AUSTROASIATIC STUDIES

Papers from ICAAL 4

Mon-Khmer Studies Journal
Special Issue No. 3

Volume editors:
Sophana Srichampa
Paul Sidwell
Kenneth Gregerson

Pacific Linguistics E-9

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.

First published 2011

National Library of Australia Cataloguing-in-Publication entry:

Author:	International Conference on Austroasiatic Linguistics (4th:, 2009 : Salaya, Thailand).
Title:	Austroasiatic studies [electronic resource]: papers from ICAAl 4, Mon-Khmer Studies Journal Special issue no. 3 / edited by Sophana Srichampa, Paul Sidwell and Kenneth Gregerson.
ISBN:	9780858836426 (Electronic document)
Subjects:	Austroasiatic languages--Congresses.
Other Authors/	Srichampa, Sophana.
Contributors:	Sidwell, Paul J., Gregerson, Kenneth 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

Special Session on Iambic Effects in Austroasiatic

Preface

This is the second volume of papers from the forth International Conference on Austroasiatic Linguistics (ICAAL4) held at Mahidol University, Thailand, October 29-30, 2009. Please consult the Preface to the first volume for a history of the ICAAL movement and details about the ICAAL4 meeting.

An important feature of this volume is the collection of five papers under the theme of the *Special Session on Iambic Effects in Austroasiatic Languages*. These papers came out the session originally coordinated by Kenneth ('Ken') Gregerson at the ICAAL4 meeting. The general problem of understanding the role of iambic rhythm in the phonology (and ultimately also the syntax) of Austroasiatic is emerging as a crucial one for progress in historical reconstruction, and understanding both internally and externally motivated restructuring that is such a prominent feature of many Austroasiatic tongues. In particular, a better understand of these phenomena will be crucial in unlocking the reasons behind the extreme typological differences between Munda and Mon-Khmer languages. Editorial responsibility for the papers in this session was principally taken by Ken, until he suffered some health problems in 2011, and Paul took over finalising the volume. Friends and colleagues will be glad to know that at the time of writing this Ken is making great strides towards recovery.

The rest of this volume contains papers from a variety of sessions and themes, indicative of the wide scope of topics discussed at the meeting. We are also pleased that both volumes have also been completed and made available in time for the ICAAL5 meeting, scheduled for November 9-11 2011, and we pledge to do our best to continue the timely publication of ICAAL papers as an important aspect of the new vigour that the field of Austroasiatic Studies is now enjoying.

Dr. Paul Sidwell
Assoc. Prof. Sophana Srichampa
Dr. Kenneth Gregerson
October 2011

Vietnamese Demonstratives *đây*, *đó*, *kia*[1]

Mayumi Adachi
The University of Tokyo, Japan

1. Introduction

It is customarily understood that Vietnamese[2] has three series of basic demonstratives, that is, proximal *đây*, medial *đó*[3] and distal *kia*. Table 1 shows their forms. Vietnamese demonstratives occur near the end of NP with noun, just before possessive expressions[4].

Table 1: Forms of Vietnamese demonstratives

	Proximal *đây*		Medial *đó*		Distal *kia*	
as full NP, local adverbial	*đây*	this, here	*đó*	that, there	*kia*	yonder, over there
in NP with noun	N + *này*	this N	N + *đó*	that N	N + *kia*	that N over there
	cái *này* CL this	this one	cái *đó* CL that	that one	cái *kia* CL yonder	that one over there
	chỗ *này* place this	this place	chỗ *đó* place that	that place	chỗ *kia* place yonder	that place over there
	người *này* person this	this person	người *đó* person that	that person	người *kia* person yonder	that person over there

According to Kinsui et al. (2002), functions of demonstratives can be classified as follows.

- **Deictic use**: "to directly refer to a recognizable referent in the objective world, typically, refer to visible one (p.218)."
- **Anaphoric use**: "to refer to a co-referential referent with an antecedent introduced into the context with a text (p.220)."
- **Memorative use**: "to refer to a referent in the speaker's knowledge based on his or her direct experience in the past (p.221)."

The present paper will examine the criteria of the classification of Vietnamese demonstratives based on this categorization.

[1] I am indebted to Tooru Hayasi for comments on earlier drafts of this paper, and to my consultant Nguyễn Dương Liễu for her time and patience.

[2] In this paper, Vietnamese examples are orthographized.

[3] In many cases, *đó* as full NP and local adverbial can be replaced with *đấy*. Similarly, *đó* in NP with noun can be replaced by *ấy*. In this paper, *đó* is used as a representative form.

[4] See Nguyễn Đình-Hòa (1997: 171-184) for illustration of Vietnamese noun phrase construction.

Adachi, Mayumi. 2011. "Vietnamese Demonstratives *đây, đó, kia*." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.1-8.

2. Previous Studies

Concerning spatial deixis, Anderson and Keenan (1985:282-286) classify three-term demonstrative systems under three subcategories: distance oriented, person oriented and the other. Which category do Vietnamese demonstratives fall into? As shown in Table 2, in the literature on Vietnamese demonstratives, the distance between the speaker and the referent on the one hand, and the addressee on the other, is usually mentioned as the criteria of classification of their deictic uses. Emeneau (1951:92) takes the former into account: *này* (this), *ấy* (that) and *kia* (that "further off than *ấy*").

Thompson (1987) makes mention of the addressee in addition to distance: While *đây* indicates something "close to speaker or newly introduced", *đấy* refers to an object "remote or already identified (p.142)" and "near hearer (p.144)". *Kia* is not included in basic demonstratives, but supplementarily explained that it is "more remote than *đấy*" (p.253).

Tomita (1988:778-779) also takes account of the addressee: proximal *đây* makes reference to what is "near the speaker in terms of space (and time)". Medial *đấy* and *đó* points out to something "far from the speaker and close to the addressee". And, distal *kia* refers to an object "remote from both the speaker and the addressee". However, he remarks that the difference between medial and distal demonstratives is relative, therefore, it remains unclear.

Nguyễn Phú Phong (2002) emphasizes on the eye direction of the addressee. When the speaker and the addressee are side by side, looking in the same direction (they are in tandem), *đây* points out to something near both people, *đó* indicates the object far from both of them, and *kia* refers to what is more remote than *đó*. Yet, if we suppose that they are face to face, *đây* makes reference to something close to the speaker, *đó* refers to an object near the addressee, *kia* indicates the referent remote from both of them (pp.110-111). In addition, he claims that both *đây* and *đó* have anaphoric use with the form '*đây* (or *đó*) + copula *là*'. *Đó* is more frequently used than *đây* in the anaphoric use. He explains that these two demonstratives are constrastive in terms of "proximité/distance", "moi/non-moi" and "appel/rappel". In his language data, the form "*Kia* là X" wasn't found (pp. 143-145).

It can be said that, basically, *đây* points out the referent close to the speaker, and *kia* indicates the referent far from the speaker. However, what *đó* refers to is difficult to define because in practice *đó* doesn't always point out a medial proximity between *đay* and *kia*, or a referent near the addressee. In addition, the referent in the memory is said to be referred to by the medial *đó*. In the following, both the position of the addressee in the deixis and the use of the memorative *kia* are reexamined.

Table 2: Previous studies on Vietnamese demonstratives

Use / Author	Series	Deictic			Anaphoric	Memorative
		Speaker	Addressee	Others		
Emeneau (1951)	*đây*	near	---	---	---	---
	đó	remote	---	---	---	---
	kia	remote	---	more remote than *đó*	---	---
Thompson (1987)	*đây*	near	---	newly introduced	---	---
	đó	remote	near	already identified	YES	---
	kia	remote	---	more remote than *đó*	---	---
Tomita (1988)	*đây*	near	---	---	---	---
	đó	remote	near	unclear distinction	---	---
	kia	remote	remote		---	---
Nguyễn (2002)	*đây*	near	near / remote	---	near	---
	đó	remote	near / remote	---	remote	---
	kia	remote	remote	more remote than *đó*	NO	---

3. Deictic use

3.1. The position of the addressee

In deixis, what matters is whether the speaker feels the referent to be near or not, while the position of the addressee is irrelevant as to the use of demonstratives. Thus, *đó* doesn't always refer to the referent close to the addressee. It can be refered to by other demonstratives. For example, *đây* can refer to the referent in the speaker's hand (1) as well as one in the addressee's hand (2). These examples are quoted from the novel of Nguyễn Nhật Ánh (2005).

(1) *Quý ròm thò tay vào ngăn bàn lấy ra một con dao găm.*
Quy skinny stretch hand in drawer table take out one CL. knife sharp
Nó huơ huơ con dao trong không khí:
he shake CL knife in air
*"Tôi sẽ dùng con dao { **này** / *đó / *kia } cắt ngón tay mình chơi!"*
I will use CL knife this that yonder cut finger hand myself play

'(Skinny Quy is a magician.) Skinny Quy put his hand in the drawer of the table, and took out a sharp knife. He shook the knife in the air: 'I'll cut my finger with this knife!'

(2) *Chị Ngần nhìn điếu thuốc trên tay Quy ròm:*
Elder sister Ngan look CL cigarette on hand Quy skinny
*"Cái { **này** / đó / *kia } để làm gì vậy?"*
CL this that yonder to do what FP

'Ms. Ngan looked at the cigarette on Skinny Quy's hand: "What are you going to do with it?" '

In some languages, demonstratives can take on the kinds of meaning; for example, indicating an emotional attitude, or personal interest, or familiarity (Dixon 2003:91). In Vietnamese *đây* and *đó* indicate psychological distance as well as physical distance between the speaker and the reference. If *đây* is replaced by *đó* in example (2), the speaker has an intuition that the referent, that is, the cigarette is unsuitable for Skinny Quy. Similarly, if *đó* is used in example (3), the referent in the addressee's hand can be considered as something new and unfamiliar to the speaker.

(3) (The speaker is sitting next to the addressee. The addressee is eating something smells really bad. The speaker can't stand the smell and so he or she asks:)
"Món { này / đó / ? kia } là gì? Mùi { này / đó / ? kia } thối nhỉ?"
CL this that yonder COP what smell this that yonder bad FP
'What's that? It smells bad.'

3.2. The difference between đó and kia

While *kia* points out a visible referent far from the speaker, *đó* indicates either a visible or invisible referent far from the speaker regardless of the existence of the addressee. Thus, *đó* doesn't always point to an intermediate proximity between *đây* and *kia*. In the case of a visible referent remote from the speaker, *đó* and *kia* are often interchangeable. Example (4) is quoted from Nguyễn Nhật Ánh (2005).

(4) *Quy ròm đứng đằng sau chiếc bàn lặng lẽ theo dõi, chẳng nói một lời.*
Quy skinny stand side back CL table quietly watch not at all say one word
Chỉ đến khi Tiểu Long thò tay định thử cào lên vệt sáng,
only come time little Long stretch hand intend try scratch up mark clear
*nó mới trầm giọng bảo: "Khỏi cào { *Đây / **Đó** / Kia } là miếng băng keo!"*
he finally lower voice tell don't scratch this that yonder COP CL tape glue

'Skinny Quy stood in the back of the table, and watched quietly with saying no words. When Little Long stretched his hand out to try to scratch up a clear mark, Skinny Quy finally told in a low voice: "Don't scratch that! It's scotch tape!"'

In example (5), both *đó* and *kia* are used to point out the same building. This is an example quoted from Japanese novel translated into Vietnamese.

(5) (Two people are looking down the street from the balcony.)

"Cái	*nhà*	*đằng*	*{ *này / đó /*	***kia***	*}*	*là*	*gì*	*thế?"*					
CL	house	side	this	that	yonder	COP	what	FP					
- Tôi	*hỏi.*	*Tôi*	*nhìn*	*thấy*	*ở*	*phía*	*chân*	*núi*	*1*	*tòa nhà*	*lớn*	*phía*	*trên*
I	ask	I	see	find	on	side	foot	mountain	1	house	big	side	upper
vẫn	*còn*	*lộ*	*ra*	*chút*	*cốt*	*sắt*	*nên*	*để*	*ý.*	*(snip)*			
still		expose	out	a little	pillar	iron	thus	pay	attention				
"Cái	*{ *này /*	***đó***	*/ kia }*	*hả,*	*cái*	*{ *này /*	***đó***	*/ kia }*	*là*	*khách sạn*	*đấy*[5]*."*		
CL	this	that	yonder	FP	CL	this	that	yonder	COP	hotel		FP	

[Yoshimoto (2008)]

'What's that building over there?" I asked. I saw a big building at the foot of the mountain, with some steel frames still exposed on the top of the building and I was a bit distracted. (snip) "You mean that one? Oh, that's a hotel."'

Then, what's the difference between *đó* and *kia*? First, as already mentioned in section 3.1, *đó* connotes that the reference is new and unfamiliar to the speaker. *Kia* doesn't have such connotation.

Second, *đó* can mention a vague referent in the form of 'wh-words + *đó*', which means 'someone', 'something' 'somewhere' and so on. Following examples are quoted from Đoàn et al. (2005: 140).

(6) *Trước khi đi ngủ, em ấy thường bắt mẹ kể*
before time go sleep (s)he always catch mother tell
*một câu chuyện nào { *này / đó / *kia }.*
one story which this that yonder
'Before going to bed, (s)he always nags her mother to tell a story.'

(7) *Chiều nay có ai { *này / đó / *kia } đến tìm anh đấy.*
afternoon this have who this that yonder come find elder brother FP
'Someone came to look for you this afternoon.'

(8) *Cậu có rỗi không? Đi đâu { *này / đó / *kia } cho mát đi.*
uncle Q free Q go where this that yonder for cool FP
'Are you free now? Let's go somewhere to enjoy the cool air.'

(9) *Chiều nay mẹ nấu món gì { *này / đó / *kia } thật ngon nhé.*
afternoon this mother cook food what this that yonder really tasty FP.
'I'll make you something really tasty this afternoon, OK?'

Third, sometimes *kia* pairs up with *đây*, which means 'one and the other'. In this case, *kia* can't be replaced by *đó*. Following examples are quoted from Hồ et al (2006: 1726).

(10) *(Tay này) giữ cái chai, tay { *này / *đó / kia } mở nút chai.*
hand this hold CL bottle hand this that yonder open cork bottle
'Hold the bottle and pull the cork out with the other hand.'

(11) *Anh có thể viết tiếp ở mặt { *này / *đó / kia } tờ giấy.*
elder brother can write continue on side this that yonder CL paper
'You may continue on the other side of paper.'

[5] Some sentence-final particles and interjections have the same appearances demonstratives as full NP, local adverbial and in NP with noun, but, having no evident referent, can be omitted from the sentence. They denote, in addition to the spatial distance, the addressing to the addressee as well as the speaker's attitudes such as intimacy, surprise and aversion. For details, see Adachi (2010).

Fourth, *đó* generally doesn't have a pair, though *đây* and *đó* represent the speaker and the addressee respectively when they are treated as personal pronouns. Nguyễn Phú Phong (1992:133, 2002:139) cites examples as shown in (12), (13) and (14). In some literary works, this usage can be seen in depiction of intimacy and dialectal features of the characters (Trần 2009). However, according to my language consultant[6], it is not used in everyday conversations.

(12) *Đây đi chợ, đấy có đi không?*
this go market that Q go Q
'I'm going to the market. What about you?'
[Nguyễn (1992: 133)]

(13) *Đó để đây làm cho.*
that let this do for
'Let me do for you.'
[Nguyễn (2002:139)]

(14) *Đấy vàng, đây cũng đồng đen. Đấy hoa Thiên lý, đây sen Tây hồ.*
that gold this also copper black that flower jasmine this lotus West Lake
'If you are worth gold, then I am brass. If you are jasmine, then I am lotus of West Lake.'
[Nguyễn (2002:139)]

4. Anaphoric use

Đó has typical anaphoric use. That is to say, it can refer to a non-specific referent and a referent determined by an antecedent in the text as shown in (15), (16) and (17).

(15) *Nếu kịp tàu siêu tốc thì chúng ta đi tàu { *này / đó / *kia} nhé.*
If in time train express then we go train this that yonder FP
'If we can catch an express train, let's go by it.'
(They haven't known which express train comes first yet.)
[Japanese example is quated from Kinsui et al (2002: 220).]

(16) *Anh đi đâu tôi đi { *đây / đấy / *kia }.*
elder borther go where I go here there yonder
'Wherever you go, I'll follow you.' [Đoàn Thiện Thuật et al (2005: 114-115)]

(17) *Sinh viên nào đến hội trường, xin đưa giấy này*
student which come hall please hand paper this
*cho sinh viên { *này / đó / *kia }.*
to student this that yonder
'Please hand this document to any student who comes to this hall.'
[Japanese example is quated from Kinsui et al (2002: 220).]

Đây also can be used in the anaphoric use. It can refer to an antecedent which is introduced into the context with a text, and can also make reference to a non-specific antecedent as shown in (18).

(18) *Tôi sẽ kết hôn với nhà doanh nghiệp trẻ có nhiều tài sản.*
I will marry with businessman young have many property
Người nào cũng được. Và, tôi sẽ đi du lịch khắp thế giới
person which also OK and I will go travel around world
*với người { này / đó / *kia }.*
with person this that yonder
'I'll marry anyone, as long as he is a rich, young entrepreneur.
And I'll travel around the world with him.'
[Japanese example is quated from Kinsui et al (2002: 224).]

[6] The language consultant is a native Vietnamese speaker (female), born in Hanoi in 1968.

5. Memorative use

Usually *đó* is used to mention the referent in the knowledge of the speaker.

(19) (Suddenly policemen rushed into the room to arrest the man who had been there a while ago.)
*Người { *này / đó / *kia } ở đâu rồi?*
person this that yonder be where already
'Where did he go?'
[Japanese example is quated from Kinsui et al (2002: 220-221)]

(20) *Lâm đợi cho Loan đi rồi, hỏi vợ:*
Lam wait let Loan go already ask wife
- *Thế nào?*
how
- *Thế nào gì?*
how what
- *Việc { *này / **đó** / kia} thành bại?*
matter this that yonder success fail

'Lâm waited until Loan was gone, and then he asked his wife:
- How was it?
- About what?
- How's the thing, I mean the outcome?'
(Mr. Lam was trying to ask his wife if Ms. Loan had accepted the proposal from Mr. Dung or not.)
[Nhất Linh (1983)]

Kia as well as *đó* also have memorative use. Adachi (2008) takes example (21) and states "*kia* refers to the secret between the speaker and the addressee or in a small group".

(21) *Chuyện{ đó / kia} thế nào rồi?*
matter that yonder how already
'How do you make out in that matter?'

According to my informant, if *kia* is used in (20), it also sounds like Mr. Lam and his wife are talking about the secret, and they don't want let it out. On the other hand, the use of *đó* emphasizes that the proposal is not Lam's business but the other people's. Thus, it can be said that, *đó* refers to the referent which the speaker feels to be unfamiliar with him, while *kia* indicates the object the speaker directly knows and feels to be familiar with him. Yoshida (1980) states that in some languages demonstratives refer to the referent mutually and tacitly understood by the speaker and the addressee although it was not mentioned in the former discourse. Vietnamese *kia* has the taxit function. There are two more examples.

(22) *Thằng {*này / **đó** / kia} là da đen hay da trắng?*
guy this that yonder COP skin black or skin white?
'Is he black or white?'
(The speaker suspects that his girlfriend is cheating on him.)
[Yamada (2008)]

In the example (22), We can also use *thằng kia*, however the speaker must hate the man who had an affair with his girlfriend, so *đó* is more appropriate in this case.

(23) *Những ngón tay thon thả kia, những cử chỉ kia,*
some finger hand slender yonder some expression yonder
và rồi cả cái điệu bộ kia nữa, mà là đàn ông ư?
and then all CL mannerism yonder more but COP man FP
'Those slender fingers, expressions and mannerisms... and yet he is a GUY??"'
[Yoshimoto (2008)].

In the example (23) the speaker can't believe that her best friend's mother is a man in fact because he totally looks like a woman. The mother is not present on that occasion, so the speaker is visualizing the finger and expressions, which is familiar to her.

The similarity and the difference between *đó* and *kia* is summarized as follows.

i. Both *đó* and *kia* have deictic and memorative uses in common, however, the difference between cannot be explained on the base of the position of the addressee.

ii. It seems to quality of the knowledge of the speaker is rather related to the classification of the uses.

iii. While *kia* refers to an object which is already determined and the speaker directly know, *đó* refers to an object which is still indetermined and the speaker indirectly knows.

6. Summary and Conclusion

The present paper examines how Vietnamese demonstratives can be classified. The results are as follows.

1. *Đây* (proximal) directly indicates a referent close to the speaker. It can refer to an antecedent which is introduced into the context with a text, and can also make reference to a non-specific antecedent.

2. *Kia* (distal) points out a visible referent far from the speaker. Sometimes *kia* pairs up with *đây*, which means 'one and the other'. While *kia* doesn't have anaphoric use, it occasionally refer to an object familiar to the speaker within the direct experience of the speaker.

3. *Đó* (medial) indicates either a visible or invisible referent far from the speaker regardless of the existence of the addressee. *Đó* doesn't always point out the place between *đây* and *kia*. The speaker often has an intuition that the referent is a new and unfamiliar thing to him or her. It can also mention a vague referent in the form of 'wh-words + *đó*'. *Đó* generally doesn't have a pair, though *đây* and *đó* represent the speaker and the addressee respectively when they are treated as personal pronouns. *Đó* has typical anaphoric use, and moreover it can make reference to an object, which the speaker feels to be distant, in direct experience of the speaker.

Thus, it can be said that Vietnamese demonstratives has three-term distance oriented system according to the definition of Anderson and Keenan (1985). As can be seen in Table 3, *đây* consistently refers to the referent physically and psycologically close to the speaker. *Đó* also makes reference to the referent far from the speaker. However, *kia* points out the object physically remote from the speaker in deictic use while it refers to the referent the speaker feels close to him or her in memorative use. At this time, I hyposesize that visiblity and concreteness in dexis are related to psycological clossness to the referent of the speaker in memorative use. Moreover, the distionction of *đó* and *kia* when they refers to a visible object remote from the speaker in deictic use as well as the relation between the addressee and *đó* still remain unclear, thus, further investigation is needed.

Table 3. Vietnamese demonstratives (modified)

	Deictic			Anaphoric	Memorative
	Speaker	Addressee	Others		
đây	near	near / remote	familiar	near	---
đó	remote	near / remote	unfamiliar vague	remote	remote
kia	remote	near / remote	visible, directly recognizable 'the other'	---	near tacit

Abbreviations

CL: Classifier COP: Copula FP: Final particle Q: Question

Quoted novels

Nguễn, Nhật Anh. (2005) *Kính Vạn hoa* [Kaleidoscope] 1. Hà Nội: Nhà xuất bản Kim Đồng.
Nhất Linh. (1983) Takeuchi Yonosuke (trans.) *Đoạn Tuyệt* [Break off]. Tokyo: Daigaku Syorin.
The original work: Nhất Linh. (1935) *Đoạn Tuyệt* [Break off].
Yamada, Amy. (2008) Lương Việt Dũng (trans.) *Đôi mắt ấy vẫn ở trên giường* [The eyes on the bed]. Hà Nội: Nhà xuất bản Hội Nhà văn.
The original work: Yamada, Eimi. (1985) *Beddo Taimu - aizu* [Bedtime eyes]., Tokyo: Kawade Shobo Shinsha.
Yoshimoto, Banana. (2008) Vũ Hoa (trans.) *Tsugumi*. Đà Nẵng: Nhà xuất bản Đà Nẵng.
The original work: Yoshimoto, Banana. (1983) *Tsugumi*. Tokyo: Chuokoron- Shinsha.

References

Adachi, Mayumi. (2008) “Betonamugo shijishi *đây*, *đó*, *kia* no chokuji youhou to shouou youhou ― nihongo shijishi to no taishou wo moto ni ― [Deictic use and anaphoric use of Vietnamese demonstratives *đây*, *đó*, *kia*: based on comparison with Japanese demonstratives]” , *Tokyo daigaku gengogaku ronshuu* [Tokyo University linguistic papers] 27, 207-216.
Adachi, Mayumi. (2010) “Betonamugo no bunmatsushi [Sentence-final particles in Vietnamese]: *đây, này, đấy, ấy* and *kia*. *Tokyo daigaku gengogaku ronshuu* [Tokyo University linguistic papers] 29, 1-7.
Anderson, Stephen R. and Edward L. Keenan. (1985) “Deixis.” Timothy Shopen (ed.) *Language typology and syntactic description: grammatical categories and the lexicon*. Cambridge: Cambridge University Press, 259-308.
Dixon, R. M. W.. (2003) “Demonstratives: a cross-linguistic typology”. *Studies in Language* 27, 61-112.
Đoàn Thiện Thuật et al. (2005) *Thực hành tiếng Việt, trình độ C - sách dùng cho người nước ngoài* [Practical Vietnamese, level C – a book for the use of foreigners], Hà Nội: Nhà xuất bản Thế Giới.
Emeneau, Murray. (1951) *Studies in Vietnamese (Annamese) grammar*. Berkeley: University of California Press.
Hồ Hải Thụy et al. (2006) *Từ điển Anh-Việt* [English-Vietnamese dictionary]. Thành Phố Hồ Chí Minh: Nhà xuất bản Thành Phố Hồ Chí Minh.
Kinsui, Satoshi., Tomoko Okazaki., Mikyon Jo.. (2002) “Shijishi no rekishiteki, taishougengogakuteki kenkyuu: nihongo, kankokugo, torukogo. [A historical and comparative study of demonstratives: Japanese, Korean, and Turkish]”, Ogoshi Naoki. (ed.) *Taishougengogaku* [Comparative Linguistics]. Tokyo: University of Tokyo Press, 217-247.
Nguyễn, Đình-Hòa. (1997) *Vietnamese – Tiếng Việt không son phấn* [Vietanamese without veneer]. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Nguyễn, Phú Phong. (1992) “Vietnamese demonstratives revisited”, *Mon-Kumer Studies* 20: 127-136.
Nguyễn, Phú Phong. (2002) *Những vấn đề ngữ pháp tiếng Việt – loài từ và chỉ thị từ* [Some problems of Vietnamese grammar – classifiers and demonstratives]. Hà Nội: Nhà xuất bản Đại Học Quốc Gia Hà Nội.
Thompson, Laurence C.. (1987) *A Vietnamese Reference Grammar*. Hawaii: University of Hawaii Press.
Tomita, Kenji. (1988) “Vetonamugo [Vietnamese].” Kamei, Takashi, Rokurou Kouno and Eiichi Chino (eds.) *Gengogaku Daijiten* [Linguistic Encyclopedia] 1, Tokyo: Sanseido, 759-787.
Trần, Thị Nhàn. (2009) “Ý nghĩa và chức năng của từ *đây* [Meaning and function of the word *đây*]”. *Ngôn ngữ* [Language] 10, 11-22.
Yoshida, Shuuji. (1980) “Shijishi ni mirareru kuukan bunkatsu to sono fuhensei [Typology and universality of the cognition of space division through the analysis of demonstratives]”. *Kokuritsu minzokugaku hakubutsukan kenkyuu houkoku* [Bulletin of the national museum of ethnology] 5, 3, 833-950.

Is Shom Pen a Distinct Branch of Austroasiatic?

Roger Blench
Kay Williamson Educational Foundation
(Cambridge)

Paul Sidwell
Center for Research in Computational Linguistics (Bangkok),
Australian National University (Canberra)

1. Introduction

The Shom Pen language is spoken on Great Nicobar by a few hundred hunter-gatherers. Until recently, the language was too poorly known to make any definitive statement about its affiliations, although it has usually been considered part of Nicobarese. However, the availability of new materials (Chattopadhyay & Mukhopadhyay 2003), have made possible a more extended analysis. One of the authors has argued in print (Blench 2007) that the Shom Pen language may well be an isolate, not a relative of Nicobarese. In this paper[1] we would like to put forward an alternative view that Shom Pen may be a distinct branch of Austroasiatic, but not part of Nicobarese. The bases for this are cognates with Austroasiatic that do not appear to be shared with Nicobarese.

2. The Shom Pen Language: previous studies

Until recently, the language of the Shom Pen had remained unknown apart from ca. 100 words recorded by De Roepstorff (1875), the scattered lexical items in Man (1886) and the comparative list in Man (1889). Although our knowledge of Nicobarese is imperfect there are several book-length sources for this group, for example Whitehead (1925), Das (1977) and Radakrishnan (1981). There is also a short ethnography by Rizvi (1990) which includes some scant linguistic data, mainly kin terms.

Although most reference books list Shom Pen as part of the Nicobarese group, evidence for this has been modest. It appears to rest on a number of isoglosses, and the assumption that being spoken in the Nicobars, it is likely to be related to the other languages of the archipelago. Rather tellingly among early sources, Man (1886:436) says; 'of words in ordinary use there are very few in the Shom Pen dialect which bear any resemblance to the equivalents in the language of the coast people'. Man (1886) also observes substantial linguistic variation between Shom Pen settlements;

> In noting down the words for common objects as spoken by these (*dakan-kat*) people I found that in most instances they differed from the equivalent used by the Shorn Pen of Lafal and Ganges Harbour.

And similar references to at least two distinct groups are made by Rizvi (1990).

However, since the 1990s a couple of new sources have appeared that have improved our knowledge somewhat, enough that we can begin to offer some analysis and hypotheses worthy of further investigation:

- Chattopadhyay & Mukhopadhyay (2003), is a lexicon of about 750 words of the Shom Pen language, and includes some discussion concerning the affiliation of Shom Pen.

Blench, Roger and Paul Sidwell. 2011. "Is Shom Pen a distinct Branch of Austroasiatic?" In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.9-18.

- Elangaiyan et al. (1995) is a Shom Pen-Hindi primer with some 70 words transcribed in Devanāgarī script, and subsequently retranscribed into Roman script by van Driem (2008).

Van Driem (2008) reports that considering the data in Elangaiyan et al. (1995) and de Roepstorff (1875) he found (with assistance from Gerard Diffloth) out of a total of 222 lexemes, 109 with Nicobarese cognates, 57 with Southern Mon-Khmer[7] cognates, while 7 of the latter group have no apparent Nicobarese etymologies. He concludes that Shom Pen does belong within the Nicobarese group. Van Driem also discussed another source, Chattopadhyay & Mukhopadhyay (2003), but concluded that problems with the data made it too difficult to work with.

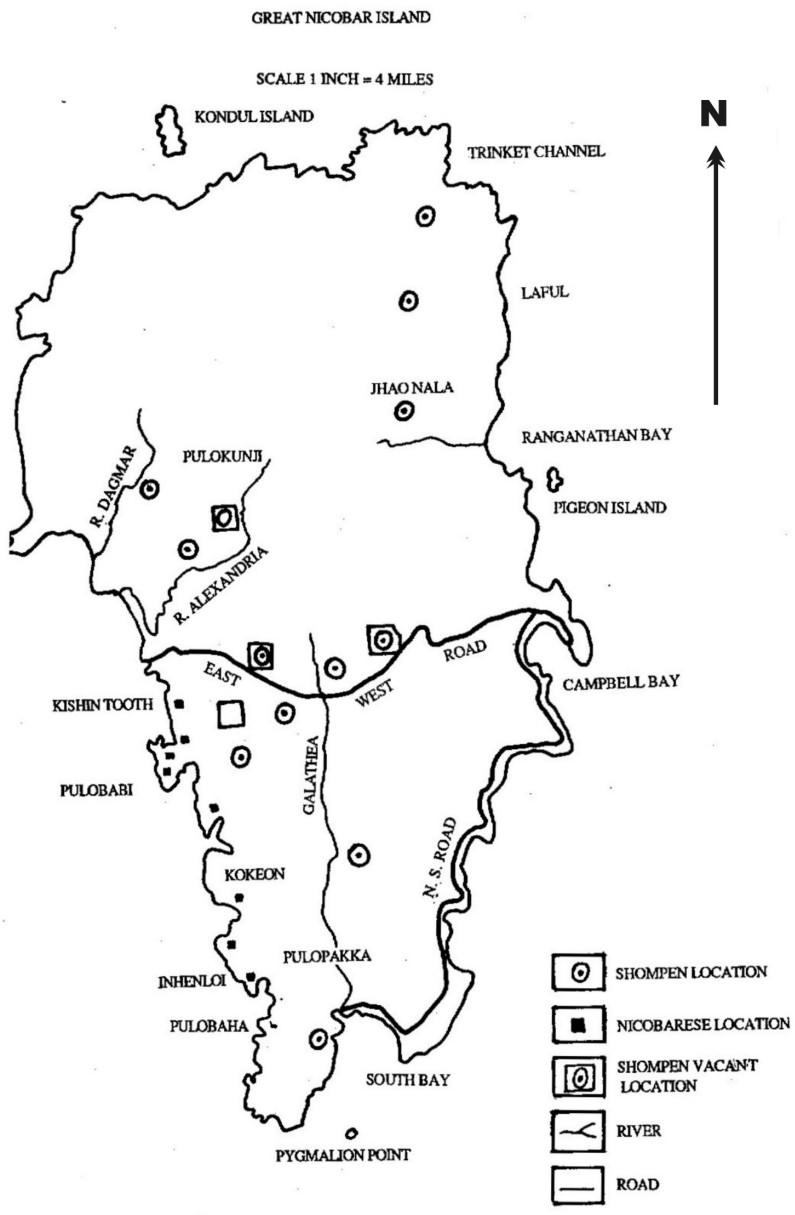


Figure 1: Map showing locations of Shom Pen Communities (Risvi 1990:8).

[7] The term "Southern Mon-Khmer" being Diffloth's (2005 and passi.) category for Monic, Aslian and Nicobarese, which is not necessarily accepted by the present writers.

3. The Chattopadhyay & Mukhopadhyay (2003) data interpretation

Chattopadhyay & Mukhopadhyay (2003, henceforth C&M), makes available a significant body of new data on the Shom Pen language, enough to construct a better argument for the affiliation of Shom Pen. Unfortunately the source is not to modern standards of presentation and analysis, and in fact contains various anomalies that suggest that it originates with an earlier - perhaps colonial era - manuscript:

- the authors make no explicit claim of gathering the linguistic data in the field, only that they "collected information" from Shom Pens at the Shom Pen Hut complex
- much of the ethnographic information is lifted directly from Rizvi (1990) without attribution,
- there is no serious attempt at segmentation into words/morphemes, although many entries are transparently phrases or compounds,
- there are serious anomalies in the explanation and use of the phonetic transcription.

The last point is most telling. We are told that there are both short and long low central vowels /a, ā/, but no other central vowels (e.g. no schwa) and no length contrast elsewhere in the vocalism. Then, oddly on p.12 we are told that, "One characteristic feature of Shompen vowel system is the pronunciation of the same vowel in succession without lengthening the vowel." This is indicated with a double vowel letter spelling, and always appears to occur within the same syllable peak. There is no indication of hiatus between such vowels, and both /h/ and /ʔ/ are discussed elsewhere unproblematically. To us this really looks like a naïve (non-linguistic) reading of a third party manuscript rather than an original piece of work.

Both internal and external evidence suggest that the a-macron is an attempt to notate a schwa or similar weak vowel, whose absence would be typologically/areally rather astounding. First of all, the /ā/ sign has a distinctive distribution:

1. in non-final syllables,
2. in final syllables apparently as on-glide/off-glide of diphthongs,
3. in Malay loan words, and close Nicobarese and Aslian cognates, /ā/ corresponds to either /ə, ɨ, ɯ/ or unstressed /a/.

Examples:

C&M Shom Pen	Malay	Other AA
bāteāu 'catch (s.t. falling)'	*bantu*	Jahai *bantuʔ* 'to catch something that is falling'
kādāb 'salt'	*garam*	Jahai *garəm* 'salt'
wānāŋ 'thread'	*benang*	Jahai *bnɛŋ* 'thread'
nāŋ 'ear'		Nancowry *naŋ* 'ear'
ɸeāo 'to strike'		Car *feːl* 'strike (with hand, fist)'
hãĩ 'nasal mucus'		Nancowry *hehɛː* 'snivel mucus from the nose'
kāi 'a skin disease'		Temiar *maŋkʌi* 'wound, scar, scabies'
toāy 'canoe'		Nancowry *ruɯe ~ duɯe* 'canoe'
ɸewāi 'measure of full span'		OldKhmer, Chrau etc. *was* 'to measure'
xɛ̃āi 'blood vessel'		Nancowry *ihɛː*, Kuy *rsai* 'vein'
poeā 'lift'		Chrau *pəːʔ*, OldKhmer *pəːk* 'to lift'

Only in some ambiguous cases does the /ā/ possibly correspond to a long /aː/ in AA etymologies. These may still be via a short vowel, and thus would not invalidate our claim, e.g. (note presyllable ā in the first example):

C&M Shom Pen	Other AA
kācām 'to wait'	Khmer, Kuy etc. *cam*, but Stieng *ca:m* 'to wait'
tām 'tooth'	Bahnar *təʔa:m* 'molar'

The double vowel notation is infrequent, but where is occurs it does appear to indicate long vowels, e.g.:

C&M Shom Pen	Other AA
peii 'full'	Nancowry *paha:e*, Jahai *bəhiʔ*, PMK **bhi:ʔ* 'sated, full'
kāʔeem 'bone'	Jahai *ɟʔeŋ*, PMK **cʔa:ŋ* 'bone'
mooijob 'fireplace'	Car *muɯiə* 'fireplace'(note: Shompen *job* 'fire')

The above observations allow us to make a number of suggestions in terms of the appropriate phonetic interpretation of the C&M data. We can also suggest that in various places the digraph *in* is intended to represent the palatal nasal *ɲ*, and *i* and *u* should be treated as glides, and some other Roman to IPA equivalents. Subsequently below we will suggest a preliminary phonemic representation of forms in curly brackets { } that converts a-macron to schwa, and double vowels with their IPA length mark.

4. Comparison of Chattopadhyay & Mukhopadhyay with Elangaiyan lists.

It is important, first of all, to determine if there is a reliable relationship between the two important recent sources discussed above. Both have notational problems, but is none-the-less clear that in a number of cases the same word is being represented. We have compiled the following 25 comparisons:

C&M Shom Pen	**Elangaiyan Shom Pen**	**Nicobarese cognates**
giob {gjob} 'nail'	*ugiyøv* 'fingernail'	
juāg {ɟuəg} 'spider'	*jøva:k* 'spider'	
kagai {kagaj} 'child'	*o:kʔay* 'child'	
kāiāi {kəjəj} parrot'	*kayayøy* 'parrot'	
kohãt {kohət} 'girl'	*ø:kʔa:t* 'girl'	
kokugāu {kokugəw} 'scorpion'	*giya:v* 'scorpion'	
komeāin {komeəŋ} 'forehead'	*kuma:ñ* 'forehead'	
komeoigo {komeoigo} 'sand'	*tyõvgo:* 'beach, sand'	
kotoõi {kotõ:j} 'lip'	*to:y* 'lip'	
luou {luow} 'necklace'	*lovvu* 'necklace'	
neāu {neəw} 'cheek'	*miyøv* 'cheek'	
meāin {meəŋ} 'eye'	*hma:ñ* 'eye'	
nuŋāi {nuŋəj} 'snake'	*hmøŋøy* 'snake'	
omeoin {omeʔõj} 'banana'	*mø:ʔøy* 'banana'	
phiŋuāin {phiŋuəŋ} 'throat'	*nuva:ñ* 'neck'	
uŋāu {uŋəw} 'bee'	*phøŋø* 'bee'	
xiug {xiug} 'sun'	*hiv ~ hi:v* 'sun'	
coāi {coəj} 'monkey'	*cyo:y* 'macaque'	Car *ʔəɲci* 'monkey'
jiāu {ɟiəw} 'crab'	*løgø:v* 'crab'	Car *ʔin-ji:w* 'a land crab'
eab {ʔeap} 'centipede'	*œyʔev* 'centipede'	Nanc. *kaʔeap* 'centipede'
guiāo {gujəw} 'coconut'	*høya:v* 'coconut'	Nanc. *ʔojaw* 'coconut'
lau {law} 'thigh'	*løv* 'thigh'	Nanc. *pulɔ:* 'thigh'
nāŋ {naŋ}	*naŋ* 'bamboo'	Nanc. Car *naŋ* 'ear'
niãi {niãj} 'house'	*ñiyo* 'house'	Nanc. *ɲi:* 'house'
ŋo {ŋo}	*ŋaŋvo* 'bamboo'	Nanc. *ŋoa* 'bamboo'

Given the notational issues, and the lack of perfect corresponding glosses, 25 out of 70 is a very respectable score. It also includes several isoglosses that appear to lack Nicobarese cognates, e.g.: 'sun',

'bee', 'fingernail', child', 'eye', 'snake'. Therefore it seems clear that the same language or two closely related languages are represented in these sources.

5. Austroasiatic etymologies.

Blench (2007) sets out the data from C & M in an IPA-like transcription and notes some obvious etymologies. In this paper we go further, and present the results of a fairly exhaustive comparison of the entire C&M corpus with the *Mon-Khmer Comparative Dictionary* of Shorto (2006). In summary, we identify 65 lexemes/morphemes that have reasonable Austroasiatic etymologies that go beyond Nicobarese or even Aslian, but have cognates in other branches of the phylum. These comparisons achieve several goals; 1) the Austroasiatic heritage of Shom Pen is convincingly demonstrated, and 2) we have a baseline for assembling the beginnings of an historical phonology of Shom Pen. Our compilation follows:

C&M Shom Pen	Short (2006) index and reconstruction
tai {taj} 'hand'	66.A **ti:ʔ* 'hand, arm'
kāināi {kəinəj} 'bat'	93.A **kn[i]ʔ* 'rat, mouse'
poeā {poeə} 'lift'	100a.A **pə:ʔ* 'lift'
mɛd-hẽu {mɛd=hẽu} 'weep'	1045.A **mat* 'eye'
apāo {ʔapəw} 'wife'	113.A **[ʔ]boʔ* 'mother'
niãi {ɲiəj} 'hut'	152.A **[j]aa[ʔ]* 'house'
lau {law} 'thigh'	223.A **blu:ʔ* 'thigh'
laʔõõy {laʔ=ʔõ:j} 'yellow leaf'	230.A **slaʔ* 'leaf'
okināu {okinəw} 'bat'	237.A **klwaʔ* 'bat'
xẽãi {xẽəj} 'blood vessel'	249a A **[ ]rsi:ʔ* 'nerve, sinew, vein'
nāɸe {nəɸe} 'yawn'	251.A **haʔ* 'to open [mouth]'
hãũ {hə̃w} 'wood'	254.A **jho:ʔ* 'wood'
peii {pe:j} 'full'	259 A **bhi:ʔ* 'sated, full'
okheāg {okheəg} 'vomit'	474 A **hɔ:k* 'to vomit'
guāg {guəg} 'knee'	486.A **kʔɔŋ* 'knee '
kāʔeem {kaʔe:m} 'bone'	488 A **cʔa:ŋ* 'bone'
kogāg {kogəg} 'branch (of tree)'	496 A **kaŋ* 'transverse, to branch'
koāuŋ {koəuŋ} 'leaf stalk'	506 A **tkuəŋ* 'stalk'
khoāg {khoəg} 'boiling'	513.A **guəŋ* 'to cook in water'
keuŋ {keuŋ} 'pillar'	518 A/B **pgɔŋ; *pgɔɔŋ* 'beam'
gināug {g<in>əug} 'beam (horizontal)'	518.A/B **pgɔŋ; *pgɔɔŋ* 'beam'
kocuoug {kocuoug} 'long'	537.A/B **jo[o]ŋ* 'long, high'
(ko) ceog {ko=cə:g} 'foot'	538(I).D **jə:ŋ* 'foot, leg'
nɛtoɔŋ {nɛtoɔŋ} 'carry on shoulder'	548.A **tu:ŋ* 'to carry on pole between two'
kouāu {kouəw} 'skin, husk'	568 A **cnkuər* 'integument.'
nāidɔŋ {nəj=dɔŋ} 'stick'	585.A **tɗɔ:ŋ* 'stalk, tail'
peāg {peəg} 'cockroach'	630.C **ɓiəŋ* 'spider'
nyɛŋ {njɛŋ} 'horn'	699.A **draŋ* 'horn.'
lāuŋ {ləuŋ} 'open (bag etc.), unfold'	720.A **la[:]ŋ* 'to unfold'
tameauŋ {tameauŋ} 'mouth'	911.D **muəɲ[ ]* 'mouth'
cuoid {cuoid} 'heavy'	1146.A **ɟən* 'heavy '
phugāi {phu=gəj} 'four'	1166.B **puən* 'four'
eab {ʔeap} 'centipede'	1226.A **kʔip; *[k]ʔiip; *kʔiəp etc.* 'centipede'
hɛkāb {hɛkəb} 'bite'	1231.B **ka:p* 'to bite'
kolheɔb {kolheɔb} 'ant'	1290.A **kla:p* 'flying white ant. '

toɔb {toɔb} 'kiss'	1296.A **thəːp* 'to sniff'
tām {təm} 'tooth'	1303.A **[t]ʔaːm* '(molar) tooth'
kāyāb {kəjəb} 'cloud'	1305.A **[m]ʔəːm* 'cloud'
toʔŋeāum {toʔ=ŋeəum} 'cheek'	1318.B **dgaːm* 'molar tooth, jaw'
kācām {kə=cəm} 'stay, rest'	1325.B **caːm* 'to wait, watch'
kāteob {kəteob} 'red'	1362.B **duːm* 'ripe, ripe-coloured'
leɔb {leɔb} 'leech'	1410.B **tləəm* 'land leech'
hoɔp {hɔːp} 'wash, clean (with water)'	1426.A **huːm* 'bathe'
tiub / teub {tiub} 'blood'	1430.A **ɟhaːm* 'blood'
koi {koj} 'head; top'	1443.A **kuj* 'head'
kɜ̃i {kɜ̃j} 'lizard (house)'	1446.A/B **[d]kuːj;*[d]kuəj* 'Calotes lizard'
kākoāy {kə=koəj} 'sit'	1448.A **ŋguj* 'to sit down'
thaāgge {thaʔəg=ge} 'thin'	1451.A **[ ]rgəj* 'thin, lean'
kojāi {koɟəj} 'fly'	1534.A **ruj* 'fly'
tĩẽu {tĩẽu} 'round'	1625.B **diər* 'to circle round'
hekāu {hekəl} 'break'	1702.A **kal* 'to break off'
kāokāo {kəw=kəw} 'bark'	1709.B **[c]kuəl* 'bark.'
kāo {kəw} 'belly'	1721.B **sgə[ ]l* 'middle'
deāu {deəw} 'heel'	1748.D **kɗəəl* 'heel'
pĩ ɜ̃i {pĩ=ɜ̃j} 'rain'	180.A **briːʔ* 'sky, rain'
oleāu {oleəw} 'man'	1857.A **klaːw* 'male sexual organs'
ikheāu {ik=heəw} 'get up'	1869.A **haːw* 'to climb, ascend'
gikoāi {gi=koəj} 'wash; ich'	1881.B **kuəs* 'to scrape, scratch, shave'
napĩɜ̃i {napĩɜ̃j} 'sweep (water)'	1916.D **tpiːs* 'to sweep'
(ko) ghiāu {ɣiaw} 'root'	1927.A **ris* 'root'
ɸewāi {ɸe=wəj} 'measure of full span'	1951.B **wəs* 'to measure'
kāhooi {kə=hoːj} 'wind'	1958.E **ha[ ]h* 'to blow'
mɛthou {mɛ=thou} 'nipple'	1999.A **tɔh* 'breast, mamma'
mhou {mhow} 'nose'	2045.A **muh* 'nose'

The above list, at 65 items, reflects approximately 10% of the lexicon represented in C&M, given a certain amount of redundancy in the original. Given the various difficulties acknowledged above, this is probably less than is really represented in the data. On the basis of this, we can make a number of observations:

1. Many, although not all, Austroasiatic word final nasals are realized as voiced stops.
2. Prevocalic rhotic /r-/ is realized as the glide /j-/.
3. Word final /-h/ is realized as zero.
4. Word final /-r, -l/ are usually realized as glides /-j, -w/.
5. Austroasiatic long vowels frequently become diphthongs.
6. Regular development of voice onset timing is not evident, perhaps reflecting unreliable transcription.

Items a) and b) above may be most significant, these are discussed in the following sections.

