¹²⁰ 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-kst-jɔŋ-ya ‘my book’, ka-kst-(jɔŋ)-ju-bon ‘Bon’s book’, ka-tuat-(jɔŋ)-u-dēŋ ‘A tree’s branch’.
Comitative PP as Head in Khasi

13. \(ka\)-\textit{khɨnnaʔ} [\(AP\) \(ba\) \(u\)-\textit{ban} \(u\)-\textit{ya}-\textit{jɨŋ} \textit{-lay} \(0_1\)]
   \(f.s\text{-child}\) \(adjr\) \(m.s\text{-Ban}\) \(3m.s\text{-gm-swim-together}\)
   ‘The girl Ban is swimming with …’

   Verbs such as \(ya\text{-}kren\) ‘converse’, \(ya\text{-}thɔʔ\) ‘marry’, \(ya\text{-}šɔʔ\) ‘fight’, \(ya\text{-}kindō\) ‘meet’ which are inherently reciprocal verbs do not require the incorporation of the adverb \textit{-lay} ‘together’, and hence, the occurrence of \textit{-lay} with the embedded verb is prohibited in gap relatives as well as sentential relatives.

14. \(ka\)-\textit{khɨnnaʔ} [\(ba\) \(u\)-\textit{ban} \(u\)-\textit{ya}-\textit{krɛn} \(0_{i,AP}\) ] …
   \(f.s\text{-child}\) \(adjr\) \(m.s\text{-Ban}\) \(3m.s\text{-Vrec-speak}\)
   ‘The girl Ban is talking with …’

15. \(ka\)-\textit{khɨnnaʔ} [ \(ba\) \(u\)-\textit{ban} \(u\)-\textit{thɔʔ} \(0_{i,AP}\) ] …
   \(f.s\text{-child}\) \(adjr\) \(m.s\text{-Ban}\) \(3m.s\text{-Vrec-write}\)
   ‘The girl Ban is married to …’

16. \(ka\)-\textit{khɨnnaʔ} [ \(ba\) \(u\)-\textit{ban} \(u\)-\textit{šɔʔ} \(0_{i,AP}\) ] …
   \(f.s\text{-child}\) \(adjr\) \(m.s\text{-Ban}\) \(3m.s\text{-Vrec-beat}\)
   ‘The girl Ban is fighting with …’

17. \(ka\)-\textit{khɨnnaʔ} [ \(ba\) \(u\)-\textit{ban} \(u\)-\textit{kindō} \(0_{i,PP}\) ] …
   \(f.s\text{-child}\) \(adjr\) \(m.s\text{-Ban}\) \(3m.s\text{-Vrec-meet}\)
   ‘The girl Ban met with …’

The other inherently reciprocal verbs that take a comitative PP are \(-ya\text{-}leǐkāi\) ‘play’, \(-ya\text{-}māi\) ‘fight (verbally)’

We shall argue that it is the incorporation of these two markers – \(ya\)-‘group marker’ and \textit{-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\(^{122}\), 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.

\(^{121}\) Note that \(thɔʔ\) means ‘write’ and \(ya\text{-}thɔʔ\) 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.

\(^{122}\) 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, in transitive , 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’, marɛʔ ‘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-krɛn ‘talk’, ya-thɔʔ ‘marry’, ya-xɔʔ ‘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 *thematic ally 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* 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

---

123 Let us consider (i) from Khasi.

(i) *ka-tari, [ ba u-ban u-ɔt ya-u-sɔʔ ɔi ] ka-ITERAL ka-ba-lɔŋ*

f.s:knife adj 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.
Comitative PP as Head in Khasi

191

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 (IHR) as in Sangtam or Tenydie, 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-kren ‘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-šida ka-ya-thɔʔ ya-/bad- u-ban
    f.s- Shida 3f.s-marry acc/com m.s-Ban
    ‘Shida married Ban.’

Comitative PP as Head in Khasi
21. ka-šida ka-ya-kindō?¹²⁴ 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 illustrative¹²⁵.

22. ka-khɨnnaʔ [ba  u-ban  u-ya-šɔʔ ] ka-dɛi
f.s-child adjr m.s-Ban 3m.s-vrec-beat 3f.s-be
ka-paralɔk jɔŋ-ŋa
f.s-friend gen-1s
‘The girl Ban is fighting with is my friend’

23. ka-khɨnnaʔ [ba  u-ban  u-ya:kren ] ka-dɛi
f.s-child adjr m.s-Ban 3m.s-vrec-speak 3f.s-be
ka-paralɔk jɔŋ-ŋa
f.s-friend gen-1s
‘The girl Ban is talking with is my friend.’