6. Aslian connection?

The hardening of final nasals to oral stops in Shom Pen, already remarked upon by van Driem (2008), is especially characteristic of Northern Aslian languages, although some degree of pre-stopping of final nasals occurs in much of Aslian. In this context it is significant that we have compiled more than 30

Shom Pen-Aslian isoglosses, lacking apparent Nicobarese cognates, and these comparisons, perhaps 90% of them, only have matches, or have their best matches, with Northern Aslian languages such as Jehai and Kensiu. This is perhaps the most difficult aspect of this question for us to investigate, because we do not have extensive Aslian data for all sub-groups, so there may be some skewing of the data. Yet it is apparent that in our data matching, the etymologies with good South Aslian matches strongly tend to have general Austroasiatic etymologies. Our list of Shom Pen-Aslian isoglosses follows (Aslian items have MKLP indices included):

C&M Shom Pen	Aslian isoglosses
ecoāu {ʔecoəw} 'blind'	Jehai *cūʔ* 'to be blind' (Bur2005:C:292)
kaiug kao {kaiug kaw} 'carry'	Jehai *gɔl* 'to carry on one's back or shoulder' (Bur2005:C:354)
gigāb {gi=gəb} 'waist'	Jehai *gɛl* 'waist' (Bur2005:C:359)
geāu {geəw} 'belching'	Jehai *gos* 'belch' (Bur2005:C:390)
okɔy {okɔj} 'slow'	Jehai *hakɨj* 'to be slow' (Bur2005:C:420)
kaiugŋhã̄ {kaiug=hə̃} 'road'	Jehai *har* 'small path, trail' (Bur2005:C:440)
tāogheu {təw=ghew} 'take'	Jehai *jɔw* 'to take' (Bur2005:C:549)
anikā {ʔanikə} 'right (side)'	Jehai *lokaʔ* 'right' (Pha2006:C:475-2)
kaugāi {kaugəj} 'feed'	Jehai *prgɛj* 'to feed' (Bur2005:C:1206)
gõāŋ {gõəŋ} 'frog (non edible)'	Jehai *riŋkoŋ* 'frog (type of)' (Bur2005:C:1275); Kensiu *ʔikegŋ* 'green frog' (Bis1994:C:1578)
opugāi {opu=gəj} 'mosquito'	Jehai *ʔagas* 'mosquito' (Pha2006:C:395-2)
job {ɟob} 'fire'	Kensiu *cəbm ʔɔs* 'to start a fire, to ignite a fire' (Bis1994:C:165)
cuāg {cuəg} 'stream'	Kensiu *ɟənọgŋ* 'stream, creek' (Bis1994:C:1447)
kāitohe {kəit=ohe} 'cut (wood etc.)'	Kensiu *kɛc* 'to cut' (Pha2006:C:687-1)
ɣiʔuŋāi {jiʔuŋ=əj} 'Adam's apple'	Kensiu *lahọgŋ* 'larynx; esophagus' (Bis1994:C:691)
tauŋhe {tauŋhe} 'kick'	Kensiu *tənagŋ* 'to kick' (Bis1994:C:1323)
hə̃̄ub {hə̃ub} 'close, shut'	Semai *dahup* 'to shut, to close (a door, a mouth, etc.)' (Mea1987:C:4503-7)
kāu {kəw} 'bend'	Semai *gʌaʊ* 'crooked, curved, bent' (Mea1987:C:1051)
kāpuāu {kəpuəw} 'soft'	Semai *ɟʌbuʔ* 'soft, pleasant to touch' (Mea1987:C:3003)
kwə̃̄o {kwə̃w} 'frog (edible)'	Semai *kaɪoʊʔ* 'edible toad, giant frog' (Mea1987:C:1315)
gigou {gi=gow} 'pour; strain off'	Semai *kɔh* 'to pour water into (something)' (Mea1987:C:4314)
heiei {hejej} 'smell'	Semelai *hajĩr* 'smell of burning flesh or bones, or postnatal blood' (Kru2004:C:1031)
giāo {ɟəw} 'snake'	Semelai *tiɟɔ* 'snake (noun)' (Kru2004:C:1039); Jehai *taɟuʔ* 'snake (generic)' (Bur2005:C:1471)
ugābeāu {ugə=beəw} 'child (male)'	Temiar *bə* 'be male (adjective)' (Mea1998:C:230)
gipẽ̄ {gi=pẽ} 'forest'	Temiar *beʔ* 'forest, jungle (noun)' (Mea1998:C:291)
diāi {diəj} 'hammer'	Temiar *dʌdal* 'to hit, to hammer, to pound' (Mea1998:C:707)
hẽ̄in {hẽj} 'move'	Temiar *hɔit* 'to go, to move along' (Mea1998:C:1071)
guice {guice} 'many'	Temiar *ɟaɟeʔ* 'many, much, abundant' (Mea1998:C:1118)
dɛo {dɛw} 'seat of canoe'	Temiar *kendeʔ* 'seat, chair (noun)' (Mea1998:C:1403)
kāi {kəj} 'skin disease sp.'	Temiar *maŋkʌi* 'wound, scar, scabies, injury (noun)' (Mea1998:C:1689)
hāneɸeāg {həne=ɸeəg} 'slap'	Temiar *paːg/* 'to slap, to spank (transitive verb)' (Mea1998:C:1956)

7. Nicobarese isoglosses

Van Driem and Diffloth also observed that, for the Shom Pen data they closely examined, they could find Nicobarese cognates for about half the items, and this is broadly in line with our findings for the C&M data. This by itself does not mean that Shom Pen is a Nicobarese language, since once could model various scenarios in which an unrelated language become relexified by contact with neighbouring tongues. We can also hypothesise that normal difficulties in elicitation could see many Nicobarese words offered and recorded among Shom Pen lexicon. Below we present 47 examples of apparent Shom Pen-Nicobarese isoglosses lacking wider external cognates (Nicobarese items have MKLP indices included):

C&M Shom Pen	Nicobarese isoglosses
hɛhɛn {hɛ=hɛn} 'annoyed (to be)'	Car *liːn-hɛn* 'to be annoyed with' (Whi1925:C:3430)
bāɸoāg {bə=ɸoəg} 'blowing nasal mucus'	Car *puŋ* 'slime' (Whi1925:C:4613)
okād {okəd} 'blunt'	Car *tətkɯt* 'blunt' (Das1977:C:1852)
tha {tha} 'catch (s.t. falling)'	Car *tae* 'catch' (something thrown)' (Das1977:C:1614)
heiaghe {hei=aghe} 'cough'	Car *ʔɛhɛ* 'cough' (Das1977:C:2100)
jiāu {ɟiəw} 'crab'	Car *ʔin-jiːw* 'a big land crab' (Whi1925:C:2052)
kagau {ka=gaw} 'cut (wood etc.)'	Car *kuəl* 'to chop; cut down' (Whi1925:C:3106)
cuou {cuow} 'dark, night'	Car *cuəl ha-təːm* 'night, midnight' (Whi1925:C:551)
okheāŋ {ok=heəŋ} 'deaf'	Car *tat haŋ* 'deaf person'' (Das1977:C:1741)
hekkao {hek=kaw} 'defecate'	Car *ʔət kɔː* 'defecate'' (Das1977:C:2143)
kokeɔ {kokew} 'father's b., mother's b.'	Car *jəŋ kikoː* 'uncle' (Das1977:C:476)
mooijob {moːj=ɟob} 'fireplace'	Car *mɯiə* 'fireplace' (Das1977:C:1150)
(nɛ) kugāu {kugəw} 'front legs (animal)'	Car *kɛl* 'hand, foreleg of animal' (Das1977:C:634)
heheu {he=hew} 'jerk'	Car *hoːk* 'to draw; pull' (Whi1925:C:1690)
tẽou {tẽow} 'jerk'	Car *tɯk* 'pull, jerk, pull out, weed' (Das1977:C:1904)
gihou {gihow} 'land'	Car *taku sɛho* 'land cleared for ploughing'' (Das1977:C:1669)
ijeʔā {ʔij=eʔə} 'little (a)'	Car *həj* 'thin' (of dimension, object, density)' (Das1977:C:387)
houou {houou} 'moon'	Car *soho* 'full moon' (Das1977:C:1580)
ɸeāu ghāu {ɸeəw ghəw} 'run'	Car *fal* 'run' (Das1977:C:110)
ĩão {jə̃w} 'sneeze'	Car *jɛː-sə* 'to sneeze' (onomatopoeic)' (Whi1925:C:6630)
ɸeāo {ɸeəw} 'strike'	Car *feːl* 'strike' (with hand, fist)' (Das1977:C:130)
gao xeuŋāo {gao xeu=ŋāo} 'swallow'	Car *ŋɔːk* 'to swallow' (Whi1925:C:4247)
ghuou {g=huow} 'tear (as of grass)'	Car *hɔːʔ* 'to make trips for broom by tearing coconut leaf' (Whi1925:C:1777)
ijeʔã̄ {ʔij=eʔə} 'thin (as thread)'	Car *həj* 'thin' (of dimension, object, density)' (Das1977:C:387)
(nɛ) tāi {təj} 'tusk of boar'	Car *kanel haeti* 'tusk' (Das1977:C:566)
muou {muow} 'collect, pile up'	Car, Nancowry *hamuːl* 'collect' (Das1977:C:255)
haguinaŋ {haguj=naŋ} 'hole of ear'	Car *naŋ* 'ear' (Das1977:C:1184); Nancowry *naŋ* 'ear' (Man1889:C:3831)
ca {ca} 'my'	Car *cu-ʔə*'I' (Whi1925:C:542); Nancowry *cɯa* 'chüa I, my.' (Man1889:C:361)
nho {n=ho} 'bark'	Nancowry *ʔohɔ* 'bark of dog.' (Man1889:C:4034)
puggāi {pug=gəj} 'carry on back'	Nancowry *ʔokaj-haŋa* 'bear, carry, take away.' (Man1889:C:4047)
thā (ki tai) {thə} 'clap'	Nancowry *tə́ʔ* 'to slap, hit' (Sho2006:C:2010-8)

hecāu {he=cəw} 'climb'	Nancowry *cuak-lare* 'ascend a ladder or stairs.' (Man1889:C:382)
lubiāu {lub=iaw} 'coconut kernel'	Nancowry *cia-ʔojaw* 'coconut tree' (Man1889:C:270)
jāg {ɟəg} 'coconut kernel'	Nancowry *juaŋ-ʔojaw* 'coconut-fruit.' (Man1889:C:5941)
kaiug {kajug} 'go'	Nancowry *kajiŋa* 'go away, leave, depart.' (Man1889:C:2420)
juou {ɟuow} 'hair of head'	Nancowry *jok* 'hair' (Man1889:C:5866)
xiug {xiug} 'sun'	Nancowry *heŋ* 'sun, day.' (Man1889:C:1683)
ẽcag {ẽcag} 'lame'	Nancowry *ʔoŋ-cuaŋ* 'cripple' (noun).' (Man1889:C:4261)
puggāi {pug=gəj} 'marriage'	Nancowry *ʔokaj-haŋa* 'bear, carry, take away.' (Man1889:C:4047)
ɛguɸɛg {ʔɛgu=ɸɛg} 'mouth'	Nancowry *ʔoal-faŋ* 'mouth.' (Man1889:C:3968)
hə̃i {hə̃j} 'nasal mucus'	Nancowry *hehe* 'snivel mucus from the nose' (Man1889:C:1625)
ɔŋān {ʔɔŋə̃} 'roast'	Nancowry *ŋɔʃ-ŋa* 'scald' (verb), burn, singe.' (Man1889:C:3902)
eiõu {ʔei=ʔõw} 'sky'	Nancowry *cu-ʔoal* 'clear sky.' (Man1889:C:459)
(nā) ŋoāin {ŋoə̃j} 'stone'	Nancowry *maŋ-ŋe* 'stone, rock, rock, sandstone.' (Man1889:C:3582)
xiug {xiug} 'sun'	Nancowry *ʃup-heŋ* 'sunset.' (Man1889:C:5111)
helein {helein} 'wrestling'	Nancowry *halina* 'wrestle in anger or in earnest.' (Man1889:C:1190)

It is difficult to make strong claims based on the above data, since at this preliminary stage it is not clear to what extent loans, cognates, and chance resemblances are contributing. But at least it is clear that some loans are represented. A clear case in point is the word *naŋ* 'ear', which is reflected without change in all three languages. It is especially telling that *naŋ* preserves the final nasal, rather than the velar stop /-g/ which is expected by regular correspondence. Other examples that look like loans include 'deaf', 'wrestling'. On the other hand, a form such as. Shom Pen *xiug {xiug}* 'sun' does appear to correspond to Nancowry *heŋ* 'sun, day', with regular hardening of the final. Others in this class include: 'mucus', 'coconut kernel', 'lame', 'go'. Multiple examples that seem to indicate a regular loss of etymological final oral stops in Shom Pen, examples include: 'jerk', 'swallow', 'tear (as of grass)', 'clap', 'climb', 'hair'.

8. Conclusion

The data above present various ambiguities that make interpretation difficult. However, we can make a number of observations:

- Shom Pen shows numerous lexical items common to Nicobarese and Aslian languages (jointly and severally).
- Comparative analysis indicates a chain shift in the history of Shom Pen in which final stops were lost, and then reintroduced to the language by hardening of final nasals.
- Superficially final nasal hardening resembles Aslian, but in Aslian languages the result was a merger of final nasals and stops, which is structurally different.
- There are Nicobarese loans in Shom Pen, but these do not appear to be enough to explain the extent of common vocabulary.

Consequently, our answer to the question "Is Shom Pen a distinct branch of Austroasiatic?" posed by Blench at the ICAAL4 meeting, is, maybe. On the present evidence Shom Pen looks to some extent like to be an

innovative Nicobarese language, but that on its own would not account for the extent of apparent Aslian cognates, especially North and Central Aslian cognates. A reasonable suggestion is that Shom Pen, Nicobarese, and Aslian form three branches of a Southern Mon-Khmer family. If such a hypothesis is adopted, then Shom Pen is a distinct - although not coordinate - branch of Austroasiatic.

References

Bishop, Nancy M. & Mary M. Peterson. 1994. *Kensiw rhyming list*. Bangkok: Thammasat University.

Blench, R.M. 2007. The language of the Shom Pen: a language isolate in the Nicobar Islands. *Mother Tongue XII*: 179-202.

Burenhult, N. 2005. *A grammar of Jahai*. Pacific linguistics, 566. Canberra: Pacific Linguistics. http://en.wikipedia.org/wiki/Special:BookSources/0858835541

Chattopadhyay, Subhash Chandra and Asok Kumar Mukhopadhyay 2003. *The Language of the Shompen of Great Nicobar : a preliminary appraisal*. Kolkata: Anthropological Survey of India.

Das, A.R. 1977. *A study on the Nicobarese language*. Calcutta: Anthropological Survey of India.

De Roëpstorff, 1875. [2nd ed.] Vocabulary of dialects spoken in the Nicobar and Andaman islands. Calcutta: [no publisher].

Elangaiyan, Rathinasabapathy, et al. 1995. *Shompen-Hindi Bilingual Primer Śompen Bhāratī 1*.Port Blair and Mysore: Tribal Welfare Department of the Andaman and Nicobar Administration, Ādivāsī Kalyāṇa Vibhāg Aṇḍamān tathā Nikobār Praśāsan, and Central Institute of Indian Languages [23 unnumbered and 192 numbered pages].

Kruspe, Nicole D. 2004. *A Grammar of Semelai.* Cambridge: Cambridge University Press (Cambridge Grammatical Descriptions).

Man, E.H. 1886. A Brief Account of the Nicobar Islanders, with Special Reference to the Inland Tribe of Great Nicobar. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:428-451.

Man, E.H. 1889. A dictionary of the Central Nicobarese language. London: W.H. Allen.

Means, N., and P. Means (1986). *Sengoi-English, English-Sengoi dictionary*. University of Toronto and York University: The Joint Centre on Modern East Asia.

Means, Nathalie, Gordon P. Means; Abuk Alang 1998. *Temiar-English, English-Temiar dictionary*. Saint Paul, Minn. : Hamline University Press

Radhakrishnan, R. 1981. The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language. Edmonton, Canada: Linguistic Research Inc.

Rizvi, S.N.H. 1990. *The Shompen*. Calcutta: Seagull Books.

Shorto, H. 2006. *A Mon-Khmer comparative dictionary*. Paul Sidwell, Doug Cooper and Christian Bauer (eds.) PL 579. Canberra: ANU.

van Driem, George. 2008. The Shompen of Great Nicobar Island: New linguistic and genetic data, and the Austroasiatic homeland revisited. *Mother Tongue* 13:227-247.

Whitehead, G. 1925. *Dictionary of the Car-Nicobarese language.* Rangoon: American Baptist Mission Press.

Language Vitality and Mon Ethnic Group Attitudes Towards Ethnic Tourism Development in the Western Region of Thailand: A Preliminary Report[8]

Sujaritlak Deepadung, Sumittra Suraratdecha, Narong Ardsmiti and Pichet Setaphong
Research Institute for Languages and Cultures of Asia
Mahidol University, Salaya, Thailand

Abstract

This paper presents preliminary findings of the research project "Ethnicity, Language, Culture, and Ethnic Tourism Development." Based on the collected data, ethno-linguistic maps show Mon ethnic groups residing in eight provinces of the western region: Kanchanaburi, Ratchaburi, Nakhon Pathom, Phetchaburi, Prachuap Khiri Khan, Suphan Buri, Sumut Sakhon and Samut Songkhram. A survey using guided questionnaires of language attitudes and language use of Mon people in these provinces was carried out using a qualitative approach. The Mons, who came to Thailand during mid Ayudhaya and early Rattanakosin periods are Thai citizens found in Ratchaburi, Nakhon Pathom, Phetchaburi, Suphan Buri, Sumut Sakhon and Samut Songkhram. Preliminary findings revealed that Mons in Ratchaburi and Nakhon Pathom show strong language vitality in everyday life, while the other four provinces show the least vitality. Those who came after 1948 (2491 B.E.), are mostly non-Thai citizens. They are found in Ban Wang Ka village, Kanchanaburi and Ban Huaykriap village, Prachuap Khiri Khan. For all of the Mon communities, the attitudes towards their language are positive, and most of them believe that ethnic tourism will be good for their communities.

1. Introduction

1.1 The research project "Ethnicity, Language, Culture and Ethnic Tourism Development"

No one community is monolingual. Thailand is no exception. Linguistic diversity can be found in all regions of Thailand. Unfortunately, with the rapid development of information technology world-wide, such diversity is being threatened. In recent years, many organizations have begun campaigning on the preservation of linguistic as well as cultural diversity in many countries. To name a few, UNESCO declared Policy on Linguistic Rights in 1997, United Nations announced the Declaration on the Rights of Indigenous People in 2007 and declared the year of 2008 as 'International Year of Languages' (UNESCO, 2009).

The Research Institute for Languages and Cultures of Asia also recognizes the importance and value of linguistic and cultural diversity in Thailand and has conducted a number of research projects on the matter. The present research project, 'Ethnicity, Language, Culture, and Ethnic Tourism Development' is developed with objectives to explore the impact of globalization on the ethnic peoples, particularly their indigenous

[8] The "Ethnicity, Language, Culture, and Ethnic Tourism Development" research team would like to thank Mahidol University for the research grant of the 2009 fiscal year. The research team includes: Principle investigator: Somsonge Burusphat; Co-researchers: Sujaritlak Deepadung, Sumittra Suraratdecha, Amon Saengmanee; Narong Ardsamiti, Pattama Patpong, and Pichet Setaphong; Research assistants: Sanparee Yokyong and Peranun Chaloekkanit.

Deepadung, Sujaritlak, Sumittra Suraratdecha, Narong Ardsmiti and Pichet Setaphong . 2011. "Language Vitality and Mon Ethnic Group Attitudes Towards Ethnic Tourism Development in the Western Region of Thailand: A Preliminary Report." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds*.) Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.19-32.

languages and cultures. The findings of the research could be used as a fundamental for national language planning as well as sustainable ethnic tourism plans. During the first year of the research project, the research team will be collecting and analyzing sociolinguistic data in eight provinces; Nakhon Pathom, Ratchaburi, Kanchanaburi, Suphan Buri, Samut Sakhon, Samut Songkhram, Phetchaburi, and Prachuap Khiri Khan, in the western region of Thailand. During the second year, the research team will select one ethnic community as a model community to develop sustainable ethnic tourism. The ultimate goals of the project are to promote diversity in unity and the preservation of indigenous languages and cultures, as well as to strengthen the ethnic communities.

The present paper,[9] "Language vitality and Mon ethnic group attitudes towards ethnic tourism development in the western region of Thailand: A Preliminary Report" is a preliminary survey of the research project, 'Ethnicity, Language, Culture, and Ethnic Tourism Development'. The paper sets its goals to (1) identify the Mon language spoken in the western region of Thailand, (2) explore the Mon's language use and attitudes toward languages of the wider community, and (3) evaluate on the development of sustainable ethnic tourism in the region. The research outcomes of these goals include:

1. An updated linguistic map of areas inhabited by Mon, overlapped with multiple layers of cultural information of Mon
2. A summary of the community leaders' language use and attitudes towards their own language and culture in Kanchanaburi, Nakhon Pathom, Ratchaburi, Phetchaburi, Prachuap Khiri Khan, Samut Sakhon, Samut Songkhram and Suphan Buri, in the western region of Thailand
3. The community leaders' perspectives on a prospective sustainable ethnic tourism plan.

1.2 Mon people and language affiliation

With a long and difficult history in Southeast Asia, at present Mon people can be found in Myanmar (Burma) and in Thailand– in lower Burma and in central Thailand. Linguistically, the Mon language belongs to Mon-Khmer sub-family within the Austroasiatic language family. The Mons, who came from Myanmar to Thailand during mid-Ayudhaya (1569-1767) and early Rattanakosin periods (1767-1782) are all Thai citizens now. They are in Ratchaburi, Nakhon Pathom, Phetchaburi, Suphan Buri, Samut Sakhon and Samut Songkhram.[10] Those, who arrived between 1948 and the present time are mostly non-Thai citizens. They are found in Kanchanaburi and Prachuap Khiri Khan. The Mon language also has a writing system which dates back to the sixth century.[11]

1.3 Attitude and language use

A speaker's attitude towards their own language is one of the important factors used to determine the linguistic situation of the language, or even the life span of the language. How the speakers view or think of their own language and the languages of the wider community and how they use their language in their daily lives; e.g., to whom, in which occasions, and where, can help determine the future of the language. According to Smalley (1994), there are many languages spoken in Thailand, all of which can be grouped into different hierarchies, with Standard Thai at the top and ethnic languages at the bottom. It can also be assumed that speakers tend to be interested in learning the languages of the higher hierarchies than the one that their language belongs to. The speakers would, accordingly, tend to ignore the languages in their hierarchy and

[9] This paper was presented at the Fourth International Conference on Austroasiatic Linguistics at the Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom, Thailand. October 29 - 30, 2009.

[10] These Mon people first came to settle in Phra Nakhon Si Ayuthaya Province and many provinces north of Bangkok, namely in Nonthaburi and Pathum Thani. Later, some of them were moved to resettle a Mon community in Samut Prakan. Also, new settlements can be found along the Maeklong River in Ratchaburi Province. There are small isolated Mon communities in Lamphun and Chiang Mai Provinces in the north of Thailand too. (Deepadung, Gajaseni, Ketwisit & Deepadung, 1995).

[11] Old Mon, from about the sixth century to the end of the twelfth; Middle Mon, from the thirteenth century to the eighteenth; and Modern Mon.

those in the lower. Previous work has shown that speaker's attitudes toward languages play an important role in their use of language(s). Such work would also consider the 'domain' of language use, or the place where the language is spoken, e.g., home, school, work, governmental offices, temple.

1.4 Sustainable ethnic tourism and ethno-linguistic vitality

Tourism is an industry that is growing very rapidly. It is now one of the fastest growing industries. With its strong commercial stand in the world economy, tourism creates employment and benefits in all nations as a whole. In addition, the tourism industry was the first to enact the United Nations' Agenda 21 concerning sustainable development. Sustainable tourism is supposed to support integrity of the place, benefit residents, conserve resources, respect local culture and traditions, strive for quality, not quantity, not to abuse its products, and this means greater satisfaction for visitors who would be willing to re-visit the place and tell their friends to share their great experience. It is only recently that there has been a growing recognition of the importance of combining the needs of traditional urban management with the need to plan for tourism. Some of the most important principles of sustainable tourism development include (Jamieson & Noble, 2000):

1. Tourism should be initiated with the help of broad-based community-inputs and the community should maintain control of tourism development.

2. Tourism should provide quality employment to its community residents and a linkage between the local businesses and tourism should be established.

3. A code of practice should be established for tourism at all levels – national, regional, and local – based on internationally accepted standards. Guidelines for tourism operations, impact assessment, monitoring of cumulative impacts, and limits to acceptable change should be established.

4. Education and training programmes to improve and manage heritage and natural resources should be established.

To become sustainable, the ethnic tourism industry should also enforce the role of indigenous peoples, Agenda 21: Chapter 26: Recognizing and strengthening the role of indigenous people and their communities (UN, 2009).

2. Methodology

2.1 Ethno-linguistic map of the west

The purpose of the linguistic map is to portray the distribution of all ethnic languages found in the western part of Thailand. In constructing the map, the research team members were assigned a particular location in the western part of Thailand. Together, we are working on eight provinces including Nakhon Pathom, Ratachaburi, Kanchanaburi, Suphan Buri, Samut Sakhon, Samut Songkhram, Phetchaburi, and Prachuap Khiri Khan. The data collected include ethnic languages spoken in each location, still pictures and videotape of the ethnic and cultural activities. To obtain the linguistic data, the researcher first conducted a library research on previous work, gathering all the information on which districts or sub-districts are the residences of relevant ethnic peoples. After the secondary data was collected and summarized, a survey form was constructed for a postal survey. The form includes a list of village numbers, village names, and ethnic names as exemplified in table 1.

Table 1: Example of a survey form (Nonglu Subdistrict, Sangkhla Buri District, Kanchanaburi Province)

Village names / Village numbers	Ethnic Names							
	Mon	Karen	Thai Song Dam	Thai Phuan	Lao Wiang	Lao Khrang	Lao Ti	Others
M. 1 Nithe		√						
M. 2 Wangka	√							
M. 3 Lai Nam	√	√			√			
M. 4 Huay Kop		√						
M. 5 Wiakhadi		√						
M. 6 Huay Malai	√	√						
M. 7 Mai Phatthana	√	√						
M. 8 Songkaria	√	√						
M. 9 Phra Chedi Sam Ong	√	√			√			
M. 10 Prarai Nok	√	√						

The survey form was sent out to all relevant district and sub-district offices. The purpose of the survey was to obtain updated information on the languages spoken in each location in details– what ethnic languages are spoken and in which village. To assure the validity of the data received, the research team also went to their field locations to collect more data on the general living environment, seeking information on the vitality of ethnic languages and cultures, as well as cultural activities.

2.2 Participants and procedures

An in-depth interview, a qualitative research technique, is selected as a method of data collection in probing and exploring the speakers' feelings and perspectives on their own ethnic language. According to Milena, Dainora and Alin (2008, 1282), "qualitative research methods are also preferable when the investigation is oriented to determine motivation, perceptions or believes". The in-depth interview is used to obtain self-reported data about speakers' attitudes towards their language as well as the usage of their language.

The result from this method of data collection will be used to complement and extend the findings of results from a quantitative method to be conducted later to achieve the goals of the research project. It is also expected that this combination of methodologies will enhance the convergent validity of this research.

3. Findings and discussion

3.1 Ethno-linguistic maps of the Mon ethic group

The survey of areas in which Mon reside reveals that they live mostly in the eight provinces in the western region of Thailand, that is, Kanchanaburi, Ratchaburi, Nakhon Pathom, Phetchaburi, Prachuap Khiri Khan, Suphan Buri, Sumut Sakhon and Samut Songkhram. The Mon communities in these eight provinces are identified using three criteria. First, the Mon communities are settled communities where Mon people have lived for a long period of time. Second, areas where Mon immigrants live are excluded. Third, other ethnic communities where some Mon people have moved to and where the Mon language is not used in everyday life, are also excluded. Based on these criteria, a number of villages are screened for display on the linguistic maps. By using the geographical information system (GIS) technique, maps of the Mon villages where the Mon language is spoken, overlapped with multiple layers of cultural information about ethnic

groups, have been created. In this section, the linguistic and cultural information of the Mon people in the eight provinces are presented.

As mentioned earlier and in Footnote 11, the first group of Mon communities were established in provinces north of Bangkok, i.e., in Nonthaburi and Pathum Thani provinces which are, according to Bauer (1990, 21), "…the largest concentration of Mon settlements." The second largest settlements of Mon are in Samut Sakhon, Samut Songkhram and Phetchaburi provinces. The third largest and "probably the densest (is) located along the Meklong River in Ratchaburi province", (Bauer, 1990, 21). The second community of Mon is in Sangkhla Buri District, Kanchanaburi province, and Mon people can be found elsewhere in Kanchanaburi as well. At present, and for the purposes of this study, Mon communities still can be found in all of the eight provinces mentioned. Table 2 lists the eight provinces where the Mon language is spoken, including the names of districts and subdistricts, and the numbers of villages.

Table 2: Data on Provinces, Districts, Sub-districts and Villages where Mon is spoken

Kanchanaburi		
Districts	*Sub districts*	*Villages*
Mueang	Ko Samrong (1), Ban Kao (2), Chong Sadao (6)	9
Tha Maka	Takhram En (2), Phra Thaen (1), Wai Niao (4), Khok Tabong (1)	8
Tha Muang	Wang Sala (2), Nong Tak Ya (2)	4
Bo Phloi	Chong Dan (3), Nong Krang (1)	4
Sai Yok	Lum Sum (2), Tha Sao (8), Wang Khrachae (7), Si Mongkhon (4)	21
Thong Pha Phum	Tha Khanun (1) , Pilok (2) , Linthin (3) , Hin Dat (6) , Chalae (1) , Huai Khayeng (5)	18
Sangkhla Buri	Nong Lu (7) , Plangphle (4)	11
Si Sawat	Khao Chot (1)	1
Dan Makham Tia	Dan Makham Tia (4) , Klondo (3)	7
Ratchaburi		
Districts	*Sub districts*	*Villages*
Mueang	Chedi Hak (1) , Ang Thong (2)	3
Chom Bueng	Dan Thap Tako (2) , Kaem On (1) , Boek Phrai (1)	4
Suan Phueng	Suan Phueng (5) , Tanao Si (6)	11
Ban Kha	Ban Kha (8)	8
Pak Tho	Wang Manao (1), Yang Hak (6)	7
Wat Phleng	Ko San Phra (6)	6
Damnoen Saduak	Khun Phithak (1) , Tha Nat (4) , Ban Rai (1)	6
Ban Pong	Ban Muang (6) , Nakhon Chom (6) , Khung Phayom (5) , Suan Kluai (2) , Khao Khlung (5)	24
Photharam	Don Krabueang (6) , Nong Pho (1) , Khlong Ta Khot (7) , Chamrae (7) , Soi Fa (5) , Tao Pun (7) , Thammasen (1) , Khao Cha-ngum (1)	35
Phetchaburi		
Districts	*Sub districts*	*Villages*
Ban Laem	Bang Khrok (1) , Bang Tabun (1)	2
Khao Yoi	Nong Pla Lai (1)	1
Nong Ya Plong	Nong Ya Plong (2)	2
Prachuap-KhiriKhan		
Districts	*Sub districts*	*Villages*
Pran Buri	Wang Phong (3)	3
Bang Saphan	Ron Thong (8) , Chai Kasem (4) , Thong Mongkho(1)	13

Bang Saphan Noi	Pak Phraek (1)	1
Sam Roi Yot	Sila Loi (3)	3
Nakhon Pathom		
Districts	*Sub districts*	*Villages*
Kamphaeng Saen	Thung Khwang (5) , Thung Luk Nok (4)	9
Bang Len	Sai Ngam (3) , Nin Phet (3) , Bua Pak Tha (3) , Bang Phasi (2)	11
Sam Phran	Talat Chinda (2)	2
Suphan Buri		
Districts	*Sub districts*	*Villages*
Mueang	Taling Chan (1)	1
U Thong	Chedi (1)	1
Song Phi Nong	Bang Phlap (2) , Si Samran (1) , Thung Khok (2) , Bo Suphan (5) , Don Manao (3)	13
Samut Sakhon		
Mueang	Chai Mongkhon (1) , Kalong (3) , Ban Ko (8) , Bang Krachao (8) , Khok Kham (1) , Ban Bo (5) , Bang Tho Rat (10) , Tha Sai (8) , Phanthai Norasing (1) , Tha Chin (2) , Bang Ya Phraek (3) , Na Di (9)	59
Ban Phaeo	Ban Phaeo (7) , Yokkrabat (4) , Amphaeng (1) , Lak Song (3) , Lak Sam (11) , Khlong Tan (2) , Chet Rio (5) , Suan Som (1) , Nong Bua (5)	39
Samut Songkhram		
Districts	*Sub districts*	*Villages*
Mueang	Laem Yai (7) , Khlong Khon (6) , Ban Prok (4) , Bang Chakreng (3)	20
Bang Khonthi	Bang Krabue (1)	1

3.2 Traditional Mon culture and transition

Traditional Mon culture has undergone to change as a result of the pressure of modernity and a creeping assimilation notwithstanding the similarities between Thai and Mon cultures. Like most Thais, Mon people are Buddhists, so they usually get together and participate in Buddhist ceremonies such as Buddhist Lent Day, the end of Buddhist Lent Day, Wisakha Bucha Day etc. There are temples in every Mon community and gold artwork of Hintha birds[12] a mythological wild goose, can be seen on pillars within the monasteries. The Mon and the Thai practice different forms of Buddhism. In Thailand, there are two Buddhist sects, Thammayut and Mahanikai. Mon Buddhism, which has strict discipline and ordination rules, belongs to the Thammayut sect while the majority of Thais belong to the Mahanikai sect. However, the Mon still maintain a ritual of paying respect to their ancestral spirits, both house spirits and village spirits. According to Guillon,

> Kalok is primarily a spirit, a demon, an immaterial being that sometimes takes on visible form, as for example the owl, kalok gacem.…..
> Also, on certain occasions they proceed to perform the kalok dance. Finally, the solemn offering to the Kalok takes place once a years, or once every three years,…..
> These practices– Kalok, turtle and snake– exist on among the Mons. They are not found among the Burmese or the Thai, at any rate; in fact, the latter identify the Mons by means of these beliefs.
> Guillon (1999, 23)

Both the Mon and the Thai celebrate Songkran – traditional New Year festival in Mid-April, but this colorful celebration is more important in traditional Mon culture than in Thai culture. Foster (1982, 10) states

[12] "haṃsa" = the Brahmany goose (Guillon, 1999, 147).

that "…and most Thai who know anything of the Mon know of the Mon Songkran. … Of the specific activities associated with Mon Songkran, only one is distinctively Mon: a courting game called sabaa." Peng Songkran or 'the Chronicle of rice in cool water' – is the most important dish during the Songkran festival for the Mon. This specially cooked rice is offered to the monks and to highly respected elders during the three-day Songkran festival.

The floating of model boats to get rid of the past year's troubles and bad omens, which is held in September, is another practice of the Mon people in Sangkhla Buri. The Mon in Nakhon Pathom also perform a ritual whereby the image of Thera Upagupta travels by boat along the canal to allow the neighborhood to pay respects.[13]

Apart from language, which is the topic of the next section, some other traditional elements of the Mon, such as houses, costumes and food are listed in Table 3 to show the changes and assimilation of the Mon communities in the eight provinces.

Table 3: Mon: changes and assimilation

Provinces	Traditional costume	House	Food	sabaa, peng Songkran	kalok Dance	Floating a model boat
Nakhon Pathom	on special occasions	contemporary style	on special occasions	during "Songkran"	every 2,3 years	-
Phetchaburi	on special occasions	contemporary style	on special occasions	during "Songkran"	-	-
Ratchaburi	only elderly people dressing a traditional style on special occasions; launching a Mon dressing campaign for children on every Friday, elderly people still wear hair in a bun	contemporary style	in daily life	no sabaa, but still have peng Songkran	every 2, 3 years	-
Samut Sakhon	only elderly people dressing a traditional style on special occasions	contemporary style	on special occasions	during "Songkran"	every 2, 3 years	-
Samut Songkhram	on special occasions	contemporary style	on special occasions	-	-	-
Suphan Buri	elderly people still wear hair in a bun	contemporary style	on special occasions	-	every 2, 3 years	-
Kanchanaburi	in daily life	mixed	in daily life	-	-	annually
Prachuap Khiri Khan*	in daily life	mixed	in daily life	-	-	-

*Rubber plantation communities.

Smalley (1994, 225) points out that "This long-standing Mon population has been heavily assimilated, with extensive intermarriage Mon and Thai." As shown in table 10, the Mon people in the six provinces, except for those in Sangkhla Buri, Kanchanaburi province and in Huay Kriap, Pachuap Khiri Khan province, do not wear traditional Mon costumes in daily life, but do wear them on special occasions, especially the elderly. The younger generations wear modern style clothes. Hair dressing in the Mon tradition, can be found only with elderly people.

"Building a house for the Mon should meet all the auspicious elements including the time and direction" (Gajaseni, 1984, 64). In this study, traditional houses were very rare to find as the Mon descendants today tend to build new houses in a contemporary style. Even some of the Mon in Sangkhla Buri, who migrated into Thailand 60 years ago, have also changed the style of their houses. The Mons in Prachuap

[13] Guillon (1999, 106) states that "This legend of Upagupta has produced a whole body of beliefs and rituals in Burma and Northern Thailand, still very popular today.

Khiri Khan, who came as laborers for the rubber plantation areas, live in temporary huts and cottages on their employers' land.

From table 3, food that is unique to the Mon can be found in daily life only in the Sangkhla Buri community. During Songkran Festival, peng songkran in Mon or khaw chae in Thai, is still an obligatory important dish to offer to monks, spirits and respected elderly people in Nakhon Pathom, Phetchaburi, Ratchaburi, Samut Sakhon and Kanchanaburi.

Sabaa which is usually played during the Songkran festival, is still a part of activities during Songkran in Nakhon Pathom, Ratchaburi, Phetchaburi and Samut Sakhon. However, sabaa, which used to be a courting game, tossing the fruit of the snuffbox bean between male and female players, is now a kind of traditional folk play which is associated with Mon people and culture. In Sangkhla Buri, sabaa turns out to be a kind of gambling, so sabaa is not allowed to be played in the Mon community anymore.

The spirit worship of the ancestors including the spirits of birds, snakes, turtles, bamboos, clothes, etc. or kalok dance is still practiced in Nakhon Pathom, Ratchaburi, Samut Sakhon, and Suphan Buri. To host a spirit worship ceremony usually costs a lot of money, so it is quite understandable why there is no such ceremony in Sangkhla Buri Mon community. Only Mon communities in Sangkhla Buri have an annual ritual-floating a model boat down the stream– to get rid of bad luck.

To sum up what is going on in the provinces described above, Smalley's statement (1994, 226) is still appropriate and accurate: "Most of this ethnic difference is gone, or nearly-gone, or partially gone, at different places and with different Mon people, to the extent that some descendants of the Mon may not even be aware of their Mon ancestry or partial ancestry."[14] Like many other ethnic communities, the Mon cultures are facing cultural assimilation. Many traditional cultures are going to disappear.

3.3 Mon attitudes towards their language

Gajaseni (1984, 64) states that "The Mon are one of the people in Southeast Asia, who have a long history and unique, outstanding culture. Some Mon cultural elements, such as language and letters of the alphabet, art, music, law, religion, etc. have become prototypes for those of neighboring peoples." Burmese scripts, Northern Thai characters and Northeastern Thai characters have the Mon alphabet as their basis. Central Thai has also borrowed a lot of words from the Mon language. However, the number of the Mon people who can speak the language is diminishing, notwithstanding of the number of people who know the literary language.

In this study, a pattern of language attitude was found across all Mon communities in all provinces. Older generations, 30 years old and above, have a good attitude towards their language-they are proud to be Mon and not shy to speak the language in public. Mon is useful as an in-group language (secret language). In some cases they use Mon to communicate among their group in front of non-Mon speakers. The in-depth interviews also reveal the same results across all communities, that is, the older generation would like the young to be proud of their ethnicity and try to preserve the language and culture of the group. The language attitudes of the Mon people in 7 provinces are shown in Table 4.

[14] For example, the Mon who came to Thailand during Ayudhaya and early Rattanakosin period and settled widely in Kanchanaburi province are Thai citizens– they are in Tha Khanun sub-district of Sai Yok district.

Table 4: Language attitude of the Mons in 7 provinces

Provinces / **Ages**	Language attitude		
	POS	NEU	NEG
Kanchanaburi			
- Age 18-35 years	✓		
- Age 36-59 years	✓		
- Age over 60 years	✓		
Ratchaburi			
- Age 18-35 years	✓		
- Age 36-59 years	✓		
- Age over 60 years	✓		
Phetchaburi			
- Age 18-35 years			✓
- Age 36-59 years		✓	
- Age over 60 years		✓	
Nakhon Pathom			
- Age 18-35 years	✓		
- Age 36-59 years	✓		
- Age over 60 years	✓		
Samut Sakhon			
- Age 18-35 years	✓		
- Age 36-59 years	✓		
- Age over 60 years	✓		
Samut Songkhram			
- Age 18-35 years	✓		
- Age 36-59 years		✓	
- Age over 60 years	✓		
Suphan Buri			
- Age 18-35 years	✓		
- Age 36-59 years	✓		
- Age over 60 years	✓		

Table 4 shows language attitude of the Mons in 7 provinces, according to age-group. Prachuap Khiri Khan province is not included in this quantitative studies. All of the Mon people in 6 provinces- Kanchanaburi, Ratchaburi, Nakhon Pathom, Samut Sakhon, Samut Songkhram and Suphan Buri, have a positive attitude towards their language, except for those in Phetchaburi where the younger generation has a negative attitude while the middle and the older generations are neutral.

However, some communities such as those in Samut Sakhon, Samut Songkhram, Phetchburi, Ratchaburi, and Suphan Buri, think that the language will disappear in the near future, while in Kanchanaburi, Prachuap Khiri Khan, and Nakhon Pathom (Bang Len) they think that the Mon language will be in use for generations. All of the community leaders think that Mon is the language most appropriate for Mon traditions and ceremonies. The languages that they want their children to learn are Thai, English, Chinese, and Mon. Despite the positive attitude towards their language, some of the community leaders think that it is inevitable that the Mon language will disappear as the society is changing and many young Mons think of the Mon language as outdated. These young people consider the Thai language as modern and refined. All of the participants in the interviews acknowledged the importance of the Thai language: they all want their children to be fluent in Thai for educational and professional purposes. On the other hand, they want their children to understand Mon to the extent that they can communicate with other Mons in the community. In Guillon's words (1999, 3), "As is the use with all minorities, the Mons who are undergoing assimilation in Burma and

in Thailand are deeply distressed by the idea that their children will no longer speak their mother tongue." The following diagram summarized the people's attitude towards their own language.

Positive: The Mon language is the most useful language in the Mon community and should be preserved for the maintenance of the Mon society.

The Mon and Thai languages are both useful in the Mon community as they function in different domains.

Negative: The Mon language is no longer needed. Mon people live in Thailand and should be fluent in Thai.

Figure 1: Mon language attitude scale.

3.4 Language vitality of Mon

3.4.1 Results

In order to bring back linguistic behavior, Bradley (2009, 13) suggests that… "There is no one answer, but I believe we can identify at least five key sociolinguistic factors: Identity, Vitality, Setting, Domains and Policy." Language vitality is the focus of this study.

Table 5: Language vitality of the Mons in 7 provinces

Provinces / **Domains**	Language use		
	S	M	W
Kanchanaburi			
- Family domain	✓		
- Community domain	✓		
- Public and governmental domain		✓	
Ratchaburi			
- Family domain		✓	
- Community domain			✓
- Public and governmental domain			✓
Phetchaburi			
- Family domain			✓
- Community domain			✓
- Public and governmental domain			✓
Nakhon Pathom			
- Family domain		✓	
- Community domain			✓
- Public and governmental domain			✓
Samut Sakhon			
- Family domain		✓	
- Community domain		✓	
- Public and governmental domain			✓
Samut Songkhram			
- Family domain			✓
- Community domain			✓
- Public and governmental domain			✓
Suphan Buri			
- Family domain			✓
- Community domain			✓
- Public and governmental domain			✓

The results of this quantitative study of the Mon language vitality as determined by language use in three domains, are summarized in Table 5. The results are as follows. 1) Language vitality is high if the language use is strong (S) in all 3 domains or at least in 2 domains- in Kanchanaburi the language is strongly used in the family and community domains while language use in the official domain is moderate. 2) Language vitality is medium if its use is moderate (M) in at least one of the 3 domains- in Nakhon Pathom and Ratchaburi the language is still used moderately in the family domain and weakly used in the community and official domains, while in Samut Sakhon, it is used moderately both in the family and community domains, with Thai being used in the public and governmental domains. 3) Mon language vitality in Phetchaburi, Samut Songkhram and Suphan Buri is very low because its use is at the weakest end of the scale in all three domains.

The results from the in-depth interviews reveal that there is no such community where Mon is spoken in all domains by all age groups. It appears that official domains, e.g., work, school, district offices, are not places for the Mon language any more. Even in Kanchanaburi which seems to be the strongest community, Mon is not spoken in official domains. The locals may speak in Mon to Mon officers with whom they are acquainted; however the default language is Thai in general. In addition, every community, strong or weak, is now facing the same problem; younger generations do not speak Mon as much or as well as the older generations; and younger generations do not take pride in being Mon, and some do not understand Mon any more. The ones who are proud of being Mon are the older generations, approximately 30 years of age and above. In general, Mon is still used at home and in the village, but not all Mon parents teach the language to their children. The active speakers of Mon are still those in their 30's and above. The continuum below shows the language vitality determined by domains where the language is spoken and age groups of Mon speakers.

Strongest: The Mon language is used by Mon people of all ages in all domains.
Medium: The Mon language is used by Mon people over 30 years old. Mon people use both Mon and Thai fluently in different domains.
Weakest: The Mon language is used by few elderly people. Most Mon people speak Thai.

Language vitality continuum

Strongest		***Medium***	***Weakest***
Ø	*Kanchanaburi*	*Ratchaburi*	*Samut Songkhram*
		Nakhon Pathom	*Suphan Buri*
		Samut Sakhon	*Phetchaburi*

Figure 2: Mon language vitality scale

3.4.2 Reasons

The results of language vitality gained from this study can be explained by a number of non-linguistic factors- geographic location, history of settlement, education and technology. With regard to location and history of settlement, the Mon community at Sangkhla Buri District, Kanchanaburi Province is located on the Thai-Myanmar border. The village is about 60 years old, it was first settled in 1948 (after the W.W. II), i.e., in the middle of the Rattanakosin period. The Mon communities in Ratchaburi, Phetchaburi, Nakhon Pathom, Samut Sakhon, Samut Songkhram and Suphan Buri, on the other hand are found mostly either in or close to municipal or urban areas in each province and all of these communities were established during the late Ayuddhaya and early Rattanakosin periods (1569-1782), i.e., they are about 250 years old (See details in Deepadung & Dumsa-ard, 2007). So length of time settled and location may be important reasons for why the Sangkhla Buri Mon have the strongest vitality of all of the Mon communities in the western region of Thailand.

Education maybe another significant factor. Most Mon living in Sangkhla Buri are non-Thai citizens, and even though their children go to Thai public schools in the area, access to higher education is still very

limited, compared to those in the other 6 provinces who are now Thai citizens with education right, the same as other Thai nationals. With an educational system that has the Thai language as its medium of instruction and a long standing education policy that emphasizes the importance of the national language, the Mon language of these Mon descendants has all but disappeared. However, currently, thanks to a developing change in policy direction together with changing attitudes within some Mon communities, the Mon language used by villagers in Sangkhla Buri should be advantageously maintained, with some of the communities in the 6 provinces reviving theirs. The results of this study will certainly facilitate this development. This second factor may be the most important determinant of the vitality of the Mon language in Thailand.

Technology, associated with development and globalization, whether it be electricity, infrastructure, transport or tele-communications, etc.- is a third factor impacting on the use of the Mon language. Thus technological advances together with the location, history and education, result in the Mon language vitality in the 3 provinces of Phetchaburi, Samut Songkhram and Suphan Buri being on the weakest axis of the continuum. However, with the Mon's change in attitude with respect to their language and the preservation of their cultural heritage, the language vitality of some of the communities in Ratchaburi, Nakhon Pathom and Samut Sakhon is now tending towards moderate.

The language deprivation issue is recognized by all communities, some accept whatever result the future holds for their language, some do not. Some have started to teach Mon at school or after school, such as in Ratchaburi, Suphan Buri, and Phetchaburi. There are students interested in taking these classes, however, the number of interested students is not impressive, and for some classes, the number seems to be fewer each year. The parents sometimes question the practicality of such language maintenance campaigns; they think it is probably better to teach a more useful language, like English, rather than Mon, an ethnic language which Mon people themselves do not speak much any more. English can help them pass exams, or get a better job. A community leader in Suphan Buri addressed the problem, "I want my children to speak some Mon because we are Mon, but I don't know with whom they can talk. Thai is used at school. Thai is used at the market. Thai is used at home. Only the elderly can still speak Mon to each other". Another community leader in Kanchanaburi commented that in the past they used to have a problem obtaining Thai ID. Now they are facing the problem of how to convince the younger generations to take pride in being Mon and preserve the language and culture from generation to generation.

3.5 Mon attitudes towards ethnic tourism development (ETD)

The preliminary survey of the Mon people's attitudes towards ethnic tourism development reveals that the Sangkhla Buri Mons in Kanchanaburi, the Ban Muang Mon in Ratchaburi, and Bang Len Mon in Nakhon Pathom and Chet Riu Mon in Samut Sakhon have the most positive attitudes towards ethnic tourism development. The Mon village at Ban Wang Ka, Sangkhla Buri district is already an attractive spot for tourism by virtue of it being a Mon village, its location at the Three Pagoda Pass and its having attractive natural resources.[15] Nowadays, there are lots of tourists visiting Sangkhla Buri regularly, especially during the winter season. Implementing ethnic tourism development may help the younger generations get to know or be proud of their Mon and help preserve their language and traditions.

The Mon of Ban Muang in Ban Pong District, Ratchaburi Province, already have a Mon cultural center and museum. The younger generations do not speak Mon but there are still a lot of elders who can speak it. At present, the bilingual education campaign is at its first stage, so implementing ethnic tourism development will help maintain Mon traditional culture and the Mon language.

Among the three Mon communities of Nakhon Pathom Province, the Mon community at Bang Len District fully supports ethnic tourism development while the other two communities have no interest. Bang Len Mon believe that ethnic tourism development will help preserve their Mon language and culture, as well as increase villagers' income.

[15] See details of the Sangkhla Buri Mon in Deepadung & Dumsa-ard (2007, pp. 53-65).

The Mon people of Chet Riu in Ban Phaeo District, Samut Sakhon have a neutral attitude towards ethnic tourism development in terms of their village life and the strength of their Mon culture and traditions. Together with other Mon communities in the province, such as in Ban Rai and Ban Ko, they are interested in ETD.

In Samut Songkhram, Phetchaburi and Suphan Buri provinces, the Mon communities believe that ethnic tourism development would be impossible in these provinces. Mon is still spoken, but mostly by the elders. Traditions and cultural activities are similar to those of Thai.

The Mon communities in Prachuap Khiri Khan are communities in rubber plantation areas, so to implement ethnic tourism development is impossible.

A summary of the Mon people's attitude towards ethnic tourism development is as follows:

***Positive**: ETD already exists in the Mon community such as home-stay tourism and Mon cultural centers.– Kanchanaburi, Ratchaburi,and Bang Len (Nakhon Pathom).*

ETD is attractive and interesting but there are some disadvantages such as no other tourist attractions (waterfalls, caves); no active community leaders; poor public relations; dis-united community.– Nakhon Pathom and Chet Riu (Samut Sakhon).

***Negative**: The Mon community is too small to develop ethnic tourism and is highly integrated into the with Thai community.– Samut Songkhram, Phetchaburi, Suphan Buri.*

Figure 3: Mon people attitude towards ETD scale

4. Conclusion

This paper presents preliminary findings of the research project entitled "Ethnicity, Language, Culture and Ethnic Tourism Development." The initial attempt is to locate Mon ethnic groups residing in the western region of Thailand. A survey form was devised for data collection on Mon inhabited areas. Based on the collected data, an ethno-linguistic map of areas inhabited by Mon, overlapped with multiple layers of linguistic and cultural information of Mon, is presented. The linguistic map shows that Mon ethnic groups reside in all of the eight provinces in the western region of Thailand.

Language attitudes among middle-aged and old-aged people are positive in all provinces. Mon people in these age groups wish their children could speak Mon because they are proud of their language and culture. With respect to language vitality as determined by language use and the language attitudes of Mon groups in the 7 provinces, the study shows that Kanchanaburi Mon have the strongest vitality whereas Phetchaburi, Samut Songkhram and Suphan Buri have the weakest Mon language vitality. Mon communities in Ratchaburi, Nakhon Pathom and Samut Sakhon are placed midway along the continuum. As for attitudes towards ethnic tourism development, the preliminary findings reveal that they are positive in Kanchanaburi (Sangkhla Buri Mon), Nakhon Pathom (Bang Len Mon) and Ratchaburi. Some Mon communities in Nakhon Pathom and Samut Sakhon have neutral attitudes forwards ETD. Based on the location and size of the communities, language vitality and the preservation of cultural heritage, the Phetchaburi, Samut Songkhram and Suphan Buri Mon have better assimilated into the Thai community, so their attitudes towards ETD are negative. Despite the Mon people's positive attitudes towards their language, linguistic and cultural transition and assimilation seem to be unavoidable for various reasons - duration of residence in Thailand, education and technology, and Thai government policy, being among the main factors.

Based on these preliminary findings, the data analysis will be followed by a Participatory Action Research (PAR). In carrying out Participatory Action Research plan or PAR plan, one model community will be selected based on a number of supporting factors such as overall community strength, availability of man power, the authenticity of the linguistic and cultural data available in the community, and most importantly, the willingness of the whole community to participate in the project. After a model community is selected,

the research team will proceed by preparing the community, providing all the relevant information. The research team will re-visit the field site and discuss the objectives of the whole project in detail with the community, looking for leaders, community developers, and coordinators for the project. Basically, the research team will be working together with the community as a consultant as opposed to a director. There will be a series of activities and discussions between the research team and the community to promote mutual understanding and cooperation, as well as to assert an active role for the community towards sustainable ethnic tourism in the community.

References

Bauer, C. (1990). Language and ethnicity: the Mon in Burma and Thailand. In G. Wijeyewardene (Ed.), *Ethnic groups across national boundaries in mainland Southeast Asia* (pp. 14-47). Singapore: Institute of Southeast Asian Studies.

Bradley, D. (2009, November). *Resilience thinking and language endangerment*. Paper presented at the 42nd international conference on Sino-Tibetan Languages and Linguistics, Payap University, Chiang Mai, Thailand.

Deepadung, S., Gajaseni, S., Ketwisit, W. & Deepadung, A. (1995). *Mon : History and Changes in social, cultural and political role during the 200 years of Rattanakosin*. Bangkok : Sahadhammik Co.Ltd.(In Thai)

Deepadung , S. & Dumsa-ard, P. (2007). From Moulmein in Myanmar to Sangkhla Buri in Thailand: An ethnographic study of a Mon village. *MKS*, *37*, 53-65.

Foster, B. L. (1982). *Commerce and ethnic differences: The case of the Mons in Thailand*. Ohio:The center for international studies, Ohio University.

Gajaseni, S. (1984). Mon culture and customs. *Muang Boran Journal*, *10*(3), 64-65.

Guillon, E. (1999). *The Mons: a civilization of Southeast Asia.* Bangkok: The Siam society under Royal Patronage.

Jamieson, W. & Noble A. (2000). *A manual for sustainable tourism destination management*. CUC-UEM Project, AIT, 2000. Retrieved October, 2009, from http://www.gdrc.org/uem/eco-tour/etour-principles.html

Milena, Z.R., Dainora, G., & Alin, S. (2008). Qualitative research methods: a comparison between focus-group and in-depth interview. *The Journal of the Faculty of Economics - Economic*, 4(1), 1279-1283. Retrieved October, 2009, from http://steconomice.uoradea.ro/anale/volume/ 2008/v4-management-marketing/235.pdf

Smalley, W. A. (1994*) Linguistic diversity and national unity: Language ecology in Thailand.* Chicago: the University of Chicago press.

UNESCO (2009). Retrieved September, 2009, from http://portal.unesco.org/culture/en/ev.phpURL_ID=34603&URL_DO=DO_TOPIC&URL_SECTION=201.html

UN (2009). Retrieved September, 2009, from http://www.un.org/esa/dsd/agenda21/ res_agenda21_26.shtml

The North Bahnaric Clade: A Computational Approach[16]

Jerold A. Edmondson
University of Texas Arlington

Ken Gregerson
SIL International

Paul Sidwell
Australian National University

1. Introduction to the North Bahnaric languages.

Languages of the North Bahnaric sub-branch were among the first Austroasiatic tongues of Indo-China to be known and documented by westerners. In the mid-1800s the Vietnam central highlands were notionally under Siamese rule, via the Prince of Bassac, based at today's Ubon Ratchatani. In practise this amounted to hardly more than occasional visits by Siamese tax collectors willing to brave the serpentine trails that wound through the heavily forested plateaus and valleys. The Vietnamese were seen even less often, being largely terrified of the 'barbarous' highlanders.

Approaching from the coastal side, French missionaries penetrated the Highlands and established a permanent presence by the 1860s. A concerted campaign of conversion of the locals, initially to Catholicism, and later also Protestantism, was a tremendous success. The missionary endeavour ensured that important North Bahnaric languages, such as Sedang and Rengao, were documented and substantial lexicons became available in the West (e.g. Bastian 1868 and others).

By the 1930s there was a substantial plantation economy, directed by the French and worked by the local montagnards, many of whom were educated in French, and living in a world quite alien to that of the Vietnamese dominate lowlands. With decolonisation the region came under effective Vietnamese control, and there began a long process of integration with the lowlands, both social and economic, that continues to this day. Along with this, Vietnamese has replaced French as the administrative and inter-language of the Highlands.

Today there are at least a dozen named North Bahnaric languages that can be distinguished on linguistic grounds (plus a number of local and dialect variants). The common names for these languages, and population figures (taken from ethnologue.com), are given with the map figure below.

Additionally, Bahnar, an important local language used extensively by the Catholic Church based in Kontum, was also classified by linguists as North Bahnaric for many decades, but since the mid-1970s it has been increasingly recognised as belonging to a different, Central Bahnaric sub-branch.

Demographically we have the following situation; while there are more than a quarter million people speaking North Bahnaric languages, this figure is dominated by *Sedang* and *Hre* with more than 200,000 speakers combined. These two languages are very closely related; in fact in the late 1960s Kenneth Smith

[16] The authors wish to acknowledge helpful comments by Simon Greenhill; the usual disclaimers of responsibility apply. Additionally, it must be noted that Paul Sidwell's contribution to this paper was made possible by support received from the National Endowment for the Humanities. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily represent those of the National Endowment for the Humanities.

Edmondson, Jerold, Ken Gregerson and Paul Sidwell. 2011. "The North Bahnaric Clade: A Computational Approach." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.33-47.

prepared a draft reconstruction of proto-Hre-Sedang which, although not published, is cited extensively in his (1972) reconstruction of proto-North Bahnaric. The smaller neighbouring language of Todrah very likely sub-groups with Hre-Sedang, based on the phonological analysis by Gregerson and Smith (1973).

Language (population)
Duan (4350 in 1981)
Halang (17500 in 2002)
Hre (113000 in 1999)
Jeh (23210 in 2002)
Kacho/Kaco' (3370 in 2007)
Katua (3000 in 1981)
Kayong (2000 in 1981)
Monom (5000 in 1973)
Rengao (16000 in 2002)
Sedang (101000 in 1999)
Takua (12800 in 2000)
Todrah (9140 in 2000)

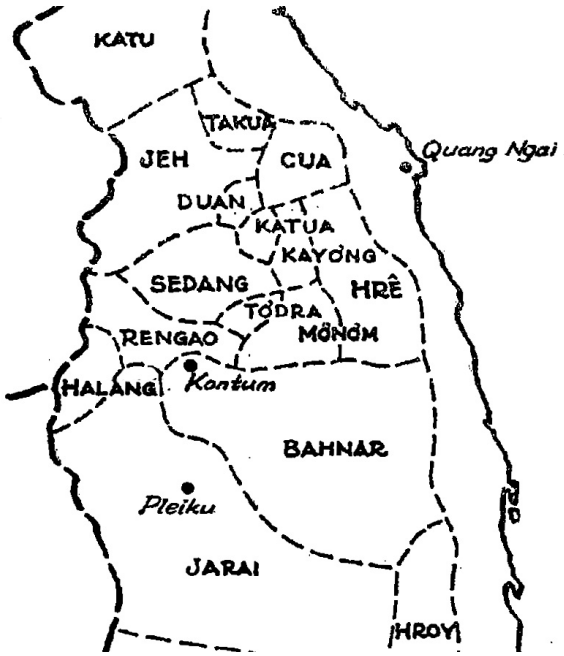


Figure 1: Fragment of SIL Map "Ethno-Linguistic Minorities of South Vietnam" (1966)

The next largest communities are the *Jeh* and *Halang*, with a combined population of around 40,000, spoken by communities to the north west and south west respectively of Sedang. Jeh and Halang are also very closely related; and a reconstruction of proto-Jeh-Halang was proposed by Thomas & Smith (967), and their comparative lexicon shows only rather trivial sound changes separating the two. The smaller groups Duan and Kayong are not well documented, but the limited available lexical data suggests a closer affinity with Jeh-Halang than to Hre-Sedang.

Then there are *Rengao* and *Kacho*. The Rengao have a long association with the Bahnar, so much so that until the 1960s French sources referred to then as the "Bahnar-Rengao", by virtue of the proximity of certain Rengao villages to the Bahnar area around Kontum City. French sources such as the Bahnar dictionary of Guilleminet (1959-63), in fact, identify Rengao as one of the 'sub-tribes' of Bahnar, however, he goes on to make clear that what he terms 'Bahnar Rengao' are not to be confused with the more numerous Rengao identified by Kemlin in 1900. Unfortunately, the confusion persists even today in some quarters that Rengao is only a dialect of Bahnar. Gregerson (1990) distinguishes three varieties of Rengao as: Central Rengao (Plei jodrap, Dak Wok), Sedang Rengao (Dak Kadem) and Bahnar Rengao (villages close to Kontum City). Thus, certain Rengao villages are heavily influenced by their Sedang or Bahnar and others (Central) represent a more authentically distinct variety. Allowing for numerous Bahnar loans, Rengao is clearly North Bahanric; furthermore it is characterised by the monophthongisation of the etymological diphthongs, and the emergence of new on gliding diphthongs in tense register vowels. Interestingly, Kacho and Ramam seem to show the same or similar vowel history to Rengao, although superficially the overall resemblance to Rengao is reduced by a later independent general devoicing of stops in Kacho/Ramam.

Monom, *Katua*, and *Takua* are of more uncertain status. Smith (1973) suggested that Katua and Cua form an Eastern sub-group of North Bahnaric, but more recently Sidwell (2009) showed that Cua constitutes another branch of Bahnaric (East-Bahnaric) while Katua apparently shares with Cua a rule that turns final nasals into stops based on later contact.

Essentially the broad pattern of one very large sub-group (Hre-Sedang) and numerous smaller groups is explained by geography: the Sedang occupy the large fertile plateau north of Kontum, and the Hre a similarly hospitable region about half way to the coast. In between and around are steep mountain sides, with small villages clinging to ridgelines and narrow river flats. Farthest to the West, Kacho speakers live in Ratanakiri Province of Cambodia, with speakers of their sister dialect *Ramam* in a small community of just a few hundred on the Vietnam side of the border.

Linguistically North Bahnaric is marked by the development of a tense-lax voice opposition or 'register' system, which remains quite mysterious as to its origins compared to analogous developments in other Austroasiatic groups. In all but Sedang, the contrast is realised as breathy verses plain voice, while in Sedang the whole system has become tenser, so that the otherwise plain voice has become creaky, and breathy syllables became plain. The striking fact is that, while generally such register systems develop in connection with devoicing of voiced stops, as is well known for Mon, Middle Khmer, and others, there was no general devoicing in North Bahnaric, in fact rather perversely Sedang alone has a general devoicing of initial stops correlating with its tensing of the register system.

Also, North Bahnaric languages show a unique correspondence of [s ~ ts] to [t] in the rest of Bahnaric, e.g.

	North Bahnaric			Other Bahnaric			
English	Sedang	Jeh	Rengao	Bahnar	Chrau	Laven	Nyaheun
'bone'	*kəsiəŋ*	*kəsi̯aŋ*	*kətsɐ̣ːŋ*	*kətiːŋ*	*ntiːŋ*	*kətɨəŋ*	*grɨəŋ*
'monitor'	*kəsom*	*kəsu̯am*	*kətsọːm*	*kətoːm*	*kətuːm*	--	--
'stem, begin'	*pəsiəm*	--	*tsɐ̣ːm*	*pətəːm*	*təːm*	*taːm*	*tɐːm*

And there was a fronting of proto-Bahnaric ***ə**: that emptied the central vowel space, e.g.:

	North Bahnaric			Other Bahnaric			
English	Sedang	Jeh	Rengao	Bahnar	Chrau	Laven	Nyaheun
'leech	*pliəm*	*plɛɛ̣m*	*plɐ̣ːm*	*pləːm*	*pləːm*	*plaːm*	*plɐːm*
'answer'	*tiəl*	*tɛɛ̣l*	*tɐ̣ːl*	*təːl*	--	*taːl*	--

North Bahnaric is also distinct, with some striking lexical innovations. For example, NB languages show [hi ~ hĩ] for 'day', replacing the AA etymon *tŋaj; and [ʔmaw ~ bạw] 'paddy' replacing AA *ɓaːː. There are also numerous loans from Chamic; Chamic speaking Jarai live immediately to the south of Kontum, and have a long history of trade and intermarriage with North Bahnaric speakers. But even more important is probably the 2000 or so years that Chams controlled the Vietnam central coast and hinterland, until the Vietnamese overran the Champa provinces of Amaravati and Vijaya in the 15th century. Cham towers still stand here and there in the central highlands, and various montagnard rituals have been observed to contain Hindu elements apparently communicated by the early Chams (most Chams adopting Islam during the second Millennium).

Above we mentioned that Smith (1972) was convinced of a subgroup Sedang-Hre and drafted a reconstruction. Thomas & Smith (1967) also published a paper on Proto Jeh-Halang within North Bahnaric. Gregerson & Smith (1973) argued that the Todrah (composed of Didrah and Modrah) subgroup has an affinity to Sedang-Hre as well. Additionally, Rengao and Kacho apparently show a similar history in their vocalism. We will want to see if the account in pursued in the following section confirms or contradicts these views. Since none of the North Bahnaric languages have a written tradition of any depth nor is there record of a parent, like Old English, etc. of the languages studied here, one cannot make use of known chronological precursor forms of today's languages to aid in deciding subgroups or for reconstructing the parent Proto-North Bahnaric. In just such cases the approach to be outlined in Section 2 is a way to move the research forward by proposing an estimate of the descent of Proto North Bahnaric with branches using a computational hypothesis of descent.

2. Reconstruction and computational solutions to the North Bahnaric proto language.

Linguists in the early 19th century discovered that languages belong to linguistic families.[17] The problem from the earliest times has been to show which languages belongs to what families, to discover then how the parent language in treelike fashion developed into subgroups that ultimately ended in daughter languages. Joseph Greenberg (1957:44) may have been the first to have noted that finding a linguistic tree of language descent involved a forest of possibilities. For eight languages, he observed 4140 trees are possible.[18] That led him to propose an approach he called mass comparison (aka mega comparison) that used a few items of basic vocabulary, some morphological features, and some higher level elements. With this method he found the largely accepted African subgroups Afro-Asiatic and Niger-Congo and, more contested Nilo-Saharan. His Amerind for the Americas, though, has been strong criticized by comparatists, especially Lyle Campbell. Critics have said that Greenberg's analysis relied on a small number of words across many languages.

In contrast to Greenburg, the comparative tradition has argued that membership and the nature of subgroups can be shown with precision by using the more restricted methods of Comparative Reconstruction.19 This approach was employed by 19th century scholars August Schleicher, Rasmus Rask, Jakob Gramm, and others, who studied a small number of languages but examined a large number of words, contra Greenberg, cf. Campbell (2004). Modern comparatists point to lists of steps or algorithms, to be used. Campbell (pp.126-142) describes several steps: *(a) assemble potential cognates, (b) establish correspondence sets, (c) discover which sets are in complementary distribution (d) reconstruct the proto-phonemes, and (e) examine the reconstructed set typologically*. The proto language that results is a "…hypothesis (or better said, a set of interconnected hypotheses…aspects of the hypothesized reconstruction can be tested and proven wrong." (p. 147).

Campbell recognizes that testing of hypothesized reconstructions should be welcomed. One method that has been developed to test hypothesized reconstructions is the computational estimations of a language's history by employing techniques and software from molecular biology. Despite Campbell's willingness to consider quantitative methods in historical linguistics, others have not viewed these techniques with favor. Consider, for example, the judgment of one Paul Ross (quoted in Embleton (1986) and McMahon & McMahon (2003), who cite from sources this quote:

> In comparison to old and established techniques, numerical methods must surely always be either inefficient or supererogatory. That is to say, on the one hand, if no solution to a problem of this kind can be reached with old methods then I would not trust a numerical solution, and on the other hand, if a solution can be reached with the old methods, then a numerical solution is unnecessary.
> (Ross 1950:59)

Not everyone has been quite so negative. Kroeber and Chrétien (1937:85) said that a statistical validation can confirm "correct insight, or, where insight judgments are in conflict, help to decide between them…" In the case of North Bahnaric, quantitative estimations of the evolutionary history can be helpful, as one is dealing with lesser-studied languages, where reconstruction must begin with contemporary data. We don't have the benefit of a written history as in the Indo-European languages, which could provide us with a depth and variety of old forms, nor is there an accepted benchmark proposal to which quantitative solution can be compared. The question is then how to begin a quantitative account of a language's history?

---

[17] Some don't, of course. There are doubtless many languages that are hard to assign to a family. Whenever a language has undergone a discontinuous period in its development, pidgins, creoles, fusions, etc., then assignment to a family is not possible. Similarly, there are cases of convergence or strong interlectal borrowing that also make a clear choice impossible.

[18] Greenberg's calculation missed some possible trees. According to Felsenstein (1978), there would be 135,135 rooted trees or 10,395 unrooted trees. Still he was aware that finding a tree of descent would involve a lot of possibilities.

[19] Despite the claim of algorithmic approaches to reconstruction, many practioners have proposed a intuitive-deductive approach, such as *retrograde reconstruction*, cf. Benedict 1994.

The most promising approach for estimating the history of a language group is a set of techniques borrowed from evolutionary biology. The approach is not intended to replace classical comparison, but to verify conflicting hypotheses or to make a hypothesis about a language whose ancestry has not been decided or where there are conflicting views.

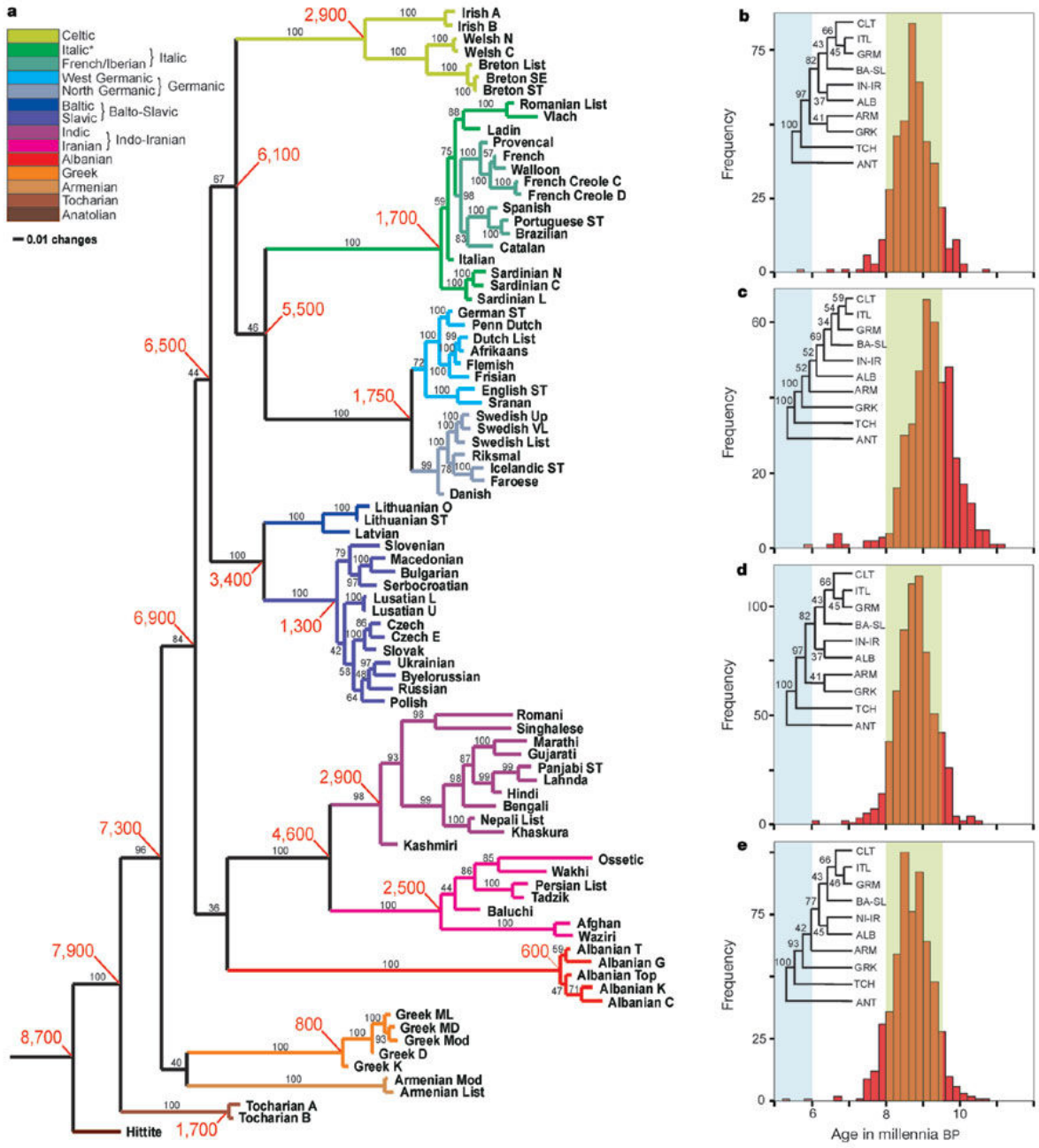


Figure 1: Indo-European tree with absolute dates for branching (Gray & Atkinson 2003).

There exist many commonalities between historical linguistics and biology. For example, one of main areas of biological study is biological systematics, which is concerned with determining the evolutionary history of organisms, including the branching topology of the tree and the length of its branches. However, in the past, biology, as language study, had a problem in the analysis of the taxonomies of hybridized species; how is one to treat these cases? A similar problem occurs in the study of language contact and convergence phenomenon. Also, languages have similarities with biological organisms because they

both live and change over time. One large difference between them, though, is that until recently linguistics has not been reliably able to determine the ages of language dead or living, lacking robust methods for putting numbers onto the points of branching in a reconstruction. That situation changed in 2003 with the publication of Gray & Atkinson (2003), which shows all the branching of IE and includes the time of the splits in red.

Let us now examine some relevant issues about the history of biological methods. Until 50 years ago or so, branching decisions were made by examining the structural and behavioral characteristics of organisms, the *morphology* (not in the sense of linguistics) of a plant or animal. This approach is still important. But as genetic structures became better understood, there has been a shift to the DNA as the ultimate decision maker. And for calculations with DNA, computation methods are imperative. When work in that direction began, biologists used a clustering method, UPGMA (an unweighted pair group method with arithmetic mean), to construct phylogenetic trees. This method was also the one used in lexicostatistics to analyze the percentages in shared cognacy matrices, cf. Greenhill & Gray (2009:4). For biologists, this technique, not surprisingly, did not always give correct results, especially when there were differences in the speed of changes among the organisms; using UPGMA, members of the same clusters all have the same branch length cf. Felsenstein 1978, Wichmann n.d. It was one of the problems that caused lexicostatistics to fall into disfavor among linguists. Thus, understandably, linguists were suspicious and some reluctant to accept a proposal they thought to be similar to lexiocostatistics, cf. the discussion on Greenhill & Gray (2009:1). But biologists turned these early failures into a wide set of computational instruments that can deal with a "non-uniform evolution across a data set."

One of the most reliable is Bayesian inference implemented in MrBayes, which was employed by Gray & Atkinson (2003). It relies on an overt model of evolution created before the computation. Gray & Atkinson (2003) for their analysis of IE data employed *a restriction site model* that allowed cognates to be gained or lost at different rates. Because Bayesian inference starts from a model of evolution and uses a stochastic method to evaluate candidate trees as it processes, it is regarded as the most powerful and robust approach available today.

Consider the results of applying MrBayes to a data set with DNA encoding of character states.

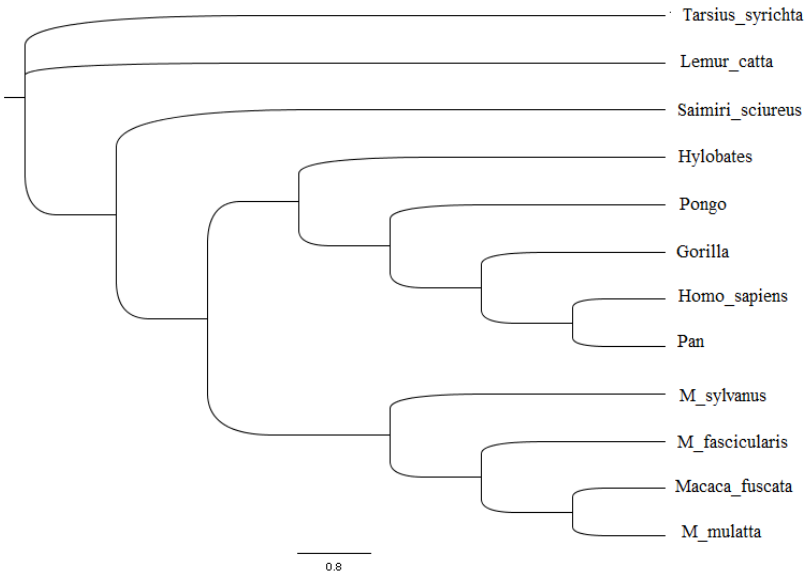


Figure 2a: A Bayesian tree of a descent for primates—Homo sapiens, Pan, Gorilla, and gibbon, macaques, etc. as computed by Mr Bayes trees and in Figure (1b)

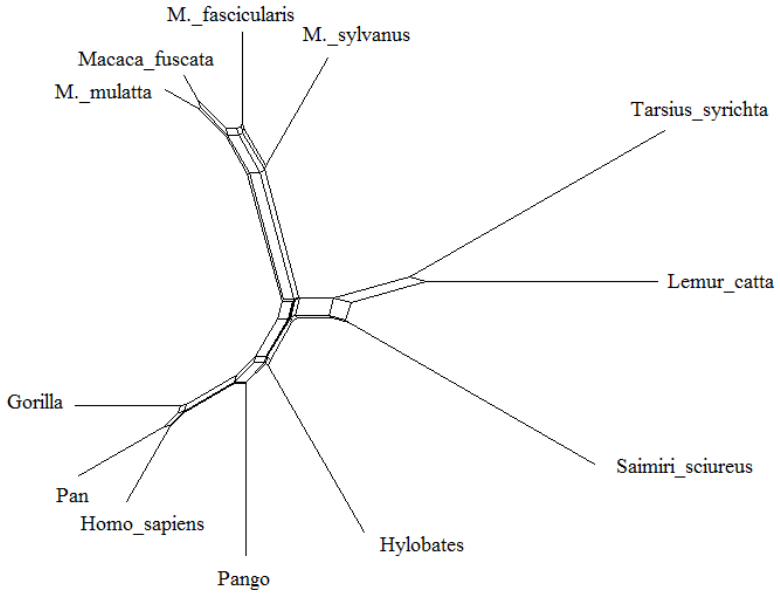


Figure 2b: A Network using the same dataset computed by Splits-Tree 4.0.
The line length in Splits-Tree indicates the distances among species

The two graphs above have been produced with a file that has a model with datatype = DNA, but for linguistic cladistics this datatype and other model types are not appropriate. If we are able to use these methods and approaches from biology, then we must choose among their tools. Admittedly, the two disciplines have similar aims, but, to use the software of the biologist, linguists need to develop machine-tractable coding of data of a type that can be processed by a computer and to construct models for the evolution of linguistic trees approach to the object, cf. Greenhill & Gray 2009.

Fernández-Baca (2000:2-4) give an excellent teaching example of how characters for this clade can be digitized. Consider the animal group (taxa) made up of the lamprey, shark, salmon, and lizard and the characters for these animals[20] *(a) paired fins, (b) jaws, (c) dermal bones, (d) fin rays (spines), (e) lungs, and (f) rasping tongue* for which a character state matrix can be created.

(3)

	a	b	c	d	e	f
lamprey	0	0	0	0	0	1
shark	1	1	0	1	0	0
salmon	1	1	1	1	0	0
lizard	1	1	1	0	1	0

In this example binary encoding for possession of the character states is used; 1 means "has the character"; 0 means "it does not". Non-binary coding is also possible, but won't be discussed here. To accommodate larger databases for coding the lexical, morphological characters of a large language group, more powerful data manipulation, as is provided by MS-Excel, and other such spread sheets, is needed. Also since glosses are more numerous than taxa, it is most convenient to transpose the matrix (3) to a form as in (3'), because editing and adding additional rows is easier than adding columns. Later the orientation can be re-transposed for the nexus file, which is the file type required by Mr Bayes and Splits-Tree 4.0.

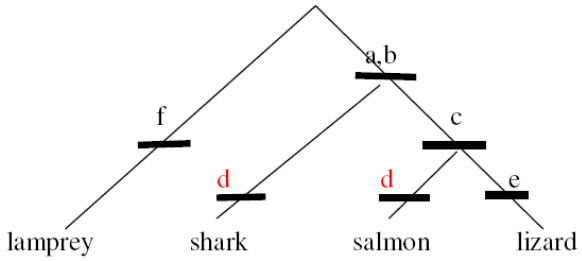
[20] In languages study "character" equates to gloss, an overlying semantic etic grid or list for which an investigator looks to find a structure or word for each semantic cell. A word for any given cell is the character state for that manifests that gloss in the language being studied, Nichols & Warnow (2008:765). Examples of gloss lists are the 100 and 200 gloss lists of Swadesh, and ones adopted for specific regions of study.

(3')

Taxa	lamprey	shark	salmon	lizard
Gloss				
paired fins	0	1	1	1
jaws	0	1	1	1
dermal bones	0	0	1	1
fin rays spines	0	1	1	0
lungs	0	0	0	1
rasping tongue	1	0	0	0

The data in matrix (3) corresponds to a tree structure with character changes corresponding to evolutionary steps in the tree, starting at the top

(4)



The phylogeny in (4) is rooted from an invisible branch above it with all its states unknown. The points in the tree where the character states emerge are indicated by a bar.

(5)

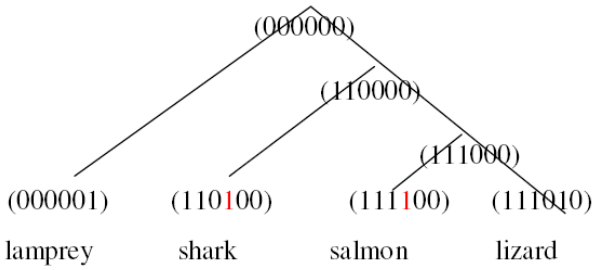


Figure 5 shows how glosses and taxa determine a character state matrix.

Another way is to represent the result of the changes beginning at the top (root) with later nodes inheriting character states from above or manifesting innovations.

2.1 Homoplasy.

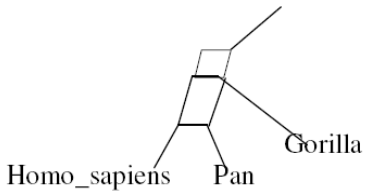
One can see among the characters in (5) that the same innovation has occurred twice (red). The character *(d) fin rays* appears to have been innovated independently two times; once by both sharks and once by salmon. In other words the development of spines (fin rays) represents a *parallel development*. Parallel development of characters, like the development of wings in birds, mammals, insects, etc., impedes finding the ancestry of this clade and is an example of homoplasy. In natural language evolution there are also examples of homoplasy. Examples might be: (a) the parallel development of *devoicing of final consonants* among several IE languages, e.g. in German and Russian and (b) *diphthongization of /ī ū/ to /ai au/*, respectively, from Middle English to Modern English and Middle High German to Modern German, respectively as in [rīd/t] to [raid/t] 'to ride' and [hūs] to [haus] 'house'.

2.2 Networks vs. Trees.

The evolutionary history of a language has been usually represented by a tree of descent to show the topology (subgrouping) and branch length. But it has been argued since the 19th century whether it would be

better to represent language change by waves or trees. Leskien insisted that rules could have not exception; Schmidt pled for a wave account as change spread across an area. Biology must also account for non-treelike phenomena such as variation, hybridization, interlectal (horizontal) borrowing, dialect continua, etc. Splits-Tree tries to do justice to both non-tree-like developments while still preserving some properties of trees with its proposal, which allows more than one path to get from one node to the next, as in (6).

(6)



Figures (2b) and (6) include information about the tree signal as well as about noise in the signal (Nichols & Warnow 2008:764). This network can represent cases where there are contact events between the parallel lines or homoplasy (parallel development or back mutation). The length of lines from one taxon to another is representative of the distance between them. Despite their advantages, it is difficult to interpret these graphs, as they do not represent the evolutionary history of the clade. Nevertheless, linguistic work has often analyzed data with Splits-Tree as well as Mr Bayes.

2.3 Characters and Character States.

To prepare linguistic data for analysis, we must create a matrix with taxa (languages) indexing the columns and characters indexing the rows. The matrix is itself filled with character states as in (3). For linguistic analysis, it is best to put the actual linguistic items into the cells of the matrix. That way one can return to check examples for errors. If one were to put in only a 1 or a 0 at this point, then it would difficult to check past work. So, we must take advantage of every aid in this bookkeeping-like task and first put in a representative of the lexical item or morphological feature and its variation across taxa. Linguistic characters can be of virtually any type; Nichols & Warnow (2008:764) say "...variously as features, traits, properties, variables, and probably other terms in linguistics." They go on to say that character can be of two types. One is *cognates*, which are individual morphological forms with meaning that are clearly descended from an ancestor form. The other kinds are *phyletic characters*, e.g. sound, morphological, or other changes within North Bahnaric clade. These must be included in the character state matrix file signaling that the changed forms should be recorded on the row below in the character state matrix, e.g. in North Bahnaric final nasals go to homorganic stops in Takua.

(7) a. Takua: ʔu**p** 'winnow', Jeh: ʔu̯**m** 'winnow', Halang: ʔu̯:**m** 'winnow', Sedang: ʔuə**m** 'to winnow', Hrê: ʔu̯m 'winnow rice', Rengao: həʔu̯:m 'winnow'

b. Takua: ʔu**t** 'fire, charcoal', Kayong: u**n** 'fire', Jeh: ʔu̯**n** 'fire', Halang: ʔu**n** fire', Sedang: ʔɔ**n** 'fire', Hrê: ʔu̯**n** 'fire', Mơnam: ŭ**n** 'fire', Rengao: ʔu̯**n** 'fire'

c. Takua: hri**k** 'hundred', Kayong: re**ng** 'hundred', Jeh: rɛ:**ŋ** 'hundred', Halang: rɛa**ŋ** 'hundred', Sedang: hri̯**ŋ** 'hundred', Hrê: hri**ɲ** 'hundred', Mơnam: hrie**ng** 'hundred', Rengao: hri:**ŋ** 'hundred'

This change must be unique in this clade and should be recorded in the row below, as is seen for the word *star* in (8b).

(8a)

Jeh	Halang	Rengao	Kachoq	Rmam	Didrah	Modrah	Hre	Sedang	Bonam	Monom	Takua	Kayong
hlɔŋ	həloŋ	həlɔŋ	hmlɑŋ	hnuoŋ	həlɔŋ	hənɔŋ	həlɔŋ	həlɔ́ŋ	solong	saloŋ		hơlong
											salɔk	

And later the filled cells are replaced with 1 (yes) with 0 (no) and ? for missing data.

(8b)

Jeh	Halang	Rengao	Kachoq	Rmam	Didrah	Modrah	Hre	Sedang	Bonam	Monom	Takua	Kayong
1	1	1	1	1	1	1	1	1	1	1	0	1
0	0	0	0	0	0	0	0	0	0	0	1	0

In order to justify a different character state for 'star' in Takua, a lexical item needs to show that the change is a recurring rule and also that the rule needs to be specific (operates in a defined environment) and be unique to this clade. *Typological characters*, for example, "natural" occurring rules (final devoicing, h → s, shortening of diphthongs, etc.), and glottalization of consonants, development of tone, etc. cannot define a different character state, since typological rules are homoplastic by definition, i.e. not unique to one clade, cf. Nichols & Warnow (2008:765).

2.4. Nexus files.

Nexus files contain the data to be processed by many of the commonly used software implementations of phylogenetic computation. There are two different formats for Nexus files. Both types are ASCII or text. The files are divided into blocks, sections with particular functions; we will illustrate only one.

The first block is the NEXUS *header* for defining the number of taxa and characters. The second, FORMAT establishes the datatype. Usually the type is STANDARD with binary encoding. After that is the MATRIX part of the file, which has the taxa and the binary encoding of the character states, here abbreviated to 5 columns of the 852 columns. Finally, the file is closed with END, to which is added an appendix that allows the very same file to be used by MrBayes. It shows the number of generations, print frequency, sample frequency, number of chains, and the instruction to save branch length.

```
NEXUS
begin data;
dimensions ntax=13 nchar=852;
FORMAT
MISSING=? [GAP=?]  Datatype=STANDARD [SYMBOLS = "0 1"];
  MATRIX
  Jeh         1     1     0     0     0 …
  Halang      1     0     0     0     0 …
  Rengao      1     0     0     0     0 …
  Kacho       0     1     0     0     0 …
  Ramam       0     1     0     0     0 …
  Didrah      1     0     0     0     0 …
  Modrah      1     0     0     0     0 …
  Hre         1     0     0     0     0 …
  Sedang      1     0     0     0     0 …
  Bonam       1     0     0     0     0 …
  Monom       1     0     0     0     0 …
  Takua       1     0     0     0     0 …
  Kayong      1     0     0     0     0 …
;
END;
begin mrbayes;
mcmcp ngen=2000000 printfreq=1000 samplefreq=100 nchains=4 savebrlens=yes;
end;
```

3. Results of the calculation.

The calculation for the North Bahnaric clade was completed using Splits-Tree 4.0. The network approach is easy to complete and the calculation time is brief, a few seconds at most. The character matrix

file consisted of the 13 taxa and 235 characters that correspond to 852 character states. The model assumed is that all possible trees are of equal value all Splits-Tree produces an estimate of the distance among taxa.

3.1 Splits Tree 4.0 Analysis

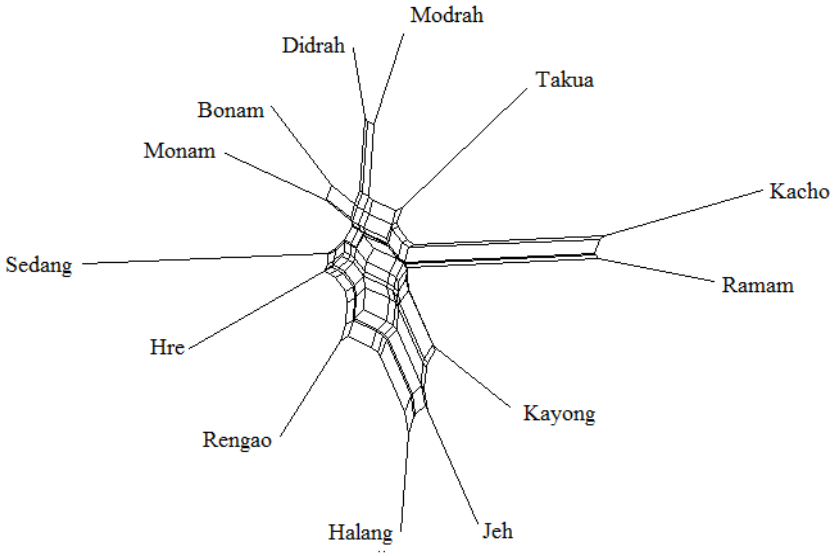


Figure (9) The North Bahnaric clade with a Splits-Tree analysis of distance

Figure (9) shows the expected sister pair Jeh and Halang and a more distantly connected taxon, Kayong, cf. section 1. This grouping had the strongest connection of all taxa in the branch calculations. The sister branches—Monam-Bonam, Didrah-Modrah, and Kachoq-Rmam—have also always appeared in a tight group pattern. Takua, in opposition to the sisters, always shows up as a singleton. Takua is on a branch that emerges from the main body of the network without forming connections to neighbors. Such branches are said to show *rapid radiation* meaning they have evolved rapidly with strong isolation from other taxa.

Another way to represent the data for the graph is to use a "greedy compatible data set" from the Split-tree network, considwer Figure (10). This kind of computation is often used as a first approximation. The underlying approach is based on a search, or better said, a problem-solving approach to a search, a search heuristic or search strategy. Greedily compatible search means making a choice on what seems most compatible (formally defined) at the moment and assume that, in subsequent decisions of choice, applying this strategy repeatedly will be globally the optimal choice. We are trusting that a fast solution, followed up by a string of speedy solutions will lead to a true solution. But greedy compatibility can be wrong, as H. L. Menken said in (The Divine Afflatus, in the New York Evening Mail (November 16, 1917), "There is always an easy solution to every human problem—neat, plausible, and wrong." Nevertheless, the greedily compatible result closely resembles the network graph starting with Jeh-Halang and Kayong, which then progresses to a rapid radiation Kacho-Ramam pair. Soon thereafter the isolate Takua is followed by Sedang-Hre-Rengao and the two sister pairs Monam-Bonam and Didrah-Modrah

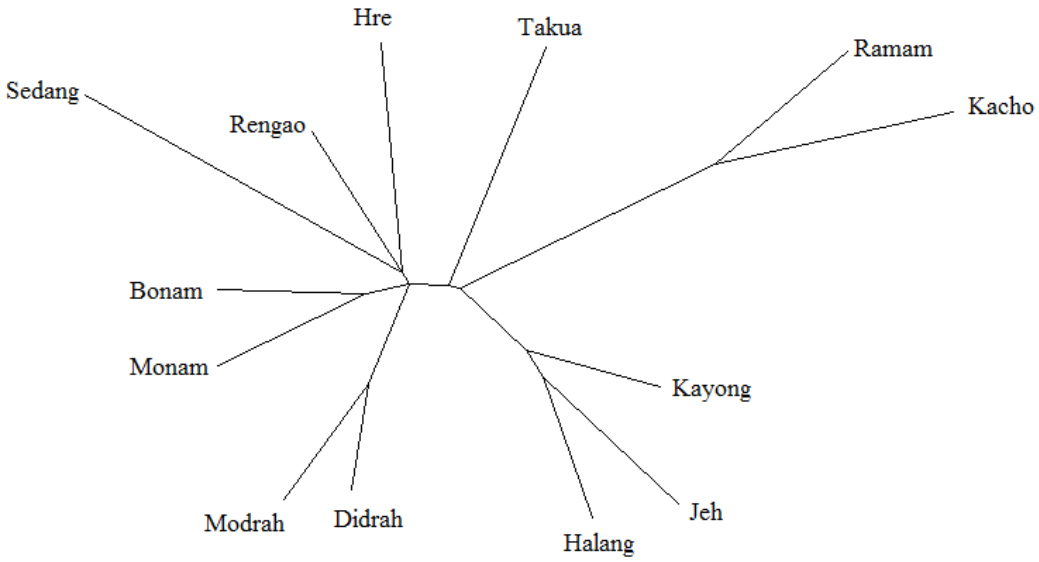


Figure (10). The Mr Bayes analysis with Mesquite

3.2 MrBayes Analysis

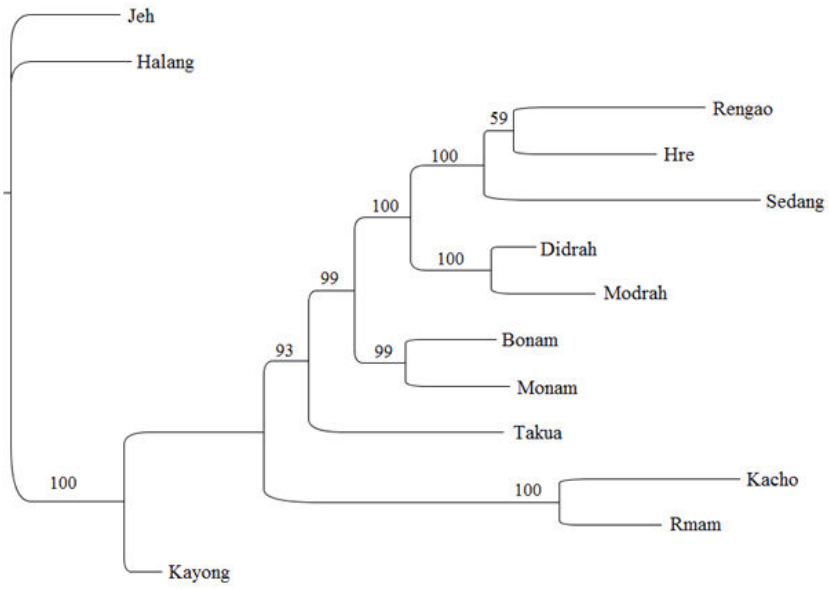


Figure (11): The MrBayes result with branch length showing distance,
extracted and processed by Fig Tree 1.3.1

MrBayes calculations often take some time. For example, when the number of generations was set to 2 million, the run time was 40 minutes. The candidate tree was created from a file with 852 character states and 13 taxa. The model used at the start was the "flat" where every possible tree was valued equally. Then the algorithm started a random walk through model tree space comparing a starting tree model with new candidates and then observing the probability of the sequences being produced by a given model tree. If the new candidate tree is evaluated as more probable, then it takes the place of the standard and the search continues in this manner until there is no longer improvement. At termination of the calculation the value of the convergence diagnostic in the computation for stationarity was 1.000 (should approach 1) for the potential scale reduction factor of Gelman & Rubin (1992); the standard deviation of splits was 0.003 (should

approach 0), Nichols & Warnow (2008:774). From the credible set of trees, 29 trees were sampled. The 90% credible set had 4 trees; the 95% credible set had 7 trees, and the 99% credible set had 14 trees.

The result in Figure (11) has the consensus calculation for each branch length, which is reported by adding a command in the MrBayes chain. One feature that is evident is that the robustness for the branch between Rengao and Hre was only 59, while the others all had credibility values of 93 plus. Still the overall consensus is still strong support for this result.

One can use another graphic representation to show the reliability of the result. As mentioned above in our discussion of Split-Tree one can input the tree in Figure (11) into Splits-Tree, which can compute a consensus network. This evaluation also has [[Rengao-Hre] Sedang] in this format.

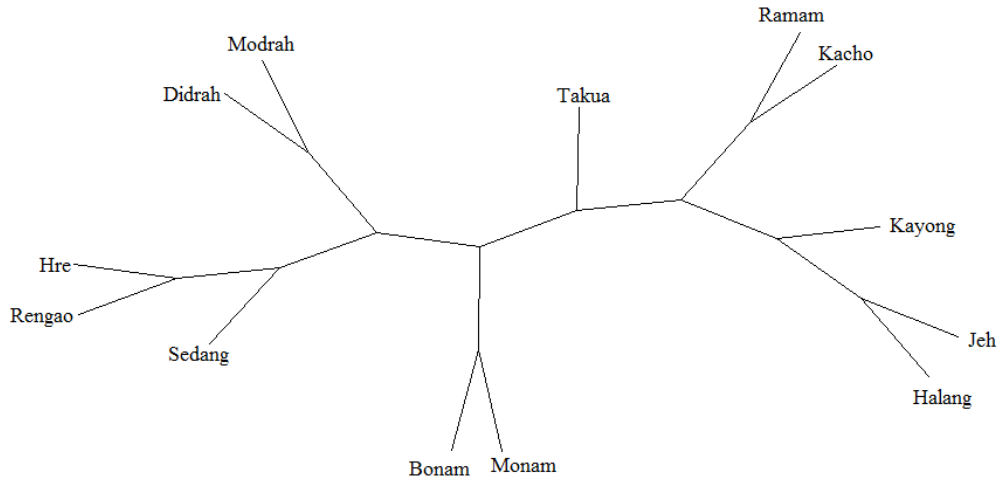


Figure (12): Consensus Network of Bayesian inference calculation for the North Bahnaric clade

This consensus network represents the overall interrelations among the taxa, showing especially that Hre and Rengao are more strongly related to each other than either is to Sedang. Since the relationship among the taxa is in focus, the branch length is not.

4. Discussion of results.

At the beginning of this paper we noted that earlier scholars saw strong connections between Hre, Sedang, and Rengao and that linkage is supported in our results. Similarly, Jeh and Halang, were so close that language workers in residence decided to reconstruct their source language. In fact all of the daughter languages that thought to exist in groups were confirm by all the evaluations: Hre-Sedang-Rengao, Jeh-Halang-Kayong, Monam-Bonam, Didrah-Modrah, as well as Kachoq-Rmam. The only singleton was Takua.

It is clear that one needs to discuss a bit the credibility of the Rengao-Hre connection. The results tell us that for this data set Mr Bayes had a lower credibility for this branch than the others. We examined the input file and found the following distribution of character states vs. languages:

Jeh	Halang	Rengao	Kachoq	Rmam	Didrah	Modrah	Hre	Sedang	Bonam	Monom	Takua	Kayong
220	230	236	197	196	180	177	215	240	90	146	122	173

Bonam, Takua, and Kayong had fewer data points than the other taxa, because our data sources for these languages had fewer items. So it does not seem to have been a factor in the calculation. But Rengao, Sedang, and Hre were richly represented and the cognates among these three totalled 127, though Rengao and Hre had slightly more cognate states in common than between Hre and Sedang.

As we stated at the beginning, phylogenetic calculation of the evolution history of a language is not a reconstruction, it is an estimate of the history of the clade. We are unable use examples from old languages and to put time points for the splits, since we have no extra linguistic data for clade members that would allow us to put a time scale on the numerical lengths.

We close in stating that this estimate is a hypothesis about the divisions in this family. It seems to confirm earlier informal work and will hopefully contribute to a comparative study by later investigators.

Data file available on request.

References

Barbançon, François, Tandy Warnow, Steven N. Evans, Donald Ringe, and Luay Nakhleh An experimental study comparing linguistic phylogenetic reconstruction methods (BWERN) (2009). Proceedings of the conference Languages and Genes, held at UC Santa Barbara (Cambridge University Press)

Bastian, Adolf. 1868. Reise durch Kambodja nach Cochinchina. *Die Voelker des oestlichen Asien: Studien und Reisen von Dr. Adolf Bastian, Vierter Band*. Jena, Herman Costenoble, reprinted 1967 by Oxford University Press, Kuala Lumpur.

Benedict, Paul. 1994. Retrograde Reconstruction in Southeast Asia. In *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, edited by Karen L. Adams and Thomas John Hudak. 15-23. Arizona State University, Program for Southeast Asian Studies

Campbell, Lyle. 2004. *Historical Linguistics: an Introduction* (2nd edition). Edinburgh: Edinburgh University Press, and Cambridge, MA: MIT Press.

Donahue, Mark, S. Musgrave, B. Whiting, and S. Wichmann. 2011. Typological feature analysis models linguistic geography. *Language* 87.2.369-81.

Embleton, Sheila M. 1986. *Statistics in Historical Linguistics*, Brockmeyer Verlag, Bochum

Felsenstein, Joseph. 1978. The number of evolutionary trees. *Systematic Zoology* 27.1.27-33.

Fernández-Baca, David. 2000. Du, D.-Z & X. Cheng (eds.) Steiner trees in industries. Kluwer Academic Publishers.

Gelman, A. and Rubin, D. B. 1992 Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Gray, Russell & Quentin Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin**.** *Nature* 426, 435-439 (27 November 2003).

Greenberg, Joseph. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press

Greenhill, S.J. & Gray, R.D. 2009. Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In A. Adelaar & A. Pawley (eds.) *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Canberra: Pacific Linguistics.

Gregerson, Kenneth & Kenneth Smith. 1973. The development of Todrah register. *Mon-Khmer Studies* 4.143-182

Gregerson, Kenneth J. 1990. Dân tộc Rongao, Bahnar và các dân tộc lân cận [Rengao, Bahnar and their neighbors]. *Tạp chí khoa học xã hội* [Journal of Social Science], No 2: 80-90.

Guilleminet, Paul et R.P. Jules Alberty. 1959. *Dictionnaire bahnar-français.* Volume 1.& 2 Hanoi/Paris, Publications de Ecôle Française d'Extrême Orient.

Hoenigswald, Henry. 1960..*Language change and linguistic reconstruction*. Chicago: Univ. of Chicago Press.

Huson, D.H. and D. Bryant. (2006)Application of Phylogenetic Networks in Evolutionary Studies, Molecular Biology and Evolution, 23(2):254-267

Kroeber, A. L. & C. D. Chrétien 1937. Quantitative Classification of Indo–European Languages. *Language*. 13:83–103

McMahon, April & Robert McMahon. 2003. Finding families: quantitative methods in language classification. *Transactions of the Philological Society* 101.7–55.

Nichols, J. & T. Warnow. 2008. Tutorial on computational linguistic phylogeny. Language and Linguistic Compass 2/5 (2008): 760–820, 10.1111/j.1749-818x.2008.00082.x.

Page, R. D. M & E. C. Holmes. (1998) Molecular evolutions: A phylogenetic approach. Blackwell Science: Oxford.

Sidwell, Paul. 2009. How many branches in a tree? Cua and East (North) Bahnaric. In Bethwyn Evans (Ed.) *Discovering History through Language: papers in honour of Malcolm Ross*. Canberra, Pacific Linguistics, pp.193-204.

Smith, Kenneth, D. 1973. Eastern North Bahnaric: Cua and Kotua. *Mon-Khmer Studies,* 4: 113-118.

Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas, Language Data Series, Summer Institute of Linguistics.Thomas David, & Marilyn Smith. 1967. Proto-Jeh-Halang. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 20: 157–175.

Online sources:

Wichmann, Søren. n. d. Quantitative approaches to language change. =wichmann+ppt+on+language+change&oq=wichmann+ppt+on+language+change&aq=f&aqi=&aql=&gs_sm=e&gs_upl=8006l27452l0l27652l51l51l1l0l36l46l0l215l815l0.4.1l5l0&bav=on.2,or.r_gc.r_pw.&fp=a1353fad56f6e656&biw=1278&bih=715 accessed on September 18, 2011.

While not mentioned in the text the works from which the data was taken are (full details or images supplied on request):

Sidwell, Paul. n.d. PNB etymologies working file. Ms.
N A. ND. Halang-dictionary Unicode unchecked. Ms
Hrê-Sedang data Unicode Ms.
Gradin & Cohen. Jeh 1979
Kayong list 2007
Smith, K. 2007. Modrah annotated. Ms
Monam 2007. Ms.
Monam list SIL unknown. Ms.
Smith, K & Cooper proto-Jah-Halang
Roman list 2007.
Sedang Dictionary Unicode. Ms.
Maier. 1964. Takua. Ms
Takua lexicon. Ms.
Todrah list 2007.