24. ka-khɨnnaʔ [ba  u-ban  u-ya-thɔʔ ] ka-dɛi
f.s-child adjr m.s-Ban 3m.s-vrec-write 3f.s-be
ka-paralɔk jɔŋ-ŋa
f.s-friend gen-1s
‘The girl Ban is married to is my friend.’

25. ka-khɨnnaʔ [ba-  u-ban  u-ya-kindōʔ ] ka-dɛi
f.s-child adjr m.s-Ban 3m.s-vrec-meet 3f.s-be
ka-paralɔk jɔŋ-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).

f.s-child adjr m.s-Ban 3m.s-gm-beat-together 3f.s-be
ka-paralɔk jɔŋ-ŋa
f.s-friend gen-1s
‘The girl who(m) Ban along with someone else is beating (someone) is my friend.’

¹²⁴ The verbal reciprocal –ya is grammaticalized and hence, ya-kindōʔ is lexicalized as one word.
¹²⁵ Note that the marker –ya-in these sentences functions as a verbal reciprocal marker, and not as a group marker.
Comitative PP as Head in Khasi

27. ka-khɨnnaʔ [ba u-ban u-ya-kren-laŋ ] ka-dɛi
   f.s-child adjr m.s-Ban 3m.s-gm-speak-together 3f.s-be
   ka-paralɔk jəŋ-ŋa
   f.s-friend gen-1s
   ‘The girl who(m) Ban together with someone else is talking to is my friend.’

28. ka-khɨnnaʔ [ba u-ban u-ya-thɔʔ-laŋ ]
   f.s-child adjr m.s-Ban 3m.s-gm-write-together
   ka-dɛi ka-paralɔk 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.]126

(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 eẓuṭṭə
   Babu come.adjr letter

Tamil (Dravidian)

30. *kumār veḷiye pōna umā
    Kumar out go:pst.adjr Uma
    ‘Uma, with whom Kumar went out.’ (Annamalai 1997: 78)

Telugu (Dravidian)

31. *nēnu veḷḷ:in:a saraḷa
    I go:pst-adjr Sarala
    ‘Sarala with whom I went.’ (Ramarao 2003: 79)

---

126 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 suśila
   boy (ø-with) came.pst.adjr 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 Ple.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.
Comitative PP as Head in Khasi

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), marɛʔ ‘run’ in (11), ya:dkāi ‘walk’ in (12), and jŋi ‘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-khɨnnaʔi [ba u-ban u-wan oš-a-ʂnɔŋ PP]
  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 jŋi ‘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-khɨnnaʔi [ba u-ban u-ya-jŋi-laŋ]
  u-ɖɛi u-paralɔk 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-khɨnnaʔi [ba u-ban u-ya-pār-laŋ]
  u-ɖɛi u-paralɔk 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\textsuperscript{127} 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-kɔt ha-lɔr-ka-mɛc
   m.s-Wan 3m.s:put acc-pl-book loc-surface-f.s:table
   ‘Wan puts the books on the table.’

43. dɛi ha-lɔr-ka- mɛc ba u-wan u-bōʔ ya-ki-kɔt
   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 -\textit{ya}-and of the adverb -\textit{laŋ} ‘together’ is obligatory as in (44).

44. dɛi bad-u-paralɔk jɔŋ-ŋa ba u-ban u-ya-jgi-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 -\textit{ya}-occurs and the adverb -\textit{laŋ} ‘together’ does not as in (45).

45. dɛi bad-ka-paralɔk jɔŋ-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 -\textit{ya} and -\textit{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 -\textit{ya} and the adverb -\textit{laŋ} with predicates that are [-reciprocal] and only the verbal reciprocal -\textit{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

\textsuperscript{127} Thanks to N. Pramodoni for a helpful discussion on this issue in Manipuri.
Comitative PP as Head in Khasi

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 -laŋ-'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 -laŋ-‘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 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-khinna?] u-dei u-paralɔk 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-)k a ba u-brēw u-la-āy
f:money acc-3fs adjr m-human 3ms-pst:give
ha-u-khinna?] 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-khinna? [(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-ŋ ŋ u ba jŋ 3ms-be 3ms advm tall
‘The boy who the man gave the money to is tall.’
Oblique object Modification

Locative

(4)  \text{ka-mē’d [ha-ka ba ya-bo? ya-ka-kot] ka-la?-kd’a?}
\quad \text{f-table loc-3fs adjr 1s-put acc-f-book 3fs-perf-break}
\quad \text{‘The table on which I put the book is broken.’}