Burmese in Mon Syntax – External Influence and Internal Development

Mathias Jenny
University of Zurich

1. Introduction

Mon and Burmese[21] have been in close political and cultural contact for at least a thousand years, with changing balances of power and dominance. While a Mon kingdom centered at Thaton (probably part of the greater Dvāravatī cultural area in central Thailand) seems to have been the dominating polity in what today is southern Burma from the second half of the first millennium up to the 11th century, when according to traditional Burmese and Mon history, a view largely followed by most Western historians, King Anawrahta of Pagán invaded Thaton and introduced, among others, Mon literacy to the Burmese empire (see e.g. Phayre 1883:34f, Taw Sein Ko 1892:49f). Though this history has recently been contested by some scholars (Aung Thwin 2005), we can gather from the Mon inscriptional corpus that Mon was one of the first and most important vernacular languages both in Thailand and Burma. While the number of Old Mon inscriptions found in Thailand is rather small, consisting mostly of short texts, we have a fair number of long narrative inscriptions from Pagán (see, among others, Coedès 1929, Duroiselle et al. 1919-1928). Mon was apparently used at the court of Pagán as superordinate literary language, probably never spoken by large segments of the population of Pagán, alongside with Pali and later Burmese.[22] Given this situation of Mon being used in a politico-culturally Burmese environment from an early period, we cannot *a priori* exclude Burmese influence already in these Old Mon texts. This influence can be excluded only in the Mon inscriptions of Dvāravatī, but their shortness does not allow much in terms of conclusions about the pre-Pagán structure of Mon. With the beginning of Middle Mon around the 14th century, Burmese influence apparently increases in Mon, indicating that Burmese by that time has become politically and perhaps economically dominant. This increase of Burmese elements in all domains of Mon has continued into the modern period, at least in the Mon dialects spoken in Burma, and is still continuing with seemingly ever increasing speed. The dialects of the Mon populations that migrated to Thailand starting in the 16th century or earlier have been under ever increasing linguistic pressure and have consequently adapted more to the Thai model. The aim of this study is to examine a few examples of apparent Burmese influences in the structure of Mon, especially in the domain of syntax. Structural replication or "pattern loan" (Sakel 2007) in language contact is generally taken to be possible only in very close contact situations with large numbers of bilingual speakers in a society (see e.g. Matras 2009, Thomason and Kaufman 1988). The question to be answered concerns the status of the features under discussion as contact induced or rather independent, language internal developments in Mon. I will concentrate in this study on the following features, which strike as typological unusual in Mon at a first glance: Fronted interrogatives, clause final subordinators, and the loss of overt relative clause marking.

[21] I use the traditional terminology for Burma and Burmese rather than the more recent Myanmar/Myanma for linguistic reasons.

[22] The oldest Pyu inscriptions found in Burma predate Mon in that county, but not much can be said about the political and social status of the Pyu language at the time relevant to the present study. It may very have become a purely ceremonial language by the time of the Pagán kingdom.

Jenny, Mathias. 2011. "Burmese in Mon Syntax - External Influence and Internal Development." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.48-64.

2. The structure of Mon and Burmese

Typologically, Mon and Burmese are very different in most respects, apart probably from semantic convergence that can be attributed to the close cultural contact and partly to the shared Theravāda Buddhist background. Many loans went in both directions at different times, and a large number of idiomatic expressions are parallel, each language using its own lexical material to build a larger semantic unit (calques). In many cases these features are also shared with other languages in the area, but to my knowledge no systematic study has been conducted or published yet in this field, including Tai, Karen and other languages in the Burma cultural-linguistic area.

2.1 Mon

As far as we can tell at the present state of comparative Austroasiatic syntax, Mon is in many respects a fairly typical member of the Mon-Khmer branch of the Austroasiatic family. It is non-tonal, but possesses two distinct phonemic registers, which probably arose relatively late in the language, according to Shorto's (1971) dictionary only after the Middle Mon period, after the devoicing of originally voiced stops. Like other Mon-Khmer languages, Mon has a rich inventory of vowels and diphthongs, further enriched by the register distinction, including the central vowel /ɤ/, the back unrounded /ɒ/ as distinct from rounded /ɔ/. Vowel length is not distinctive in Mon, and probably has not been so since Old Mon. Initial consonants include voiceless/aspirated nasals (except */hŋ/) and the lateral /hl/ (but not /hr/ and /hy/), the implosives /ɓ/ and /ɗ/, as well as an opposition between palatal /c/ and palatalised /ky/ (the two latter sounds merge in some northern dialects in Burma). Initial clusters are restricted to stop+liquid, the old clusters stop+h being better analyzed as aspirated stops in Modern Mon. A restricted set of consonants occurs in final position, all unreleased, including the palatals /c/ and /ɲ/ as well as /ʔ/ and /h/. (see Jenny 2005:23-37 for a more complete account of Mon phonology).

Modern Mon has retained traces of Old Mon morphology, though not much of it is productive anymore. Old Mon morphology consisted mainly of derivative and a few inflectional pre- and infixes. The only affix still productive seems to be the universal prefix *hə-*, which forms any kind of derivative lexeme from a base verb and also tends to replace other prefixes (such as the causative *pə-*) in some dialects. The contrast base verb - causative verb is expressed consistently, but the forms involved are lexicalized, though the formation is still transparent in many cases. Modern Mon is for most practical purposes an isolating language.

Unlike its big neighbors, Mon does not make regular use of nominal classifiers, which are obligatory in both Burmese and Thai when counting or, in Thai, individualizing entities. Mon does make a difference between measure words, including units of weight, time, etc., and other nouns. The order in a numeral expression with the former is numeral - noun, as in *ɓa ŋuə* 'two days', while with the latter it is noun - numeral, as in *rɔ̀ə ɓa* 'two friends', lit. 'friend two'. The plural of nouns can be expressed by post-nominal *təʔ*, which resembles the Burmese post-nominal plural marker *tó* (cf. Bauer 2006:41).

Syntactically, Mon is predominantly SVO, with frequent fronting of topical or focal elements for contrast or emphasis. Known or contextually retrievable arguments are often omitted. In ditransitive expressions, the recipient precedes the theme. Modifiers generally follow the modified element, that is, attributive expressions (relative clauses, possessors, attributive verbs, demonstratives) follow the noun. Subordinate clauses usually follow the matrix clause, but this order can be inverted for pragmatic and sometimes semantic reasons. Markers of subordination occur in clause initial position in the subordinate clause. Interrogative elements generally occur *in situ*, that is in the position where they syntactically belong, but some are more or less regularly fronted (s. section 3.1). As other Southeast Asian languages, Mon exhibits a strong preference for topic - comment sentence structure. Verb serialization in Mon is mostly of the nuclear type, that is, verbs in a serial construction are adjacent with no intervening elements.

2.2 Burmese

Burmese is a typical Tibeto-Burman language in many respects, while it also shows typical Southeast Asian features. Burmese distinguishes 3 tones (4 according to some authors), which are realized in

a number of allotones each, depending on the context. The vowel inventory is of medium size, with four height distinctions and rounded back versus unrounded front vowels, except for the low vowel /a/, which is central. All vowels except /ɛ/ and /ɔ/ can occur nasalized. The diphthongs (ai, au, ei, ou) occur only in nasalized or glottalized syllables, spelled here with final /n/ and /ʔ/ respectively. Initial consonants can be voiced or voiceless, with phonemic aspiration distinction also in the nasals, /l ~ hl/ and /w ~ hw/. Initial clusters occur with /y/ or /w/ as second element only.

Burmese has some productive morphological processes, most importantly the deverbalizing prefix /ʔə-/ and what Okell and Allott (2001:273ff) call the "induced creaky tone", marking, among other functions, some kind of dependence of the so marked word or phrase on the following unit (word, phrase or clause), i.e. it can be seen as attributive marker. The inherited Tibeto-Burman causativizing prefix **s*- is found in a number of verbs as aspiration of the initial consonant of the base verb. In a few rare cases it is retained as orthographic <s->, now pronounced /θ/, as in the pair *ʔeiʔ* 'sleep' and *θeiʔ* 'put to bed'. These forms are regarded lexical today, though transparent in their formation, and new causative forms are built periphrastically.

Numeral classifiers are used obligatorily in Burmese whenever a noun is combined with a numeral, which in the case of *tiʔ* 'one', *hniʔ* 'two' and *bɛhniʔ* 'how many' is cliticized to the classifier as *tə=, hnə=* and *bɛhnə=* respectively. The word order is noun-number-classifier. No special classifier is used with measure words such as *né* 'day', *kilo* 'kilogram, *main* 'mile' etc..

The constituent order of Burmese is strictly SOV or OSV, with the post-verbal position available to anti-topics or afterthoughts, which are not strictly part of the clause and usually separated from it by an intonational pause. The order of preverbal elements (arguments and peripheral phrases) is determined by pragmatics, with the main focus position being the one immediately preceding the verb. Arguments are often omitted if present in the discourse context. Grammatical and semantic roles are marked by clitic postpositions, which, in the case of subject and object, also have pragmatic functions. Modifying elements can precede or follow the modified, but possessors, demonstratives, and relative clauses always precede the noun they modify. Subordinate clauses, marked with a clause final subordinator, precede the main clause. Only main clauses with verbal predicates are marked as finite by one of a small set of finite markers, expressing tense or status (future/irrealis, non-future/realis, changed state. negative). Interrogative elements usually occur in the immediately preverbal high-focus position, which omitted arguments, often becomes clause initial. For more complete descriptions of Burmese grammar, see e.g. Okell 1969 or Wheatley 2003.

2.3 Mon and Burmese compared

The following examples illustrate the similarities and differences between Mon and Burmese. The most obvious similarities are rooted in a long common cultural background shared by Mon and Burmese and can be seen in common semantics. Compare the following lexemes in Mon and Burmese with the closest corresponding words in Thai. It is not possible to tell with any certainty whether the common semantics reflect Burmese influence in Mon or the other way round, as the historical data are not conclusive. It is obvious, though, that Burmese and Mon are more close to each other in this respect than either of them is to Thai, which none the less shares a similar cultural background. Especially interesting from a historical-cultural perspective is the word for rice, the main staple food in Southeast Asian societies, for which Thai has only a single word *khâaw* (covering also other kinds of cereal crops). This generic term in Thai can be specified by a modifying element like ~ *săan* 'husked, uncooked rice', ~ *sŭəy* 'cooked rice', etc.

(1) Semantic similarities

Bumese	**Mon**	**Gloss**	**Thai**
lɛʔ	*tɔə*	‘arm’	*khɛ̌ɛn*
		‘hand’	*muɯ*
ʧhe	*càŋ*	‘leg’	*khǎa*
		‘foot’	*tiin/tháaw*
zəbà	*sɒʔ*	‘rice (plant/grain)’	*khâaw*
shan	*haoʔ*	‘uncooked rice’	
thəmìn	*pɤŋ*	‘cooked rice’	
θauʔ	*sɤŋ*	‘smoke’	*sùup*
		‘drink’	*kin*
		‘take medicine’	
sà	*ɕiəʔ*	‘eat’	
ɕí	*nùm*	‘have’	*mii*
		‘be at’	*yùu*
ne	*mɔ̀ŋ*	‘stay, remain’	

The structural differences are numerous and the following examples are given here to illustrate some of the main points, with the Mon sentences given in (a) and the corresponding Burmese expressions in (b).

(2) Possessive expression

a. *hɒəʔ* ***ʔuə***
house 1s
‘my house’

b. ***ʧɛənɔ́*** *ʔein*
1m.ATTR house
‘my house’

(3) Conditional clause

a. ***yɔ̀.raʔ*** *dɛ̀h* *ʔa* *ʔuə* *pèk* *noŋ.*
if 3 go 1s follow ASRT
‘If he goes, I’ll go along.’

b. *θu* *θwà* ***yin*** *ʧɛənɔ* *laiʔ* *mɛ.*
3 go if 1m follow FUT
‘If he goes, I’ll go along.’

(4) Adnominal interrogative

a. *pèh* *ket* *lòc* ***lɒ*** *rao.*
2 take text which Q
‘which book do you want (to take)?’

b. *mìn* ***bɛ*** *sa.ʔouʔ* *yu* *mə=lɛ̀.*
2 which book take FUT=Q
‘which book do you want (to take)?’

(5) Adverbial interrogative

a. *dɛ̀h* *ʔa* ***ʔəlɒ*** *rao.*
3 go where Q
‘Where is he going?’

b. *θu* ***bɛ*** *θwà* *mə=lɛ̀.*
3 where go FUT=Q
‘Where is he going?’

(6) Relative clause

a. *mənìh* **[***(mə=)hɒm* *ʔərè***]REL**
person (REL=)speak language
‘the person who speaks’

b. **[***səgà* *pyɔ̀* *tɛ́***]REL** *lu*
language speak NFUT.ATTR person
‘the person who is speaking’

(7) Adposition/demonstrative

a.	***dɔ̀ə***	***(kɒ)***	*hɒəʔ*	***tɤʔ***		b.	***ho***	*ʔein*	***thɛ̀***	***hma***
	LOC	OBL	house	DIST			DIST	house	inside	LOC
	‘in that house (over there)’						‘in that house’			

The differences illustrated in examples (2-7) are consistent with the overall typological structure of Mon as VO and Burmese as OV languages. The optional use of the oblique preposition *kɒ* in combination with the locative *dɔ̀ə* ‘in, at’ in (7a) may be seen as an attempt to replicate the complex postposition structure of Burmese *thɛ̀ hma* ‘at the inside’ seen in (7b).[23]

3. Burmese looking patterns in Mon

In spite of the profound typological differences between Mon and Burmese, there are a umber of structures in Mon that look strikingly similar to the corresponding Burmese patterns. Given the intimate contact in which the two languages stood for centuries, it seems reasonable to explain these constructions as the result of influence by one language over the other, that is as instances of structural borrowing or replication. As seen above (2.3), it is not always easy or possible to determine the direction of influence, especially where historical documents are missing or indecisive. In these cases we have to resort to either typological tendencies or to comparison with related languages of both sides involved, hoping to find hints as to the origin of a structure. While in the case of common semantics as seen above this does not lead to conclusive results, the undertaking might be more promising in the case of syntactic patterns.

Generally, intensive lexical replication is seen as prerequisite to structural replication. In the case of Mon and Burmese, it is difficult to establish lexical replication to a large extent for the Middle Mon period, where converging patterns become more common in inscriptions. Large numbers of Burmese lexemes are found in spoken Mon in Burma, much less in Mon varieties in Thailand. The latter separated from the bulk of Mon speakers at different periods, starting probably from the 16th century. The same is true for some of the structural similarities between Mon and Burmese, which are more prominent in Burma Mon than in Thailand Mon. As the main subject of this study is Burma Mon, reference to the Thailand varieties will be made only sporadically, as more research is needed in this respect.

3.1 Interrogatives - fronted and in situ

As seen in examples (4) and (5) above, interrogatives in Mon and Burmese mostly occur *in situ*, that is in the position that is occupied by the questioned constituent in the answer. Possible answers to the above questions would be (4′) and (5′).

(4′)	a.	*ʔuə*	*ket*	*lòc*	***nɔʔ***	*raʔ.*		b.	*ʨənɔ*	***di***	*sa.ʔouʔ*	*yu*	*mɛ.*
		1s	take	text	PROX	FOC			1m	PROX	book	take	FUT
		‘I (want to) take this book.’							‘I (will) take this book.’				

(5′)	a.	*dɛh*	*ʔa*	***phɛ̀ə***	*raʔ.*		b.	*θu*	***ʨàun***	*θwà*	*mɛ.*
		3	go	school	FOC			3	school	go	FUT
		‘He is going to school.’						‘He is going to school.’			

There is a tendency in Mon, though, to front interrogatives, both on the phrase and clause levels. This can be seen already in Old Mon, as illustrated in example (8) from a Pagán period inscription.

(8)	***mu***	*het*	*man*	*tirlaʔ*	*gruṅ*	*yo.*
	what	reason	REL	lord	laugh	Q
	‘Why did you laugh, lord?’					

[23] Shan, a Tai language in close contact with Burmese for centuries, also shows this tendency: *ti néɯ hɤ́n* ‘in the house’, lit. ‘at in house’, versus standard Thai *nay bâan* ‘in the house’ or *thîi bâan* ‘at home’ (cf. **thîi nay bâan*).

In this sentence, the adnominal interrogative *mu* 'what' occurs in prenominal position. This is still the case in Spoken Mon, where the interrogative *mùʔ* usually precedes the noun it modifies, unlike the interrogative *lʊ*, which always follows the noun (see (4) above). The answer to *mu het* 'what reason' could be *het woʔ* 'this reason', that is the position occupied by the interrogative is not the same as the demonstrative in the answer. The interrogative adverbial expression *mu het* '(for) what reason, why' is itself fronted in a cleft construction, overtly marked by the relativizer *man*. Another conceivable interpretation of the Old Mon sentence is as elliptical clause, that is the copula *das* after the subject *mu* has been dropped. In this case, a literal translation would be 'What (is) the reason that you laughed, lord?'. If this analysis is correct, we would have here an instance of reanalysis in the further development of Mon resulting in prenominal *mùʔ* also in contexts where an elliptical explanation is not possible, as in *mùʔ hwaʔ* 'what food', *mùʔ lòc* 'what book', etc. This pattern has been extended over other interrogatives, especially *ɲèh.kɔ̀h* 'who', as in *ɲèh.kɔ̀h kon* 'whose child', *ɲèh.kɔ̀h ka* 'whose car', etc., a type of expression not found in the Old Mon corpus. They find almost exact parallels in Burmese: *bəθú θà* 'whose child' and *bəθú kà* 'whose car'. The only difference is that Burmese marks the attributive function of the interrogative pronoun overtly by the high tone ("induced creaky tone" according to Okell and Allott 2001:273), while Mon does not have any morphological device to mark a possessor. It is interesting to note that the possessor marking is much less frequent in southern Burmese dialects, which are in close contact with Mon, than in standard Burmese.[24]

Fronting of interrogatives occurs in Spoken Mon also on clause level, similar to Old Mon. The following examples illustrate fronted interrogatives functioning as object and adverbial.

(9)	a.	*mùʔ*	*paʔ*	*(rao).*	(10)	a.	*chəlɔʔ*	*cao*	*(rao).*
		what	do	(Q)			when	return	(Q)
		'What are you doing?'					'When are you coming back?'		

Here again Burmese has parallel constructions, as seen in (9b) and (10b).

(9)	b.	*ba*	*louʔ*	*lɛ̀.*	(10)	b.	*bɛtɔ́*	*pyan*	*mə=lɛ̀.*
		what	do	Q			when	return	FUT=Q
		'What are you doing?'					'When are you coming back?'		

The parallelism is not perfect, though, as a closer look reveals. If there is an overt subject in sentences (9) and (10), it is placed after the interrogative in Mon, but before it in Burmese, with the inverse order being ungrammatical in both languages:

(9′)	a.	*mùʔ*	*pèh*	*paʔ.*	(10′)	a.	*chəlɔʔ pèh*		*cao.*
		what	2	do			when	2	return
(9′)	b.	*mìn*	*ba*	*louʔ*	*lɛ̀.*	(10′)	b.	*mìn*	*bɛtɔ́*	*pyan*	*mə=lɛ̀.*
		2	what	do	Q			2	when	return	FUT=Q

In sentence (9′), the interrogative pronoun can be repeated, either as reduplication of the fronted interrogative or *in situ*, giving *mùʔ mùʔ pèh paʔ rao* or *mùʔ pèh paʔ mùʔ rao*. In the latter case, the sentence final question particle is usually not omitted. This repetition of the interrogative element is common only with the monosyllabic *mù*, though some other forms may be repeated as well. Some speakers of Mon add the otherwise all but obsolete relativizer in immediately before the verb, marking the clause as cleft as in Old Mon.

(9′′)	*mùʔ*	*mùʔ*	*pèh*	*mə=paʔ.*	(10′′)	*chəlɔʔ pèh*		*mə=cao.*	
	what	what	2	REL=do		when	2	REL=return	

With the gradual loss of the relativizer, the cleft construction can be reanalysed as fronted interrogative. See section 3.3 for details of the loss of the relativizer OM *mun/man/ma* > Literary Mon *mɛ̀ʔ*, Spoken Mon *mə=*.

[24] While the process is productive in standard Burmese, is seems to be resticted to a few conventionalized forms, especially pronouns and some kinship terms, in southern Burmese varieties spoken in Mon and Karen States.

In Mon, *ɲèh.kɔ̀h* ‘who’ occurs in phrase initial position as adnominal interrogative and in clause initial position as subject, never as object, while *mùʔ* ‘what’ occurs as a prenominal interrogative in phrase initial position or as subject or object in clause initial position. This is markedly different from the situation in Burmese, where all interrogatives occur in prenominal or preverbal position, the grammatical relations being marked by postpositional markers where deemed necessary. Compare the expressions in examples (11) and (12).

(11) a. *chan ɲèh.kɔ̀h.*
love who
‘Who do (you) love?’

(11) b. *bəθú ko ʨhiʔ lɛ̀.*
who.ATTR OBJ love Q
‘Who do (you) love?’

(12) a. *ɲèh.kɔ̀h chan.*
who love
‘Who loves (you)?’

(12) b. *bəθu (ká) ʨhiʔ lɛ̀.*
who (SBJ) love Q
‘Who loves (you)?’

While the similarities between the Mon and Burmese are striking at a first glance, they are rather superficial when the details are taken into account. In Mon, there is interrogative fronting arising from original cleft sentences, while in Burmese interrogative elements occur in the high-focus preverbal position. The question arises therefore whether the fronting of interrogatives is due to Burmese influence, or rather an independent development in Mon. Unfortunately, comparative work on Mon-Khmer syntax is still a big desideratum in areal studies of Southeast Asia. We therefore have to content ourselves with a rather random look at some more or less closely related languages within the Mon-Khmer family for which adequate descriptive material is available. Similar constructions are indeed found in a number of Mon-Khmer languages, such as Chrau (Thomas 1971:195ff):

(12) a. ***mŏq*** *ănh ôp.*
what 1s do
‘What can I do?’

b. ***păch.n’hya*** *mai ôp.*
what 2 do
‘What are you doing?’

c. ***păch*** *daq*
what water
‘What water.’

These constructions appear parallel to the Mon constructions seen above, both on the clause and phrase levels. Also the Aslian languages spoken in Malaysia exhibit similar constructions, such as Jahai (Burenhult 2005:89f):

(13) a. ***makɛn*** *hajɛ̃ʔ*
who house
‘whose house’

b. ***mamɛj*** *paj dʔ-deʔ.*
what 2s.DIS IMPF-do
‘What are you doing?’

c. ***mamɛj*** *mɔh ɟim*
what 2s.FAM cry
‘Why do you cry?’

Semelai interrogative expressions more closely resemble Old Mon cleft constructions, sometimes containing an overt relativizer. The following examples are taken from Kruspe (2004:330).

(14) a. ***hmɔh mə****=ma=lən.*
what REL=IRR=want
‘What (is it) that (you) want?’

b. ***kadeh mə****=ga=yɛ=jon.*
who REL=IMM=1A=give
‘To whom (is it) that I am going to give (it)?’

c. ***dɔ kadeh****,nʔ-gɔʔ naʔ-hɛʔ.*
OF who NML-fell.tree DEM-LOC:above
‘Whose (is that) tree-felling up there?’

d. ***tɔm haʔ hɔ̃n****, ns-dɔs ji nɛŋ.*
SRC LOC where NML-arrive 2 before
‘From where (was) your arrival before?’

The possibility to front interrogatives and other focal elements in cleft sentences is not restricted to Mon-Khmer languages, of course, though it is rare in other Southeast Asian languages such as Thai. The Burmese influence seems to be an indirect one, reinforcing a pre-existing minor pattern and triggering or facilitating subsequent reanalysis, at least in the case of prenominal interrogatives. There is thus a convergence on the surface of Mon towards Burmese, but with very different underlying structures. The

parallelism does not include all interrogatives, and it is not extended to non-interrogative nominal modifiers such as possessors or demonstratives.

3.2 Clause final subordinators

Subordinators in Old Mon regularly occur in clause initial position, as is expected in a VO language with subordinate clauses after the matrix clause. Fronting is possible for pragmatic reasons in the case of adverbial and complement clauses. In this and the following section we will look at the development of conditional and complement clauses, which seem to have switched to pre-matrix position with clause final subordinators, at least superficially.

3.2.1 Conditional clauses

In Old Mon, conditional clauses are introduced by the subordinator *yal* 'when, if', which in Spoken Mon becomes *yɔ̀* or *yɔ̀.raʔ*, the latter form being a lexicalized phrase Middle Mon *yar dah* 'if (it) is (the case that)'. Conditional clauses normally occur before the matrix clause, consistent with the iconic structure of Mon syntax. The following Old Mon example is from the 12th century Kubyaukgyi inscription at Pagán (Luce, G. H. and Bohmu Ba Shin 1961).

(15) ***yal*** *kcit* *sak* *ñaḥ* *ma* *yām.*
if die not.exist person REL weep
'If they die, there is no one to weep for them.'

In Middle Mon, conditional clauses are introduced by *yar/yal/yaw/yow* and frequently end in *ma gaḥ* 'which is said, speaking of which', or *mā*, as in (16) and (17). As both forms can function as topic markers, their occurrence in conditional clauses is consistent with the close functional relationship between conditional clauses and topics (cf. Haiman 1978). The first marker, *ma gaḥ* has close parallels in Burmese *sho ta* 'the said, saying', *sho yin* 'if (you) say' and *sho* 'say'. The second is given in Shorto 1971 (291) as "particle marking introduction of new subj. of discourse" and is connected to the Burmese topic marker *hma* 'as regards', with the direction of borrowing not entirely clear.[25]

(16) ***yar*** *tḍek* ***ma*** ***gaḥ*** ...
if wet REL say
'if it is wetted ...'

(17) *pḍay* *paṭisandhi* *ey* *kəp.tlā gah* *kwon* *truh* *mway* *ey* *goʔ* ***mā*** ...
LOC confinement 1s first that child male one 1s get if
'If in my first confinement I get a son, ...'

In Literary and Spoken Mon, the original conditional subordinator can be omitted, with the clause final topic marker being enough to get the conditional reading, as the development illustrated in (18) shows. The two varieties differ only in the choice of the clause final topic marker, *məkɛ̀h* in Literary Mon and *teh* in Spoken Mon. The origin of the latter, which does not appear in texts until very recently, is unknown. It's phonetic shape suggests an indigenous origin, though it is not attested in any period of the language before the modern Spoken Mon. A connection with the Old Burmese emphatic particle *teh* would make sense functionally, but the latter form is found only in very formal styles of Modern Burmese as *dì*. It is not readily conceivable how the literary Burmese form could have entered colloquial Mon, though this is not *a priori* impossible, of course, as we do not have any colloquial documents of Mon and Burmese of pre-modern periods. This particle in Mon also occurs as topic marker on the phrase level, as in *hɒəʔ ʔuə teh* 'as for my house'.

[25] Based on the presence of the form in the Old Mon but not Old Burmese data, Bauer (2006:42) suggests a Mon origin of this word, which was borrowed into Middle Burmese. Given the limited corpus we have of Old Mon and Old Burmese, non-occurrence of a form in in the corpus cannot necessarily be taken as proof of non-existence in the language.

(18) ***yɔ̀.raʔ*** *dɛ̀h* *hùʔ* *klɤŋ* ...
if 3 NEG come
>
yɔ̀.raʔ *dɛ̀h* *hùʔ* *klɤŋ* ***məkɛ̀h/teh*** ...
if 3 NEG come TOP
>
dɛ̀h *hùʔ* *klɤŋ* ***məkɛ̀h/teh*** ...
3 NEG come (TOP >) COND

'if he isn't coming ...'

Similar constructions with topic markers or topic-comment linkers (TCL) between protasis and apodosis exists in other Southeast Asian languages, as illustrated in the following examples.

(19) Thai
(thâa) *fŏn* *tòk* *raw* ***kɔ̂*** *khâw* *pay* *khâaŋ* *nay.*
if rain fall 1p TCL enter go side in
'if it rains, we will go inside.'

(20) Khmer (Jacob 1968:100)
lòːk *ʔaoy* *khŋom* *mərɔ̀ːy* *rìəl* ***kɔː*** *khŋom* *mìːən* *prak* *doː(r)* *cùːn* *lòːk.*
2 give 1 hundred riel TCL 1 have money change offer 2
'If you give me a hundred riels, I have change for you.'

(21) Nyahkur (Payau 1979:251)
chur *dɛː* *heːw* *poːŋ* *wəj* ***san*** *kul* *poːŋ* *kul* *dɛː* *caːʔ.*
dog it hungryrice I thus give rice give it eat
'The dog is hungry, thus I give it rice.' or 'If the dog is hungry, I give it rice.'

Note that in Thai and Nyahkur the linker *kɔ̂/san* occurs after the subject of the apodosis. Alternative constructions are usually available with a clause initial conditional marker in the protasis, like *thâa* in Thai, *baə* in Khmer, and *khan* in Nyahkur, the latter being a loan from Thai/Lao *khrán* 'when'. The relation between the two clauses is not overtly marked, the conditional reading being purely inferential, based on the semantics of the two clauses. If there is no conditional match, the same construction yields concessive or sequential readings. The topic-comment linker is therefore clearly not a conditional marker, nor a subordinator of some kind, but rather an indicator of a logical link between two parts of an utterance. In Mon on the other hand, the topicalizer *teh* is restricted to conditional contexts.[26] Also after phrasal elements, *teh* seems to be mostly used as a specialized (contrastive) topic marker with a conditional connotation along the lines of 'if it comes to X'. There is thus a marked difference between Mon and other Southeast Asian languages, that could be due to Burmese influence. Conditional clauses in Burmese are normally placed before the apodosis and end in *hlyin* or *yin*, the former being the literary, the latter the colloquial form. Apart from being a conditional marker, *hlyin/yin* is in Literary Burmese also used as distributive and emphatic marker. In the latter function this marker has been borrowed into Middle Mon as *heṅ*, Literary Mon *hiəŋ* (Shorto 1971:403). Examples (22) - (23) illustrate the different uses of Burmese *hlyin/yin* (from Okell and Allott 2001:228f).

(22) *mə=ʔaun* ***yin*** *thaʔ* *phye* *yá* *ʔòun* *hma* *pɔ́.*
NEG=succeed COND pile.up answer get again FUT.NML INSIST
'If you don't pass (the exam), you will have to take it again of course.'

[26] The particle *teh* is listed in Sakomoto's Mon-Japanese dictionary of Thailand Mon under the form *tih* with various functions, some paralleling Thai *kɔ̂* rather than Burmese *hlyin/yin* 'if' (Sakomoto 1994:361f).

(23) *luʔ.laʔ.yè* *ʔəthèin.ʔəhmaʔ* *ko* *tə=hniʔ* ***hlyin*** *tə=tɕein* *tɕìn.pá* *θi.*
freedom memorial OBJ one=year DISTR one=time celebrate NFUT
'Independence celebrations are held once a year.'

These Burmese structures can be expressed in Mon in exactly parallel constructions using the (topic turned) conditional marker *teh*:

(22′) *hùʔ* *ʔɔŋ* ***teh*** *tèh* *thɔp* *sɔh* *noŋ.*
NEG succeed COND TOUCH PILE.UP answer ASRT

(23′) *mùə* *hnam* ***teh*** *kɤ̀ʔ* *paʔ* *puə* *thaŋ.saʔ* *lòt.làt.yey* *mùə* *wèə* *raʔ.*
one year COND get do festival celebrate independence one time FOC

The Mon constructions as closely as possible imitate the corresponding Burmese expressions using matter and patterns already available in Mon. As in the case of fronted interrogatives, Burmese influence seems to have enhanced pre-existing possibilities in Mon, resulting in similar surface structures, without replicating Burmese constructions exactly or borrowing phonetic/lexical material from Burmese. The resulting parallelism between Mon and Burmese is closer than with Thai and other Southeast Asian languages, a state of affairs that has already been observed above in section 2.3 in the case of shared semantics. The parallelism between Mon and Burmese allows also idiomatic expressions to be calqued directly, like the frequent Burmese *da sho (yin)* 'in this case', literally '(if) you say this', which is rendered in Mon as *hɒm kɔ̀h teh* with the same literal meaning.

3.2.2 Complement clauses

Complement clauses function as object arguments of verbs of speaking ('say, ask, answer') or mental activities ('think, dream, know'), including verbs of perception like seeing and hearing, or as subject arguments of sentences.[27] A common pattern in Southeast Asian languages is to link the complement clause to a preceding matrix clause by a grammaticalized verb meaning 'to say'. This is found in Mon-Khmer languages such as Chrau *panh* (Thomas 1971:73ff), Khmer *tha:* (Jacob 1968:115ff) as well as Tai and other language families. The following examples illustrate the use of the complementizer in Khmer (24) and Thai (25).

(24) *khŋom* *mùn* *ba:n* *dɤŋ* ***tha:*** *kɔ̀ət* *kɔmpùŋ.tae* *səse:(r)* *sɔmbɔt(r).*
1s NEG get know SAY 3 hum CONT write letter
'I didn't know he was in the middle of writing a letter.' (Jacob 1968:115)

(25) *mêɛ* *mây* *rúu* ***wâa*** *lûuk* *pay* *rooŋ.riən* *léɛw.*
mother NEG know say son go school NSIT
'The mother didn't know that her son had already gone to school.'

A similar construction is found in Literary Mon, using the obsolete verb *kɛ̀h* 'say' (from Old Mon/Middle Mon <gaḥ>, cf. 3.2.1 above) in complement clause initial position, as seen in (26), taken from a contemporary publication (*The legend of Kyaikhtiyo* by Pālita Thera).

(26) *thiəŋ* ***kɛ̀h*** *dɛh* *kəliəŋ* *cao* *ʔa* *phɤh* *raʔ.*
think say 3 return return go still FOC
'I though that he would still come back home.'

Other languages prefer paratactic constructions, with the complement clause preceding or following the matrix clause, as in example (27) from Nyahkur (Payau 1979:266) and (28) from Katu (Costello and Sulavan 1993:373).

[27] There are other functions of complement clauses, for example as complements of nominal expressions, but these are beyond the present discussion.

(27) *ɲin kasɔː kul ʔaɲah ɲin bɔk cɔw doːŋ.*
he tell to who he want return home
'Whom did he tell that he wanted to return home?'

(28) *móón mabô đââng ravaai.*
speak priestess bring soul
'They say the priestess brings back the person's soul.'

This construction is also available in Spoken Mon, usually with the complement clause preceding the matrix clause, as in (29).

(29) *ɗɛh hùʔ klɤŋ raʔ, ɗɛh hɒm.*
3 NEG come FOC 3 speak
'He said he wasn't coming any more.'

Burmese prefers nominalized complement clauses, though a clause final subordinator *ló* (*hú* in Literary Burmese) can also be used in some cases, depending on the semantics of the matrix verb. Example (30) illustrates the use of a nominalized complement clause. In colloquial Burmese, paratactic constructions are common, with the complement clause usually preceding the matrix, as seen in (31). Nominalization of a complete clause in Burmese is achieved by changing the sentence final predicate markers *tɛ* and *mɛ* into the nominal forms *ta* and *hma* (from *té ha* nad *mé ha*, originally attributive forms with the semantically empty nominal head *ha* 'thing'), or by adding *sho ta* (lit. 'the saying') to a finite clause. Verb phrases are nominalized by adding one of a number of postposed nominal heads such as *ʨhìn, hmú, yè*. Optionally the specific, topical (referential) object marker *ko* can be added to these forms if the complement clause functions as object of the matrix verb. Literary Burmese adds the object marker *ko* to a finite object complement clause with no overt nominalizer, as seen in (32).

(30) *θu la* ***ta*** *ʨənɔ θí tɛ.*
3 come NFUT.NML 1m know NFUT
'I know hat he came.'

(31) *mìn bɛtɔ́ la mə=lɛ̀ θu mè tɛ.*
2 when come FUT=Q 3 ask NFUT
'He asked when you are coming.'

(32) *θu la θi* ***ko*** *ʨənouʔ θí θi.*
3 come NFUT OBJ 1 know NFUT
'I know that he came.'

Apart from the complement clause structures seen above, there is in Mon a frequent pattern used in cases where the complement clause is presupposed. In this case the complement precedes the matrix and is marked as non-predicative, topical or referential by the medial demonstrative *kɔ̀h* (see Jenny 2009 for a detailed discussion). A relevant example taken from a Mon novel (*The emaciated bitch* by Cantimācāra) is given in (33). Sentence (34) immediately precedes the sentence given above in (26).

(33) *ɲèh hùʔ kɒ məkɛ̀h hùʔ kɤ̀ʔ ɕiəʔ* ***kɔ̀h,*** *ʔuə tɛm mɔ̀ŋ raʔ.*
person NEG give COND NEG get eat TOP 1s know stay FOC
'I know that I cannot eat anything if they don't give me [food].'

(34) *ʔəlɒ ɗɛh ʔa* ***kɔ̀h*** *ʔuə hùʔ tɛm, ʔuə hùʔ paʔ cɒt.*
where 3 go TOP 1s NEG know 1s NEG do heart
'I don't know where he went, I didn't care.'

Unlike Burmese CLAUSE *ko*, Mon CLAUSE *kòh* is also used to mark subject complement clauses, as in (35a). In this case Burmese uses the nominalized form, with or without a subject marker, as seen in sentence (35b).

(35) a. *tɛm ket yɔ̀ə **kɔ̀h** tɔ̀h wì kɤ̀ʔ kəwɛk.*
know take sick TOP be cure get half

b. *yɔ̀ga ko θí-**tɕhìn** **θi** tə=wɛʔ kú.θá-**tɕhìn** phyiʔ θi.*
illness OBJ know-NML SBJ one=half cure-NML be NFUT
'To know the disease is half the cure.' (Ketumati 1965:193)

Mon *kɔ̀h* and Burmese *ko*, though clearly different in meaning and origin, do nevertheless show some overlap in function. Both are used to mark topical elements, in Burmese most frequently (but not exclusively) object NPs. The phonetic similarity has probably enhanced the use of Mon *kɔ̀h* in some of the functions of Burmese *ko*, a case of "lookalike" as described in Aikhenvald and Dixon (2006:33). Functionally, Mon *kɔ̀h* corresponds more closely to nominalization in Burmese in that it renders an expression non-predicative. This is not to say that *kɔ̀h* is a nominalizer (*pace* Bauer 1982:331), though, but rather that it covers some functions of a nominalizing device.

3.3 Loss of relativizer

Old Mon possessed a relativizer *mun, min, man, ma* which was formally and functionally distinct from the morphological attributive form of verbs. The relativized NP can have any grammatical function in the relative clause introduced with the relativizer, including peripheral. The following examples from the Old Mon inscriptional corpus illustrate the different usages (taken from Shorto 1971:281f, 297). In (36) the relativized function is that of subject, *ʔan* being used as a transitive verb 'have few'. In (37), the relativized function is that of object, with the subject in the relative clause omitted. In (38) again the realivized function is that of object, this time with an overt subject in the relative clause. In the first part of sentence (39), *ma* introduces a relativized subject, in the second part *mun* introduces a relativized object. In (40), the relativized function introduced by *min* is oblique ('the manner in which ...'). In all these cases the relativizer occurs between the head noun and the relative clause.

(36) *ñaḥ **ma** ʔan ḍīk*
person REL few slave
'those who have few slaves'

(37) *ḍek **mun** jun ta kyāk*
servant REL make.over BEN sacred
'the servants which he made over to the shrine'

(38) *pun dān **ma** smiṅ pa*
merit donation REL king do
'the acts of merit and charity which the king performed'

(39) *ḍey sthān **ma** kaḥ sak kṣīw **mun** tarley gawaṁpatither croḥ.ptāw ci*
LOC place REL NEG NEG shake REL lord.1s Thera.Gavampati maintain EMPH
'in a place which cannot be shaken, which the lord Thera Gavampati maintains'

(40) *row **min** kyek buddha tirley byādes goḥ*
manner REL sacred Buddha lord.1s foretell that
'as the Lord Buddha had foretold'

While the normal structure in Old Mon is HEAD - REL - CLAUSE, as seen in the preceding examples, the relativizer is "rarely following [the] subject of [the] relative clause" (Shorto 1971:281), as in sentence (41).

(41) *smiṅ daddharāja dewatāw guṁloṅ **ma** siw*
king Daddharaja god ATTR.many REL attend
'King Daddharaja whom the gods attend'

The semantic (or pragmatic?) difference between the two construction types is not clear. There is no discernible difference between sentences (38) and (41), though they use different patterns. With the beginning of Middle Mon, the less frequent pattern with the relativizer occurring after the subject within the relative clause, becomes more common. In Shorto's (1971:282) words: "When antecedent noun denotes goal or locus of action, *ma* usually follows subject of relative clause." The changing pattern is evident in pleonastic constructions with the relativizer occuring both in clause initial position in its full form, and after the subject of the relative clause in its weak form, as seen in (42) from the 15th century Shwedagon inscription.

(42) *cetī dhāt swok kyāk tray* ***min*** *tapussa bhallika* ***ma*** *thāpanā lar*
stupa relic hair sacred holy REL Tapussa Bhallika REL enshrine deposit
'the stupa of the hair relics of the Buddha which tapussa and Bhallika built'

More frequent is the occurrence of the relativizer only in the preverbal position of the relative clause, as in (43) and (44). In this case, it is always the short form *ma* that is used, presumably pronounced as proclitic *mə* already in Middle Mon, and encroaching on the functions of the attributive marker *ma*.

(43) *galān dewatau* ***ma*** *həm*[28]
word god REL speak
'the words that the gods spoke'

(44) *piṇḍapāt ñaḥ* ***ma*** *kəw dān*
almsfood person REL give donation
'the almsfood which the people had given'

The regular pattern of Old Mon is still used, but appears to be receding. In sentence (45) the realtivized function is that of object, while in (46) it is oblique (locative).

(45) *dhar* ***ma*** *ey goʔ graṅ ket woʔ*
doctrine REL 1s get understand take this
'the doctrine which I came to understand'

(46) *pḍay kalyāṇīgaṅ* ***ma*** *kyāk tray laḅoh jaku*
LOC Kalyani.river REL sacred holy sprinkle body
'on the Kalyani river, where the Buddha bathed'

In Literary Mon, the situation remains largely unchanged from Middle Mon. While usually the relativizer is adjacent to the verb of the relative clause, it can occur in clause initial position. The pronunciation is regularly *mə*. This phonetic weakening of the marker has gone even further in Spoken Mon, where in most cases it disappears altogether. For some speakers there is still a felt trace of *mə-* or *m-*, phonetically prefixed to the verb, in some contexts, though this is barely audible at best. There is no clause initial relativizer in Spoken Mon. Relevant examples are given in (47) and (48).

(47) *mùʔ dɛh* ***mə=****hɒm.* *(mᵊ=, m=)* **mùʔ mə dɛh hɒm.*
what 3 REL=speak
'What did he say?'

48) *hənày pèh* ***mə=****ʔa (mᵊ=, m=)* **hənày mə pèh ʔa.*
place 2 REL=go
'the place you are going'

More common are constructions without any overt relative marking, usually with the topic marker *kɔ̀h* added to the expression to mark it as non-predicative and referential (see 3.2.2 above). The following examples illustrate the relative clauses in colloquial Spoken Mon.

[28] I use the symbol *ə* to transliterate the graphic complex <ui~iu> in Mon and Burmese. For argumentation on this issue see Dempsey 2001:210-211.

(49) *ʔərè* *pèh* *hɒm* *(kɔ̀h)*
language 2 speak (TOP)
'the things you said'

(50) *kɤ̀ʔ* *tɛ̀h* *ɓɛ̀ʔ* *kon* *ŋèə* *həkaoʔklày* *kɔ̀h.*
get tough REF child frog body seek TOP
'He got the little frog he was looking for.'

The development from Old Mon to Spoken Mon can be summarized as follows (italics = minor use patterns):

Old Mon	HEAD - **REL** - $\text{CLAUSE}_{\text{REL}}$ ~ *HEAD - [S - **REL** - V]$_{REL}$*
Middle Mon	HEAD - [S - **REL** - V]$_{\text{REL}}$ ~ *HEAD - **REL** - CLAUSE_{REL}*
Literary Mon	HEAD - [S - **REL** - V]$_{\text{REL}}$ ~ *(HEAD - **REL** - CLAUSE_{REL})*
Spoken Mon	HEAD - $\text{CLAUSE}_{\text{REL}}$ (*kɔ̀h*) ~ *(HEAD - [S - **REL** - V]$_{REL}$)*

There is a gradual move from clause initial to preverbal position from Old Mon via Middle Mon to Literary Mon, and a subsequent marginalization and finally drop of the relative marker in Spoken Mon. In a related process, the attributive form, originally distinct from the relative, has become non-productive in Middle and Literary Mon. In Old Mon, there is some overlap in use of attributive and relative forms, such as ***ma*** *gloṅ* - *g**um**loṅ* '(which are) many', ***ma*** *yās* - *y**im**ās* '(which is) shining', and ***ma*** *nom* - *l**m**om* '(who is) having'. Attributive forms are restricted to subject function, that is *k**am**lon* is the person 'who works' (*klon*), never the work 'done', similarly *l**m**om* is the person 'who has', not the thing 'had'. Morphological attributive forms became lexicalized, while new attributives were built periphrastically. The word *kəmɒəh* which is part of *pərɔ̀c kəmɒəh* 'pepper' (lit. 'good chilli', compare Burmese *ŋəyouʔ kàun*), for example, is derived from Old Mon *khis* 'good', Spoken Mon *khɒh*. The original attributive form, spelt <khmih> in modern Mon, is used only in this combination, the present attributive being unmarked or the periphrastic *mə=khah*. With the loss of the productivity of the attributive infix *-m-*, the relative marker *ma* took over its function and was attached to the verb as proclitic *mə=*. This lead to a formal merger of the attributivizer and the relativizer, which at some point obviously became more attributivizer-like and inseparable from the verb, while it still covered the whole range of functions of the clause initial relativizer. From the data it is evident that this process had already begun to some degree in Old Mon, but was accelerated in Middle Mon. That the relativizer as such was lost early in the spoken language is further evidenced by the corresponding constructions in Nyahkur, the closest relative of Mon. Nyahkur makes use of two relativizers, both loans from Thai, namely *ʔan* and *thîː* (Payau 1979:154ff), as seen in examples (51) and (52).

(51) *phanih**(ʔan)*** *juːn* *ʔəl* *tɛʔ* *thɛ* *muː* *wəj.*
person (REL) stand keep that be friend I
'the person standing there is my friend.'

(52) *hmiəw **thiː*** *pheh* *kul* *wəj* *kacɛt* *ʔəːj.*
cat REL you give I die already
'The cat which you gave me has already died.'

Interestingly, Nyahkur also allows zero marked relative clauses with relativized functions other than subject. This is clearly different from Thai, which allows only relativized subjects to occur in unmarked relative clauses. Example (53) illustrates this usage, with a gap between the head noun and the relative clause.

(53) *hmiəw* *baːr* *tuh* ***Ø*** *pheh* *khamaj* *khɔːŋ* *wəj.*
cat two CL (REL) you see of I
'The two cats (which) you see are mine.'

This construction is similar to Mon, as seen in (53´) and different from Thai, as seen in (53´´).

(53′) *həkɔə ɓa **Ø** pèh chɤ kɔ̀h hmɛk ʔuə raʔ.*
cat two (REL) 2 see TOP POSS 1s FOC

(53′′) *mɛɛw sɔ̌ɔŋ tuə ***(thîi)** khun hěn pen khɔ̌ɔŋ phǒm.*
cat two CL *(REL) 2 see be POSS 1m

While Nyahkur clearly borrowed its overt relative markers from Thai (or Lao), it is obvious that the relativization strategy with a gap cannot be due to Thai influence and is likely to be inherited from Old Mon. There has been no proven contact between Nyahkur and Mon since the break up of Dvāravatī around the 10th or 11th century, which suggests that the loss of the relative marker or at least the beginning of this process dates back to that early time.

The position change and subsequent merger with the attributive marker of the relativizer in Mon temporally coincides with the period of increasing Burmese influence in Mon, especially in terms of pattern replication. It is therefore sensible to ask whether Burmese influence played a part in the development in Mon.

In Burmese, relative clauses always precede their heads and are marked with the high tone ("induced creaky tone") on the finite marker *tɛ* 'non-future/realis' or *mɛ* 'future/irrealis'. The induced creaky tone is used to indicate some kind of dependence of the element following it to the element thus marked (see 3.1 above). In the case of verbal expressions this can be interpreted as attributive or relative. In other words, Burmese does not make a distinction between attributive and relative expressions. As the verb is in clause final position, followed only by the finite marker in its attributive form, the visible marker of the relative clause is always adjacent to the verb (as well as to the head noun). This is similar to the attributive form in Old Mon with the infix *-um-* on the verb or the alternative proclitic *ma=*. The following sentences illustrate the structure of relative clauses in (colloquial) Burmese.

(54) *la **té** lu twe*
come NFUT.ATTR person PL
'the people who came'

(55) *tɛəanɔ θú ko pè **té** sa.ʔouʔ*
1m 3.ATTR OBJ give NFUT.ATTR book
'the book which I gave him'

(56) *θu ɛa ne **té** phà-khəlè ko pyan twé tɛ.*
3 seek stay NFUT.ATTR frog-DIM OBJ return find NFUT
'He found the little frog he had been looking for.'

It seems plausible that convergence with the Burmese structure led to a reanalysis of the attributive marker in Mon, so that it came to take over the functions of the relativizer, in other words relative clauses came to be marked by the periphrastic attributive form of the verb. The Burmese relative construction [S - V+ATTR]$_{REL}$ - HEAD was replicated in Mon as HEAD - [S - ATTR+V]$_{REL}$. This replication did not come out of nothing, but was rather a strengthening or enhancement of a pre-existing minor use pattern in Old Mon (cf. Heine and Kuteva 2005:44ff) with subsequent extension to new syntactic contexts. Later the weakened attributive marker further lost phonetic material (*mə=* > *mᵊ=* > *m=*) in close phonetic juncture with the following verb and was finally dropped, leaving Spoken Mon with no overt relativizer. While in Middle Mon the proclitic negation marker *ha* intervenes between the relative/attributive clitic as seen in (57), this is not possible in Spoken Mon, showing the passage from clitic to affix.[29] In these contexts the relative/attributive marker is regularly omitted.

[29] It should be noted that the negation marker in Spoken Mon, going back to a strengthened form *hùʔ* from Middle Mon *ha*, is itself in close juncture with the verb and in some cases realized as affix (pre-aspiation and labialization of the initial consonant, see Jenny 2003). The presence of *hùʔ* precludes the occurrence of the relative marker in Spoken Mon.

(57) *pḍay thān ma dah cnih kəm lar, thān* ***ma*** ***ha*** *seṅ* ...
LOC place REL be pier also ADD place REL NEG be.so ...
'whether in a place which is a bathing place or one which is not ...'

The development from a free word *ma* to a phonetically reduced clitic to a prefix and finally loss is an example of secondary grammaticalizatzion (Norde 2009:), that can be observed in the development of Mon. The corresponding Burmese form is a morphologically marked clitic (induced creaky tone) always adjacent to a verb (or the final element in a verb phrase), thus having some affix-like properties (choice of host).

	free marker			**clitic**		**(? affix)**		**loss**
Mon	*ma*	+ CLAUSE	>	*ma=/mə=*	>	*mə*-V/*mˀ*-V	>	*Ø*
Burmese				V=*tɛ́/=mɛ́*				

This development is not only due to convergence with Burmese structure, but also to the fact that the productivity and functionality of the original attributive forms was drastically reduced since Middle Mon, often leading to lexicalized forms with often idiomatic meanings. The new periphrastic construction was not used obligatorily, so that there was no systemic gap in Mon with the loss of the relativizer. The discourse topic marker *kɔ̀h*, which marks non-predicative expressions, became frequently used after relative expressions, partly taking over the function of the old relativizer. This function is not fully grammaticalized, though, and rather implicational. Nyahkur probably initially saw a similar development with the loss of the inherited relativizer, which under heavy Thai/Lao influence led to a felt gap in the system, which was filled by borrowed markers.

4. Conclusions

Besides a seemingly ever increasing number of Burmese lexemes in Mon, there are a number superficially similar constructions in Mon and Burmese. In many cases, a more detailed study reveals important differences in the syntax of these constructions, though. This fact suggests that the Burmese influence was rather indirect, often as strengthening of preexisting usage patterns in Mon itself. Due to areal convergence towards Burmese, for a few centuries the politically and often economically dominant language, Mon has undergone some degree of remodeling and come to be typologically closer to Burmese in some respects. This restructuring of Mon syntax has stopped at a rather superficial level of syntax and did not affect the structure of Mon as such, i.e. there is no sign of metatypy, as can for example be seen in many instances in Nyahkur, which has become very Thai-like in its syntactic structure.

As been pointed out in this paper, much more work has to be done, especially in terms of comparative Austroasiatic (and other language families) morphosyntax in order to gain better insights into the structure of these languages, which, together with a thorough study of the development of Mon as documented in the inscriptions and other texts, should put us in a better position to single out instances of contact induced changes and achieve broader understanding of the Southeast Asian linguistic area.

The examples given in this paper are not meant to be exhaustive in any way, but are rather seen as starting point for further investigation in the field of areal convergence in Southeast Asia, including potential instances of Mon influence in Burmese, as well as mutual influence among other languages in the area.

References

Aikhenvald, A. Y. and R. M. W. Dixon. 2006. *Grammars in contact. A cross-linguistic typology.* Oxford: OUP.

Aung Thwin, M. 2005. *The mists of Rāmaññа. The legend that was lower Burma.* Honolulu: University of Hawai'i Press.

Bauer, C. 1982. *Morphology and syntax of Spoken Mon.* [unpublished PhD thesis]. London: SOAS.

Bauer, C. 2006. Reflections on early Mon-Burmese grammar. In Beckwith, C. I. *Medieval Tibeto-Burman languages 2*. Leiden/Boston: Brill, 39-43.
Burenhult, N. 2005. *A grammar of Jahai.* Canberra: Pacific Linguistics.
Coedès, G. 1929. *Recueil des inscriptions du Siam.* Bangkok: Bangkok Times Press.
Costello, N. and K. Sulavan. 1993. *Katu folktales and society.* (Katu-Lao-English). Vientiane: Ministry of Informatuion and Culture.
Dempsey, J. 2001. Remarks on the vowel system of Old Burmese. *LTBA* 24.2, 205-234.
Duroiselle, C., Taw Sein Ko and C. O. Blagden (eds.). 1919-1928. *Epigraphia Birmanica* Vol. I-III. Rangoon: Government Printing.
Haiman, J. 1978. *Conditionals* are topics. *Language* 54, 564-589.
Heine, B. and T. Kuteva. 2005. *Language contact and grammatical change.* Cambridge: CUP.
Jacob, J. 1968. *Introduction to Cambodian.* London: Oxford UP.
Jenny, M. 2003. New Infixes in Spoken Mon. In *Mon-Khmer Studies* 33, 183-194.
Jenny, M. 2005. *The verb system of Mon.* Zurich: ASAS.
Jenny, M. 2009. Deixis and information structure in Mon. The multifunctional particle *kòh*. In *JSEALS* 2.
Ketumati, Bh. 1965. *4 languages: English, Mon, Burmese and Pali.* Ye: Kharaik Sam Monastery.
Kruspe, N. 2004. *A grammar of Semelai*. Cambridge: CUP.
Luce, G. H. and Bohmu Ba Shin. 1961. *Pagan Myinkaba Kubyauk-Gyi temple of Rājākumār.* Rangoon: Bulletin of the Burma Historical Commission.
Matras, Y. 2009. *Language contact.* Cambridge: CUP.
Matras, Y. and J. Sakel (eds.) 2007. *Grammatical borrowing in cross-linguistic perspective.* Berlin/New York: Mouton de Gruyter.
Norde, M. 2009. *Degrammaticalization.* Oxford: OUP.
Okell, J. and A. Allott. 2001. *Burmese/Myanmar dictionary of grammatical forms.* Richmond: Curzon.
Payau, M. 1979. A description of Chaobon (ŋahkur) : An Austroasiatic language in Thailand. (Master thesis). Nakhon Pathom: Mahidol University.
Phayre, Sir A. P. 1998 [1883]. History of Burma including Burma Proper, Pegu, Taungu, Tenasserim, and Arakan.From the earliest time to the end of the first war with British India. Reprinted in Bangkok: Orchid Press.
Sakel, J. 2007. Types of loan: Matter and pattern. In Matras and Sakel (eds.) 2007, 15-29.
Sakomoto, Y. 1994. *Mon-Japanese dictionary.* Tokyo: Tokyo University of Foreign Studies.
Sapir, E. 1921. *Language. An introduction to the study of speech.* New York: Harcourt, Brace and Company.
Shorto, H. L. 1971. A dictionary of the Mon inscriptions from the sixth to the sixteenth centuries. London: Oxford University Press.
Taw Sein Ko. 1892. The Kalyānī inscriptions erected by King Dhammacetī at Pegu in 1476 A.D. Text and translation. Rangoon: Government Printing.
Thomas, D. D. 1971. *Chrau grammar.* Honolulu: University of Hawaii Press.
Thomason, S. G. 2001. *Language contact. An introduction.* Washington D.C.: Georgetown UP.
Thomason, S. G. and T. Kaufman. 1988. Language contact, creolization, and genetic linguistics. Berkeley: UCP.

An Analysis of Generic Structure Potential of Some Selected Austroasiatic Folktales[30]

Pattama Patpong
Research Institute for Languages and Cultures of Asia, Mahidol University

Abstract

This paper is concerned with a linguistic exploration of some selected Austroasiatic folktales. It draws on Systemic Functional Linguistics to investigate the discourse structure of some selected Austroasiatic folktales. In this current research, a Generic Structural Potential or text structure of some selected Austroasiatic folktales is proposed. The Generic Structure Potential is a generic (or schematic) structure of texts as they unfold through time (Hasan, 1984a/1996). In this paper, some linguistic features (i.e., lexicogrammatical patterns) involved in constituting the selected Austroasiatic folktales' Generic Structure Potential are discussed. The data for this study were drawn from a corpus of twenty Austroasiatic folktales ranging from Northern sub-branch, Eastern sub-branch and Southern sub-branch of Austroasiatic language family. They were taken from secondary sources (e.g., theses appendices, field notes). This paper opens with a brief of theoretical concepts of Generic Structure Potential. The Generic Structural Potential of the selected Austroasiatic folktales is discussed. The findings reveal that the Generic Structure Potential of Austroasiatic folktales are analysed into five generic stages: Placement Event, Initiating Event, Sequent Event, Final Event and Moral. A discussion on semantic properties of Placement Event is provided. Finally, some lexicogrammatical patterns that constitute the semantic properties of the Placement Event are discussed.

1. Introduction

Folktales are a variety of traditional narrative. In terms of text structure, folktales are texts that are created as a form of traditional story that tries to explain or understand the world. Imagination is vividly woven around talking animals, mythical creatures, supernatural beings, and magical objects (cf. Thompson, 1946). Folktales usually have no identified author and are orally passed down from generation to generation. In all societies, folktales are considered treasures of mankind. Telling folktales is a naturally clever method of teaching valuable lessons and mirror the values and culture of a society. In terms of medium, they can be either spoken or written. The typical mode of communicating folktales is spoken mode known as oral traditional folktales. Later these traditional folktales are recorded and printed out in the form of folktale collections.

Based on Systemic Functional Linguistics, folktales can be viewed according to three contextual variables: field, tenor, and mode. **Field** refers to a social action taking place in context. Field of folktales is to explore personal or vicarious (real or imagined) experiences of a flow of events forming an episode. In addition, social actions in folktales are used to construe a world-view and to constitute "a discursive medium for collective probing and problem-solving" (Ochs, 1997: 202). **Tenor** refers to kinds of role relationship.

[30] This research project is supported by Mahidol University. I would like to express my thankfulness to Professor Somsonge Burusphat for her inspiration on narrative discourse studies. I am greatly indebted to Professor Christian Matthiessen and Dr Kazuhiro Teruya for their continued support and encouragement. Special thanks go to Mr Richard Hiam for his editing support.

Patpong, Pattama. 2011. "An Analysis of Generic Structure Potential of Some Selected Austroasiatic Folktales." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.65-80.

Folktales enact social roles, relationships, obligation of values, and calibrate interactant models of shared experience. Social order in society is enacted through folktales. Folktales are also actualized as "a tool for instantiating social and personal identities" (Ochs, 1997: 202). **Mode** refers to the role of language in the interaction. The mode of folktales can be divided into two orientations: form and function. Based on form orientation, folktales present field and tenor contextual values in spoken and written forms. In terms of function orientation, folktales have various purposes: entertaining, teaching, informing, updating, and inspiring listeners and readers.

This paper aims to extend the study of Generic Structure Potential (text structure) of Thai folktales (Patpong, 2006, 2009) to ethnic folktales in general, and to explore and identify the Generic Structure Potential of Austroasiatic folktales in particular. The Generic Structure Potential of Austroasiatic folktales is explored in terms of text organization. Typically, the nature of the narrative is a combination of character orientation and a series of events that unfold through chronological succession (Longacre 1974; Martin 1992; Ochs 1997).

2. Text selection

In this paper, a corpus of twenty Austroasiatic folktales is explored. These twenty folktales are taken from secondary sources (e.g., theses appendices, field notes). Based on an index of folktale type (Thompson, 1946) folktales can be classified into two main types — simple, and complex. Simple tales are tales with one motif, while complex tales are tales with multiple motifs which involve many protagonists. In this research, all folktales selected are simple folktales. Table 1 presents Austroasiatic languages used in this research, their sub language classifications and habitats. Table 2 lists folktales used in this research. Each folktale is broken into clause complexes and clause simplexes.

No.	Language	Sub language family	Habitat
1	Lawa (2 texts)	Northern Mon-Khmer: Palaungic	Ban Phae Village, Amphoe Mae sariang, Mae Hong Son Province (La-Up Dialect)
2	Plang (2 texts)	Northern Mon-Khmer: Palaungic	Huay Namkhun village, Tambon Mae Fa Luang, Amphoe Mae Fa Luang, Chiang Rai Province
3	Khmu (2 texts)	Northern Mon-Khmer: Khmuic	Nam Sod Village, Tambon Lae, Amphoe Thung Chang, Nan Province
4	Pray (2 texts)	Northern Mon-Khmer: Khmuic	Phaeklang Village, Thung Chang District, Nan Province
5	Samre (2 texts)	Eastern Mon-Khmer: Pearic	Ban Ma-muang and Ban Nonsi, Bo Rai District, Trat province
6	Kasong (2 texts)	Eastern Mon-Khmer: Pearic	Ban Khlong Saen, Ban Padaw and Ban Damchumphon, Bo Rai District, Tart Province
7	Chong (2 texts)	Eastern Mon-Khmer: Pearic	Ban Bo rai, Trat Province
8	Sô (2 texts)	Eastern Mon-Khmer: Katuic	Kusuman District, Sakon Nakhon Province
			Kusuman District, Sakon Nakhon Province
9	Mon (2 texts)	Southern Mon-Khmer: Monic	Pathumthani Province
			Literary texts
10	Nyah Kur (2 texts)	Southern Mon-Khmer: Monic	Wangkampaeng Village, Tambon Cheebon, Bankwaw District, Chaiyapoom Province
			Ban Tha Duang, Nong Phai District, Petchabun Province

Table 1: A corpus of twenty Austroasiatic folktales used in this research

No.	Branch	Language	Folktales/Motif	Clause complexes	Clause simplexes
1	Northern Mon-Khmer	Lawa	The Orphan	32	90
			The Barking Deer and the Ghost (formula tale)	14	55
2	Northern Mon-Khmer	Plang	The Gibbon (animal tale)	6	20
			The Monkey (animal tale)	7	28
3	Northern Mon-Khmer	Khmu	The Lucky Man	12	34
			The Big Tree and the Small Tree	17	32
4	Northern Mon-Khmer	Pray	The Horned Owl (formula tale)	32	67
			The Two Cow Traders	39	141
5	Eastern Mon-Khmer	Samre	The Ancient Man	20	60
			The Rice Bone	11	49
6	Eastern Mon-Khmer	Kasong	Pla Lot (A kind of fish) (animal tale)	43	147
			The Dog Having a Human Child	40	157
7	Eastern Mon-Khmer	Chong	The Cow and the Tiger	14	35
			The Cicada and the Ant	15	28
8	Eastern Mon-Khmer	Sô	The Monkey and the Turtle	57	165
			The Daughter-in-law and the Mother-in-law	25	77
9	Southern Mon-Khmer	Mon	The Rich Man's Son	27	76
			The Intelligent Monkey	11	33
10	Southern Mon-Khmer	Nyah Kur	The Rainbow	8	24
			The Four Men	8	17
			Total	**438**	**1,335**

Table 2: Number of clause complexes and clause simplexes

3. Theoretical framework : Hasan's Generic Structure Potential

As mentioned in the Introduction Section, at the contextual stratum, folktales can be viewed in terms of variables of field, tenor, and mode. These contextual values can be realized structurally by a basic element known as "Generic Structure Potential" (GSP). The interest in the generic (or schematic) structure of texts has been greatly influenced by Hasan (Halliday and Hasan, 1985; Hasan, 1978, 1984b, 1984c, and 1996) and Martin (1985, 1992). The Generic Structure Potential (GSP) is an abstract category; it is descriptive of the total range of textual structures available within a genre. The GSP represents the total potential of structural resources for a genre.

The notion of Generic Structure Potential proposed by Hasan rests on the same principle of discourse analysis as that proposed by Longacre (1974, 1976, and 1983). Longacre proposed a distinction between deep and surface structures. Longacre's framework has served as the basis for research into narratives in Thai by Burusphat (1990, 1991, and 1992).

The approaches proposed by Hasan and Longacre, especially the deep structure idea for analysing narrative discourse, appear to be similar to one another, in the way they are concerned with text organization as it unfolds through time.

Longacre's deep structure

[1] Exposition [2] Inciting moment [3] Development conflict [4] Climax

[5] Denouement [6] Final suspense [7] Conclusion

Hasan's Generic Structure Potential of nursery tales

[(<Placement> ^) Initiating Event^] Sequent Event ^ Final Event [^ (Finale) • (Moral)]

Based on Hasan's Generic Structure Potential of nursery tales, one can find an opening stage in which the tale is oriented to time, place and the persona. The opening or **Placement Event** in text structure has certain semantic requirements. At the beginning of the tale, Placement Event, a character has to be introduced or 'particularised'. The character's circumstances must involve 'distance in time'. Later in the course of folktale development, the impact coming with the **Initiating Event** requires that the Placement Event set up a habit or pattern of events that is disturbed by a new and strange event. This intruding event carries the tale on through the **Sequent Event**. This in turn can be repeated until the tale's complication is resolved with the **Final Event**, along with the possibility of comments of what we learn **(Moral)** or the closure of the tale **(Coda)**.

4. Generic Structure Potential Analysis of Austroasiatic folktales

In this section, the Generic Structure Potential of three sub-branches of Austroasiatic folktales will be explored one by one. This is followed by the Generic Structure Potential the twenty Austroasiatic folktales.

4.1. Generic Structure Potential of three sub-branches of Austroasiatic folktales

In terms of folktale development, folktale text is constitutive of its situation unfolding through time. Every stage of folktale development is realized by a selected segment of text which is correspondent to each generic stage of the folktale.

Table 3, Table 4 and Table 5 show the generic segmentation of the three sub-branches of Austroasiatic folktales, each segment being indicated by numbered clause complexes.

GSP / Folktales	**Placement Event**	**Initiating Event**	**Sequent Event**			**Final Event**
			Sequent: Emerging	**Sequent: Intensifying**	**Sequent: Solving**	
Lawa 1	1	2	3-12	13-21	22-30	31-32
Lawa 2	1	2	3-4	5-6	7-13	14
Plang 1	1	2	3	4	5	6
Plang 2	1	2	3	4-5	6	7
Khmu 1	1-2	3	4-6	7-10	11	12
Khmu 2	1-6	7	8-11	12-15	16	17
Pray 1	1	2	3-14	15-30	31	32
Pray 2	1-2	3	4-12	13-34	35-38	39

Table 3: Segment of the Northern sub-branch folktales based on generic stage (4 languages)

GSP / Folktales	Placement Event	Initiating Event	Sequent Event			Final Event	Moral
			Sequent: Emerging	Sequent: Intensifying	Sequent: Solving		
Samre 1	1	2	3-7	8-16	17-18	19-20	—
Samre 2	1-2	3	4-5	6-8	9	10-11	—
Kasong 1	1-3	4	5-11	12-35	36-39	40-43	—
Kasong 2	1-5	6	7-14	15-28	29-37	38	39-40
Chong 1	1-2	3	4-10	11	12-13	14	
Chong 2	1-3	4	5-8	9-10	11-12	13	14-15
Sô 1	1-3	4	5-15	16-35	36-52	53-57	—
Sô 2	1-5	6	7-9	10-17	18-21	22-25	—

Table 4: Segment of the Eastern sub-branch folktales based on generic stage (4 languages)

GSP / Folktales	Placement Event	Initiating Event	Sequent Event			Final Event
			Sequent: Emerging	Sequent: Intensifying	Sequent: Solving	
Mon 1	1-6	7	8-14	15-24	25-26	27
Mon 2	1	2	3-6	7-9	10	11
Nyah Kur 1	1	2	3-4	5-6	7	8
Nyah Kur 2	1	2	3-4	5-6	7	8

Table 5: Segment of the Southern sub-branch folktales based on generic stage (2 languages)

As can be seen from the above tables, some generic stages are obligatory, whereas one stage is optional. In the Austroasiatic folktale corpus, Placement Event, Initiating Event, Sequent Event and Final Event are obligatory. Moral is optional. It is interesting to note that the Sequent Event is the most lengthy and complicated stage. It is further categorized into three sub-stages: Emerging Event, Intensifying Event and Solving Event. These three sub-stages correspond to Longacre's deep structure (1974, 1976 and 1983): Developing Conflict, Climax and Denouement respectively.

4.2. Generic Structure Potential of Austroasiatic folktales

As mentioned earlier, this analysis of Austroasiatic folktales is based on Hasan's framework "Generic Structure Potential" (1984b, 1996). Her framework for describing nursery tales is used, supplemented by a few extra customized for Austroasiatic folktales. Thus, Austroasiatic folktales can be analysed into five generic stages, as shown below.

[(<Placement Event>) ^ Initiating Event ^] Sequent Event ^ Final Event ^ (Moral)

( ) round brackets: enclose optional elements such as Placement and Moral

No round brackets: indicate obligatory elements

< > angled brackets: enclose elements whose lexicogrammatical realization may be included or interspersed with the lexicogrammatical realization of some other element(s)

[ ] square bracket: enclose the boundaries of a limitation of sequence. Indicated by enclosing the relevant elements, because mobile elements are mobile within certain limits

^ carat sign: indicate relative sequence

curved arrow: the possibility of iteration for that element

Based on the Generic Structure Potential, there are five generic stages: Placement Event, Initiating Event, Sequent Event, Final Event and Moral. Among these key five generic stages, Moral is an optional

stage. This stage is found in Kasong 2 'The Dog Having a Human Child' and Chong 2 'The Cicada and the Ant'. These two folktales are animal tales which end the folktales by a moral closure.

Each generic stage of Austroasiatic folktales mentioned above is realized by different semantic properties which are in turn achieved by lexicogrammatical realization. The following section is concerned with the semantic properties of the generic stage of Placement Event. These semantic properties are presented together with their lexicogrammatical realizations.

5. Semantic properties and lexicogrammatical realizations of the Placement Event

The semantic properties of the Placement Event are explored in this paper, as this generic stage is the starting point of all folktales. In terms of macro thematicalization, it is an obligatory stage and the most important stage of the folktales. The Placement Event is the first point of the folktale departure where the setting of the folktale is foregrounded and key protagonists are introduced into the scene.

Placement Event

Placement is composed of two semantic categories — nuclear and elaborative. The semantic nucleus comprises two categories — crucial element, and associated element. The elaborative properties consist of an obligatory element and an optional element. Table 6 summarizes the semantic properties and their lexicogrammatical realizations.

Semantic Property	**Sub-category**	**Possible Features**	**Realization**
nuclear property	crucial element	person particularization	***person particularization*** • group: nominal group Existent/Actor/Sayer: nom. group
		(1) established person particularization	***established person particularization*** • group: nominal group Actor/Senser/Carrier: nom. group
		(2) referencing of person particularization	***referencing of person particularization*** • group: nominal group: (non-elided/elided) pronoun
		introduction of protagonist	• clause: ↘ PROCESS TYPE Participant: ↘ Existent/Sayer: nom. group Process: existential/verbal: ↘ verbal group
	associated element	framing (1) temporal framing (2) spatial framing	***framing*** • clause: ↘ circumstance of Time and Space

Table 6: Semantic properties of generic structure of the Placement Event

Semantic Property	**Sub-category**	**Possible Features**	**Realization**
elaborative property	obligatory element	habitude (1) habitual event	habitual event • clause: ↘ PROCESS TYPE Participant: ↘ Actor/Behaviour/Senser/Carrier: nom. gp. Process: ↘ verbal group
		(2) extent specification	extent specification (frequency and duration) • clause: ↘ circumstance of Extent: Time: frequency and duration • group: adverbial group
	optional element	attribution	• clause: ↘ PROCESS TYPE Participant: ↘ Carrier/Identifier: nom. group Process: relational: ↘ verbal group • group: nominal group: Epithet/Possession/Ordinal/Classifier/Qualifier

Table 6 (cont.): Semantic properties of generic structure of the Placement Event

In this paper, some semantic properties of generic structure of the Placement Event are discussed further. The presentation starts with a discussion of nuclear semantic properties. This is followed by elaborative properties.

5.1. Nuclear semantic properties: Crucial element

5.1.1. Person particularization

The crucial element relevant in realizing Placement is that of "character particularization", which may be realized lexicogrammatically is a semantic property of person particularization.The Placement Event can be realized by formulaic features of the character particularization. In transitivity terms, it is realized by an existential process clause with a non-specific nominal group as Existent.

Lexicogrammatical realization of person particularization

group: nominal group

Existent: nominal group

The person particularization is realized through nominal groups as follows:

• an animate or quasi-animate noun as Thing followed by a cardinal Numeral and Classifier

(1) Pray 2: [1.1]

mi:	*khram*	*ʔuʔ*	*piaʔ*	*lɔŋ*
exist/have	people	be	two	CL.
[verbal group: Existential process]	[nominal group]			

'There were two people.'

(2) Lawa 2: [1.1]

kai	*pɔih*	*tiʔ*	*tuə*
exist/have	barking deer	one	CL.
[verbal group: Existential process]	[nominal group]		

'There was a barking deer.'

5.1.1.1. Established person particularization

The established person particularization refers to a protagonist who has already been introduced into the tale. It is expressed by a specific nominal group which is realized by an animate or quasi-animate noun as Thing. It is followed by a Classifier and definite determiner 'this, these'.

group: nominal group

nominal group: Noun as Thing ^ (Classifier) ^ determinate Determiner "this"

(3) Pray 2: [2.2]

khram	***piaʔ***	***lɔŋ***	***neː***	*mi*	*lɔŋ*	*ɕɯː*	*bɔʔ*	*taːw*
people	two	CL.	this	one	CL.	name	Mr.	taw

'(Among) these two persons, one person named Mr Taw.'

(4) Khmu 1: [9.1]

táʔ tháːw	***niː***	*kɔː*	*laːw*	*ʔún*	*Ṇkóʔ*	*sáːm*	*wan*	*sáːn*	*khɯːn*
old man	**this**	conj.	tell	give	paddy	three	day	three	night

'This old man told (the kind woman) to grow rice for three days and three nights.'

(5) Khmu 2: [2.1]

Ṇphɔʔ niʔ	*jɛt*	*naːj*	*kúŋ phiʔ*
man	**this**	stay	in	forest

'This man stayed in a forest.'

5.1.1.2. Referencing of person particularization

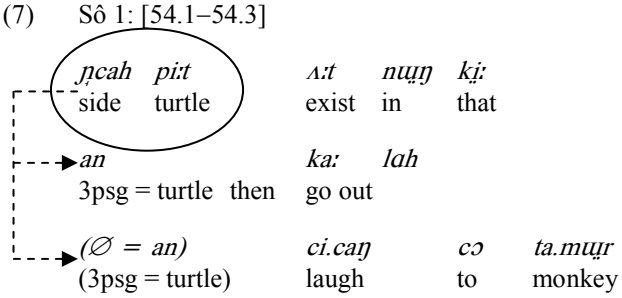
The referencing of person particularization is one of the textual resources which are used to track an identifiable, and continuous referent. This textual strategy, of retaining a referent in the unfolding discourse, is called "referential presumption" (Matthiessen, 2004a: 780). In terms of the textual metafunction, many unmarked topical Themes are implicit. For example, this happens when a narrator refers anaphorically to a referent that has already been established. Sequences of such implicit references create reference chains. The established character is realized by an elided personal pronoun, as shown in the following examples.

(6) Sô 1: [19.1–19.4]

ta.mɯ̱r	*kaː*	*bic*		
monkey	then	lie_down		
(∅ = an)	*krɑ*			
(3psg = monkey)	guard			
(∅ = an)	*lo̱ːn*	*ta.nʌːm*		
(3psg = monkey)	examine	tree		
priat	*an*	*pa.lay*	*lɛwʔ*	
banana	3psg	fruit	PRT_emphasis	

'Then he lay down by the tree, guarded, and examined it, until it bore fruit already.'

(7) Sô 1: [54.1–54.3]



ɲcah	*piːt*	*ʌːt*	*nuɻŋ*	*kɻː*	
side	turtle	exist	in	that	
an		*kaː*	*lɑh*		
3psg = turtle	then	go out			
(∅ = an)		*ci.caŋ*	*cɔ*	*ta.muɻr*	
(3psg = turtle)		laugh	to	monkey	

'As for the turtle, he stayed in there. He then came out laughing at the monkey.'

5.1.1.3. Introduction of protagonist

"Character particularization" or "protagonist" is introduced by either an existential or a verbal process clause. In the unmarked case, the protagonist is made known by an existential process clause expressed by a lexical verb, for example, Lawa 'kai'; Khmu 'ʔáh'; Pray 'miː'; Kasong 'ʔiːn'; Sô 'mɻː'; Mon 'nùm'; Nyah Kur 'nɻ̈ːm' meaning 'exist/have'. Examples of some folktales are given below.

- Protagonist introduction achieved by "existential process" clause

clause — PROCESS TYPE

Process: existential process: verbal group

(8) Khmu 2 [1]

hɔ́ːŋ hɔ́ːŋ	*niː*	***ʔáh***	*tút saʔɔːŋ*	*paːj*	*tút*
be very long	this	**exist/have**	tree	two	CL.

'A very long time ago, there were two trees.'

(9) Sô 1 [2.1]

mɻː	*ta.muɻr*	*kap*	*piːt*
exist/have	monkey	and	turtle

'There was a monkey and a turtle.'

(10) Mon 1 [1.1]

nùm	*sethi*	*mòa*
exist/have	rich man	one

'There was a rich man.'

(11) Nyah Kur 2 [1.1]

nɻ̈ːm	*phaɲɻ*	*pan*	*ɲah*
exist/have	man	four	CL.

'There were four men.'

5.2. Nuclear semantic properties: Associated element

5.2.1. Framing

Framing specifies the temporal and spatial frame of a folktale. The events and protagonists are placed at a point in time far removed from that of the folktale's creation or reception (Hasan, 1984b, 1996: 59). The spatio-temporal distance is an associated element which is realized by a circumstantial element, either by temporal or spatial circumstances.

5.2.1.1. Temporal framing

Temporal framing is achieved by a congruent realization of the circumstance of Time with the feature "far" and "uncertain". The temporal setting can be realized both at clause rank and group/phrase rank.

• Temporal framing at group/phrase rank: it is realized (↘) by a circumstantial element.

clause: ↘ Time

group: adverbial group: adverb of time

phrase: prepositional phrase

• Temporal framing expressed by an adverbial group

(12) Lawa 1: [1.1]

goʔ ʔlaŋ	*kai*	*ciə*	*kuən dɔi*	*tiʔ*
long ago	exist/have	tale	orphan	one

'A long time ago, there was a tale of a couple.'

• Temporal framing expressed by a prepositional phrase

(13) Sô 2: [1]

tæːray	*miˈː*	*kuʔyaː*	*nɨˈŋ*	*kuʔman*
in the old day	exist/have	mother-in-law	and	daughter-in-law

'In the old days, there were a daughter-in-law and a mother-in-law.'

5.2.1.2. Spatial Framing

Spatial framing is realized by: (a) a circumstance of Place with a feature "certain" or "uncertain"; and/or (b) a participant representing a mythical being (Hasan, 1984b, 1996: 63). In this current research, only instances of the circumstance of Place with a feature, certain or uncertain, are found. An example is given below.

• Spatial framing is expressed by a circumstance of 'uncertain' Place (e.g., a forest, a temple): prepositional phrase

(14) Sô 1: [1.1–1.2]

mi̯ː	*mu̯aj*	*tʰɯɤ*	*tɛː.a.raj*	***nu̯ŋ***	***tru̯ŋ.tra***	***mu̯ay***	***bɑːn***	*mi̯ː*
exist/have	one	time	old_days	**in**	**forest**	**one**	**place**	exist/have
ta.mu̯r	*kap*	*piːt*						
monkey	with	turtle						

'In the old days, there was one place in the forest. There was a monkey and a turtle.'

5.3. Elaborative semantic properties: Obligatory element

The elaborative semantic property is composed of two categories — obligatory element and optional element. The difference between obligatory elements and optional elements is that, while obligatory elements are present in all the selected tales, the optional ones are not necessarily present.

5.3.1. Habitude

Habitude refers to habitual acts/states of the particularized character(s) (Hasan, 1984b, 1996: 61). There are two subtypes of habitude semantic property: (i) habitual event; and (ii) extent specification.

5.3.1.1. Habitual event

Habitual event is realized lexicogrammatically by a clause with a particular combination of transitivity features. It can be expressed both at clause rank and group rank. For example,

(a) **clause** — PROCESS TYPE

Process: verbal group (e.g. verbal group with a range of different process types, mainly material process)

(i) existential process

The existence of person particularization functions as the background to the set of events which constitute the various stages of the story. This existence is realized by an existential process (e.g. *there was a king, there was a husband and wife*).

(ii) material and behavioural processes

Material and behavioural processes specify the protagonist's activities of doing and happening. The specification includes the protagonist's occupation and routine.

(15) Khmu: 2 [2.1–2.2]

Ṇphɔʔ	*niʔ*	***jet***	*naːj*	*kúŋ phiʔ*
man	this	**stay**	in	forest

'This man stayed in a forest.'

(∅)	*nəːŋ*	*kam*	*sát*	*mót*	*ʔaʔ*
(he)	know	language	animal	all	Fp.

'(He) knew all animal languages.'

• protagonist's occupation

(16) Nyah Kur 1: [1.1–1.3]

nɔːm	*baːr*	*ɲah*	*maːŋ*	*nthal*
exist /have	two	CL.	husband	wife

'There were two people, husband and wife.'

maːŋ	*ʔar*	*paːʔ*	*hŋaːn*
husband	go	**do**	**work**

'The husband went to work.'

phraw	***thɔŋ***	*hĩːʔ*
wife	**stay**	house

'The wife stayed at home.'

• protagonist's routine (e.g. go fishing)

(17) Kasong 1: [1.1–1.2]

pàː	*nak*	*nàːŋ*	*lamɔ̀ːŋ*	*kiː*	*naː*	***ceːw***	***kʰrîːt***	*méːl*
two	CL.	grandma	grandpa	that	P.	**go**	**fishing**	fish

'Grandma and grandpa went fishing,

(Ø = pʰuək nak)	***ceːw***	***kʰrîːt***	*Ø*	*tʰuk ʔaːw*	*tʰuk ʔaːw*
(they)	**go**	**net**	(fish)	everyday	everyday

(they) went fishing everyday.'

(iii) mental process

Mental processes represent the protagonist's feeling.

• protagonist's feeling: emotive

(18) Chong 1: [1–2]

ŋuə	*mɔŋ*	*wạːj*	*thɔː*	*klɔːsaː*
cow	and	tiger	do	friend together

'A cow and a tiger were friends.'

kʰeːnŋuə *mɔŋ* *kʰeːn waːj* *raksaː*
cow cub and tiger cub love together
'A cow cub and a tiger cub loved one another.'

(iv) relational process

Relational processes express the protagonist's characteristics by stating quality and possessive status.

(19) Sô 2: [1–2.3]
tæːray *miˈː* *kuʔjaː* *nɪˈŋ* *kuʔman*
in the old day exist/have mother-in-law and daughter-in-law
'In the old days, there was a daughter-in-law and a mother-in-law.'

kuʔjaː *pan* *kuʔjaː* *maˈt* *suːt*
mother-in-law be mother-in-law eye blind
'The mother-in-law was blind.'

(∅ = naˈw) ***thaw*** *cəˈː*
(she) **old** ASP.: Pfv.
'and (she was) old.'

kuʔman *kiˈː* *hunhuː* *huhhaːj*
daughter-in-law that easily angry and annoyed
'The daughter-in-law was easily angry and annoyed.'

(b) **group**: verbal group: ASPECT

Many ethnic folktales model time grammatically as ASPECT rather than TENSE, as in Thai. Using aspect makers, for example, a perfective aspect (e.g. *cəˈː* 'ASP: Pfv.: already' as in Sô tale) indicates that a set of events has already taken place (as in Example 19). Some events may happen as part of a habitude cycle.

5.3.1.2 Extent specification

Extent specification is one resource which is used to characterize the habitude semantic property. Habitude can be achieved by the use of circumstantial resources, particularly the circumstance of Extent of the subtype frequency.

clause: ↘ Extent: frequency

group: adverbial group

Extent of frequency expresses regularity of activity. It emphasizes the iterative actions of the main character(s). It is realized by an adverbial group as in Samre example.

(20) Samre 1: [1.1–1.4]
paaɹ^{C} *nak*^{B} *chanɯɯn*^{A} *kluəŋ*^{B} *ʔiinˀ*^{A} *khiinˀ*^{A} *muuj*^{C} *nak*^{B}
two CL. wife husband have child one class.
'A wife and husband had a child.'

∅ *ʔiinˀ*^{A} *khiinˀ*^{A}
(they) have child
'They had a child.'

khiinˀ^{A} *kɯɯɹ*^{A} *nɔɔŋ*^{B} *minˀ*^{A}
child be with mother
'The child was looked after by the wife,'

∅ *kɔ̀ʔ*^{B} *jaam*^{B} ***tuk***^{B} ***ʔaaw***^{A} ***tuk***^{B} ***ʔaaw***^{A} ***tuk***^{B} ***ʔaaw***^{A}
(he = child) then cry **every day** **every day** **every day**
'and (he) always cried every day.'

(21) Kasong 1: [1.1–1.2]
pàː *nak* *nàːŋ* *lamɔ̀ːŋ* *kiː* *naː* *ceːw* *kʰríːt* *méːl*
two CL. grandma grandpa that P. go fishingfish
'Grandma and grandpa went fishing,'

(∅ = pʰuək nak) *ceːw* *kʰríːt* *∅* ***tʰuk ʔaːw*** ***tʰuk ʔaːw***
(they) go net (fish) **every day** **every day**
'(they) went fishing everyday.'

5.4. Elaborative semantic properties: Optional element

5.4.1. Attribution

Attribution is a characteristic (e.g. quality, status, possessive, or relationship) which is assigned to the particularized character. Attribution is realized by relational process clauses — intensive, and possessive. In the relational clause, the particularized character is the Carrier/Identifer, which is represented by a nominal group, with Epithet and Numerative modification. Lexicogramatical realization of Attribute is presented below:

(a) **clause** — PROCESS TYPE

Process: ascriptive/identifying relational process: verbal group

• attribution realized by a relational process: intensive ascriptive process

(22) Chong 2: [3]
cakcan ***cɔŋhɔŋ***
cicada **be arrogant**
'The cicada was arrogant.'

• attribution realized by relational process: possessive ascriptive process

(23) Mon 1: [1.1–1.2]
nùm *sethi* *mòa*
exist/have rich man one
'There was a rich man.'

(∅ = dèh) ***nùm*** *kon* *plài* *ɓa* *(mòa)*
(he) **have** child male child two one
'(he) had a son.'

(24) Samre 1: [1.1–1.2]
paaɪ^{C} *nak*^{B} *chanɯɯn*^{A} *kluəŋ*^{B} ***ʔiin***^{A} *khiin*^{A} *muuj*^{C} *nak*^{B}
two CL. wife husband **have** child one CL.
'A wife and husband had a child.'

∅ ***ʔiin***^{A} *khiin*^{A}
(they) **have** child
'They had a child.'

• attribution realized by a relational process: intensive identifying process

(25) Kasong 2: [3.1–3.4]

k^{h}e:n	*rà:j pa:l*	*nak*	*ʔi:n*	*klɔ̀:ŋ*	*mot*	*hɔ́:j*
child	twelve	CL.	have	husband	all	Fp.

'The twelve children had husbands.'

klɔ̀:ŋ	*nak*	***pen***	*taha:n*
husband	he	**be**	solider

'The husbands were soliders,'

(Ø = p^{h}uək nak)	***pen***	*cá:w*
(they)	**be**	royal family member

'(they) were royal family members,'

(Ø = p^{h}uək nak)	***pen***	*na:j*	*mot*
(they)	**be**	noble	all

'(they) were nobles.'

(b) **group**: nominal group: Qualifier/Ordinal/Classifier/

(i) Qualifier

(26) Khmu 1: [2–3]

ʔɯŋ	*kon*	*ʔák*	*lɔ́:ʔ*
one	person	mind	good

'One person was kind.'

ʔɯŋ	*kon*	*ʔák*	***ʔaj***	***lɔ́:ʔ***	*hó:c*
one	person	mind	**NEG.**	**good**	ASP.:Pfv.

'The other person was mean.'

(ii) Ordinal and Classifier

(27) Samre 1: [1.1–1.2]

paaɪ^{C}	***nak^{B}***	*chanɯɯn^{A}*	*kluəŋ^{B}*	*ʔiin^{A}*	*khiin^{A}*	***muuj^{C}***	***nak^{B}***
two	**CL.**	wife	husband	have	child	**one**	**CL.**

'A wife and husband had a child.'

Conclusion

Folktales are constructed as an ongoing process of selection of features in simultaneous systems. By tracking the local selections, the pattern of lexicogrammar in folktales unfolds in the texts. In this paper, the Generic Structure Potential of twenty Austroasiatic folktales is explored. The paper explores three strata of Austroasiatic folktales: contextual stratum, semantic properties of the Placement Event, and lexicogrammatical stratum realizing the Placement Event.

At the contextual stratum, the Generic Structure Potential of the Asutroasiatic folktales is also investigated. The analysis reveals that Austroasiatic folktales are analysed into five generic stages — Placement Event, Initiating Event, Sequent Event, Final Event, and Moral. These stages vary across the twenty Austroasiatic folktales. At the semantics stratum, all five generic stages were realized by semantic properties. The semantic properties comprise nuclear and elaborative semantic repertoires. At the lexicogrammatical stratum, the folktales realized as systems of meaning are in turn realized by systems of wordings that are expressed at clause and group/phrase rank.

Based on the current research findings, it is suggested that there remains further study to be undertaken such as: exploration of Generic Structure Potential of folktales of other language families and a comparative study of Generic Structure Potential of folktales taken from different language families.

References

Burusphat, Somsonge. (1990). "The biparties structure of information in Thai narrative discourse." *Journal of Language and Linguistics*. 9.1: 65-78

Burusphat, Somsonge. (1991). *The structure of Thai narrative*. Summer Institute of Linguistics and The University of Texas at Arlington. Publication 98.

Burusphat, Somsonge. (1992). "The Identification of storyline in Thai narrative discourse." In Shin Hwang, J. Ja, and William R. Merrifield (eds*.)*, *Language in context: essays for Robert E. Longacre*.

Burusphat, Somsonge. (2001). "Foregounding in some selected Southeast Asian narratives." Paper presented at Southeast Asian Linguistics Society XI Conference. Institute of Language and Culture for Rural Development. Mahidol University, Thailand. May 16-18, 2001.

Caffarel, Alice, James R. Martin and Christian M.I.M. Matthiessen (eds.), (2004). Language typology: a functional perspective. Amsterdam: John Benjamins.

Connelly, F. Michael and D. Jean Clandinin. (1990). "Stories of experience and narrative enquiry." *Educational Researcher*. 19.5: 2-11.

Eggins, S. and Di Slade. (1997). *Analysing casual conversation*. London: Cassell.

Halliday, M.A.K. (1985/1994). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M.A.K. and Ruquiya Hasan. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K. and Christian M.I.M. Matthiessen. (2004). Third edition. *An introduction to functional grammar*. London: Edward Arnold.

Hasan, Ruquiya. (1978). "Text in the systemic-functional model." In Wolfgang. Dressler. (ed.), *Current trends in textlinguistics*. Berlin: de Gruyter. 228-246.

Hasan, Ruquiya. (1979). "On the notion of text." In Petöfi, J.S. (ed.), *Text vs. sentence: Basic questions of textlinguistics*. (Papers in Textlinguistics 20.2). Buske: Hamburg.

Hasan, Ruquiya. (1984a). "Coherence and cohesive harmony." In James Flood (ed.), *Understanding reading comprehension: cognition, language and the structure of prose*. Newark: International Reading Association.

Hasan, Ruquiya. (1984b/1996). "The nursery tales as a genre." *Nottingham Linguistic Circular*. 13: 71-102. (Reprinted in Hasan, Ruquiya. 1996. Ways of Saying: Ways of Meaning. London: Cassell.)

Hasan, Ruquiya. (1984c). "The structure of the nursery tale: an essay in text typology." In L. Coveri (ed.), *Linguistica Testuale*. Rome: Bulzoni.

Katpathum, Vichian. (1999). *Nithan Phuenban (Thai Folktales)*. Bangkok: Patthanasuksa. (in Thai)

Longacre, Robert E. (1974). "Narrative versus other discourse genre." In R.M. Brend (ed.) *Advances in Tagmemics*. Amsterdam: North-Holland. 357-367.

Longacre, Robert E., (1981). "A spectrum and profile approach to discourse analysis." *Text*.1.4: 337-359.

Longacre, Robert E. (1990). "Storyline concerns and word order typology in East and West Africa." *Studies in African Linguistics*. Supplement 10: 1-10.

Martin, James R. and Plum Guenter A. (1997). "Construing experience: some story genres." *Journal of Narrative and Life History*. 7.1-4: 299-308.

Martin, James R, Christian M.I.M. Matthiessen and Clare Painter. (1997). *Working with functional grammar*. London: Edward Arnold.

Matthiessen, Christian M.I.M. (1995a). "*Lexicogrammatical cartography: English systems*." Tokyo: International Language Sciences Publishers.

Matthiessen, Christian M.I.M. (2000). "Narrative register — lexicogrammatical patterns." MS. Department of Linguistics, Macquarie University.

Matthiessen, Christian, M.I.M. (2004). "Description motifs and generalization." In Caffarel, Alice, James R. Martin and Christian M.I.M. Matthiessen (eds.), *Language typology: a functional perspective*. Amsterdam: John Benjamins.

Matthiessen, Christian, M.I.M. (n.d.c). "Text Noah's Ark." MS. Department of Linguistics, Macquarie University.

Matthiessen, Christian M.I.M. and M.A.K. Halliday (1997). *Systemic functional grammar: a first step into the theory*. Sydney.

Ochs, Elinor. (1997). "Narrative." In Teun van Dijk (ed.), *Discourse as structure and process*. Discourse Studies 1: A Multidisciplinary Introduction. London: Sage. 185-207.

Polkinghorne, Donald E. (1988). *Narrative Knowing and the Human Sciences*. New York: State University of New York Press.

Patpong, Pattama. (2006). A Systemic Functional Interpretation of Thai Grammar: An Analysis of Narrative Discourse. Ph.D. thesis, Macquarie University.

Patpong, Pattama (2009). An Analysis of Gerneric Structure Potential of Thai Folktales: Systemic Functional Linguistics. A paper presented at the International Conference on Asian Folklore 2009 at the Mirical Grand Hotel, Bangkok, Thailand, July 2nd-3rd 2009.

Premsrirat, Suwilai. 1995. "On language maintenance and language shift in minority languages on Thailand: a case study of So (Thavung)." Paper presented at the International Symposium on Endangered Languages, University of Tokyo, November 18-20, 1995.

Premsrirat, Suwilai. 2007. Endangered Languages of Thailand. *Int'l J. Soc. Lang*. 186: 75-93.

Teruya, Kazuhiro. (1998). An exploration into the world of experience: a systemic functional interpretation of the grammar of Japanese. Ph.D. thesis, Macquarie University

Thai, Minh Duc. (2004). "Metafunctional profile of Vietnamese." In Caffarel, Alice, James R. Martin and

Christian M.I.M. Matthiessen (eds.), *Language typology: a functional perspective.* Amsterdam: John Benjamins,

Thompson, Stith. (1946). *The Folktale*. Holt, Rinehart and Winston. (Reprinted 1977. Berkeley: University of California Press.)

Toolan, Michael J. (1988). *Narrative: a critical linguistic introduction*. London: Routledge.

Ure, Jean (1989). "Text Typology." MS.

Other Sources Consulted

Chuasuwan, Sudsawad. 1990. *A Phonology of Nyah Kur at Ban Tha Duang, Petchabun Province*. M.A. Thesis. Institute of Language and Culture for Rural Development, Mahidol University.

Giaphong, Suchada. 2004. Plang Grammar as spoken in Huay Namkhun Village, Chiang Rai Province. M.A.thesis, Institute of Language and Culture for Rural Development, Mahidol University.

Kamnuansin, Sunee. 2002. *Kasong Syntax*. M.A.thesis, Institute of Language and Culture for Rural Development, Mahidol University.

Komonkitiskun, Jiranan. 1985. *Some General Characteristics of Lawa Grammar (La-Up Dialect)*. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Lerthirunwong, Malai. 1980. A Syntactic Description of So: An Austroasiatic Language in Thailand. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Malapol, Mingkwan. 1989. Pray Grammar at Ban Pae Klang, Thung Chang District, Nan Province. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Memamas, Payau. 1979. A Description of Chaobon (ɲahkur): An Austroasiatic Language in Thailand. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Ploykaew, Pornsawan. 2001. *Samre Grammar*. Ph.D thesis, Institute of Language and Culture for Rural Development, Mahidol University.

Rotchanan, Nattamon. 2009. *Chong Syntax*. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Wongnoppharalert, Saowalak. 1993. *Nam Sod Khmu Syntactic Structure: A Study in Tagmemics, Transformation and Case Grammar*. M.A. Thesis. Institute of Language and Culture for Rural Development. Mahidol University.

Kantrum: the Status of Folk Music Education within the Thai Khmer Cultural Zone

Saowapa Pornsiripongse and Anan Soplerk
Research Institute for Languages and Cultures of Asia, Mahidol University

Abstract

Kantrum, Thai-Khmer folk music written in Khmer language, is very popular within the Thai-Khmer cultural zone in the northeast of Thailand, namely in Burirum, Surin and Srisaket provinces. Because of the influence of western and commercial music on the back of globalization, *Kantrum* has gradually become less important to the Thai-Khmer way of life. Many *Kantrum* songs and rhythms were played without identification which is the cause of the loss. The way to rebuild recognition of these songs is to recover their names and include them as part of the formal education curriculum. This article aims to survey the situation of folk music at all levels of the formal education system; primary, secondary and higher within the Thai-Khmer cultural zone and to recover and identify *Kantrum* rhythms and names of songs regularly recorded and sung, using focus group discussion methodology. Many well known *Kantrum* singers, musicians, composers, and *Kantrum* radio announcers were invited to brainstorm and discover the original names. They were then classified into three levels, from simple to complex.

Introduction

Thai-Khmer ethnic groups living in the northeast of Thailand, are crowded into three provinces, namely Burirum, Surin and Srisaket and scattered throughout other provinces such as in some parts of Nakonrachaseema, Ubonrachathanee, Roiet, and Mahasarakam. (Itsara Choosri, 2552: 28). It is estimated that there are 1.4 million people who speak Khmer in Thailand (Suwilai Premsrirat, 1997: 130). Almost all of them live in Surin, Burirum and Srisaket. However, their Khmer language is different from that used in Cambodia.

Kantrum, Thai-Khmer ethnic identity, is the traditional music of the Thai-Khmer ethnic group and very popular within the Thai-Khmer cultural zone. In the past it was only performed as part of ritual therapy but has now been extended to be played at wedding ceremonies and as entertainment. The pleasure of *Kantrum* music is its melody and rhythm for dancing. The traditional *Kantrum* instruments are *Pi-or*[31], *Saw-Kantrum*[32], *Klong-Kantrum*[33], *Chings*[34] and *Chabs*[35]. These days, western musical instruments such as guitars, drums and violins may be used depending on the fancy of the audience. [36] *Kantrum* lyrics are in Khmer, recounting Khmer way of life, worldviews, social values, domestic roles and etc. So, we can say that *Kantrum* is a medium for the cultural reproduction of the Thai-Khmer ethnic people.

[31] *Pi-or* is a double reed woodwind instrument. It consists of two parts, the reed and the body.

[32] *Saw-Kantrum* is a bowed stringed instrument with sound box made of wood, the head of which is covered with snake skin.

[33] *Klong-Kantrum* has a characteristic shape similar to the Thon of the Central Thailand

[34] *Ching*s are a pair of cymbals made of thick metal alloy and shape like a small cup.

[35] *Chabs* are shaped like plates made of thinner metal than that of the *Ching*

[36] See *www.mapculture.org* Accessed on 16 October 2009.

Pornsiripongse, Saowapa and Anan Soplerk. 2011. "Kantrum: the Status of Folk Music Education within the Thai-Khmer Cultural Zone." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.81-88.

The usual way of passing on *Kantrum* knowledge from generation to generation, the same as other indigenous wisdom, is by recital, not by documentation or formal education. This kind of learning was suitable in the past but the influence of western and commercial music on the back of globalization has endangered *Kantrum* music. So, it has gradually become less vital to the Thai-Khmer way of life. Leaving the traditional of *Kantrum* to die is risky because of the loss of many *Kantrum* songs and rhythms which were played without ever being properly identified. Therefore, one method to preserve *Kantrum* is to rebuild recognition of these songs, recover their names and include them as a part of the formal education curriculum for the Thai-Khmer cultural zone.

The objectives of this article are to survey the status of the folk music, especially *Kantrum*, at all levels of the formal education system; primary, secondary and higher within the Thai-Khmer cultural zone and to recover and identify *Kantrum* rhythms and the names of songs regularly recorded and sung. (See an example of *Kantrum* for secondary level at http://www.youtube.com/watch?v=B9OdBbs8ygI)

Figure 1: Map of North East Thailand showing Burirum, Surin and Srisaket Provinces (in blue)

The methodologies used for data collection are (1) a survey (2) workshops and (3) the recording of all *Kantrum* rhythms and notations. The 2,856 questionnaires were sent out to all levels of government schools, primary, secondary and higher education, in Burirum, Surin and Srisaket to find out the status of folk music education. In the first workshop, many well known *Kantrum* singers, musicians, composers, and *Kantrum* radio announcers were invited to brainstorm and discover the original names of the songs and rhythms. These people were invited again to the second workshop to check the names, rhythms and notations.

The situation of teaching folk music in the Thai-Khmer cultural zone

Number of schools teaching folk music

The movement to preserve *Kantrum* was started by individuals and groups in Surin, the province with the strongest Khmer culture- *Kantrum* musicians, teachers, and those people who recognize the value of local cultures. However, their effort were not matched in equal measure by government support, especially for the education authorities who were not ever able to provide data on how many schools taught *Kantrum*. Therefore, the questionnaires sent to the government schools were important for the study, but only 535

copies of the 2,865 copies or 18.73% were returned. From those that were completed, it appears that 193 schools or 36.27% have at some time taught folk music.

The types of folk music being taught in this cultural zone are, in descending order, *Pong lang*[37], (60 schools or 27.7%) *Kantrum*, (49 schools or 22.68) *Mo lam*[38], *Mahoree*[39] *and Jarieng*[40] as shown in table 1. When considering each province, it was found that *Ponglang* is taught in 31 schools in Srisaket, 21 schools in Burirum, and only 8 schools in Sirin. *Kantrum* on the other hand is taught in 25 schools in Surin, 15 schools in Burirum and 9 schools in Srisaket. This data confirms that Thai-Khmer culture is most influencial in Surin which is thus considered the center of *Kantrum*.