Ablative

(5)  \text{ka-jaka [na-ka ba u-(la)-wan] ka-lŋŋŋ ka-ba įŋāi}
\quad \text{f-place abl-3fs adjr 3ms-pst-cut 3fs-be 3fs-adjm far}
\quad \text{‘The place which he came from is far.’}

Instrumental

(6)  \text{ka-tari [da-ka ba u-(la)-ɔt ya-u-sɔ?] ka-lŋŋŋ ka-ba lōn}
\quad \text{f-knife instl-3fs adjr 3ms-pst-cut acc-m-fruit 3fs-be 3fs-adjr blunt}
\quad \text{‘The knife with which he cut the fruit is blunt.’}

Comitative

(7)  \text{ka-khinna? [bad-ka ba ya-ya-wan-lag]}
\quad \text{f-child com-3fs adjr 1s-vr:come:together}
\quad \text{ka-dei ka-paral (jɔŋ)u-meban}
\quad \text{3fs-be f-younger sibling gen-m-Meban}
\quad \text{‘The girl who came with me is Meban’s sister.’}

Object of the Genitive Modified

(8)  \text{u-brēw [u ba ya-ka-yēŋ jɔŋ-u la-pinjɔt da-ka-ēr yɔŋ] u-daŋ-yām}
\quad \text{m-human 3ms adjr acc-f-house gen-3ms pst-destroy instl-f-storm 3ms-prog-cry}
\quad \text{‘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)  \text{u-brēw [u ba ka-yēŋ}
\quad \text{m-human 3ms adjr f-house}
\quad \text{jɔŋ-u ka-(la)-shaʔ-pinjɔt ha-ka-ēr yɔŋ] u-daŋ-yām}
\quad \text{gen-3ms 3fs-pst-SA-destroy loc-f-storm 3ms-prog-cry}
\quad \text{‘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)  \text{u-brēw [u ba u-lan u-kham-jrɔŋ (ban}
\quad \text{m-human 3ms adjr m-Lan 3ms-more-tall than}
\quad \text{ya-u)] u-dei u-paralɔk jɔŋ-ŋa}
\quad \text{acc-3ms 3ms-be m-friend gen-1s}

Karumuri Subbarao, Gracious Temsen
‘The man who Lan is taller than is my friend.’

(11)  
\[ u\-br\text{ē}w \ [ ya\-u\-ba \ u\-lan \ u\-kham\-jr\text{ŋ}] \ u\-dei \ u\-paral\text{ɔ}k \ j\text{ŋ}-\text{ŋ}a \]
\[ m\text{-human} \ \text{acc-3ms-adjr} \ \text{m-Lan} \ \text{3ms-more-tall} \ \text{3ms-be} \ \text{m-friend} \ \text{gen-1s} \]
‘The man who Lan is taller than is my friend.’

Appendix II

Participles in Khasi

Subject Modification

(1)  
\[ u\-br\text{ē}w \ [ ba \ āy \ ya\-ka\-pisa \ ha\-u\-k\text{hınna}?] \ u\-dei \ u\-paral\text{ɔ}k \ j\text{ŋ}-\text{ŋ}a \]
\[ m\text{-human} \ \text{adjr} \ \text{give} \ \text{acc-f-money} \ \text{dat-m-child} \ \text{3ms-be} \ \text{m-friend} \ \text{gen-1s} \]
‘The man who gave the money to the boy is my friend.’

Direct and Indirect objects Modification

Direct object

(2)  
\[ ka\-pisa \ [ ba \ u\-br\text{ē}w \ u\-āy \ ha\-u\-k\text{hınna}?] \]
\[ f\text{-money} \ \text{adjr} \ m\text{-human} \ \text{3ms-give} \ \text{dat-m-child} \]
\[ ka\-dei \ ka \ ba \ la\?-\text{jɔt} \]
\[ \text{3ms-be} \ \text{3fs} \ \text{adjr} \ \text{perf-be torn} \]
‘The money which the man gave to the boy is torn.’

Indirect object

(3)  
\[ u\-k\text{hınna}? \ [ ba \ u\-br\text{ē}w \ u\-āy \ ya\-ka\-pisa ] \]
\[ m\text{-child} \ \text{adjr} \ m\text{-human} \ \text{3ms-give} \ \text{acc-f-money} \]
\[ u\-dei \ u \ ba \ jr\text{ŋ} \]
\[ \text{3ms-be} \ \text{3ms} \ \text{adjm} \ \text{tall} \]
‘The boy who the man gave the money to is tall.’