Why is folk music taught in so few schools in the Thai-Khmer cultural zone? The reasons are that (1) There is no folk music in their communities, (2) the schools are located in the Thai cultural zone or the cultural zones of ethnic groups such as *Kui* or *Lao* and (3) there are no folk music teachers.

Levels and frequency of folk music teaching

According to the Ministry of Education, the educational system is separated into 4 levels; the first level, grades 1-3[41], the second level, grades 4-6, the third level, grades 7-9 and the fourth level, grades 10-12. The folk music has been or is being taught at the second (128 schools), third (79 schools), first (55) and the fourth (18) as shown in table 2. Students aged 10-15 years are the most suited to folk music studies while those 7-9 years are too young to appreciate the social and cultural relevance. Fourth level students are generally too busy preparing for higher education to have time for folk music. However, the number of the secondary schools teaching folk music is less than primary schools.

The time allocated to folk music studies at all levels is, once a week for 75.50% or 151 schools; twice a week for 10% and less than once a week for 3.5% or 7 schools. About 22 schools or 11.40% taught it as part of musical club activities after school time or only leading up to special celebrations. (See details in table 2)

Patterns of folk music teaching in schools

Almost all of the schools used various methods of teaching. It is found that 85 schools or 31.48% taught folk music using folk musicians themselves, 59 schools or 21.85%, taught through music teachers, 55 schools or 20.37% relied on teachers of other subjects and 31 schools or 11.48% took students out to learn from folk music experts in the communities. (Details shown in table 3) The number of schools arranging musical courses within the communities is the lowest because it is difficult to organize together with other subjects, is expensive and very few communities have folk music experts available. However, many schools encourage students to join community ceremonies at which folk music is played on all occasions.

Lacking music teachers, folk music is taught by other teachers so they teach what is in the text, but can not demonstrate how to play the instruments so the students have to learn from CD. One interesting method discovered is called “Older teach younger”. This method can be used in other schools with no music teachers. The student who can play folk music teaches and shares knowledge with other students. This method is used at all levels of educational system. At Arts and Culture Preservation Clubs at the higher education level, it is found that students who join the clubs can usually play folk music since their parents are

---

[37] The *Pong Lang* is a log xylophone from the northeast or Isan. It has 12 wooden bars, with a strong rope, tied together in a row at each end. The instrument can be played solo or in an ensemble. The *Pong Lang* is used for all kinds of occasions, especially for festivals and ceremonies.

[38] *Mo lam* is a traditional Isan song. It uses flexible melodies which are tailored to the tones of the words in the text. The words usually reflect the difficulties of rural Isan life and unrequited love.

[39] *Mahoree* is the folk ensemble of the northeast. In the past it was very popular. It is used in the parades.

[40] *Jarieng*, Khmer language, means singing. It is used for telling folk stories and describing customs, and peoples way of life. It looks like *Mo Lam*, a traditional folk music of Isan. The instrument used with *Jarieng* is *Kaen.*

[41] Level 1 and 2 are primary school and level 3 and 4 are secondary school

folk musicians. Sometimes they can earn special income from playing folk music. These students also have the chance to go abroad for folk music festivals.

Activities supporting students and promoting folk music teaching in the schools

From the 535 returned questionnaires, 260 schools or 48.59% have activities developing students' potential in folk music. Student clubs or associations are the most popular activities according to 64.23% of respondents. Music clubs are preferred at 25 schools, classical dance clubs at 14 schools, classical Thai clubs at 14 schools and *Ponglang* clubs at 10 schools. (Details shown in table 4)

The problems of folk music teaching in the schools

The problems of teaching, arranging activities to promote students' potential and promoting folk music teaching in the schools are as follows:

Lack of music teachers
The lacking music teachers[42] is the major problem found at both schools currently teaching folk music and those not teaching. Four hundred and forty four schools or 78.00% raised this problem. Many schools relied on teachers with no music teaching skill or those who could play only 1 or 2 musical instruments so the quality of teaching is in doubt.

Lack of folk musicians
Many schools identified problems because of a lack of folk musicians in their communities or vicinity and noted that folk musicians in their communities had passed away. Some schools have neither music teachers nor local teachers. Non-local teachers are usually unable to speak Khmer and are not familiar with the local people.

Insufficient budget
Insufficient budget is the second biggest problem faced by 405 schools or 77.14%. They do not have enough money to buy various kinds of musical instruments so the numbers of interested students and availability of musical instruments do not match. They can not afford to engage the services of folk musicians or buy instruction media. They can not send their students to music competitions, performances or music camps either. Furthermore, music teachers or teachers responsible for teaching music can not be funded to undergo further music training.

Insufficient musical knowledge and others
Nearly a half of schools (230 schools or 43.8%) raised the issue of a lack of folk musical experience and knowledge to teach the students. Many schools said that the Ministry of Education does not promote local curricula[43]and many schools and students are not interested in folk music education. (See details in table 5)

Unidentified names of the songs and rhythms
This problem is not mentioned in the questionnaires but it arose from a workshop that was conducted with well known *Kantrum* singers, musicians, composers, and *Kantrum* radio announcers. Of 150 *Kantrum* songs and rhythms, there were about 70 or less than half with identifiable names so it is rather difficult to refer to some songs when teaching. The only way to identify these songs is to play them for recognition. So, if we can find the names and the rhythms of *Kantrum*, it will make it easier to teach students.

Un-prioritize the simple and complex Kantrum songs
As the result of the workshop, 150 *Kantrum* songs were listed with their names and rhythms. If we leave this folk music to be passed on by local people only, it may be loss in the near future. One way to ensure the survival of *Kantrum* is to push for it to become part of the formal education curriculum by arranging the songs according to their simple or complex rhythms suitable for different education levels. This task began

[42] A music teacher here means a teacher with a degree in music or dance, or has musical knowledge, or can teach how to play musical instruments.

[43] Local curriculum is a curriculum with local content of those areas such as local wisdom, local intellectuals, local arts, customs and ceremonies, local productions, etc.

at the workshop and will be implemented into some schools next semester

Higher education and Kantrum inheritance

Surin, the center of Khmer culture in South Isan, has two higher education institutes, Surin Ratchapat University and Surin Rachamongkol University. Both institutes have big and famous music and dance associations with many students. Only Surin Ratchapat University offers an optional course in folk music and about 100 students are currently registered. (Interview with a music lecturer, 11 June, 2009) A folk music curriculum is in the process of being designed. Even though it is a selective course, the supporting activities are very interesting such as inviting many *Kantrum* and folk musician experts to teach, the older teaching the younger, folk music performances in communities and teaching folk music to the interested local people, folk music contests and *Kantrum* contests. Surin Ratchamongkol University also has a folk culture preservation association with many students who can play all kinds of folk music. Their supporting activities are the same as Surin Ratchapat University. The mains problems in providing folk music courses, like at the primary and secondary levels, are insufficient budget and time to practice.

Burirum Ratchapat University is a higher education institute in Burirum. Folk music courses are not offered to students but they have managed to establish a *Kantrum* and *Ponglang* band.

Conclusions

Culturalists and UNESCO accepts that the cultural diversity is valuable and important to the world. Thailand is a country of great cultural diversity because it has more than 60 ethnic groups. (Suwilai Premsrirat, 2004). Thai-Khmer is an ethnic group living in South Isan. Folk music, especially *Kantrum* is the value identity of the Thai-Khmer that not only reflects their beliefs, value systems, taboos and ways of life but also reproduces the Thai- Khmer cultures. If *Kantrum* disappears, the Thai-Khmer people will lose their identity as well. Nowadays, *Kantrum* has gradually become less vital because of both external and internal factors; the influence of pop and commercial music, the loss and a lack of folk musicians and, most disappointingly, no real support from the government. The survival of *Kantrum* and other forms of folk music is possible if it were to be added as an integral part of the formal education curriculum. So, survey to determine the situation of folk music teaching in schools within the Thai-Khmer cultural zone has been well overdue. The result found that the situation of folk music teaching, especially *Kantrum* has been in critical condition. Only one third of the schools have folk music currently being taught. *Ponglang*, Isan or Lao music, is the most popular music taught at schools, *Kantrum* is the second but the most popular in Surin.

The school levels at which folk music is taught are, in descending order, the third, the first, the second and the forth. All levels are taught one period (fifty minutes) a week. There are many methods of teaching folk music but most commonly it is taught by folk musicians (42.96%, both inviting them from the communities and taking students out to learn in the communities where the musicians live). Many schools use a variety of methods depending on availability. One worrying finding is that many schools use texts without demonstrating the musical instruments involved due to lack of skills on the part of teachers. Promotional activities are also available at those schools that do not offer folk music classes.

The problems of folk music teaching found at every level are a lack of qualified music teachers and folk musicians, insufficient budget, a lack of sufficient knowledge of the subject, unidentified titles of songs and rhythms, and the mixing of simple and complex *Kantrum* rhythms atdifferent levels. In higher education, even though they do not have major courses in folk music, some universities have optional courses for interested students. And most of them have a lot of supporting activities in the form of cultural clubs and associations.

Suggestions

1. Folk music, as with other cultural subjects is as important as mainstream subjects in developing critical thinking, creativity and recognition of locality which form the base to further studies and forming of the ‘complete’ human being. So, the Ministry of Education should emphasize the importance of folk music as an integral part of cultural studies within the local curriculum.

2. One problem for the preservation of folk music is the lack of folk musicians. So the Ministry of Culture, in collaboration with the Ministry of Education, should establish mechanisms to encourage the young to learn and promote an appreciation of folk music through such things as folk music competitions and workshops so that musicians will have social space to demonstrate their ability and showcase their cultural identity.

References

Itsara Choosri. (2009). "Thai/Khmer identity: Historical and cultural perspectives" In the Cultural Mapping of the Ethnic groups along Thai-Cambodian Border. Nakonpathom: Institute of Language and Culture for Rural Development, Mahidol University.

Office of the National Culture Commission, Ministry of Education. (1984). *Folk Music and Traditional Performing Art of Thailand*. Report of a study on Thai Folk Music and Traditional performing Arts under the auspices of UNESCO.

Suwilai Premsrirat. (1997). "Linguistic contribution to the study of the Northern Khmer language of Thailand in the last two decades" In the Mon-Khmer Studies Journal, 27: 129-136.

Suwilai Premsrirat. (2004). Ethnolinguistic Maps in Thailand. Bangkok: Office of the National Culture Commission.

Cultural Mapping of the Ethnic groups along the Thai-Cambodain Border *" Charieng* Surin*"* (http:// www.mapculture.org)

Map of Thailand. (http://www.novabizz.com/Map/)

Appendix: Tables

Table 1 Folk music styles taught in Burerum, Surin, and Srisaket.

Province	Question-naires sent	Questionnaires completed		Not taught		Taught		*Ponglang*		*Kantrum*		*Mo Lam*		*Mahoree*		*Jarieng*		others	
		Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Burerum	956	182	19.03	120	67.03	56	32.96	21	31.34	15	22.38	4	5.97	9	13.4	2	2.98	16	23.88
Surin	865	169	19.53	103	60.94	66	39.05	8	13.11	25	40.98	5	8.19	6	9.83	0	0	17	27.86
Srisaket	1005	184	18.30	116	63.04	68	36.95	31	35.22	9	10.22	17	19.31	5	5.68	2	5.68	24	27.27
Total	2865	535	18.73	339	63.72	193	36.27	60	27.27	49	22.68	26	12.03	20	9.25	4	1.85	57	26.38

Table 2 Levels and frequency of classes in schools in Burerum, Surin, and Srisaket.

Province	Level 1					Level 2					Level 3					Level				
	1/1 w	2/w.	1/2 w.	others[44]	Total	1/1 w	2/w.	1/2 w.	others	Total	1/1 w	2/w.	1/2 w.	others	Total	1/1 w	2/w.	1/2 w.	others	Total
Burirum					14										23					4
Surin	15	20\	1	2	20	20	0	0	5	39	16	3	3	5	27	3	2	1	0	6
Srisaket	18	0	1	2	21	43	1	1	6	39	21	6	0	2	29	5	3	0	0	8
Total	33	2	2	4	55	73	1	1	11	50	37	9	3	7	79	8	5	1	0	18

[44] Teaching after school every other day or once a week or by arrangement.

Table 3 Methods of folk music taught in Burirum, Surin and Srisaket

Province	Methods of folk music taught[45]				
	Music teachers	Other teachers	Invited folk music experts	Folk music experts in the communities	others
Burirum	24	14	27	7	9
Surin	15	17	29	12	21
Srisaket	20	24	29	12	10
Total	59	55	85	31	40

Table 4 Activities supporting students and promoting folk music teaching in Burirum, Surin and Srisaket

Province	With activities		No activities		Activities					
					Clubs/ associations		Other activities		Not identify	
	No.	%	No	%	No	%	No	%	No	%
Burirum	102	55.73	81	44.26	49	62.82	29	37.17	0	-
Surin	80	47.90	87	52.09	56	60.21	35	37.63	2	.77
Srisaket	93	49.72	92	49.72	62	65.95	32	34.04	0	0
Total	275	51.4	260	48.59	167	64.23	96	36.22	2	.77

Table 5 Problems of folk music education in Burirum, Surin and Srisaket

Province	Problems		No problems		No qualified teachers	Insufficient budget	No coordinators	No body of knowledge	No promotion	Others
	Total	%	Total	%						
Burirum	11	6.25	165	93.5	137	133	52	70	45	23
Surin	9	5.42	157	94.57	132	127	45	70	38	21
Srisaket	6	3.27	177	96.72	145	145	74	90	58	22
Total	26	4.95	499	95.4	414	405	171	230	141	66
%					78.00	77.14	32.57	43.8	26.85	12.57

[45] Many schools have various methods of folk music taught.

Demonstratives in Muöt (Nancowry)

V. R. Rajasingh
Central Institute of Indian Languages, Mysore, India

1. Aim

The aim of this paper is to attempt for a theoretical framework for the definition, identification, classification and analysis of demonstratives in Muöt. The paper is data bound. The data are drawn from the Andaman Commissioned Project data base which was collected from Nancowry Island between September and December of 2004, just before the killer tsunami.

2. Muöt

Muöt here refers to that variety of language spoken by the ethnic Nicobarese inhabiting the four central Nicobar Islands, namely, Nancowry, Katchal, Kamorta and Trinket of the Nicobar archipelago, India.[46] The present day language of the ethnic Nicobarese inhabiting the islands of Nancowry, Katchal, Kamorta and Trinket attests the word Muöt as referring to the land mass which is widely known as Nancowry Island. Kloss (1971) mentions this land mass as Nankauri and Radhakrishnan (1981) as Nancowry, while Röepstorff (1987) appears to refer to it as Nancowry. However, Man (1975) uses the word Nâng kauri, as a native term of the language, to refer to the land mass covered by both the Nancowry and Kamorta islands. Besides, the language also attests similar indigenous terms to refer to the remaining three islands that define its linguistic area; they are Tehnyu for Katchal, Kinlaka for Kamorta and Laful for Trinket. Man (1975) and Lewis (2009) term this language as Central Nicobarese language, but Röepstorff (1987) and Radhakrishnan (1981) call it as Nancowry dialect and Nancowry language respectively, presumably, taking the geographical name of the land mass which is referred to as Muöt. This paper would have no hesitation in taking the view of Radhakrishnan and use the term Nancowry language, but, respecting the sentiments of the native speakers, it is prefered to use the indigenous name Muöt.

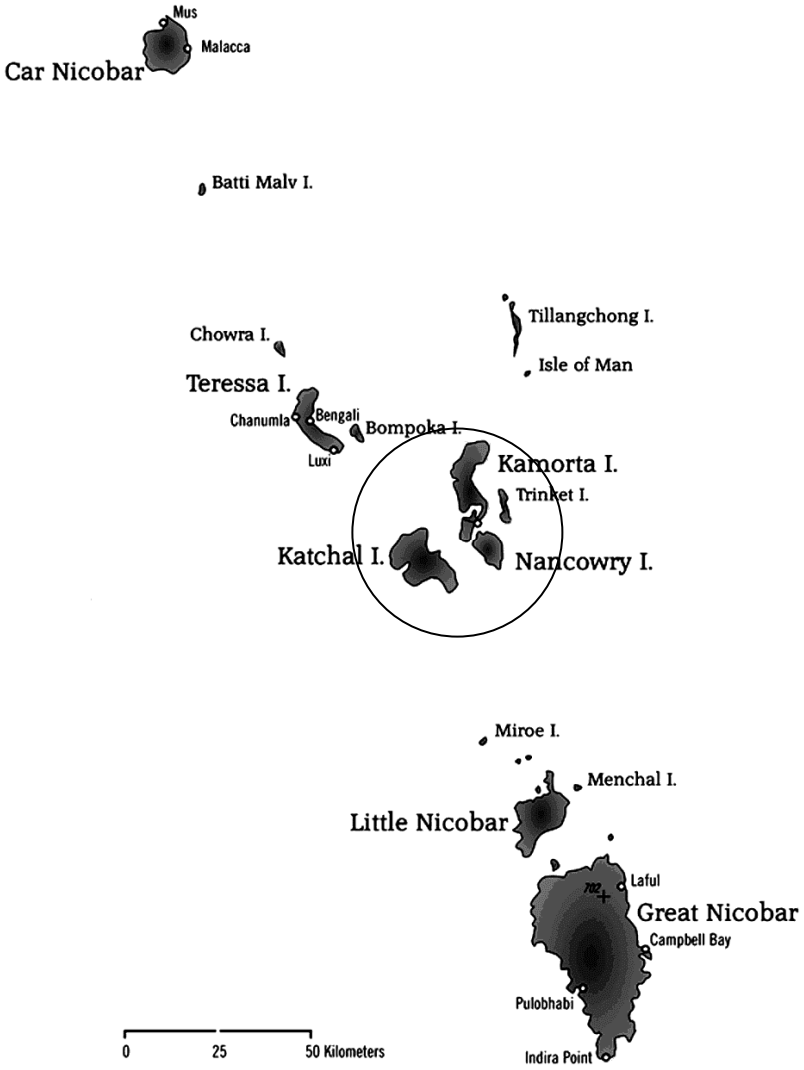
As per 2001 census, population of the four islands together is 10,083.[47] The language is said to belong to the major Austro-Asiatic family of languages through its Mon-Khmer sub-family (Lewis 2009). Morphologically it is believed to be agglutinative and syntactically of VOS pattern.[48]

[46] Lewis (2009) terms this language as Central Nicobarese and Radhakrishnan (1981) as Nancowry language. Both refer to one and the same speech habit of the Nicobarese inhabiting the four islands, Nancowry, Katchal, Kamorta and Trinket.

[47] The population figure includes a certain percentage of non-tribal population who resides in these islands due to various reasons.

[48] Detailed analysis of the language for a descriptive grammar is in progress for arriving at a definite statement.

Rajasingh, V. R.. 2011. "Demonstratives in Muöt." In Sophana Srichampa, Paul Sidwell and Ken Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.89-99.



Map: Nicobar Archipelago with Muöt area circled

3. Demonstratives

According to (Crystal 2008), 'demonstrative is a term used in grammar and semantics to refer to a class of items whose function is to point to an entity in the situation or elsewhere in a sentence.' In its further discussion it states 'depending on their grammatical role they are called demonstrative determiners or demonstrative pronouns…..Demonstratives fall within the general class of deictic expressions and are sometimes contrasted with pure indexicals.' The two sentences the work attests for illustrating the grammatical roles of demonstrative determiner and demonstrative pronoun respectively are:

That book is interesting
That is interesting

According to (Brown 2006), 'demonstratives are deictic expressions…. These indicate the relative distance of a referent in the speech situation vis-à-vis the deictic centre…. The deictic center is identified by the speaker's location at the time of utterance (p430).' It further states that 'all languages have at least two demonstratives that indicate deictic contrast…. Moreover many languages employ more than two distance terms…. Languages with more than three distance terms occur, but they are uncommon (p431).' The figure

which the work arrived at concerning the percentage of the world languages in terms of the number of demonstratives they employ, projects those with more than three demonstratives standing at 5.1 percentages (ibid). The work, in its discussion, takes into account the syntactic properties of demonstratives and also the pragmatic functions of demonstratives. As syntactic properties, it speaks about their occurrence in the pronominal, adnominal and adverbial syntactic positions. The work considers pronominal property as the substitutability of the demonstratives for full noun phrases, the adnominal property as the accompaniment of the demonstratives with the co-occurring noun and the adverbial property as the capability of the demonstratives in denoting the locational deictics. It lists the following as examples for the three said functions:

Give me *that* one
That book
Here and *there*

As pragmatic functions of demonstratives, the work tells about exophoric, anaphoric, discourse-deictic and recognitional functions (ibid).

4. Definition of demonstratives

Both (Crystal 2008) and (Brown 2006) are, implicitly, of the view that demonstratives are a class of linguistic items with morphological, syntactic and semantic functions. But, (Brown 2006) has gone a step further and has elaborated the concept of demonstrativeness to the extent of grouping languages into three types based on the number of demonstratives they posses. Accordingly, it has grouped those languages which make use of a system of two-point demonstrative as one type; those which make use of a system of three-point demonstrative as another type and those endowed with a system of more than three as the third one. It also states that syntactically demonstratives would manifest pronominal, adnominal and adverbial properties.

Muöt, the language of our discussion, appears to come under the last of the type mentioned by (Brown 2006), presumably taking its place among the 5.1 percentage of the world languages. The language exposes a four-point system of demonstratives. And, all the demonstratives are found to be determiners in their root forms which as per (Crystal 2008) can be called demonstrative determiners. They inflect for number. In the derivational morphology, they give rise to demonstrative nouns which can be ascribed to the demonstrative pronouns of (Crystal 2008). They yield adverbs also by way of derivation. Syntactically, they function as subjects and objects of sentences which can be ascribed to pronominal property of (Brown 2006). They also function as adjectival and adverbial modifiers in phrasal structures which respectively can be ascribed to the adnominal and adverbial properties of (Brown 2006).

In the light of these, the present paper would define demonstratives of Muöt as ‘a set of determiners with morphological functions of inflection and derivation coupled with the syntactic function of modifying and the semantic function of pointing to objects considering them in any one of the four points of the four-point system of relative distance measurement.’

5. Identification of demonstratives

Maintaining the definition of demonstratives arrived at above as such; the paper identifies demonstratives of Muöt diachronically.

5.1. Treatment of demonstratives in (De Röepstorff 1987)

In this classic, ‘Vocabulary of Dialects Spoken in the Nicobar and Andaman Isles’ the following lexical forms which seem related to the concept of pointing are identified. They are found along with their glosses in English.

anné	'he'
an	'it'
ánnæ	'she'
anné	'that'
aná	'there'
anné	'they'
ninné, néé	'this'

The book doesn't mention anything about their morphological or syntactic functions. However, the forms identified in it appear to reveal a two-point system of demonstrativeness with proximate and remote distances.

5.2. Treatment of demonstratives in (De Röepstorff 1884)

This monumental work 'A Dictionary of the Nancowry Dialect of the Nicobarese Language' in its introductory discussion to the grammar makes use of forms which are found related to the concept of pointing. They are discussed under three headings, the article, the demonstrative pronoun and the adverb of place.

5.2.1. The Article

Under the heading, the article, it includes, among others, the following:

ānæh
ninne or *an*

And, it states that they are demonstrative pronouns functioning as definite articles, thus manifesting one of their functions.

5.2.2. The demonstrative pronoun

Under the heading, the demonstrative pronoun, it mentions the following:

nine, nœh	'this'
an, nœh, anœh	'that'

And, it states that they are of all genders and numbers and one of the five forms, *nine* 'this' is used when the person or thing intended is near and pointed to.

5.2.3. Adverb of place

Under the heading, adverb of place, it mentions only one word,

tanæh	'there'

Although the book goes one step further to (De Röepstorff 1875) in providing the different functions of the forms identified, it also appears to reveal a two-point system of demonstratives with proximate and remote distances.

5.3. Treatment of demonstratives in (Man 1975)

This equally important work, 'A Dictionary of the Central Nicobarese Language' in its introductory discussion to grammar makes use of items which are found related to the concept of demonstrativeness. They are discussed chiefly under two headings, the article and the demonstrative pronoun.

5.3.1. The Article

Under the heading, the article, it mentions about the following:

nēe, nīna	'this'
ane	'that'

And, it states that they are demonstrative pronouns that represent definite articles.

5.3.2. The demonstrative pronoun

Here, it mentions about the same items that are listed under the heading article (cf.5.3.1) and terms them as demonstrative pronoun. They are:

nēe, nīna	'this'
ane	'that'

And it states, like adjectives, they are uninflected for number, gender or case, and they can either precede or follow the substantive with which they are employed. The work further lists the following which are substitutable for *ane* while referring to objects at some little or great distance:

ngâle	'if in northerly direction, or higher position'
ngange	'if in southerly direction'
ngâhae	'if in easterly direction'
ngaiche or *ngâshe*	'if in westerly direction or lower position'
ngaiñe	'if at or in the direction of the landing-place'

Like (De Röepstorff 1987) and (De Röepstorff 1884), this book also appears to reveal a two-point system of demonstrativeness with proximate and remote distances, but with a difference. It spells out the semantic extendibility of *ane*, the supposed remote demonstrative, to some more points of references.

5.4. Demonstratives of present day Muöt

Demonstratives of present day Muöt are determiners in their root forms. The data under analysis identify the following four forms:

nɛʔ	'this'
ʔaⁿn	'that'
ŋaⁿŋ	'that'
kəʔ	'that'

Of these, *nɛʔ* points to objects that are visible and also in close proximity to the deictic centre, while *kəʔ* to those that are invisible and also in the remotest location. *ʔaⁿn* and *ŋaⁿŋ* point to objects which are visible and are away from the deictic centre. But, the relative distance which *ŋaⁿŋ* covers from the deictic centre is greater than that covered by *ʔaⁿn*.

6. Classification of demonstratives

On the basis of the visible-invisible dichotomy and on the relative distance between the deictic centre and the object of pointing the demonstratives of present day Muöt can be glossed as,

nɛʔ	'this (proximate-visible)'
ʔaⁿn	'that (remote1-visible)'
ŋaⁿŋ	'that (remote2-visible)'
kəʔ	'that (remote-invisible)'

This can be understood clearly from the following table:

Demonstratives			
Visible			Invisible
Proximate	Remote		Remote
	Remote1	Remote2	
nɛʔ	*ʔaⁿn*	*ŋaⁿŋ*	*kəʔ*

Syntactically, they are adjectival modifiers and occur in the Det + Head phrasal structure preceding the head noun. The sentences listed below would illustrate it:[49]

1) *umpiⁿc* *nɛʔ* *ɲi:* *cəⁿ*
small prox-vis house 1-sg
'My house is small'

2) *ʈɐⁿkŋəɟən* *ʔaⁿn* *xa:ⁿs*
light rem-vis air

'Air is light'

3) *uxuv* *nen* *ŋaⁿŋ* *ʔam*
bark pa rem2-vis dog
'The dog barked'

4) *reanŋə* *kəʔ* *nɔt* *sqɟux*
run away rem-invis pig fut
'The pig will run away'

In the above sentences i.e. from 1 to 4, all the four forms *nɛʔ, ʔaⁿn, ŋaⁿŋ and kəʔ* occur in the Det + Head phrasal structures *nɛʔ ɲi: cəⁿ, ʔaⁿn xa:ⁿs, ŋaⁿŋ ʔam* and *kəʔ nɔt*[50] preceding the respective head nouns *ɲi: cəⁿ* 'my house', *xa:ⁿs* 'air', *ʔam* 'dog' and *nɔt* 'pig'.

7. Analysis of demonstratives

For analytical purposes, all the four demonstratives are looked at from their morphological, syntactic and semantic properties. Morphologically, they become inflectional as well as derivational stems in the word formation processes.

7.1. Demonstratives as inflectional stems

These demonstratives become stems to prefix the plural inflectional marker *ki-* and give rise to the respective number inflected forms, *kinɛʔ, kiʔaⁿn, kiŋaⁿŋ* and *kikəʔ*. The following would illustrate the process:

ki- + *-nɛʔ* > *kinɛʔ*
pl + prox-vis > 'pl-prox-vis'

ki- + *-ʔaⁿn* > *kiʔaⁿn*
pl + rem1-vis > 'pl-rem1-vis'

ki- + *-ŋaⁿŋ* > *kiŋaⁿŋ*
pl + rem2-vis > 'pl-rem2-vis'

[49] In sentences 1 and 2, in the absence of finite verbs, the adjectival forms *umpiⁿc* 'small' and *ʈɐⁿkŋəɟən* 'light' that form part of the verb phrase occur sentence initially.

[50] *sqɟux* at the end of the phrase is the future tense marker which is at the end of the sentence due to the free syntactic word order found in the language.

ki-	+	*-kəʔ*	>	*kikəʔ*
pl	+	rem-vis	>	'pl-rem-invis'

These number inflected forms also function as adjectival modifiers by occurring in the Det + Head phrasal structure preceding the head noun. The sentences listed below would illustrate such a function:[51]

5) *piⁿcse kinɛʔ sanex Moris*
few pl-prox-vis spear per-n
'Morris has few spears'

6) *fuksɐ kiʔaⁿn rai uɲiːxə*
fall pl-rem1-vis leaf
'The leaves fall'

7) *ɟuaŋsise nə xeh kiŋaⁿŋ sicuɑː nen*
cont sub fly pl-rem2-vis bird past
'Birds were flying'

8) *aⁿ nuaŋ kikəʔ rɯːəɨ ən ʈə ʈijəx*
two num-cl pl-rem-invis canoe 3-sg sub new
'He has two new canoes'

In sentences between 5 and 8, the number inflected forms of the four demonstratives, *kinɛʔ, kiʔaⁿn, kiŋaⁿŋ* and *kikəʔ* occur in the Det + Head phrasal structures *kinɛʔ sanex*, *kiʔaⁿn rai uɲiːxə*, *kiŋaⁿŋ sicuɑː*[52] and *kikəʔ rɯːəɨ ən*[53] preceding the respective head nouns *sanex* 'spear', *rai uɲiːxə* 'leaf', *sicuɑː* 'bird' and *rɯːəɨ ən* 'his canoe' and function as adjectival modifiers.

7. 2. Demonstratives as derivational stems

These four demonstratives become stems to affix nominal markers and adverbial markers and give rise to demonstrative nouns and demonstrative adverbs respectively.

7.2.1. Demonstratives as stems for demonstrative nouns

Two nominal suffixes are identified in the language. They are *–ɛⁿx* and *–ɛ*. Of these, the former is found occurring with *nɛʔ-* while the latter with *ʔaⁿn-*, *ŋaⁿŋ-* and *kəʔ-*. By the suffixation of the two suffixes the four demonstrative stems give rise to respective demonstrative nouns, *nɛʔɛⁿx, ʔaⁿnɛ, ŋaⁿŋɛ* and *kəʔɛ*. Such a process of derivation can be seen from the following:

nɛʔ-	+	*-ɛⁿx*	>	*nɛʔɛⁿx*
prox-vis	+	noml	>	'drd-n-prox-vis'

ʔaⁿn-	+	*-ɛ*	>	*ʔaⁿnɛ*
rem1-vis	+	noml	>	'drd-n-rem1-vis'

[51] In sentences 5 and 8, in the absence of finite verb the adjectival forms *piⁿc se* 'few' and *aⁿ nuaŋ* 'two' that form part of the verb phrase occur sentence initially.

[52] *nen* is the past tense marker which is at the end of the sentence owing to the free syntactic word order found in the language.

[53] It seems that if two adjectives occur in a phrase, according to the pattern found in the language, the numeral adjective would precede the other in the sentence initial position and the other would follow the numeral immediately or distantly. The occurrence of the descriptive adjective *ʈijəx* 'new' at the sentence end is due to the free syntactic word order found in the language. The form *ʈə* found preceding *ʈijəx*, this paper has, for the time being, taken as a subordination marker found in the language to subordinate the adjectival form with the head noun, because each aspect of either the noun or the verb are found to be subordinated with the head noun or head verb with a subordinator.

ŋaⁿŋ-	+	*-ɛ*	>	*ŋaⁿŋɛ*
rem2-vis	+	noml	>	‘drd-n-rem2-vis’

kəʔ-	+	*-ɛ*	>	*kəʔɛ*
rem-invis	+	noml	>	‘drd-n-rem-invis’

The demonstrative nouns derived thus, function as subjects of sentences. Such a function can be seen from the following sentences:[54]

9) *ɲi:* *mɛⁿ* *nɛʔɛⁿx*
house 2-sg drd-n-prox-vis
‘This is your house’

10) *ɲi:* *cəⁿ* *ʔaⁿnɛ*
house 1-sg drd-n-rem1-vis
‘That is my house’

11) *cukɛ* *əⁿn* *ŋaⁿŋɛ*
basket 3-sg drd-n-rem2-vis
‘That is his basket’

12) *karuʔ* *kəʔ* *ɲi:* *kəʔɛ*
big rem-invis house drd-n-rem-invis
‘That is a big house’

In the sentences between 9 and 12, the demonstrative nouns *nɛʔɛⁿx, ʔaⁿnɛ, ŋaⁿŋɛ* and *kəʔɛ* function as subject of the respective sentences. Besides, in sentence 12 the root form *kəʔ* function as an adjectival modifier also by occurring in the Det + Head phrasal structure.

In addition to the subjective function, it is seen that all these four derived demonstrative nouns function as objects of sentences. The sentences that follow would illustrate such a function.

13) *xinɔⁿxŋə* *nɛʔɛⁿx*
break drd-n-prox-vis
‘Break this’

14) *xinɔⁿxŋə* *ʔaⁿnɛ*
break drd-n-rem1-vis
‘Break that’

15) *xɑʔɔːʈəse* *ŋaⁿŋɛ*
call drd-n-rem2-vis
‘Call that’

16) *xiɲjɐŋ* *kəʔ* *kəʔɛ* *kəʔ* *ʔəⁿn* *nen*
chase rem-invis drd-n-rem-invis rem-invis 3-sg pa
‘He chased that’

Sentences 13 to 15 are imperatives unlike the 16th. But in all the cases the demonstrative nouns *nɛʔɛⁿx, ʔaⁿnɛ, ŋaⁿŋɛ* and *kəʔɛ* manifest their objective function. Apart from these, in sentence 16 the root form *kəʔ* function as adjectival modifiers also by occurring in the Det + Head phrasal structure in two instances.

[54] In all the four sentences, in the absence of finite verb the adjectival forms *ɲi:* mɛⁿ ‘your house’, *ɲi:* cəⁿ ‘my house’, *cukɛ əⁿn* ‘his house’ and *karuʔ kəʔ ɲi:* ‘that house’ that form part of the verb phrase occur sentence initially.

Besides functioning as subjects and objects of sentences, all these derived demonstrative nouns function as adjectival modifiers also by occurring in the Det + Head + Adj phrasal structure following the head noun. The sentences of the following can be taken for illustrating such a syntactic function:[55]

17) *sanaːⁿvəɟən* *nɛʔ* *ɟuaŋuɲiːxə* *nɛʔɛⁿx*
eatable prox-vis fruit drd-n-prox-vis
'This fruit is eatable'

18) *ɟanopəɟən* *ʔaⁿn* *reak* *ʔaⁿnɛ*
drinkable rem1-vis water drd-n-rem1-vis
'That water is drinkable'

19) *cɔŋ* *ŋaⁿŋ* *uɲiːxə* *ŋaⁿŋɛ*
tall rem2-vis tree drd-n-rem2-vis
'That tree is tall'

20) *leaⁿŋŋəse* *kəʔ* *kiɲɔⁿm* *kəʔɛ* *inkanə*[56]
lean rem-invis girl drd-n-rem-invis fem
'That girl is lean'

It can be seen that in the sentences between 17 and 20 all the four derived demonstrative nouns *nɛʔɛⁿx, ʔaⁿnɛ, ŋaⁿŋɛ* and *kəʔɛ* function as adjectival modifiers by occurring in the Det + Head + Adj phrasal structures *nɛʔ ɟuaŋ uɲiːxə nɛʔɛⁿx* 'this fruit', *ʔaⁿn reak ʔaⁿnɛ* 'that water', *ŋaⁿŋ uɲiːxə ŋaⁿŋɛ* 'that tree', and *kəʔ kiɲɔⁿm kəʔɛ inkanə* 'that girl' respectively. In all these instances the demonstrative derived nouns follow the head noun.

7.2.2. Demonstratives as stems for demonstrative adverb

Of the four demonstratives, *nɛʔ* becomes stem for the prefixation of the adverbial marker *niː-* and gives rise to the derived adverb *niːnɛʔ*. Such a process can be illustrated by the following:

niː- + *-nɛʔ* > *niːnɛʔ*
adv + prox-vis > 'drd-adv'

Syntactically, this derived adverb would function as locational adverb being in the verb phrase. The following sentence can be taken for illustrating such a function:[57]

21) *niːnɛʔ* *ufɛː* *nɛʔ* *koːən* *əⁿn* *inkɔŋə*
drd-adv pl prox-vis son 3-sg mas
'His sons are here'

As seen in 21, the derived adverb *niːnɛʔ* denote the location of the object.

7.2.3. Demonstrative noun as stem for demonstrative adverb

Of the four demonstrative nouns derived by means of the process discussed in 7.2.1., *ŋaⁿŋɛ* functions as stem for the derivation of a demonstrative adverb *ɟəŋaⁿŋɛ* 'there' by prefixing the adverbial marker *ɟə-*.[58] The following would illustrate such a process:

[55] In the absence of finite verb in all the four sentences the adjectival forms *sanaːⁿvəɟən* 'eatable', *ɟanopəɟən* 'drinkable', *cɔŋ* 'tall' and *leaⁿŋŋəse* 'lean' occur sentence initially.

[56] According to the typology found in the language, if the head noun in an adjectival phrase happened to be of human gender the gender indicator occurs at the end of the phrasal structure.

[57] In the absence of finite verb, the adverbial form occurs sentence initially.

ʈə-	+	*-ŋa^{n}ŋɛ*	>	*ʈəŋa^{n}ŋɛ*
adv	+	drd-n-rem2-vis	>	there

Syntactically, this derived adverb functions as adverbial modifier by occurring in the Head + Adv phrasal structure following the main verb. The following sentence can be taken for the illustration of such a function:

22) *ukɯ^{n}* *ʈəŋa^{n}ŋɛ* *ʔa^{n}n* *kiɲɔ^{n}m*
sit drd-adv rem1-vis child
'The child sits there'

As is seen, in sentence 22 the derived demonstrative adverb *ʈəŋa^{n}ŋɛ* 'there' occurs in the Head + Adv phrasal structure following the main verb *ukɯ^{n}* 'sit' and functions as the adverbial modifier.

8. Findings

1) Muöt of modern day exhibits a four-point distance marking demonstrative system while those of the earlier periods appear to exhibit a demonstrative system of two-point distance marking.

2) Unlike the earlier works, the present paper views demonstratives as demonstrative determiners in their root forms which then give rise to demonstrative nouns and demonstrative adverbs by way of derivation.

3) The demonstrative forms identified in (De Röepstorff 1987) appear pointing nouns in the third person also. The data of present day Muöt do not exhibit such a function.

4) (Man 1975) states that the demonstratives are uninflected for number (cf.5.3.2), while (De Röepstorff 1884) states that the demonstratives are of all numbers (cf.5.2.2). But the present paper treats number inflection as the morphological function of demonstratives and that too while occurring in the Det + Head phrasal structure.

5) Man (1975) states that demonstratives can either precede or follow the substantive with which they are employed (cf.5.3.2). But, this paper holds that demonstratives precede the substantive when they are demonstrative determiners and follow the same when they are demonstrative nouns.

6) It seems what (De Röepstorff 1884) terms as article (cf.5.2.1), demonstrative pronouns (cf.5.2.2), and adverb of place (cf.5.2.3) is a conglomeration of what the present paper views as demonstrative determiner, derived demonstrative nouns and derived demonstrative adverbs. The form *an* that has been listed as an article and as a demonstrative pronoun is being dealt with as a demonstrative determiner with the meaning, rem1-vis. Likewise the form *ānæh* which has been listed as an article and *nœh* and *anœh* which have been listed as demonstrative pronouns are dealt with as derived demonstrative nouns with the following meanings:

ānæh	'rem1-vis'
nœh	'prox-vis'
anœh	'rem1-vis'

7) But the forms *ninne* which has been listed both as article and demonstrative pronoun and *tanæh* which has been listed as adverb of place are dealt with as derived demonstrative adverbs with the meaning, prox-vis.

8) What (Man 1975) terms as article (cf.5.3.1) and demonstrative pronoun (cf.5.3.2) seems to be derived demonstrative nouns and derived demonstrative adverbs for the present paper. The forms *nēe* and *ane* which are listed both as article and demonstrative pronoun are treated as derived demonstrative nouns with the meaning, prox-vis and rem1-vis respectively. In the same way the form *nīna* which has been listed both as article and demonstrative pronoun is being treated as derived demonstrative adverb with the meaning, prox-vis.

[58] Though (De Röepstorff 1884) mentions *tanœh* 'there' in 5.2.3, it does not speak about the derivational process which this paper is concerned about.

9) Although (Man 1975) mentions *ngange* under the sub heading demonstrative pronoun (cf.5.3.2), it speaks only about its substitutability to *ane* when it is in the adverbial function of denoting southern direction. Modern Muöt employs *ŋaⁿŋɛ* as the derived demonstrative noun with the meaning, remote2-vis and *ʔaⁿnɛ* as the derived demonstrative noun with the meaning, remote1-vis. And for marking the adverbial function of denoting southern direction, modern Muöt makes use of the form *ʔafəŋ* as shown in the following sentence:

jɔ	*kaji:ŋə*	*cəⁿ*	*ʔafəŋ*	*kempəlpe:*
fut	go	I	toward south	Campbell Bay

'I go to Campbell Bay'

References

Brown Keith 2006. *Encyclopedia of Language and Linguistics* (second edition Vol.III.) Elsevier.

Crystal David 2008. *A Dictionary of Linguistics and Phonetics* (sixth edition). Blackwell Publishers.

De Röepstorff Frederick Adolph 1987 (1875). *Vocabulary of Dialects spoken in the Nicobar and Andaman Isles* (second edition) Asian Educational Services New Delhi, Madras.

_____. 1884. *A Dictionary of the Nancowry Dialect of the Nicobarese Language*. Home Department Press, Calcutta.

Gerard Diffloth 1974. Austro-Asiatic Languages. *Encyclopedia of Britannica*, Volume 2.

Justin, Anstice 1990. *The Nicobarese*. Seagull Books, Calcutta.

Kloss Boden C. 1971 (1902). *Andamans and Nicobars*. Vivek Publishing House, 18-D, Kamla Nagar, Delhi-7 (India).

Lewis, M. Paul (ed.), 2009, *Ethnologue: Languages of the world, Sixteenth edition*. Dallas: SIL International

Man, Edward Horace 1975 (1889). *A Dictionary of the Central Nicobarese Language*. Sanskaran Prakashak, Delhi.

Radhakrishnan, R. 1981. *The Nancowry word, Phonology, Affixal morphology and Roots of A Nicobarese Language.* Linguistic Research Inc. P.O. Box 5677, Station 'L' Edmonton, Alberta, Canada T6C 4G1.

Rajasingh, V.R. 2007. *Monograph on Language situation of the Nicobars, Structure of Muöt, Structure of Saněnyö and Dictionary of Nicobarese–Hindi-English* (unpublished report submitted to the A & N Administration). Central Institute of Indian Languages, Mysore & Andaman and Nicobar Administration, Port Blair.

Symbols & Abbreviations

>	becomes	mas	masculine gender
+	join together	noml	nominal marker
adj	adjective	num-cl	numeral classifier
adv	adverbial marker	pa	past tense marker
cont	continuous marker	per-n	personal noun
dem	demonstrative	pl	plural marker
det	determiner	prox	proximate
drd-adv	derived adverb	sub	subordination marker
drd-n	derived noun	vis	visible
fem	feminine gender	1-sg	first person singular
fut	future tense marker	2-sg	second person singular
invis	invisible	3-sg	third person singular
loc	locative marker		

Katuic-Bahnaric: Austroasiatic Sub-Family or Convergence Area?[59]

Paul Sidwell
Center for Research in Computational Linguistics (Bangkok),
Australian National University (Canberra)

1. Introduction

Consensus among scholars holds that the Austroasiatic languages form a dozen or 13 branch level groupings that (with or without intermediate nested branching) descend from proto-Austroasiatic. The identification of these branches, and hypotheses regarding how they might coordinate withing the Austroasiatic phylum, have largely been advanced by lexicostatistical studies, in the face of widely held views that a valid classification should be based upon at least a detailed phonological and lexical reconstruction. But progress in reconstruction of proto-Austroasiatic has been sporadic to say the least, and for the past five decades scholars have variously failed (or refused) to discuss the issue of nested branching within the terms of an historical reconstruction.

Intriguingly one of the consistent findings lexicostatistics is that the Katuic and Bahnaric branches, located adjacent to one another in central Indo-China, appear to be much closer to each other than to any other Austroasiatic branch. Even more recent phylogenetic computational studies (reported below) also consistently point to a special lexical relationship between Katuic and Bahnaric. Not withstanding these results, it is also apparent that there is a lack of support from comparative historical reconstruction to demonstrate a real nested branching relation between these two groups, and no serious suggestion of a Katuic-Bahnaric sub-family has been advanced.

Is it not a contradiction that the findings of lexicostatistics are accepted in relation to the rest of the phylum, but rejected – without explicit discussion – in respect of Katuic-Bahnaric? It would appear that the lexical studies may be telling us more about geography and contact than geneology. This was suggested by Franklin Huffman (1978), who proposed that Austroasiatic speakers radiated from a homeland located approximately in the present Katuic-Bahnaric speaking area. The model implies that the two branches in question have enjoyed a prolonged period of contact and mutual influence, thus moderating their divergence.

2. Lexicostatistics

Southeast Asian linguistics has frequently embraced the controversial lexicostatistical method, as a crude but useful huristic. The isolating typology typical of the area not only renders lexiocstistics reletively easy to apply, it also makes other classificatory methods more problematic. In this paper I will not argue the merits of lexicostatistics for deriving a classification, but I am suggesting that particular consistent features of the results of such studies do point to real linguistic phenomenon, the investigation of which can usefully contribute to our understanding of linguistic prehistory.

[59] Work on this paper was made possible by support from the National Endowment for the Humanities (Washington). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily represent those of the National Endowment for the Humanities.

Sidwell, Paul. 2011. "Katuic-Bahnaric: Austroasiatic Sub-Family or Convergence Area?." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.100-111.

From the mid 1960s onward a series of lexicostatistical studies (e.g. Thomas 1966, Thomas & Headley 1970, Headley 1976, Huffman 1978, Smith 1981, Peiros 1998)[60] consistently delineated a dozen Austroasiatic branches, neatly sorting out the earlier classificatory chaos that had treated scores of languages as a huge dialect continuum stretching from the lower Mekong to Yunnan (see, for example, Lebar et al. 1964). Yet at the same time, these various studies gave contradictory and/or weak indications of nested sub-grouping among the branches.

In fact, despite their various conclusions, all lexicostatistical studies have pointed to a consistent pattern that Huffman discussed in his presentation at the second ICAAL meeting (held in Mysore, India, 1978). That pattern is seen in the lexicostatistical matrix, reproduced here at Figure 1.[61] Huffman points out that the highest interbranch percentages are between Katuic and Bahnaric, plus there is a strong tendency for other branches to show a higher percentage between themselves and Katuic-Bahnaric than with other branches; this effect is clearly seen by looking along the top row, or down the leftmost column in his matrix. A point raised by Huffman is that the lexical relations suggested by these figures are internally contradictory, and thus difficult to represent in a cladogram.

Chart 1: Interbranch Cognate Percentages and Averages

	K-B	Kme	Mon	Pal	V-M	Pea	Asl	Kmu	Kha	Mun	Nic
Katuic-Bahnaric		47	42	34	33	35	31	32	24	27	20
Khmer	47		33	31	30	39	28	25	24	24	23
Mon	42	33		33	30	22	24	28	22	20	19
Palaungic	34	31	33		26	25	22	26	26	18	23
Viet-Muong	33	30	30	26		24	22	24	23	20	18
Pearic	35	39	22	25	24		24	22	17	19	16
Aslian	30	28	24	22	22	24		24	20	17	19
Khmuic	32	25	28	26	24	22	24		24	20	17
Khasi	24	24	22	26	23	17	20	24		18	14
Munda	27	24	20	18	20	19	17	20	18		17
Nicobarese	20	23	19	23	18	16	19	17	14	17	
Interbranch totals	324	304	273	264	250	243	231	222	212	200	186
Interbranch averages	32.4	30.4	27.3	26.4	25.0	24.3	23.1	22.2	21.2	20.0	18.6

Figure 1: *Austroasiatic lexicostatistical matrix by Huffman (1978)*

Smith also conducted several lexicostatistical studies (1974, 1978, 1981), culminating in an analysis of 45 languages, reflecting 9 branches. Figure 2 reproduces his summary results redrawn for clarity here.

	Bah.	Kat.	Pearic	Khm.	Mon	Sen.	Vietic	Khmu	Pal.
Bahnaric		24-39	24-31	24-36	26-36	30-38	21-35	20-26	19-25
Katuic	24-39		20-34	22-32	23-32	31-37	17-32	20-28	16-26
Pearic	24-31	20-34		33	24	28	22-24	23	23
Khmer	24-36	22-32	33		30	26	22-23	21	18
Monic	26-36	23-32	24	30		31	23-24	21	19
Senoic	30-38	31-37	28	26	31		22-25	26	29
Vietic	21-35	17-32	22-24	22-23	23-24	22-25		24	19
Khmuic	20-26	20-28	23	21	21	26	24		21
Palaungic	19-25	16-26	23	18	19	29	19	21	

Figure 2: *Austroasiatic lexicostatistical matrix by Smith (1981:202), redrawn for clarity (note that Senoic represents Aslian)*

Where Smith used more than one language per branch there are two figures in the relevant fields indicating the range of percentages counted. With the exception of Senoic (Aslian), where the figure is

[60] One may add several unpublished trials conducted by Shorto, discussed briefly in Sidwell (2009).

[61] Huffman's paper was not formally published, but was circulated widely informally. The primary data, and some related analyses, are available online at http://sealang.net/archives/huffman/.

skewed upwards to some extent by the use of a short list,[62] the same pattern discussed by Huffman is evident. Although interbranch percentages generally decline as one moves geographically away from the Katuic-Bahnaric area, anomalously high percentages are shared with Katuic-Bahnaric.

And a similar general pattern is apparent also in Peiros' (1998) lexicostatistics (see Figure 3).