Oblique object Modification

Locative

(4)  
\[ ka\-mē\text{d} \ [ ba \ ga\-ba? \ ya\-ka\-kɔt] \ ka\-la\?-\text{kɗ̱a}? \]
\[ f\text{-table} \ \text{adjr} \ 1\text{s-put} \ \text{acc-f-book} \ \text{3fs-perf-break} \]
‘The table on which I put the book is broken.’

Ablative

(5)  
\[ ka\-jaka \ [ ba \ u\-wan] \ ka\-lŋ \ ka\-ba \ jŋāi \]
\[ f\text{-place} \ \text{adjr} \ \text{3ms-come} \ \text{3fs-be} \ \text{3fs-adjm} \ \text{far} \]
‘The place which he came from is far.’
Instrumental

(6) ka-\textit{tari} \[ba \; u-\textit{ot} \; ya-u-s\textit{O}] \; ka-\textit{tou} \; ka-ha \; loo\textit{n} \\\n\textit{f-knife} \; \textit{adjr} \; 3\textit{ms}-cut \; \textit{acc-m-fruit} \; 3\textit{fs}-be \; 3\textit{fs}-adjm \; \textit{blunt} \\\n‘The knife with which he cut the fruit is blunt.’

Comitative

(7a) ka-\textit{khinna}? \[ba \; \eta-ya-wan-la\textit{y}] \; ka-\textit{dei} \\\n\textit{f-child} \; \textit{adjr} \; 1\textit{s-vr-come-together} \; 3\textit{fs}-be \\\n\textit{ka-para} \; (\textit{jo}\textit{ŋ})-\textit{u-meban} \\\n\textit{f-younger sibling} \; \textit{gen-m-Meban} \\\n‘The girl who came with me is Meban’s sister.’

(7b) ka-\textit{khinna}? \[ba \; ya-wan-lan \; ba-d\textit{n-a}] \\\n\textit{f-child} \; \textit{adjr} \; \textit{vr-come-together} \; \textit{com-1s} \\\n\textit{ka-dei} \; \textit{ka-para} \; (\textit{jo}\textit{ŋ})-\textit{u-meban} \\\n3\textit{fs}-be \; \textit{f-younger sibling} \; \textit{gen-m-Meban} \\\n‘The girl who came with me is Meban’s sister.’

Object of the Genitive Modified

(8) *\textit{u-brēw} \[ba \; ya-ka-yēŋ \; \textit{jøŋ-u}] \\\n\textit{m-human} \; \textit{adjr} \; \textit{acc-f-house} \; \textit{gen-3ms} \\\nla-\textit{pinjat} \; da-\textit{ka-ër} \textit{yøŋ} \; u-dan-\textit{yām} \\\n\textit{pst-destroy} \; \textit{instl-f-storm} \; 3\textit{ms-prog-cry} \\\n‘The man whose house was destroyed by the storm is crying.’ \textit{Lit: ‘The man he that his house was destroyed by the storm is crying.’}

(9) *\textit{u-brēw} \[ba \; ka-yēŋ \; \textit{jøŋ-u} \; ka-(la)-sha?-\textit{pinjat}] \\\n\textit{m-human} \; \textit{adjr} \; \textit{f-house} \; \textit{gen-3ms} \; 3\textit{fs-pst-SA-destroy} \\\nha-\textit{ka-ër} \textit{yøŋ}] \; u-dan-\textit{yām} \\\n\textit{loc-f-storm} \; 3\textit{ms-prog-cry} \\\n‘The man whose house got destroyed in the storm is crying.’ \textit{Lit: ‘The man he that his house got destroyed in the storm is crying.’}

Object of comparison

(10) *\textit{u-brēw} \[ba \; u-lan \; u-kham-\textit{xeŋ} \; (\textit{ba-n} \; ya-u)] \\\n\textit{m-human} \; \textit{adjr} \; \textit{m-Lan} \; 3\textit{ms-more-tall} \; \textit{than} \; \textit{acc-3ms} \\\nu-dei \; u-paral\textit{sk} \; jøŋ-\textit{ya} \\\n3\textit{ms-be} \; \textit{m-friend} \; \textit{gen-1s} \\\n‘The man who Lan is taller than is my friend.’

(11) *\textit{u-brēw} \[ba \; u-lan \; u-kham-\textit{xeŋ} \; u-dei \; u-paral\textit{sk} \; jøŋ-\textit{ya}] \\\n\textit{m-human} \; \textit{adjr} \; \textit{m-Lan} \; 3\textit{ms-more-tall} \; 3\textit{ms-be} \; \textit{m-friend} \; \textit{gen-1s} \\\n‘The man who Lan is taller than is my friend.’
Comitative PP as Head in Khasi

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;
infl  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

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.
This is the first of two volumes of papers from the forth 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 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