	Jeh	Bah	Chr	Kui	Smi	Mon	Nya	Vie	Ruc	Wa	Dea	Kmu	Kml	Kmr	Kha	Mun
Jeh	x	59	54	40	34	28	28	32	33	27	22	25	25	22	23	25
Bahnar	59	x	55	41	34	30	31	29	30	24	20	26	26	24	20	27
Chrau	54	55	x	43	32	37	37	30	35	30	25	30	33	27	26	27
Kui	40	41	43	x	33	28	33	30	32	27	23	27	28	29	23	26
Semai	34	34	32	33	x	27	29	24	24	24	24	22	26	22	19	16
Mon	28	30	37	28	27	x	73	26	26	28	24	25	26	24	22	24
Nyakur	28	31	37	33	29	73	x	24	26	31	28	23	30	24	24	24
Vietnamese	32	29	30	30	24	26	24	x	58	20	22	27	21	19	22	24
Ruc	33	30	35	32	24	26	26	58	x	26	25	28	25	23	26	29
Wa	27	24	30	27	24	28	31	20	26	x	51	32	34	22	23	21
Deang	22	20	25	23	24	24	28	22	25	51	x	31	26	23	21	19
Khmu	25	26	30	27	22	25	23	27	28	32	31	x	40	19	22	21
Ksinmul	25	26	33	28	26	26	30	21	25	34	26	40	x	17	21	21
Khmer	22	24	27	29	22	24	24	19	23	22	23	19	17	x	23	17
Khasi	23	20	26	23	19	22	24	22	26	23	21	22	21	23	x	23
Mundari	25	27	27	26	16	24	24	24	29	21	19	21	21	17	23	x

Figure 3: Austroasiatic lexicostatistical matrix by Peiros (1998)

	Katuic			Bahnaric			Khm.		Monic		Pal.		Viet.		Pearic		Aslian			Khmu.		Khasic		Mund.		Nico.	
Bru		57	57	41	40	43	23	24	28	31	19	26	27	30	30	34	29	28	20	27	31	24	19	25	23	17	21
Nge'	57		61	38	36	38	27	26	30	33	22	26	23	29	30	34	25	31	23	26	31	24	21	24	19	18	20
Katu	57	61		43	39	40	26	26	29	30	22	26	25	27	30	34	28	28	20	24	28	23	17	22	19	16	19
Bahnar	41	38	43		52	51	24	22	23	26	19	20	24	26	30	34	29	30	18	20	22	21	13	17	17	13	19
Jeh	40	36	39	52		55	22	23	23	24	24	26	22	27	27	27	31	31	23	21	27	23	16	20	18	14	18
Jru'	43	38	40	51	55		27	26	27	29	17	21	25	30	29	33	27	30	21	21	22	22	16	23	18	15	19
Surin	23	27	26	24	22	27		89	24	24	17	17	14	19	22	30	22	21	15	18	21	20	17	16	12	12	14
Cambodian	24	26	26	22	23	26	89		23	23	18	18	12	20	20	28	20	19	17	19	21	21	17	16	13	13	15
Mon	28	30	29	23	23	27	24	23		84	23	23	22	23	23	28	27	23	18	21	21	21	16	24	21	15	18
Nyahkur	31	33	30	26	24	29	24	23	84		24	26	22	23	23	28	27	24	19	23	24	20	15	23	20	15	17
De'ang	19	22	22	19	24	17	17	18	23	24		53	15	18	16	19	20	14	14	26	20	23	17	15	13	12	14
Wa	26	26	26	20	26	21	17	18	23	26	53		16	19	22	27	21	18	16	26	22	20	16	17	14	13	18
Muong	27	23	25	24	22	25	14	14	22	22	15	16		50	15	18	20	16	14	17	18	22	14	21	21	11	14
Ruc	30	29	27	26	27	30	19	20	23	23	18	19	50		17	20	20	18	14	22	23	21	14	23	24	14	17
Chong	30	30	30	30	27	29	22	20	23	23	16	22	15	17		61	22	22	11	21	19	19	16	18	11	14	18
Pear	34	34	34	34	27	33	30	28	28	28	19	27	18	20	61		27	29	18	25	25	20	18	20	15	16	22
Semai	29	25	28	29	31	27	22	20	27	27	20	21	20	20	22	27		38	36	21	22	20	17	16	17	14	19
Semelai	28	31	28	30	31	30	21	19	23	24	14	18	16	18	22	29	38		28	16	20	15	14	16	13	11	16
Jahai	20	23	20	18	23	21	15	17	18	19	14	16	14	14	11	18	36	28		15	17	16	10	11	12	12	14
Khmu'	27	26	24	20	21	21	18	19	21	23	26	26	17	22	21	25	21	16	15		35	22	16	18	20	12	14
Mal	31	31	28	22	27	22	21	21	21	24	20	22	18	23	19	25	22	20	17	35		26	24	22	20	12	13
Khasi	24	24	23	21	23	22	20	21	21	20	23	20	22	21	19	20	20	15	16	22	26		60	18	14	13	16
War	19	21	17	13	16	16	17	17	16	15	17	16	14	14	16	18	17	14	10	16	24	60		13	10	8	10
Mundari	25	24	22	17	20	23	16	16	24	23	15	17	21	23	18	20	16	16	11	18	22	18	13		37	10	15
Sora	23	19	19	17	18	18	12	13	21	20	13	14	21	24	11	15	17	13	12	20	20	14	10	37		8	12
Car	17	18	16	13	14	15	12	13	15	15	12	13	11	14	14	16	14	11	12	12	12	13	8	10	8		49
Nancowri	21	20	19	19	18	19	14	15	18	17	14	18	14	17	18	22	19	16	14	14	13	16	10	16	12	49	

Figure 4: *A lexicostatistical matrix of 27 Austroasiatic languages (Sidwell ms. trial conducted April 2009).*

[62] Probably this is not the only explanation: there is a tradition, going back to Skeats & Blagden (1906) of crediting Mon and Khmer influence on Aslian languages.

Intrigued by these results, I conducted several lexicostatistical trials, the results of which were discussed at the 2009 ICAAL meeting in Bangkok.[63] A matrix of results for 27 Austroasiatic languages (variously 2 or 3 languages per branch) is reproduced here at Figure 4. The languages are arranged in the same order as in Huffman's presentation, the difference being that figures for specific language pairs are given rather than only inter-branch averages.

Remarkably, although my cognate counts are generally somewhat lower than other's, perhaps because I eliminated various previously unidentified loans from the data, the figures still broadly follow the same overall pattern: shared percentages between Katuic and Bahnaric peak in the low 40s, while other groups tend to show roughly 5~10% higher agreement with Katuic-Bahnaric than with their immediate neighbours or otherwise closest relatives after consideration of Katuic-Bahnaric. Perhaps the most striking example of this is Munda. It is quite striking that the Munda languages in the sample are counted as showing between 17% and 25% cognates with Katuic-Bahnaric; this is not only the highest percentage shared with Munda, but it is shared with branches that are geographically furthest from Munda (specifically 25% between Bru and Mundari).

Analyses

The above scholars have interpreted these broadly consistent results in very different ways. Huffman did not attempt to draw a tree, but did offer this commentary:

> Now what one would logically expect, in the abstract, is a chain effect whereby Munda, for example, would be most closely related to one of its closest neighbors, such as Khasi or Nicobarese, then to Palaungic, then to Mon, and so on in an easterly direction; what we in fact have, on the contrary, is that Munda is most closely related to Katuic-Bahnaric, next to Khmer, next to Mon, and so on in a westwardly direction, and shows its lowest cognate percentages with its closest geographical neighbors, Khasi and Nicobarese! This would seem to argue for an eastern (Central Vietnam) center of dispersal and a separate westward migration for each branch of Austroasiatic.
> (Huffman 1978:5)

So Huffman suggested that the lexical relations implied by these figures are perhaps indicative of the dispersal pattern of Austroasiatic, although he did not go into specific details behind this reasoning. Smith dealt with the contradictory indications by drawing a flat tree - with branches radiating equally from a common root - reproduced here at Figure 5. He did not sub-group Katuic and Bahnaric, having taken the lowest percentages within each range as more indicative of the likely relations.

[63] All my data and analyses can be accessed at: people.anu.edu.au/~u9907217/lexico/AAclassification.html.

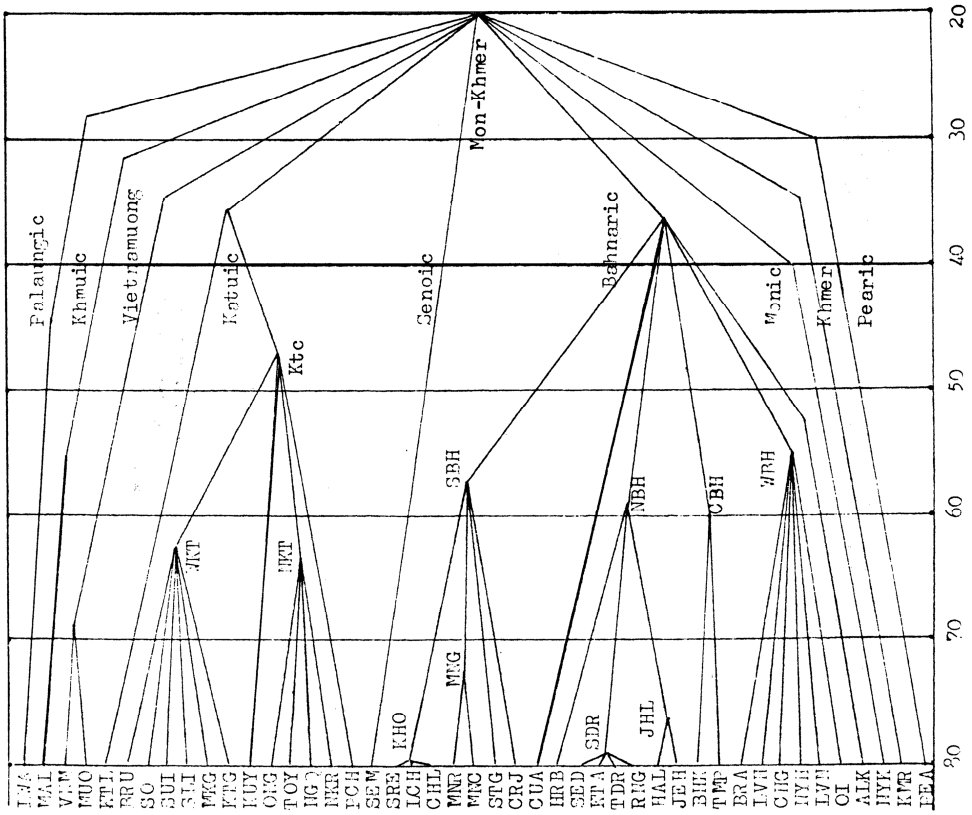


Figure 5: *Austroasiatic(Mon-Khmer) tree by Smith (1981)*
(apologies for poor image quality)

Peiros, in marked contrast, has interpreted the figure by sticking to a grouping algorithm that genetates nested branching relations (by use of the Starling software package). In both his 1998 and larger 2004 studies (the latter is not accompanied by a matrix of percentages) Peiros draws a strongly nested tree with Khasi, Munda, and Nicobarese respectively stemming from the highest branching nodes. Figure 6 reproduces his 2004 tree.

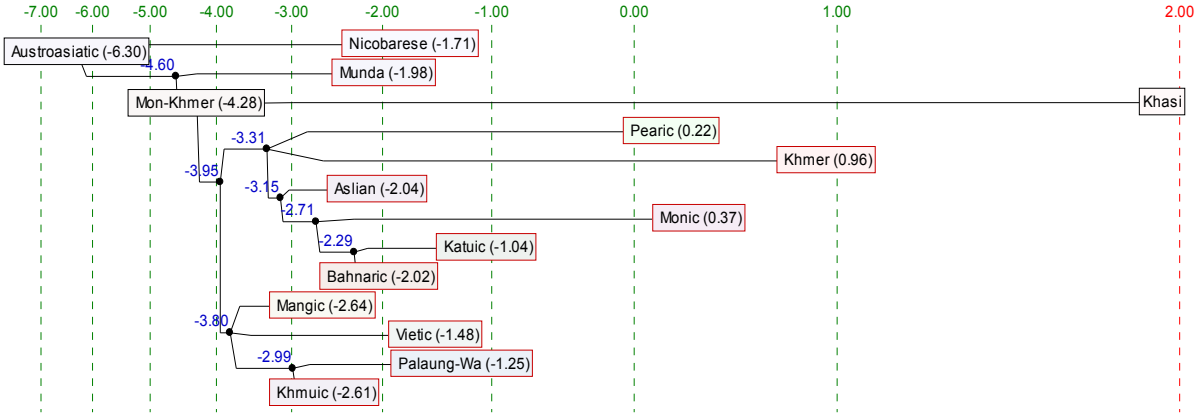


Figure 6: *Austroasiatic tree by Peiros (2004)*

All of the above studies, including my own, confirm the "Huffman Effect" as we may call it, but interpreatations vary. Superficially all the relevant studies suggesta a single Katuic-Bahnaric branch, plus

some additional mysterious factor or factors such that cognate scores generally are higher with Katuic-Bahnaric.

It is generally held that two main factors affect lexicostatical results are: 1) unrecognised borrowing, and 2) variation in rates of replacement. The first of these is probably not entirely eliminated, given the differences in cognate counts between various studies, and the insufficiently developed state of Austroasiatic comparative linguistics generally. Remarkably though, the Huffman Effect remains prominent in all relevant studies. On the other hand, the effects of varying rates of change are more problematic. These are neatly summarized by Holm (2003) as the 'Proportionality Trap': if there are three languages, A, B, C, separating at the same time, but C changes at a faster rate, A and B will retain more in common, and appear to form a sub-group. Similarly, if A and B exchange words, or borrow the same words from another source, the effect will the same, all other things being equal. Consequently, we have the real possibility that the anomalous patterns of inter-branch percentages discussed above are related to differential rates of change. The present question then, is how best to explain the very high cognate scores counted between Katuic and Bahnaric? What do they really reflect?

3. Katuic and Bahnaric innovations

It is evident from comparative historical reconstruction that Katuic and Bahanric do not form a single branch within Austroasiatic, whereas scholars are these days confident that each can be demonstrated to be a coordinate branch in its own right. Following on the heals of some 4 decades of comparative work, we now know a considerable amount about the history of Katuic and Bahnaric (e.g. Blood 1966, Diffloth 1982, Efimov 1990, Efimov, 1983, Gainey 1985, Peiros 1996, Shorto 2006, Shorto ms., Sidwell & Jacq 2003, Sidwell 1999, 2000, 2005, Smith 1972, Theraphan L-Thongkum 2001, Thomas & Smith 1967, Thomas 1967). Thanks to these works, one can compile a set of lexical and phonological criteria that are diagnostic of Katuic and Bahnaric respectively.

3.1 Katuic innovations

Katuic languages consistently reflect a number of lexical innovations; including a specific subset of numerals, e.g.:

**kdial, *k(m)paj* 'wife'	displaced **kdɔːr* (Shorto #1628) which shifted to mean 'young girl'
**trɨa* 'mushroom'	replaced **ptiːs* (Shorto #1903) (also discussed above)
**plɔː* 'head'	replaced **kuːj* (Shorto #1443), which shifted to 'person'; generally 'head' is unstable throughout AA, compare Bahnaric **gɨl, *ɓoːk*, the latter possibly continuing an AA etymon.
**kmɔː* 'year'	replaced **cnam*, which is retained in Bahnaric, although **kmɔː* was borrowed into West Bahnaric
**sɔːj* 'tail'	replaced **staʔ* (Shorto #73), Bahnaric innovated **tiəŋ*, although **sɔːj* was borrowed into West Bahnaric.
**bntuor* 'star'	loaned into West Bahnaric as **ptɔːr*; Katuic retains AA *smaŋ in the meaning 'small star'. Bahnar retains *smaŋ, plus an innovative form **s(n)lɔŋ*
**ktiak* 'earth'	replaced **tiʔ* (Shorto #63), Bahnaric languages have variously *tɛh ~ tnɛh*, which look like a confusion of Cham *tanuuh* and Pali *padesa* via Khmer
**tbat* 'six'	replaced **tpraw* (Shorto #1851) which is continued in Bahnaric, e.g. Bahnar *tədrəw*
**tgɔːl* 'eight'	replaced **thaːm* (Shorto #1431) which is continued in Bahnaric, e.g. Bahnar *təhŋaːm*. Katuic is possibly cognate with Khmuic **guːl* 'seven', Pearic **gnuːl* 'seven'
**tgias* 'nine'	replaced **dciːn[ʔ]* (Shorto #1144) which continued in Bahnaric e.g. Bahnar *sin*. Appears to be present in Mlabri (Khmuic) *gaijh* 'nine'

The significance of the Katuic numerals given that 'eight' and 'nine' is ambiguous, more are the forms for the number 'ten', for which there is a Katuic-Bahnaric isogloss:

pKatuic	pBahnaric
**ɟit*	**ɟət* (but West Bahnaric **cit*)

Curiously the vowels do not quite agree, and the West Bahnaric form looks suspiciously like a more recent borrowing from one of the various Katuic languages that have devoiced initial stops (e.g. Ong, Bru etc. *cit* 'ten'). This is not so surprising, given that West Bahnaric borrowed **səːŋ* 'five' from Katuic, and by coincidence South Bahnaric borrowed **pram* 'five' from Khmer. In fact, the presence of a central vowel after a palatal initial is somewhat marked in Bahnaric (see discussion in Sidwell 1999) but not so in Katuic, including even reflexes of 'ten' (e.g. Kui *ɲcət*). Given that Bahnaric is otherwise more conservative in its numerals, with the exception of a couple of well known but isolated loans, one can suggest that Katuic **ɟit* is an innovation, some of whose daughter reflexes diffused through Bahnaric displacing an earlier term (perhaps the etymon reflected in, e.g. Palaungic **kəəl*, Khmu *gal*, Munda *gɛl* etc. 'ten').

Katuic historical phonology is quite distinctive; Sidwell (2005) reconstructs an implosive palatal stop **ƒ*, which has clearly distinguished reflexes in all Katuic subgroups, and is well represented in Austroasiatic vocabulary, e.g.: **ƒiiŋ* 'foot, leg', **-ƒoŋ* 'high, tall', **ʃuut* 'wipe, rub', **ƒuaʔ* 'sour', **ƒɔh* 'peck, strike (as snake)' and others. Interestingly, other Austroasiatic branches appear to unambiguously indicate a plain voiced palatal stop /ɟ/, so this is either a split we do not understand, or an archaism in Katuic. Also, there are the six proto-Katuic diphthongs *(*iə, *ɨə, *uə, *ie, *ɨa, *uo*); these frequently correspond to monophthongs elsewhere in Austroasiatic, where cognates can be found, such that they appear to be phonologically innovative; a few examples:

**ʔasuom* 'shrimp'	< pAA **suːm* (Shorto #1419a)
**huor* 'to singe'	< pAA **hur* (Shorto #1685)
**pɨah* 'to launder'	< pAA **piːh* (Shorto #2021)
**mɨat* 'vulture'	< pAA **tmaːt[]* (Shorto #1051)
**ɗɨəs* 'break'	< pAA **ɗaːs* (Shorto #1903a)

3.2 Bahnaric innovations

Innovations that mark the Bahnaric branch are not so straightforward. Bahnaric is more diverse than Katuic, and it is difficult to find many specific lexemes that appear to be both innovative and restricted to all and only Bahnaric languages. Perhaps the single best example is Bahnaric **lpiət* 'tongue', which replaced pAA **l(n)taːk* (Shorto #320), apparently by infixation of **liət* 'to lick'. There are a couple other - rather well known - Bahnaric lexical innovations but they are somewhat problematic.

**ktsɨːŋ* 'bone'	replaced **cʔaːŋ* (Shorto #488), although it may well be a loan from Vietic **tʃɨəŋ* (as reconstructed by Ferlus 2007) or both Vietic and Bahnaric may have borrowed from an unknown source.
**ʔuɲ* 'fire'	replaced **[ ]ʔu(ː)s* (Shorto #1872), perhaps by sandhi or expressive deformation of the final. However, at the northern and southern ends of Bahnaric respectively, Sre /ʔous/ and Cua /ʔolh/ are regular reflexes of the AA root, suggesting perhaps that **ʔuɲ* is a latter coinage that diffused internally.

Phonologically proto-Bahnaric is unexciting; Sidwell (2002, 2010) reconstructs a mostly unmarked phonemic inventory, with a typologically common 3x3 vowel inventory plus two diphthongs (*iə, *uə). The only notable phonemes are presence of high-central vowels **ɨ, *ɨː*, which generally appear to be raised reflexes of pAA **a, *aː*, e.g.:

**ɲɨːm* 'weep'	< pAA **jaːm* (Shorto #1381)
**ʔmɨː* 'rain'	< pAA **gmaʔ* (Shorto #141)
**prɨːt* 'banana'	cf. OldMon *brat*

And the affricate **ʦ* is reconstructed to account for a curious correspondence of /s ~ ʦ/ in North Bahnaric to /t/ elsewhere in Bahnaric (in the 'bone' etymon and perhaps 4 other etyma).

Another feature of Bahnaric phonology (discussed in Sidwell 1999 and Diffloth ms.) which has been suggested as diagnostic of the branch, is an odd correspondence of glottal final rhymes. It is generally agreed that proto-Austroasiatic admitted no open syllables, such that words lacking a final consonant had a prosodically specified glottal stop, and there are various branches that continue the pattern (e.g. Khmuic, Palaungic, Aslian). Bahnaric seems to be a special case since South Bahnaric lost all final glottal stops,[64] while in the rest of the branch the reflexes are split more or less half and half between glottal and zero finals, e.g.

pAA	South Bahnaric	Bahnar	Khmer	Khmu
**briːʔ* 'forest' (Shorto #181)	**briː*	*briː*	*prej*	*briʔ*
**bluːʔ* 'thigh' (Shorto #223)	**bluː*	*bluː*	*plɨw*	*bluʔ*
**kn[i]ʔ* 'rat' (Shorto #93)	**kneː*	*kənɛː*	--	*kneʔ*
**ŋɟiːʔ* 'sick' (Shorto #55)	**ɟiː*	*ɟiʔ*	*chɨː*	--
**ɟruːʔ* 'deep' (Shorto #172)	**ɟruː*	*ɟruʔ*	*crɨw*	*ɟruʔ*
**trɔːʔ* 'hit; correct' (Shorto #174)	*crɔː* (Mnong 'pretty')	*trɔʔ*	*trow*	*tʰrɔʔ*

Sidwell (1999) dealt with this issue by reconstructing a length distinction before glottals, but that analysis was confused by unrecognized Chamic loans and cannot be sustained. It is now clear that final glottal stop was simply lost in South Bahnaric, and one must reconstruct a phonological distinction of open versus closed rhymes in proto-Bahnaric which may have various explantions. Is it a trace of some lost suprasegmental, the outcome of some early language mixing, or a regular change that simply ran out of steam before it could diffuse through the entire lexicon? It is not clear.

In summary, we see that both Katuic and Bahnaric are clearly identified on both phonological and lexical criteria. However, there are no clear examples of phonological innovations that compel us to reconstruct a common phase of development after pAA had begun to break up. Consequently we are forced to take a closer look at the lexicon.

3.3 Katuic-Bahnaric isoglosses

Analyzing the significance of isoglosses shared between branches is a difficult question, and three possibilities come to mind:

- there will be words that by chance are retained in only Katuic and Bahnaric (or we don't know of their presence elsewhere)
- there will be borrowings between these neighboring branches, or mutual borrowings from a third source, or
- there may be common innovations from a period of unique common ancestry

We may regard the first possibility as a minor factor only. Any unique retention in Katuic and Bahnaric implies its loss without trace in the other 10 or 11 branches, and on the apparent average retention rates indicated by lexicostatistics, this probability approaches a vanishingly figure, and is thus unlikely to have a detectable effect.[65]

---

[64] Final glottal stops were reintroduced to South Bahnaric largely by later shift of **-k* to **-ʔ*, and via loans.

[65] We can reasonbably estimate that each branch preserves approximately 40% ~ 50% of the Austroasiatic basic lexicon, given typical inter-branch cognate counts from high teens to around 25%. This suggests that the chance of retention in just two branches is crudely in the order of less than 0.5^{10} (by multiplying the chance of loss in any branch by the number of branches in which loss is assumed, if branches are roughly equidistant - it remains to be modelled for different nested branching senarios).

One the other hand, this writer has indentified, on a list of 1000 meanings, at least 60 Katuic-Bahnaric lexical agreements that do appear to be restricted to these two branches,[66] and thus probably variously fall in to the second and third categories above. Broadly consistent with this proportionality, six of these include items on the Swadesh 100 list, although some are not all well distributed within Bahnaric:

		pKatuic	pBahnaric	Bahnaric examples
1)	claw, nail	**krnias*	**krniəh*	*nias* Sre, *kaniah* Cua, *krnieh* Cheng
2)	dry (in the sun)	**tiːŋ/*tɨaŋ*	**tɨːŋ*	*tuəŋ* Alak, *tiak* Cua, *t̤iəŋ* Halang
3)	to stand, ascend	**dɨk*	**dɨk*	*duk* Alak, *dɨk* Cua, *dɨk* Cheng
4)	bark (of tree)	**ʔndɔ̀h*	-	*kənduh* Sre, *kadùh* Cua
5)	mountain/hill	**gɔɔr*	-	*gɔɔr* Alak, Bahnar
6)	what/where?	**-lɛʔ*	-	*lɛʔ lɛʔ* Mnong 'when', *lɛʔ* Bahnar 'final emphatic'

We can offer the following commentary:

1) This appears to be a nominalization of the root represented in Khmer *kiəs* 'to scrape'
2) Reflexes are present in Tariang, Alak, and North Bahnaric, but vowel correspondences are not quite regular in Katuic, suggestive of interference by borrowing
3) This may be from Chamic, c.f. Written Cham *dik*, which Thurgood (1999) derives from PMP **pa-nahik*.
4) The lack of agreement between vowels in the Katuic and Bahnaric forms is curious; the distribution suggests borrowing into Bahnaric from a third source such as Chamic.
5) Apparently a loan from Katuic has replaced **bnɨm* in Alak and Bahnar.
6) Also used as a base for interrogatives in North Bahnaric, *lɛʔ* may be reconstructable to pBahnaric; but in many languages interrogative have been replaced with loans from Lao, Viet etc.

Of these six isoglosses, five invite explanations involving borrowing, and one (6) is ambiguous. A similar pattern applies across the full list of such isoglosses, strongly indicating longstanding contact and lexical borrowing between these branches. This should not be surprising; it is evident that these two branches have been geographically adjacent for a long period of time, probably some thousands of years. Also, both are known to have been under common external influences, especially from Khmer and Chamic (see Thurgood 1999, Sidwell & Jacq 2003 for some discussion of borrowing between these groups).

The West Bahnaric languages, and neighboring Taliang/Kasseng, being immediately in contact with Katuic speakers, are even more heavily influenced. Sidwell & Jacq (2003) found that some 6% of the proto-West Bahnaric lexicon consists of direct borrowings from Katuic, and this writer's most recent analysis of the Taliang/Kasseng lexicon (as represented in L-Thongkum 2001) indentifies 25% (320 out of 1320) as borrowed from Katuic. It is hardly surprising that early classifications, such as Thomas & Headley (1970) mistakenly identified Kasseng as Katuic!

It is clear that much of the lexical diversification distinguishing various Bahnaric sub-groups has arisen from their extensive adoption of loans from their respective neighbors. Although this includes Khmer and Chamic, whose influence we expect to find since historically Angkor and Champa repeated fought for control of these territories (see Briggs 1951, Coedes 1968, etc.), it appears likely that it is the longstanding neighbour Katuic which has exerted the most prolonged influence, and this is the most likely explanation for the bulk of the unique lexical isoglosses between these two branches. A small residue of unexplained isoglosses would be unremarkable; constituting no more than a loose end in what is otherwise a compelling narrative of intimate language contact.

[66] Data available on request.

4. Synthesis

It seems fairly clear that Katuic and Bahnaric do not form a single branch within Austroasiatic, but rather are coordinate branches that have exchanged a substantial proportion of loans. Holm characterises this kind of relation well:

> […] the relative **location** of a language may affect the results. A central position with long lasting close neighborhood ("Sprachbund") leads to different forms of concealed borrowing, simulating a false closer relationship, whereas a peripheral location ("Saumlage") often causes lower contacts, thus leading to very conservative habits and false earlier separation (cf. e.g. Pennsylvanian Dutch).
> (Holm 2008:11)

The later part of Holm's remark also points to a possible explanation for our Huffman Effect. Holm points out that peripheral languages can be more conservative - they undergo a slower rake of change as speakers maintain their identity clinging to old habits (a well known émigré phenomenon). Of course, other peripheral groups - perhaps those especially small in number - may find themselves in a weakended position and be more susceptible to accelerated change. The former may apply to groups such as Munda, while the Nicobarese may fall in the the latter catergory.

Consequently we can now propose a history by which the anomalous lexical relations, indicated by the lexicostatical scores, have come to be. Katuic and Bahnaric, long in contact, have formed a kind of lexical Sprachbund, simulating a "false closer relationship". Not so far away geographically, and perhaps even closer to each other in the past, concealed borrowing also accounts for somewhat higher cognated scores with, for example, Mon, Khmu, Pearic. Further afield, out of obvious range of contact, the peripheral Mundas conserved AA lexicon. Between these extremes there are ambiguous cases: are the high scores - up to 31% - between Katuic-Bahnaric and South Aslian due to conservatism in South Aslian or borrowing in ancient times? Finally, the tiny Nicobarese speaking community experienced such a combination of demographic bottleneck and external pressures that their language manifested excellerated change.

References

Anderson, Gregory D.S. 2003. Dravidian influence on Munda. *International Journal of Dravidian Linguistics* 32/1:27-48

Bailey, Charles-James. 2002 (reprint). *Essays on Time-Based Linguistic Analysis*. Oxford, Clarendon Press.

Bergsland, K., & H. Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3: 111-153.

Blood, Henry F. 1966. *A Reconstruction of Proto-Mnong (Including Tentative Reconstruction of Proto-South-Bahnaric)*. M.A. Thesis, Department of Linguistics Indiana University.

Briggs, Lawrence Palmer. 1951. The ancient Khmer empire. Philadelphia, American Philosophical Society

Chaubey, Gyaneshwer; Mait Metspalu; Ying Choi; Reedik Magi; Irene Gallego Romero; Pedro Soares; Mannis van Oven; Doron M. Behar; Siiri Rootsi; Georgi Hudjashov; Chandana Basu Mallick; Monika Karmin; Mari Nelis; Juri Parik; Alla Goverdhana Reddy; Ene Metspalu; George van Driem; Yali Xue; Chris Tyler-Smith; Kumarasamy Thangaraj; Lalji Singh; Maido Remm; Martin B. Richards; Marta Mirazon Lahr; Manfred Kayser; Richard Villems; Toomas Kivisild. 2010. Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. *Molecular Biology and Evolution* Online ISSN 1537-1719 - Print ISSN 0737-4038

Coedes, Georges. 1968. *The Indianized States of Southeast Asia*. Translated by S.B. Cowing. Edited by W.F. Vella. Honolulu: East West Press.

Diffloth, Gérard. 1976. *Jah-Hut, an Austroasiatic language of Malaysia*. in N.D. Liem (ed.). *Southeast Asian Linguistic Studies* Vol.2. Pacific Linguistics Series C-No.42. Canberra, Australian National University, pp. 73-118.

Diffloth, Gérard. 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouique. *Mon-Khmer Studies* 11:47-82.

Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies* 15:139-154.

Diffloth, Gérard. 1991. Palaungic vowels in Mon-Khmer perspective. In Davison ed. Austroasiatic *Languages, essays in honour of H. L. Shorto*. SOAS, University of London.

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In: Sagart, Laurent , Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Routledge/Curzon. pp79-82.

Diffloth, Gérard. 2008. Proto-Aslian diphthongs and historical parallels in other Austroasiatic languages. Talk given at the 18th Annual Meeting of the Southeast Asian Linguistics Society, Universiti Kebangsaan Malaysia, Bangi, 22 May 2008

Donegan, Patricia and David Stampe. 2004. Rhythm and the Synthetic Drift of Munda, The Yearbook of South Asian Languages and Linguistics. Berlin and New York, De Gruyter. pp 3-36.

Efimov, A. Ju. 1990. *Istoricheskaja Fonologija Juzhnobaxnaricheskix Jazykov*. Moskva, Nauka.

Efimov, Aleksandr. 1983. *Problemy fonologicheskoj rekonstrukcii proto-katuicheskogo jazyka*. Kandidat Dissertation, Institute of Far Eastern Studies Moscow.

Ferlus, Michel. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15:1.1-27.

Ferlus, Michel. 2007. Lexique de racines Proto Viet-Muong (Proto Vietic Lexicon). Unpublished manuscript, May 2007.

Gainey, Jerry. 1985. *A comparative study of Kui, Bruu and So phonology from a genetic point of view*. Bangkok, Thailand: Chulalongkorn University thesis.

Gray, Russell D.; Atkinson, Quentin D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426 (6965): pp. 435–439.

Greenhill SJ**,** Drummond AJ, & Gray RD (2010) How Accurate and Robust Are the Phylogenetic Estimates of Austronesian Language Relationships? *PLoS ONE*, 5(3): e9573

Greenhill, Simon and Gray Russel (2009) Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In Austronesian historical linguistics and culture history: a festschrift for Robert Blust (Pp 375-397). A. Adelaar & A. Pawley (Eds). Canberra: Pacific Linguistics.

Gregerson, Kenneth J. 1976. "Tongue-Root and Register in Mon-Khmer." Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies, Two Volumes.* Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13), I:323-369

Headley, Robert K. 1976. Some Considerations on the Classification of Khmer. In P.N. Jenner, L.C. Thompson and S. Starosta (Ed.), *Austroasiatic Studies*, Honolulu, (pp. 431-452). The University Press of Hawaii.

Headley, Robert K. Jr. 1985. Proto-Pearic and the Classification of Pearic. In Ratanakul, Suriya, David Thomas and Suwilai Premsrirat (eds.) *Southeast Asian Linguistic Studies presented to Andre-G. Haudricourt*. Bangkok, Mahidol University. pp 428-478.

Holm, Hans J. 2003. The Proportionality Trap. Or: What is wrong with lexicostatistical Subgrouping.Indogermanische Forschungen108:38-46.

Holm, Hans J.. 2008 The Distribution of Data in Word Lists and its Impact on the Subgrouping of Languages. In Studies in Classification, Data Analysis, and Knowledge Organization downloaded http://www.hjholm.de/SLRD.pdf

Huffman, Franklin E. 1978. *On the centrality of Katuic-Bahnaric to Austroasiatic*. Unpublished handout distributed at the 3rd ICAAL meeting in Mysore.

Huffman, Franklin E. 1985. Vowel Permutations in Austroasiatic Languages. *Linguistics of the Sino-Tibetan Area: The State of the Art.* Pacific Linguistics Series C-No.87. Canberra: Australian National University, pp141-45.

Lebar, Frank M., Gerald C. Hickey, and John K. Musgrave. 1964. *Ethnic groups of mainland Southeast Asia*: New Haven Conneticut: Human Relations Area Files Press.

Peiros, Ilia. 1996. *Katuic Comparative Dictionary*. Pacific Linguistics Series C-132. Canberra, Australian National University.

Peiros, Ilia. 1998. *Comparative Linguistics in Southeast Asia*. Pacific Linguistics Series C-142. Canberra, Australian National University.

Peiros, Ilia. 2004. *Genetičeskaja klassifikacija avstroaziatskix jazykov*. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (dissertacija).

Pinnow, Heinz-Jürgen. 1959. *Versuch Einer Historischen Lautlehre Der Kharia-Sprache*. Wiesbaden, Otto Harrassowitz.

Schmidt, Pater Wilhelm. 1905. Grundzüge einer Lautlehrer Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayrischen Akademie der Wissenschaft* 1 (22.3):677-810.

Schmidt, Pater Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie* 5:59-109.

Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*: Canberra Pacific Linguistics 579.

Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra, Pacific Linguistics 551.

Sidwell, Paul 2005. Proto Katuic Phonology and implications for sub-grouping. In *SEALS XV: papers from the 15th meeting of the Southeast Asian Linguistics Society*, edited by Paul Sidwell, Canberra, Pacific Linguistics PL E1. pp193-204.

Sidwell, Paul. 1999. A Reconstruction of ProtoBahnaric. PhD thesis, University of Melbourne.

Sidwell, Paul. 2000, *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra, Pacific Linguistics 501.

Sidwell, Paul. 2002. Genetic classification of the Bahnaric languages: a comprehensive review. *Mon-Khmer Studies* 32:1-24.

Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: history and state of the art*. Munich, Lincom Europa.

Sidwell, Paul. 2010. Preliminary Notes on Proto Palaungic. http://sealang.net/monkhmer/database/

Skeats, William & Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsular, Volume 2*. London, MacMillian & Co.

Smith, Kenneth D. 1974. "A computer analysis of Vietnam language relationships." Work Papers of the Summer Institute of Linguistics, University of North Dakota 18: 99-113.

Smith, Kenneth D. 1978. An automatic typewriter assist to comparative linguistics, with application to 30 Mon-Khmer languages. Paper presented at the 2nd ICAAL meeting,, Mysore, India, December 19-21, 1978.

Smith, Kenneth D. 1981. "A lexico-statistical study of 45 Mon-Khmer languages." In Andrew Gonzalez and David Thomas (eds.), Linguistics across continents: Studies in honor of Richard S. Pittman, 180-205. Linguistic Society of the Philippines Monograph, 2. Manila: Summer Institute of Linguistics and Linguistic Society of the Philippines.

Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas, Language Data Series, Summer Institute of Linguistics.

Theraphan Luang-Thongkum. 2001. *Languages of the Tribes in Xekong* Province Southern Laos. The Thailand Reseach Fund.

Thomas David, & Marilyn Smith. 1967. Proto-Jeh-Halang. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung,* 20: 157–175.

Thomas David. 1966. *Mon-Khmer subgroupings in Vietnam*. In Norman Herbert Zide, ed. *Studies in Comparative Austroasiatic Linguistics.* Indo-Iranian Monographs, V. The Hague, pp 194-202.

Thomas David. 1973. A Note on the Branches of Mon-Khmer. *Mon-Khmer Studies* 4:139-40.

Thomas, David & Robert Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25:398-418.

Thomas, Dorothy M. 1967. *A Phonological Reconstruction of Proto-East-Katuic.* MA thesis, University of North Dakota.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: two thousand years of language contact and change.* Oceanic Linguistics special Publications No. 28. Honolulu, University of Hawaii Press.

Zide, Norman Herbert. 1965. Gutob-Remo vocalism and glottalized vowels in Proto-Munda" Milner et al. (eds.). Indo-Pacific Linguistic Studies (Lingua 14), 43-53

Vietnamese Slogans in the 21st Century[67]

Sophana Srichampa
Research Institute for Languages and Cultures of Asia
Mahidol University, Thailand

Abstract

A Vietnamese slogan is a sentence created by government offices through which, social practices are reflected and key concepts promoted to encourage people what to do and remind them what not to do. Moreover, social issues are reflected in slogans such as sustaining the environment, campaigning for HIV/AIDS protection, reducing road accidents and collecting tax. The themes of these slogans are expressed using vocabulary and syntactic structures at a micro level. Two conceptual structures are mainly found from the slogans, represented with imperatives, and as equations: **A=B**, **A≠B** and **A(New)⇔B(Existing)**. Slogans consist of simple, compound and complex sentences. Conceptual structure is reflected in the very syntax of the slogans. Eighty slogans are classified according to their meanings reflected in the social current issues.

Intoduction

Vietnam has developed considerably in terms of its society and economy since the implementation of the Doi Moi Policy (economic opening) from 1986 onwards. This modernization affects the Vietnamese society both positively and negatively, and the social situation is affected by both internal and external factors. Consequently many new social issues arise which need to be identified, confronted and dealt with, and the government uses various methods to identify and deal with these, including creating awareness and modify social practices through language.

The Propaganda Organizing Committee under the Communist Party of Vietnam meets with the heads of all media weekly for slogan concepts assignment, and distribution of slogans widely among the public, given the events and social issues current at that time. People also use relevant slogans in their speeches. In this way, the centralized administration attempts to provoke awareness among the Vietnamese people through all media channels. It is remarkable that the basic structure of Vietnamese slogans do not change; rather it is merely the content that is modified to reflect the particulars being communicated.

The structural properties of these slogans closely match their conceptual structures. They can be classified into two types as follows:

- imperatives: **do A** (for result **B**)
- equations: **A=B**, **A≠B** and **A(New)⇔B(Existing)**.

Imperatives are straightforward directives exhalting the people to believe or act accordingly, where **A** is the social prescription, and an expected result may or may not be stated. Equations are more complex, in

[67] I would like to express my gratitude to Dr. Paul Sidwell who helped me to improve a draft of this paper. My thanks also go to Mr. Nguyen Ngoc Binh who helped me to find the data including meaning discussion. Last, my thanks also go to Mr. Richard Hiam, my colleague who corrected a previous draft of paper.

Srichampa, Sophana. 2011. "Vietnamese Slogans in the 21st Century?." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.112-126.

these a new idea or practice **A** is introduced but it is licenced by be being conceptually equated with (=), in contrast to (≠), or consistent with (⇔) an existing idea or practice **B** whose authority is given.

The data used for this study are from Vietnamese slogans used for various social purposes to marshal the people's suggestions and collaboration for an ostensibly healthy, peaceful, wealthy and secure future for the country. The data are taken from the following sources:

1) National Committee of HIV-AIDS Prevention

2) Ministry of Natural Resources and Environment

3) Ministry of Culture and Information

These organizations take charge of the hot social issues at present. The third organization listed above is most responsible for taking care of the slogans produced by the three organizations. Therefore, they are the sources of data for this paper.

I. Slogan

The Vietnamese slogan is intended to attract people's attention and/or to suggest an idea quickly. The slogan is a kind of discourse which helps to manage social relationships between people. Social change affecting language practices is mentioned by Fairclough:

> Many of these social changes do not just involve language, but are constituted to a significant extent by changes in language practices; and it is perhaps one indication of the growing importance of language in social and cultural change that attempts to engineer the direction of change increasingly include attempts to change language practices. (Fairclough 1993:6)

Consistent with Fairclough's idea above, the Vietnamese government begins by writing a long descriptive text voiced the social problem such as a problem of tax collection from the private sector through media, for example:

> Những năm gần đây, nhờ tập trung chỉ đạo tháo gỡ khó khăn, vướng mắc về thủ tục hành chính, giao đất, cho thuê đất, giải phóng mặt bằng, cải thiện môi trường đầu tư nên huy động vốn đầu tư xây dựng cơ bản (XDCB) ở tỉnh ta tăng nhanh, số thu trong lĩnh vực này năm sau cao hơn năm trước, chiếm tỷ trọng ngày càng lớn trong tổng số thu thuế, phí hàng năm. Dù vậy, thu thuế trong lĩnh vực XDCB chưa tương xứng với nguồn vốn đầu tư phát triển. Ngoài khó thu thuế XDCB ở các thành phần kinh tế ngoài quốc doanh nhất là thu xây dựng nhà ở tư nhân, kết quả thu thuế đầu tư XDCB từ nguồn ngân sách nhà nước còn nhiều khe hở để nhà thầu trốn tránh trách nhiệm, nghĩa vụ liên quan…[68]
>
> In recent years, thanks to focused and directed solutions to difficulties and obstacles of related to administrative procedures, land allocation, land lease, site clearance and improvements in the investment environment, bond mobilization in capital construction (CC) in the provinces has increased with revenue and the proportion of total tax and fee revenue in this sector is growing annually. However, tax collection in the field of CC does not match capital investment and development. Apart from collecting CC taxes in the non-state economic sectors, especially private housing construction, state CC tax collection procedures still contain loopholes through which contractors are able to avoid responsibilities and related obligations...

Although the issue is raised in detail, such a long text can not effectively attract people's interest sufficiently to persuade them to collaborate with the government. Therefore, language practices are required that can adequately communicate and motivate to achieve the people's collaboration. Thus, the slogan is a linguistic tool for campaigning social policy among the people.

[68] From:
http://www.gdt.gov.vn/gdtLive/Trang-chu/Tin-tuc/News?contentId=135926&location=tct&location=tct

Some Vietnamese slogans focus on ideology such as constructive effects related to Communism which can be considered as part of the Vietnamese identity. Some focus on social change which causes social problems and calls for awareness in order to change people's negative behaviors. The stated purposes for the government implementing slogans are for the unity, better life and happiness of people.

Language analysis of slogan text produced by the government of Vietnam in the 21st century will interpret in terms of social practices first, followed by text analysis of vocabulary and syntactic structure at a micro level. Analysis of text, interaction and discourse context will be made at a macro level. Hence some concepts of Critical Discourse Analysis (CDA) are applied in analysis of slogans.

II. Slogan interpretations

According to the slogans launched by responsible offices, official prescription of social practices is a key concept for telling people what to do and reminding people what not to do.

2.1 Praise of the former President Hồ Chí Minh

Former president Hồ Chí Minh was a Vietnamese Communist revolutionary and National Liberation Hero who was prime minister from 1946–1955 and president from 1945–1969. He was a founder of the Democratic Republic of Vietnam (North Vietnam).

Hồ Chí Minh has remained in the hearts of the Vietnamese people from that time up to the present. He is referred to as Bác Hồ (Uncle Ho) and praised by the Communist government to the point of god-like status in Vietnam. His picture is on the Vietnamese currency, in front of public buildings, in meeting rooms, in class-rooms, houses and temples. UNESCO granted Hồ Chi Minh the title "National Liberation Hero" and "Great Man of Culture" (Vietnamnet.vn, 2007) for devoting his life to the liberation, independence and social improvement of Vietnam. Not only was he praised in the 20th century, but he continues to be respected widely around the world today. The vocabulary used in slogans reflects respect and praise for Hồ Chí Minh in several ways as follows:

2.1.1 Figure of speech[69]

Although Hồ Chí Minh passed away many years ago, the Communist government continues the effort to retain his legacy in the people's memory as if he were still alive, evidenced by the verb phrase [sống mãi], 'live forever', as in (1).

(1)	*Chủ tịch*	*Hồ Chí Minh*	***sống***	***mãi***	*trong*	*sự nghiệp*	*của*	*chúng ta.*
	President	Ho Chi Minh	**live**	**forever**	in	the task	of	us

"President Ho Chi Minh lives forever in our work."

2.1.1.1 The former president Hồ Chí Minh as an "example"

Using a word [tấm gương /gương] 'example' only

Two out of 11 slogans use the word [tấm gương /gương] 'example', to emphasise his qualities as a leading example or role model for Vietnamese people in the 21st century, as in the following examples:

(2)	*Sống,*	*chiến đấu,*	*lao động*	*và*	*học tập*	*theo* ***gương***	*Bác Hồ*	*vĩ đại.*
	Live	fight	work	and	study	follow **example**	uncle Ho	great

"Live, fight, work and study according to the example of Uncle Ho."

(3)	*Toàn*	*Đảng,*	*toàn*	*dân,*	*thực hiện*	*cần kiệm,*	*liêm chính,*	*chí công,*
	Entire	Party	entire	people	practice	economic	integrity	fair
	vô tư	*theo*	***tấm gương***	*của*	*Chủ tịch*	*Hồ Chí Minh*	*vĩ đại.*	

[69] Expression of language by which literal meaning of word is not directly employed (Knight, Lorna and others 1994: 283).

unselfish according **example** of President Ho Chi Minh great
"The entire Party and all the people serve the economy with integrity, fairness and selflessness according to the example of former President Hồ Chí Minh."

Using the word [tấm gương] + other modifier

Six slogans use the word [tấm gương] 'example' plus a modifier [đạo đức] 'moral', to emphasize his morality as a role model for the Vietnamese people as in (4).

(4) *Học tập và làm theo* ***tấm gương đạo đức*** *Hồ Chí Minh,*
Study and do according **example** **moral** Ho Chi Minh
nói đi đôi với làm, xây đi đôi với chống.
speak go pair with do construct go pair with destroy
"Study and act according to the moral example of Hồ Chí Minh: Speaking goes with acting, constructing goes with destroying."

One slogan uses the word **[tấm gương]** 'example' with the juxtaposed noun phrases **[trung với nước]** 'faithful example for the country' and [**hiếu với dân**] 'gratitude' to highlight Hồ Chí Minh as a role model, as in (5).

(5) *Học tập và làm theo* ***tấm gương trung với nước,*** **gratitude**
Study and do according **example** **faithful with** **country** ***hiếu***
với dân *của Chủ tịch Hồ Chí Minh,* ***toàn Đảng,*** ***toàn dân,*** ***toàn quân***
with people of President Ho Chi Minh, **entire party** **entire people** **entire army**
phấn đấu vì độc lập, tự do của Tổ quốc, vì chủ nghĩa xã hội.
strive for independent freedom of native land for socialism
"Study and act according to the faithful example of former President Hồ Chí Minh for the country, show gratitude towards people, the entire Party, all the people, the entire army strive for independence, freedom of the motherland for Socialism."

In (5), three juxtaposed sentences are conjoined. The first two sentences are imperative sentences **[Học tập và làm theo tấm gương trung với nước]** and **[hiếu với dân** của **Chủ tịch Hồ Chí Minh**]. The third one starts with the juxtaposed NPs [**toàn Đảng, toàn dân**, **toàn quân**] 'the entire Party, all the people, the entire army' as subject. These three NPs cover all Vietnamese people. It is predictable that although the first two sentences are subjectless, it assumes that it has the same NPs agent (subject) as the third sentence because repetition is not considered good style.

2) Address term

Three terms of address for former president Hồ Chí Minh were found in the data, [Chủ tịch Hồ Chí Minh] 'President Hồ Chí Minh' in (6), the informal [Bác Hồ] 'Uncle Hồ' as in (7) and simply "Hồ Chí Minh", as in (8).

(6) ***Chủ tịch Hồ Chí Minh*** *sống mãi trong sự nghiệp của chúng ta.*
President Ho Chi Minh live forever in the task of us
"President Ho Chi Minh lives forever in our work."

(7) *Sống, chiến đấu, lao động và học tập theo gương* ***Bác Hồ vĩ đại.***
Live fight work and study follow example **uncle Ho great**
"Live, fight, work and study according to the examples of great Uncle Ho."

(8) ***Học tập*** $_{1}$ ***và làm theo tấm gương đạo đức Hồ Chí Minh*** $_{2}$***, kiên quyết***
Study and do according example moral Ho Chi Minh decide
ngăn chặn $_{3}$ *và đẩy lùi quan liêu, tham nhũng, lãng phí, tiêu cực* $_{4}$*.*
stop and push back bureaucratic corruption waste negative
"Study and act according to the moral example of **Hồ Chí Minh** by deciding to avoid and get rid of corruption which is a negative waste."

In (7), when the slogan addresses Hồ Chí Minh as 'uncle', then the adverb following [vĩ đại] 'great', praises Uncle Hồ forever.

It is noticeable that the Vietnamese people feel familiar with Hồ Chí Minh by addressing him as bác 'uncle', a close relative. This, as well as the simple use of his common name reflects the Communist ideology of equality (in principle).

Regarding the syntactic structure of the above slogans, (7) is an imperative sentence starting by the cojoined VPs [Sống, chiến đấu, lao động và học tập] 'live, fight, work and study' as subject of the sentence - [theo] 'follow' is the verb. In (8), four coordinated VPs form a sentence shown by co-indexing numbers in the example, with a conjunction [và] 'and' which conjoins the VPs. This structure is the most popular for praising former president Hồ Chí Minh. This use of the imperative clearly suggests that the government expectes to have authority over the people.

2.1.2 Nominalisation

Nominalisation in Vietnamese is a process of forming a noun from some other word-class (Crystal 1992: 233-234). Nominalised classifiers such as **sự** or **cuộc,** are placed in front of nouns [**sự nghiệp công nghiệp hóa, hiện đại hóa]** 'the task of industrialization, modernization' as in (9) which is one of the key concepts supporting the development of Vietnam after the Doi Moi Policy. In slogans, several nominalizations are commonly found.

(9) *Học tập và làm theo tấm gương đạo đức Hồ Chí Minh, đẩy mạnh*
Study and do according example moral Ho Chi Minh push forward
sự nghiệp công nghiệp hóa, hiện đại hóa, *sớm đưa nước ta*
task industrialization modernization early bring country our
ra khỏi tình trạng kém phát triển.
go out situation less develop
"Study and act according to the moral example of Hồ Chí Minh by pushing the tasks of industrialization and modernization to develop our country beyond under-development."

2.1.3 Former President Hồ Chí Minh's ideology

Hồ Chí Minh expressed his ideology as **[cần kiệm]** 'economic'**, [liêm chính]** 'integrity'**, [chí công]** 'fair', **vô tư** 'unselfish' while he was alive. It has become a motto for the Vietnamese to behave accordingly (depending on the individual) as in (10).

10) *Toàn Đảng, toàn dân, thực hiện* ***cần kiệm, liêm chính, chí công,***
entire party entire people practice **economic integrity fair**
vô tư *theo tấm gương của Chủ tịch Hồ Chí Minh vĩ đại.*
unselfish according example of President Ho Chi Minh great
"The entire Party and all people serve the economy with integrity, fairness and unselfishness according to the example of former President Hồ Chí Minh."

According to the structure of the slogans, after instructing people to follow former president Hồ Chí Minh's example, the following purposive sentence explains the objective of the actions, [**vì mục tiêu dân giàu, nước mạnh, xã hội công bằng, dân chủ văn minh**] 'to make the people rich, the country strong and create a fair, democratic and civilized society', as in (11).

(11) *Học tập và làm theo tấm gương đạo đức Hồ Chí Minh* ***vì***
Study and act according example moral Ho Chi Minh **for**
mục tiêu dân giàu, nước mạnh, xã hội công bằng, dân chủ, văn minh.
aim people rich country strong society fair democracy civilization
"Study and act according to the moral example of Ho Chi Minh (because he wanted) to make the people be rich, the country strong and to have a fair, democratic and civilized society."

The above examples can be conceptualized as imperatives.

2.2 Mobilizing the people with revolutionary concepts

The government emphasizes that the revolution is the business of the people, by people and for people. There are key words expressing this slogan: revolution [**Cách mạng**], business of the people [**sự nghiệp của dân**], by people [**do dân**] and for people [**vì dân**]. People are always used as a target for any common property and ideology references as in (12)

(12) ***Cách mạng*** *là* ***sự nghiệp*** ***của*** ***dân,*** ***do*** ***dân*** *và* ***vì*** ***dân.***
revolution be **business** **of** **people** **by** **people** and **for** **people**
"Revolution is the business of the people, by the people and for the people."

A ⇔ B
NEW (revolution) ⇔ EXISTING VALUES and PRACTICES (of the people)

2.3 Encouraging people for the National Assembly election

Vietnam is a socialist republic with a single party called the Communist Party of Vietnam. The President functions as head of state, armed forces commander and chairman of the National Council for Defense and Security. The Prime Minister heads the cabinet with twenty-six Ministries and Commissions, confirmed by the National Assembly. Eighty percent of the deputies in the National Assembly are Communist Party members. Elections for the National Assembly are held every five years. There are 498 members. In the last election in 2007, only the Vietnamese Fatherland Front, a front of the Communist Party of Vietnam, mass organizations and affiliates, and some non-partisan groups, were allowed to participate. Three members are self-nominated and do not belong to the VVF[70]. Fifty-one seats were won by non-party candidates.[71]

Slogans were used to encourage people to vote in 2007 as a duty of all Vietnamese citizens. Moreover, slogans reminded people to consider and vote for virtuous people. The vocabulary and structures of seven slogans referring to the XIIth National Assembly election are discussed below.

2.3.1 Theme of the slogan 'The XII National Assembly election'

The common theme [**đại biểu Quốc hội (khóa XII)]** '(The XII) National Assembly election' is emphasized the election campaign as in example:

(13) *Nhiệt liệt* *chào mừng* *ngày* *bầu cử* ***đại biểu Quốc hội khóa XII***
Warmly welcome day vote **member of National Assembly XII**
ngày *20* *tháng* *5 năm 2007!*
day 20 month 5 year 2007
"Warmly welcome the XIIth National Assembly election- 20 May 2007."

2.3.2 Synonym

A group of words with similar meanings supports the Vietnamese government's election campaign as follows:

Synonyms for 'elect'

The government of Vietnam tried to launch a campaign welcoming people to the XII National Assembly Election. Two synonyms for 'elect' were found in the seven slogans: **[bầu cử]** 'vote' as in (14) and **[lựa chọn]** 'select' and [**bầu**] 'vote' as in (15).

[70] The Vietnam Veterans Federation
[71] See: www.Vietnam-ACE Electoral Knowledge Network, 24 October 2009.

(14) *Nhiệt liệt chào mừng ngày* ***bầu cử*** *đại biểu Quốc hội khóa XII –*
Warmly welcome day vote **member of** National Assembly XII
ngày 20 tháng 5 năm 2007!
day 20 month 5 year 2007
"Warmly welcome the XII[th] National Assembly election- 20 May 2007."

(15) *Sáng suốt* ***lựa chọn*** *những người có đủ đức, tài* ***bầu***
wise **select** **pl.** person have enough moral ability **vote**
vào Quốc hội khóa XII!
into national Assembly XII
"Select carefully a good representative for the XII[th] National Assembly."

Synonyms for 'voter'

The word with similar meanings to 'voter' are found in the [**Cử tri cả nước**] 'all voters' as in (16) and juxtaposed NPs [**Toàn Đảng, toàn dân, toàn quân] '**All the Party, all the people and all the army' in (17) showing a collective Vietnamese.

(16) ***Cử tri cả nước*** *tích cực tham gia bầu cử đại biểu Quốc hội khóa XII!*
voter all country active participate election member of parliament XII
"Voters throughout the country, participate in parliamentary election XII actively."

(17) ***Toàn Đảng, toàn dân, toàn quân*** *thi đua lập thành tích*
entire Party entire people entire army compete establish achieve
chào mừng cuộc bầu cử đại biểu Quốc hội khóa XII!
welcome CL. election member of National Assembly session XII
"All the Party, all the people and all the army, compete to welcome a successful the XII[th] National Assembly election."

Võ Đại Lược (2011:179) states that "The government must take steps to remove those who get rich illegally, for example, corrupt officials and smugglers." This statement in the year 2011 highlights how corruption remains a problem in Vietnam despite election campaign slogans since 2007 emphasizing the need to select members who are good [**đức]** and with ability [**tài],** as in (18).

(18) *Sáng suốt lựa chọn những người có đủ* ***đức, tài bầu***
wise select PL. person have enough **moral ability vote**
vào Quốc hội khóa XII!
into national Assembly XII
"Select carefully a good representative for the XII[th] National Assembly."

Synonyms for 'people' and 'duty'

Two words with similar meanings to 'people' are **[công dân**] 'citizen' as in (19) and [**nhân dân**] 'people' as in (20). Moreover, duty is represented by two words of similar meaning 'duty' [**nghĩa vụ**] as in (19) and [**trực tiếp]** as in (20).

(19) *Bầu cử đại biểu Quốc hội là quyền và* ***[nghĩa vụ]***
vote member of National Assembly be right and **duty**
của mỗi ***công dân****!*
of each **citizen**
"Voting in parliamentary elections is the right and duty of each person."

(20) *Bầu cử đại biểu Quốc hội là* ***[trực tiếp]*** *góp phần xây dựng*
Vote member of National Assembly be **duty** contribute construct
Nhà nước pháp quyền XHCN của ***nhân dân****, do* ***nhân dân****, vì* ***nhân dân!***
nation jurisdiction Socialist Republic of **people** by **people** for **people**
"Electing the National Assembly is a duty in order to contribute to and build the country. It is the

legal duty of the Socialist Republic's people, by people and for the people."

The above two examples can be schematized as A⇔B.

2.3.2 Figurative words for "welcome'

Vietnamese elections are organized by the Communist Party. In the one party system, they are not truly representative of the people, who vote only out of duty. Therefore, it is significant that the government uses slogans with the figurative word 'welcome' as if this National Assembly election were a personification, to be dutifully welcomed, as follows:

(21) ***[Nhiệt liệt]*** ***chào mừng*** *ngày* *bầu cử* *đại biểu Quốc hội khóa XII*
Warmly **welcome** day vote member of National Assembly XII
– ngày *20* *tháng* *5 năm 2007!*
– day 20 month 5 year 2007
"Warmly welcome the XII[th] National Assembly election- 20 May 2007."

In (21), there is an adverb **[nhiệt liệt**] 'warmly' preceding the VP [**chào mừng**] 'welcome' as if the election were a guest (personification) that the Vietnamese people should welcome warmly.

(22) *Toàn Đảng, toàn* *dân,* *toàn* *quân* ***[thi đua]*** *lập* *thành tích*
entire Party entire people entire army **compete** establish achieve
chào mừng *cuộc* *bầu cử* *đại biểu Quốc hội khóa XII!*
welcome CL. vote member of National Assembly session XII
"All the Party, all the people and all the army should compete to welcome a successful XII[th] National Assembly election."

In (22), it is interesting that the welcoming of the successful XII[th] National Assembly election should be competed for by all sectors of society, as expressed by a VP [**thi đua]** 'compete'. It encourages awareness by the Vietnamese people of the importance of the election.

2.4 Praising the Communist Party and the country

The Communist Party is the core institution of Vietnam. Although Vietnam has implemented free economy market reforms, politics remains based on Communism. In previous decades, propaganda was directed towards this issue and in this century, praising the Communist Party is still the focus in order to remind the Vietnamese to be nationalistic, as follows:

2.4.1 Figurative word for 'long live'

In two slogans praising the Communist Part and the country, a figurative word 'long live' [**muôn năm]** as in (23) and (24) is used, with a preceding adverb **[quang vinh]** 'glorious'. This reflects recognition by the Vietnamese that the Communist Party as well as the Socialist Republic of Vietnam should live long in same way as leaders of other countries are praised.

(23) *Đảng* *cộng sản* *Việt Nam* ***[quang vinh]*** ***muôn năm****!*
Party Communism Viet Nam **glorious** **long live**
"Long live the glorious Communist Party of Viet Nam"

(24) *Nước* *Cộng hòa xã hội chủ nghĩa Việt Nam* ***muôn năm****!*
country Socialist Republic of Vietnam **long live**
"Long live the Socialist Republic of Viet Nam"

It seems that the concept of revolution in 2.2 contradicts the election ideal in 2.3. Although there is an electoral system in Vietnam, the single party does not provide a real democratic choice for people to apply their voting rights. Revolution is a memory harbored by the older generation who occupy the Assembly. These two concepts are able to co-exist so far in the Vietnamese political context.

2.5 Sustaining the environment

Due to rapid economic development, population growth and urbanization, Vietnam faces pollution from solid waste, water and dirty air. Only 60 percent of the population has access to clean water. The Government issued a development plan for water supply with the objective of providing clean water for 80 percent of the population by the year 2015.[72]

For example, Industrial wastewater is also having a negative impact on the environment. According to a 2004 Government report, only 15 of the nation's 74 operating industrial zones (IZs) have centralized wastewater treatment plants. This has resulted in the direct discharge of untreated industrial wastewater into the environment. Also, a large accumulation of domestic solid waste - more than 20,000 tons - is generated in Vietnam each day. To date, burying domestic solid waste in landfills has been the only treatment method used in Vietnam.[73]

Consequently, there are slogans instructing the people to take care of the environment, which is the responsibility and future of all, in order to keep the environment clean, green, beautiful and hygienic according to Vietnamese's "Environmental Sustainability" development goals 2001-2010 (as stated by Nguyễn Hồng Sơn 2011:81).

2.5.1 'Environment' slogan theme

Six slogans use the word [**môi trường**] 'environment' as a theme for campaigns encouraging responsibility such as the NP **[trách nhiệm của toàn xã hội**] 'responsibility of all people' as in (25).

(25) ***[Bảo vệ]*** ***môi trường*** *là* ***[trách nhiệm*** ***của*** ***toàn*** ***xã hội].***
protect **environment** be **responsibility** **of** **all** **society**
"Protecting the environment is the responsibility of all society."

The above example can be schematized as A⇔B.

2.5.2 Definition of environment 'Green-clean-beautiful'

Two slogans define the targeted environment as [**Xanh – Sạch – Đẹp**] 'Green-clean-beautiful' as in (26).

(26) *Hãy hành động vì môi trường* ***Xanh – Sạch – Đẹp.***
let action for environment **green- clean- beautiful**
"Let's act for a green, clean and beautiful environment."

2.5.3 Repetition of a verb 'add'

The verb **[them]** 'add' is repeated twice in a slogan describing the positive effects for the environment of growing more trees as in (27).

(27) *Trồng* ***thêm*** *một cây xanh là* ***thêm*** *hành động vì môi trường.*
grow **add** one tree green be **add** take action for environment
"To grow one more tree is added action for the environment."

The above example can be schematized as A⇔B.

2.5.4 Verbs about environmental sustainability

Several verbs are used in environmental sustainability slogans, e.g. [**Bảo vệ**] 'protect' as in (28), [**hành động]** 'act', followed by a juxtaposed NPs as in (29), and [**giữ]** 'take care' followed by an NP as in (30).

[72] See: http://www.vietwater.com/market.html
[73] See: www. Vietnam - Environment and Pollution Control -- US Commercial Service Asia, 24 October 2009).

(28) ***Bảo vệ*** *môi trường là trách nhiệm của toàn xã hội.*
Protect environment be responsibility of all society
"Protecting the environment is the responsibility of all society."

In (28), its structure is a simple sentence.

(29) ***[Hãy] hành động*** *vì môi trường Xanh – Sạch – Đẹp.*
Let **act** for environment green- clean- beautiful
"Let's act for a green, clean and beautiful environment."

(30) ***[Hãy]*** ***giữ*** *vệ sinh, không vứt rác ra đường.*
Let **take care** hygiene not throw garbage out to road
"Let's look out for pollution and not throw garbage out into the road."

It is interesting that the slogan in (30) reminds people not to throw garbage out into the road. Slogan language reflects the societal situation. When negative actions affect society, the offices responsible will launch instructions to remind or forbid people to do or not to do something. The sentence structure of (29) and (30), starts with **an imperative marker [Hãy]** 'let' followed by a verb which sounds like advice.

2.6 Campaigns about HIV/AIDS protection

Like other countries, Vietnam must deal with the problems of HIV/ AIDS, drugs and prostitution. More than 254,000 people are living with HIV in Vietnam.[74] One hundred people a day are estimated to become infected. The majority of new infections are from sexual transmission. The numbers increase every year. Although the government has developed national policies and programs progressively, vigorous action is still needed in local areas.[75]

The use of slogans is one method in the government's campaign efforts to reduce the number of the HIV/AIDS infected people. Twenty-five slogans were found relating to HIV/AIDS campaigns and protection. The style of such slogans is descriptive and informative offering details to serve the main purpose of each slogan to help people better understand the HIV/AIDS problem.

2.6.1 Responding to the annual HIV Protection campaign

Each year, the State Committee Anti-HIV campaigns for the HIV/AIDS national protection. The slogan uses the word [**hưởng ứng**] 'respond' with a modified preceding adverb [**nhiệt liệt**] 'warmly' as in (31). The government seeks to have the people respond warmly to the HIV/AIDS protection campaign.

(31) ***[Nhiệt liệt]*** ***hưởng ứng*** *Tháng hành động quốc gia phòng, chống HIV/AIDS năm 2008!*
warmly **respond to** month action nation protection anti HIV/AIDS year 2008
"Respond warmly to the action month for national HIV/AIDS protection in the year 2008."

2.6.2 Non- discrimination against and care for HIV infected persons

There are words in the slogans instructing not to 'stigamtize' [**Không kỳ thị**] or 'discriminate' [**phân biệt**], linked by the conjunction [**và**] as in (32). The government wants people to sympathize with HIV infected persons and care [**Chăm sóc**] for them with [**hỗ trợ]** 'mutual aid', as in (33).

(32) ***Không kỳ thị*** ***và*** ***phân biệt*** *đối xử với người nhiễm HIV/AIDS!*
not stigamtize and discriminate treat with person infect HIV/AIDS
"Don't examine or discriminate against HIV/AIDS infected persons."

---

[74] See: http://www.unaids.org.vn/sitee/index.php?option=com_content&task=blogcategory&id=13&Itemid=27

[75] See: www. I:\UNAIDS Viet Nam - HIV-AIDS, Drug, Prostitute Situation 2009 to Develop Complicatedly and www. I:\HIV-AIDS in Vietnam The Current Situation, the National Response, the Emerging Challenges - Population Reference Bureau.mht

In (32), the structure of the sentence is a coordinated negative imperative. The VP start with a negative word [**Không**] 'not' and instructs people not to treat HIV patients that way.

(33) ***Chăm sóc, hỗ trợ*** *người nhiễm HIV/AIDS là góp phần phòng,*
Take care mutual aid person infect HIV/AIDS be contribute protect
chống HIV/AIDS!
against HIV/AIDS
"To take care of and offer mutual aid to HIV/AIDS infected persons is the contribution and protection against HIV/AIDS."

2.6.3 Encouraging infected pregnant women to consult the Health Care Center

Although infants may not easily become infected with HIV/AIDS, there is a slogan suggesting that pregnant women consult [**tư vấn**] the Health Care Center for an HIV examination **[xét nghiệm HIV]**, as in (34)

(34) ***[Vì]*** *những đứa con không nhiễm HIV - Phụ nữ mang thai hãy*
Because PL. CL. child not infect HIV lady pregnant should
đến ngay các cơ sở y tế để được ***tư vấn***
come immediately PL. center public health for gain **consultative**
và ***xét nghiệm HIV!***
and **examine HIV**
"In order to have a child free of HIV, a pregnant woman should consult the public health care center for an HIV examination"

2.6.4 Encouraging infected pregnant women to follow advice

Even if their infants are not infected with HIV/AIDS, there are slogans advising infected pregnant women to follow health advice as follows:

(i) Be taken care of [**chăm sóc]** and receive medical treatment for transmission protection from mother to infant [**điều trị dự phòng lây truyền HIV**] as in (35).

(35) ***[Vì]*** *những đứa con không nhiễm HIV - Phụ nữ nhiễm HIV,*
Because PL. CL. child not infect HIV- lady infect HIV,
mang thai hãy đến các cơ sở y tế để được ***chăm sóc***
pregnant should come PL. office medicine for gain **take care**
và ***điều trị dự phòng lây truyền HIV từ mẹ*** sang con!
and **receive medical treatment transmission HIV from mother** pass child
"To avoid HIV infected children, infected pregnant women should visit the public health care center to be looked after and receive medical treatment to prevent HIV transmission from mother to her infant."

(ii) Bringing a new born baby to the doctor to receive proper medical treatment, [**theo dõi**] 'follow up', [**chăm sóc**] 'take care of' and [**điều trị] '**receive medical treatment' as in (36).

(36) ***[Vì]*** *những đứa con không nhiễm HIV - Hãy đưa trẻ sinh ra*
Because PL. CL. child not infect HIV Should bring infant born
từ bà mẹ nhiễm HIV đến các cơ sở y tế để được ***theo dõi,***
from mother infect HIV come PL. office public health for gain **follow up**
chăm sóc *và* ***điều trị*** *thích hợp!*
take care and **receive medical treatment** suitable
"In order to avoid having an HIV infected child, bring the new born baby of an HIV mother to the public health care center for a check up and to receive suitable medical treatment."

(iii) Instructing people not to use [**Không dụng]** syringes [**bơm kim tiêm**] or skin piercing instruments [**các dụng cụ xuyên chích qua da**], using the adverb [**chung**] 'in common', to emphase prohibition, as in (37).

(37) ***[Vì]*** *những đứa con không nhiễm HIV -* ***Không dụng chung***
Because PL. CL. child not infect HIV **not use in common**
bơm kim tiêm *và* ***các dụng cụ xuyên chích qua da***
pump injection and **PL. instrument pierce pass skin**
để phòng lây nhiễm HIV!
for defend infectious HIV
"In order to avoid having an HIV infected child, do not share syringes and skin piercing instruments- for HIV prevention,"

(iv) Instructing people to use [**sử dụng**] condoms [**bao cao su**] in sexual relations [**trong quan hệ tình dục]**, as in (38).

(38) ***[Vì]*** *những đứa con không nhiễm HIV -* *Hãy* ***sử dụng***
Because PL. CL. child not infect HIV- should **use**
bao cao su trong quan hệ tình dục *để phòng lây nhiễm HIV!*
condom in relation sex for defend infectius HIV
"In order to avoid having an HIV infected child - use a condom in sexual relations - for HIV prevention."

(v) Reminding men [**Nam giới**] to avoid HIV infection [**dự phòng lây nhiễm HIV]** of themselves [**chính bản thân mình**] and those close to them [**người thân]**. In accordance with Confucian tradition, men in Vietnamese society are dominant over women and are generally active whereas women are expected to be passive. Some husbands become infected with HIV outside the home and subsequently transmit it to their wives.This slogan warns men specifically, as in (39).

(39) ***[Vì]*** *những đứa con không nhiễm HIV -* ***Nam giới*** *hãy chủ động*
Because PL CL. child not infect HIV- **male** should active
dự phòng lây nhiễm HIV *cho* ***chính bản thân mình*** *và* ***người thân****!*
prevent infectious HIV for **main oneself** and **person intimate**
"In order to avoid having an HIV infected child, males should actively avoid becoming infected themselves and infecting those close to them."

In terms of the structure of the above topic, all are cause-result sentences starting with the conjunction **[Vì]** 'because'. According to the above examples, they can be conceptualized as **because A therefore B.**

2.7 Reducing road accidents

Due to economic development, Vietnamese statistics show traffic accidents are the cause of 12,000 deaths a year : 90% of the accidents are from motorbikes (the main form of transportation in Vietnam).[76] There is a slogan reminding people that safety is a friend of which they should be aware, and accidents are the enemy which should be destroyed or avoided. Opposite words are used such as [**An toàn**] 'safe' vs [**tai nạn**] 'accident', and [**bạn**] 'friend' vs [**thù**] 'enemy', as in (40).

(40) ***An toàn*** *là* ***bạn, tai nạn*** *là* ***thù****.*
safe be **friend accident** be **enemy**
"Safety is a friend. Accidents are the enemy."

In (40), the structure is a compound sentence with two juxtaposed simple sentences conjoined together.

[76] See: www. I:\Taipei Times - archives.mht.

2.8 Tax collection

Vietnam, like all countries, collects personal income and private sector income tax for the country's development needs. There are 16 slogans regarding the payment of taxes as in the following examples:

2.8.1 Emphasizing the cooperation of people in tax collection

The government uses a familiarity strategy in tax collection by focusing on 'people's belonging and responsibility' which is schematized as: T(opic) be P(ossessive) for P(urpose), as in (41)

(41) ***Tiền thuế*** ***là*** ***của dân,*** ***do*** ***dân,*** ***do dân*** ***đóng góp*** ***để***
Tax **be** **of** **people** **by** **people** by **people's** **collection** **for**
[phục vụ ***lợi ích*** ***của*** ***nhân dân]***
service **usefulness** **of** **people**
"Tax is of the people, by the people and collection is for the service of the people."

The above example can be schematized as A⇔B.

2.8.2 Emphasizing the importance of private organizations paying tax

Many private companies try to avoid paying tax despite their importance in underpinning the economic development of the country. Therefore, the government reminds this group to pay its taxes for the country's benefit as the scheme:

D(esirable action) be D(esirable purpose), as in (42).

(42) ***Tổ chức*** ***cá nhân*** ***nộp thuế*** *là* ***nhân tố*** ***thức đẩy sự*** ***tăng trưởng*** ***kinh tế***
organize **individual** **pay tax** be **factor** **push** **cls.** **grow** **economy**
của ***đất nước*** *và* ***ngân sách*** ***quốc gi***
of **country** and **budget** **national**
"Pay private organization tax, to push economic growth and provide a budget for the country."

2.8.3 Informing people how important tax is to the development of the country

The strategy of comparing the economy or state to a living body that needs to be supported and kept in good health and so forth maps onto traditional values of obligation to family, or even obligation to more senior family members as equivalent to the good of the whole family. The slogan is schematized as T(opic) be M(etaphor)1 be M(etaphor)2 of D(esirable process), as in (43).

(43) ***Tiền thuế*** ***là*** ***nguồn lực,*** *là* ***mạch máu*** ***của*** ***sự phát triển***
money tax **be** **source power** be **blood-vessel** **of** **development**
kinh tế ***xã hội.***
economy **society**
"Tax is the power source and blood-vessel of social economic development."

In (43), the surface structure of the sentence is composed of two juxtaposed complements of the verb **[là]** 'to be', deleting the repeated subject [**Tiền thuế**] 'tax'.

The above examples can be schematized as A⇔B.

3. Macro-level

The macro-level analysis explores the broadest frame of the relationship between aspects at a micro level, the context of the situation.

The key social conditions are those emerging from Vietnam's opening up to modernization whilst concurrently maintaining its communist ideology. The government incorporates social ideology, recognition of social problems and social awareness in the production of slogan texts.

From the addressee's side, the process of interpretation is made by the Vietnamese people who interpret the message of the slogans and act accordingly in society. The Vietnamese people interact with government slogans in the 21st century context. The macro-level explanation is summarized in the following diagram.

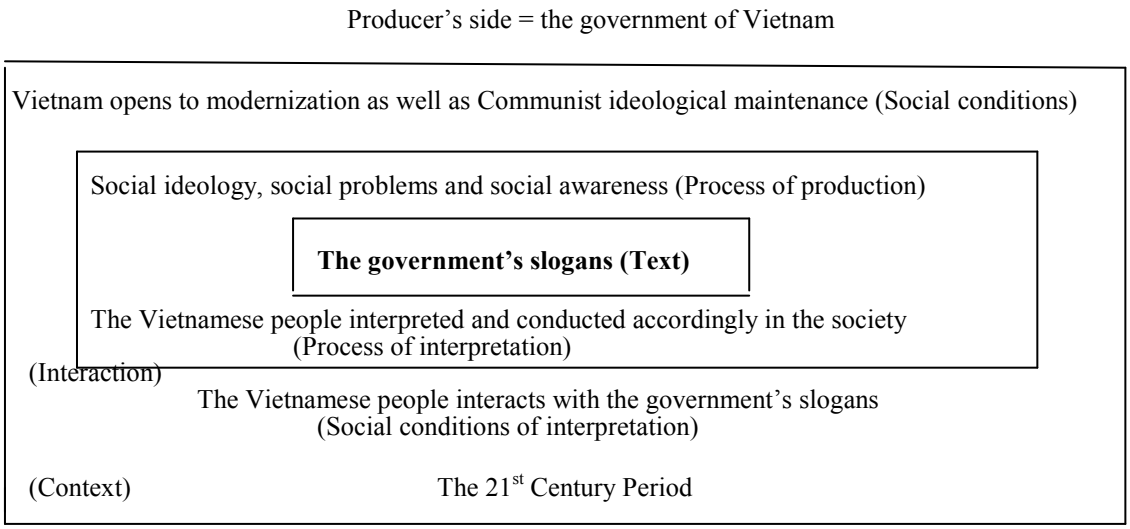


Diagram 1: Vietnamese government slogans, interaction and contexts applied using Fairclough's concept: Discourse as text, interaction and context (1989:25)

6. Summary

In the 21st century, there are new social problems confronting Vietnam resulting from the economic growth and development of the country, not encountered in the past. New ideas or policies must be incorporated into the ongoing process of development, and the government meets these in part with an ongoing campaign of public slogans propagated through various media. A survey of recent slogans finds examples relating to HIV/AIDS protection, to reduce road accidents and to tax collection among others. Moreover, the Vietnamese national political ideology is found in slogans with themes of praise the former President Hồ Chí Minh, to mobilize the people with revolutionary concepts, to encourage people for National Assembly election and to praise the Communist Party and the country

The themes of these slogans are expressed using formulaic vocabulary and simple syntactic structures at a micro level. These are generally of two types, imperative commands, and equational statements that license new ideas and practices by linking them positively with establish norms or contrasting them to undesirable states. Thus there is effectively a repertoire of simple syntactic templates that map neatly onto the conceptual structures in the formation of slogans.

References

Anderson, Benedict R.O'G. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism.* London: Verso.

Baepler, Paul. 2003. "White slaves, African masters." *Annals of the American Academy of Political and Social Science* 588(1):90-111.

Crystal, David. 1992. *A Dictionary of Linguistics and Phonetics.* The third Edition. Oxford: Blackwell Publishers.

Duranti, Alessandro. (1997). *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Fairclough, Norman. (1989). *Language and Power.* New York: Longman.

Fairclough, Norman. (1993). *Discourse and Social Cghange.* Cambridge: Polity Press.

Hornby, A. S. 1995. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
Knight, Lorna and others (ed.). 1994. *Collins Concise Dictionary and Thesaurus.* Glasgow: Harper Collins Publisher.
Lakoff, George and Mark Johnson. 1980. *Metaphors we live by.* London: University of Chicago Press.
Mannheim, Bruce & Dennis Tedlock 1995: *Dialogic Emergence Culture.* Illinois: University of Illinois.
Phumplab, Morragotwong. 2007. *Ho Chi Minh's Rhetoric: The power of language in the first Vietnamese War of Liberation (1946-1954).* Unpublished M.A. thesis in Linguistics. Faculty of Graduate Studies, Mahidol University, Thailand..
Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and company.
Srichampa, Sophana. 2007. "Vietnamese propaganda reflections from 1945-2000", *Mon-Khmer Studies* 37: 87-116.

Websites

www.businessweek.com/news/2010-09-15/vietnam-steps-up-investigations-into-companies-tax-avoidance.html, searched on 9 May 2011.
www.chinhphu.vn
www.cpv.org.vn
www.gdt.gov.vn/gdtLive/Trang-chu/Tintuc/News?contentId=135926&location=tct&location=tct, searched on 5 June 2011)
www. HIV-AIDS in Vietnam- the Current Situation, the National Response, the Emerging Challenges - Population Reference Bureau.mht.
www. Taipei Times - archives.mht
www. UNAIDS Viet Nam - HIV-AIDS, Drug, Prostitute Situation 2009 to Develop Complicatedly, searched on 24 October 2009.
www.unaids.org.vn/sitee/index.php?option=com_content&task=blogcategory&id=13&Itemid=27, searched
www.Vietnam-ACE Electoral Knowledge Network, 24 October 2009.
www.Vietnamnet.vn, searched on 5 May 2011
www.vietwater.com/market.html, searched on 21 August 2011.

The Vietnamese Expression of BODY and SOUL: A Cognitive and Cultural Linguistic Study

Ly Toan Thang
Vietnam Institute of Lexicography and Encyclopedia, Hanoi

1. Introduction

The BODY and SOUL are the basic concepts or categories of human beings. In linguistics they link to the trinity: Semantics – Cognition – Culture. Particularly, they are the key concepts of cultures and are expressed in natural languages by lexical items. The concept SOUL reflects "naïve" folk in nature believes of speakers of and at the same time usually "encyclopedic" philosophical influences of religious doctrines. As considered by some Vietnamese scholars, sometimes it is difficult to know what "naïve" is and what is "encyclopedic" embodied within them (see Nguyễn Văn Huyên 1934, Trần Quốc Vượng 2000).

However, the BODY and SOUL concepts are not in focus of attention of linguistic studies, especially in culturally and cognitively oriented linguistic researches: we can find, to my knowledge, only some publications on this topic of English, Russian and Chinese languages (see: Wiezbicka 1991, 1992; Tan Aoshuan 1999; Uryson 1999). It is not adequate, because *body* (*soul*, *mind*) are superordinate/hyperonyms, while *face, heart, finger* etc,… are subordinate/hyponyms. Cognitively speaking, we can not understand a such concept as HAND ("concept profile" – by Langacker 1987) without the concept BODY ("concept base/domain" – by Langacker 1987). It is the reason why the only concept BODY is chosen to investigate in this paper.

Moreover, if we compare an Indo-European language (like English or Russian) and a non-Indo-European, Austroasiatic one (like Vietnamese) some interesting differences will be found in the "world-view" embodied in conceptualizing physical and spiritual entities (as body and soul) of these language communities. Whereas in Western languages and cultures BODY, MIND, and SOUL are popular trilateral opposition and relationship; but in Vietnamese another picture that is the "diptych" of BODY-SOUL and BODY-MIND. In such situation, a case study of the lexical items expressing the concepts BODY and SOUL in an Austroasiatic language – Vietnamese – can examine and highlight the universal and specific characteristics of these concepts.

2. Vietnamese concepts of body and soul

In Vietnamese there are, at least, eight basic, monosyllabic terms denoting human body and soul:

+ For body: *người*, *mình*, *thân*, *xác*…
+ For soul: *hồn, vía, phách, vong*…

From these basic terms are formed some compounds like:

+ For body: *thân thể, thể xác, thân xác, thân mình, mình mẩy*…
+ For soul: *linh hồn, vong hồn, âm hồn*…

Ly Toan Thang. 2011. "The Vietnamese Expression of BODY and Soul: A Cognitive and Cultural Linguistic Study." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.127-133.

It is very interesting that in the lexical body-soul opposition a monosyllabic term can only be opposite to another monosyllabic one and the compounds-terms are in the same situation: there is no case in which a monosyllabic term is in opposition to a compound-term. Look at, for instance, following Vietnamese expressions:

(i) *Con người ta có hai phần: **hồn** và **xác***
Human being we have two parts: **soul1/spirit** and **body3**
'A person has two parts: the soul and the body'

(ii) *Con người ta có hai phần: **linh hồn** và **thể xác***
Human *being* we have two parts: **spirit/soul6** and **body5**
'A person has two parts: the spirit and the body'

The difference between these expression is that in Vietnamese, as a rule, a polysyllabic word (a compound in this case) is often more academic, more solemn, more precise than a monosyllabic one of the similar meaning. Moreover, in the (ii) the polysyllabic word *linh hồn* prototypically has the meaning 'spirit', and in the (i) the prototypical meaning of the monosyllabic word *hồn* is 'soul'.

2.1. The BODY concept

Unlike English or Russian, in Vietnamese there are a lot of lexical items which denote different aspects of the cognitive category of human body:

thân 'body1' (Sino-Vietnamese element, from Chinese *shen*)
người 'body2' (canonical Vietnamese)
xác 'body3'(canonical Vietnamese)
thân thể 'body4' (Sino-Vietnamese elements, from Chinese *shen ti*)
thể xác 'body5'/ *thân xác* 'body6'/ *thân mình* 'body7' (a Sino-Vietnamese element + a canonical Vietnamese element); *mình mẩy* 'body8'(canonical Vietnamese)

Those terms mentioned above are synonyms in the sense that they are physical/material and mortal part of person; they are real thing as opposite to invisible soul/spirit. On the other hand, they are also differentiated each from other by some criteria reflecting our world-view.

2.1.1. Body as physical structure of person, living or dead:

a) Living body:

In Vietnamese there are two monosyllabic words designating human living body: *thân* 'body1', *người* 'body2'. *Thân* 'body1' is used in many idioms, phraseologies and folk poems:

*Thương người như thể thương **thân***
love people as love **body1**
'Love people as love your **body**'

***Thân** em như hạt mưa sa*
body1 girl as drop rain fall
'The fate of a girl like a fallen drop of rain'

When a person bears seriously unhappy things, hardships or miseries, Vietnamese people refer to these events using the idiom *khổ **thân*** :

*Nó còn bé mà bố mẹ thì mất rồi, khổ **thân** nó!*
she still/yet small but parent then die already poor **body1** she
'She is a baby but her parent have died, poor her!'

The term *người* 'body2' also is used to denoting a living body. We can say in Vietnamese:

Nó *đau* *khắp* ***người***
he/she hurt all **body2**
'He felt a pain in his whole body'

Người *nó* *mệt* *mỏi* *rã rời*
body2 he/she tired be worn out
'His whole body is deadly tired'

It should be noted that this word *người* has the first sense as 'man/person/human' and the sense 'body' is only a derivation from it. In this respect Vietnamese is opposite to English because we can speak in English *everybody* (*anybody, nobody*) instead of *every man*.

b) Dead body:

For a dead body one should used the term *xác* **'**body3**',** for example:

Công an *tìm thấy* *một* *cái* ***xác***
police find one CL **body3** (of a died man)
'The police found a (died) body'

Just this term *xác* together with the term *hồn* 'soul1' composes the core opposition of BODY – SOUL concepts in Vietnamese: *xác - hồn*" 'body3 - soul1'.

2.1.2. Body as the part of person, opposite to his/her mind/spiritual part:

In this case in Vietnamese there is a term *thân* 'body1'. Speaking of Buddhism or philosophical foundations and practical exercises of meditation, Yoga or Zen, Vietnamese people often say about that we have, first of all, to relax *thân* 'our body' (from Chinese *shen*) and only after that we can relax *tâm* 'our mind, spirit'(from Chinese *xin*). Just this term *thân* together with the term *tâm* 'mind' composes the core opposition of BODY – MIND concepts in Vietnamese: *thân - tâm*" 'body3 - mind'.

2.1.3. Body viewed from the anatomical perspective.

In school biology textbooks, for example, to refer the human body in general the compound *thân thể* (Sino-Vietnamese, from Chinese: *shen ti* 'body') must be used:

Thân thể *người ta* *chia làm* *ba* *phần:* *đầu,* *mình* *và* *chân tay*
body4 people divide three part: head, torso, and leg/feet hand/arm
'Human body is divided into three parts: head, torso and (two) arms, (two) legs'

It should be stated that in this case there is no compound denoting non-physical, spiritual part of person that comes into the opposition to *thân thể*.

2.2. The SOUL concept

Unlike English and Russian, in Vietnamese there are a lot of lexical items which encode different aspects of the cognitive category of human soul:

hồn 'soul1, spirit1'(Sino-Vietnamese element, from Chinese *hun*)
vong 'soul2, spirit2'(Sino-Vietnamese element, from Chinese *vun*)
viá 'spirit3' (canonical Vietnamese)
phách 'spirit4'(Sino-Vietnamese element, from Chinese *po/phun*)
hồn viá 'soul5, spirit5'(a Sino-Vietnamese element + a canonical Vietnamese element)
bóng vía 'spirit5, heart'(both canonical Vietnamese elements)
linh hồn 'soul6'(both Sino-Vietnamese elements)
âm hồn 'soul of a dead'(both Sino-Vietnamese elements)
vong hồn 'dead person's soul'(both Sino-Vietnamese elements)

Four of the first monosyllabic words mentioned above are synonyms in the sense that as they are opposite to body, they are non-physical/immaterial spiritual part of a person. On the other hand, they are also differentiated by various features, discussed below.

2.2.1. Soul as non-physical, spiritual part in general of person, in opposition to his/her body.

Three lexical items listed above specifically concern the SOUL concept - *hồn* 'soul1, spirit1', *vía* 'spirit3', *phách* 'spirit4'. Over the world many people often think that the soul or spirit of a person is something immaterial which is not his/her body and which is going away from body to the another world. It is the same for Vietnamese culture: when a person dies Vietnamese people believe that for him/her it must be that *hồn lìa khỏi xác* 'the soul has parted from the (dead) body'. In funeral orations speakers often desire and pray that the soul of the dead person can escape this world and move back to the eternal world.

At the same time, Vietnamese people also think that even in a living body there exists the soul and spirit. A good linguistic evidence here is the use of some idioms expressing extremely strong emotions/feelings of a person, for example, when he/she is frightened or saddened too much:

Từ	*ngày*	*vợ*	*chết,*	*trông*	*nó như*	*kẻ*	*mất*	***hồn***
from	day	wife	die	look	he as	person	lose	**soul1**

'From the day of his wife's death, he looks like a person being still stunned/being out of his mind'

Lúc	*ấy*	*tôi*	*sợ*	*quá,*	***hồn vía***	*lên*	*mây*
moment	that	I	fear	too much,	**soul1 spirit3**	go up to	clouds

'That moment I am frightened out of my wits'

Nhìn thấy	*ma,*	*ai mà*	*không*	***hồn***	*xiêu,*	***phách***	*lạc*
see	ghost ,	anyone	not	**soul1**	slope,	**spirit4**	get lost

'Facing ghosts, anyone is frightened out of his senses/is scared to death'

2.2.2. Soul as the coming back of a dead person to the world.

Just as we use three terms *hồn, vía, phách* to refer to the soul or spirit as the non-physical, spiritual and immortal part of person, for the soul/spirit of a dead person who comes back to the world we can use only the term *vong* 'soul4, spirit4'(from Chinese *vun*). In the folk belief of Vietnamese people the soul after the person's death can continue to exist in another world - *cõi âm* 'the world of the dead', and our universe is constructed by the principle:

Trần	*sao,*	*âm*	*vậy*
Earth/world of the living	how,	Hades/nether world	so

'Like the land of the living, like the land of the dead'

Consequently many people believe that souls/spirits of dead persons can come back into the real world and we can raise (call up) their souls/spirits and communicate with them. In this case people often call those souls/spirits by the term *vong*, for example:

Vong	*đã*	*về*	*chưa ?*
soul4	already	return	not yet?

'The **soul4** has come back?'

If the dead person's soul has retuned, there can occur the phenomenon "*vong nhập*/*nhập hồn*" that is the soul moves into the body of a close relative who will speak and behave as the dead person revived.

2.2.3. Soul as immortal part in opposition to other mortal parts of human spirituality.

As in Western cultures, many Vietnamese people believe that *hồn* - the soul of a person - continues to exist immortally after their body dies. But, on the other hand, Vietnamese also consider a further difference between:

a) Spiritual, immortal part - *hồn* 'soul1': this spiritual "entity" is **independent** of body; entering the body it creates the life and psychology of a person; and goes away immediately after a person's body has died. Many Vietnamese people believe that souls/spirits of our ancestors are normally "living" in their graves, but that by by praying to them on their death anniversaries or at New Year festival they can come back to their altar in the house.

b) Spiritual, immortal parts - *vía* 'spirit3' and *phách* 'spirit4': these two invisible "entities" are **dependent** on body; they are parts creating the spirit of a person; and when a person dies they continue to stay in his/her body and vanish together the body.

In this sense they are more "material" than *hồn* 'soul1'. Some Vietnamese scholars note that *vía* 'spirit3' and *phách* 'spirit4' have the function to link soul and body (when the body dies they go away respectively). If the soul flies up to the Heaven, the mortal spirits move down into the Earth with the dead body.

Vía 'spirit3' and *phách* 'spirit4' can be used as synonyms. But *vía* 'spirit3' is more near to *hồn* 'soul1' than *phách* 'spirit4': Vietnamese people avoid saying about person's soul when he/she is living, so they can call his/her birthday by the term *ngày vía* 'day of spirit' that is the day this person (as a baby) receives his/her soul.

In Vietnamese, speaking about an extremely strong fear we can say:

*sợ hết **hồn*** (fear lose **soul1**) 'be frightened out of one's wits'
*sợ hết **vía*** (fear lose **spirit3**) 'be frightened out of one's wits'

but we can not say:

* *sợ hết **phách*** (fear lose **spirit4**) 'be frightened out of one's wits'

A young man can bully a boy so:

Liệu	***hồn**!*
think about/be suitable	**soul1**

'Mind!/Look out!/Beware of what you do!'

And to avoid unhappiness for babies Vietnamese people often say:

Nói	*trộm*	***vía**,*	*dạo*	*này*	*cháu*	*lớn*	*rất*	*nhanh*
speak	secretly/stealthily	**spirit3**,	time	this	little boy	grow	very	quickly

'Be stealthy for his spirit, he grows very fast these days'

But in both two cases we can not us the term *phách* 'spirit4'.

More important is that in comparison with *phách* 'spirit4', *vía* 'spirit3' has many variants. In the Vietnamese folk conception of soul and spirit some people believe that:

1) Every person has three *hồn* 'soul1'.

2) Man has seven *viá* 'spirit3', but a woman has nine. When a person passed out or lost consciousness, some people believe that his/her soul and spirit have gone away and we have to call *ba hồn bảy vía* 'three souls, seven spirits' (for men) or *ba hồn chín vía* 'three souls, nine spirits' (for women) to help soul and spirit coming back into person.

3) A person can have *vía lành* 'good spirit3' or *vía dữ* 'bad spirit3', so it can bring unhappiness. A baby cries when he/she meets a man/woman, and some people can think that this man/woman has a bad mortal spirit. In the morning, when the shop has just been opened if a first customer is bargaining very long time but finally does not buy any thing, the owner can burn a paper piece after customer's going out off the shop, and it means that the owner likes the unhappiness going away together with the mortal spirit of customer.

Finally, *phách* 'spirit4' combines with *khí* 'vigour' in the compound *khí phách* 'spirit, character' as in:

Đó là một người rất có ***khí phách***
that be one man very have **spirit/ character**
'That is a man of great character/of spirit'

While *viá* 'spirit3' combines with *bóng* 'silhouette' in the compound *bóng vía* 'spirit, heart' as in:

Nó là người yếu ***bóng vía***
he be man weak **spirit/heart**
'He is a weak-hearted man'.

3. A contrastive view of Vietnamese, English and Russian

From the above mentioned facts and arguments, we can consider that there is a very visible difference between Vietnamese and English or Russian. This is not only the question of a number (big or small) of lexical items referring to human body and soul in every language, but it is the way of conceptualizing and encoding these parts of person in those languages. It is clear that in English and Russian, for instance, there is no difference (encoded by monosyllabic words) between living or dead body, between immortal soul and mortal spirits as in Vietnamese. The difference also is embedded in the senses of semantic structure of word as in the case of *hồn* 'soul1/spirit1'.

Just like the terms *soul* in English or *dusha* in Russian, the Vietnamese *hồn* 'soul1/spirit1' is polysemantic. Its semantic extension and its particular senses are as follows:

hồn1 : soul/spirit of person.

hồn2 : expressing strong feelings/emotions in a piece of art (picture, poem, song, film); for example: *Bài hát hay phải có* ***hồn*** "Any good song must have soul'.

hồn3 : typical qualities or real meaning of something make it special; for example: *Lá cờ chiến thắng mang theo* ***hồn*** *nước* 'The flag of victory brings the soul of the country'

However, unlike English or Russian the Vietnamese term *hồn* 'soul1/spirit' has no the sense 'person' as in:

Eng.: *I won't tell a* ***soul***

Rus.: *Tut net ne odnoi* ***dushi***
here not no one **soul**
'Here nobody'

In this case Vietnamese people can use a very interesting term *ma* 'ghost' as in:

Chả có ***ma*** *nào ở đây cả*
not have **ghost** which/every at here all
'Nobody here'

Moreover, in comparison with English *soul* or Russian *dusha*, the Vietnamese term *hồn* 'soul1' does not have the sense 'moral or emotional nature of person' or 'inner, psychological world of person' as in:

Eng.: *He is a man without a* ***soul***

Rus.: *U nego chistaja* ***dusha***
at he pure soul
'He has a pure soul'

To denote this invisible "entity", in Vietnamese there is a particular compound-term:

tâm hồn (heart soul) : Sino-Vietnamese elements, from Chinese: *xin hun*)

Also, in comparison with English *spirit* or Russian *dukh*, the Vietnamese *hồn* 'soul1/ spirit' does not have the sense 'imaginary creature, ghost' as in:

Eng.: *the **spirit** world, evil **spirits***

Rus.: *dobryi **dukh*** 'good spirit'*, zloi **dukh*** 'evil spirit'

For this powerful creature in Vietnamese there is a specific compound-term:

hồn ma (soul ghost) 'imaginary creature, ghost'

3. Concluding remarks

Showing up the meanings and uses of Vietnamese lexical items encoding human body and soul, it is hoped that this paper is useful for researches in terms of semantic, cultural and cognitive orientation. Especially, being done in comparison with a language of Indo-European type as English or Russian, it is very clear that in different natural languages such concepts as SOUL (and BODY) are not universal but "ethnopsychological", "ethnophilosophical" or/and "culture-specific"(Wierzbicka 1992). It has been suggested that: "such conceptualizations of experience may spark off from someone's imagination, and then be subjected to negotiation by people over thousands of years... these culturally constructed ways of conceptualizing experience as *cultural conceptualizations* "(Sharifian 2007). They should be investigated not only in frames of "traditional" semantics but in cognitive and cultural perspective, because human languages are systems through which we express the ways we conceptualise experiences of different kinds (e.g., Palmer 1996) and the semantic content of lexical items denoting those concepts depends on how speakers of a language categorize their experiences (e.g., Sharifian 2007).

References

Langacker R.W. 1987. *Foundation of cognitive grammar,* Vol. 1 : *Theoretical prerequisites.* Stanford, CA : Standford University Press.

Nguyễn Văn Huyên 1934. *La Civilisation annamite*. Collection de la Direction de L'Introduction Publique de l'Indochine.

Palmer, G. B. 1996. *Toward a Theory of Cultural Linguistics*. Austin: University of Texas Press.

Sharifian F. 2007. L1 cultural conceptualisations in L2 learning. The case of Persian-speaking *learners of English* // Sharifian F. & Palmer G. B. (eds.) *Applied Cultural Linguistics. Implications for second language learning and intercultural communication*. John Benjamins Publishing Company - Amsterdam/Philadelphia.

Tan Aoshuan 1999. *Kitaiskii kontsept dushi, ili istorija o zabytoi dushe* (The Chinese concept of soul, or the story of a forgotten soul) // *Loghicheskii analiz jazkya. Obraz cheloveka v kul'ture I jazyke* (Logical analysis of languages. Image of man in culture and language). Moskva: Izdatel'stvo "Indrik".

Trần Quốc Vượng 2000. *Văn hóa Việt Nam: Tìm tòi và suy ngẫm*. Hà Nội: Nxb Văn hóa Dân tộc – Tạp chí văn hóa Nghệ thuật.

Uryson E.V. 1999. *Dukh i Dusha: k rekonstruktsii arkhaichnykh predstavlenii o cheloveke* (Spirit and Soul: a reconstruction of archaic Russian concepts) // *Loghicheskii analiz jazkya. Obraz cheloveka v kul'ture I jazyke* (Logical analysis of languages. Image of man in culture and language). Moskva: Izdatel'stvo "Indrik".

Wierzbicka A. 1991. *Dusha* (= soul), *Toska* (= yearning), *Sud'ba* (= fate): three key concepts in Russian language and Russian culture *// Metody formalne w opsie jezykow slowianskich*. Bialystok.

_____. 1992. *Soul, Mind, and Heart // Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford University Press.. Inc.

Development of Northern Khmer Primer using Thai Alphabet: Opportunities and Challenges[77]

Siripen Ungsitipoonporn
Mahidol University, Thailand

Abstract

The purpose of this paper is to report the opportunities and challenges with creating a Northern Khmer primer based on the Thai alphabet. The Northern Khmer language spoken in Thailand is an unwritten language. So far, elderly Northern Khmer speakers in some areas have realized that use of their language has been declining, so they are conducting a project for the development and learning management of Northern Khmer in order to both maintain the local language and improve Thai learning at the formal primary school. The teachers who are also Khmer speakers together with linguists at Mahidol University have developed the local language curriculum (Northern Khmer) which will teach grade 4 students as part of one subject in formal education.

In creating the Northern Khmer curriculum, there are many components such as orthography development, literature production, and reading and writing materials for teaching language acquisition. A primer - an alphabet-teaching book - is one component of literature production. In the case of Northern Khmer they use the Thai alphabet because the children are already familiar with Thai orthography. However, the development of the Northern Khmer primer has not been easy because the Northern Khmer language has more vowels and clusters more than in the Thai language, creating some challenges for designing the content of each lesson. The most difficult task is how to select and arrange the lessons of the Northern Khmer primer, introducing the letters from easiest to more difficult.

Introduction

Northern Khmer belongs to the Austroasiatic family, in the sub-branch of Eastern Mon-Khmer (e.g. Diffloth 2005). There are about 1.4 million Northern Khmer speakers in Thailand (Suwilai and others, 2004). The Northern Khmer speakers are primarily concentrated in Northeastern and Eastern areas of Thailand, such as the provinces of Surin, Burirum, Srisaket, and some districts of Nakorn Ratchasima, Ubol Ratchathani, Roi Et, Mahasarakham, Sakaew, Prachinburi, Chanthaburi, and Chachoengsaw.

Northern Khmer is significantly different from the Standard Khmer spoken in Cambodia as follows:

Language situation

Northern Khmer is rather a big ethnic group in Thailand but now the speakers realize that their language has gradually declined in use especially among the new generation. Along with rapid change in

[77] This paper was presented at the Fourth International Conference on Austroasiatic Linguistics (ICAAL 4), October 29-30, 2009, Bangkok, Thailand. I would like to acknowledge the community research team at Ban Pho Koang school for their attempt to design their transfer primer under the project financially supported by the Thailand Research Fund (TRF).

Ungsitipoonporn, Siripen. 2011. "Development of Northern Khmer Primer using Thai alphabet: Opportunities and Challenges." In Sophana Srichampa, Paul Sidwell & Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.134-145.

socioeconomics, language policy and the power of the majority language (Thai), several ethnic minority languages in Thailand seem to be declining faster than in the past. The younger generations are increasingly becoming monolingual in Standard Thai. Although Khmer has a written language, the Northern Khmer as distinguished from the Khmer language in Cambodia (Smalley, 1994: 137), is an unwritten language. As Smalley (1976: 45) wrote that *"Virtually none of the Khmer-speaking people in Thailand are able to read the Khmer script. Those who can read, read Thai. The scripts are related but not identical. Furthermore, even they could read the script there is a great deal of the language from Cambodia they cannot understand. They have a difficult time understanding Phnom Penh radio."* Even if the Northern Khmer speakers would like to record their traditional literature using Khmer characters, it would still be problematic because there are significant differences between the sound systems of the two Khmer varieties. At present, most of Northern Khmer speakers become bilingual or speak mix Thai and Northern Khmer languages. Therefore, if the Northern Khmer children able to read and write their own mother tongue in Thai script first, it would be help them to be literate in Thai easier than do not use Thai script.

A Thai-based writing system was developed for Northern Khmer in 1964 and was modified later in 1987 (Smalley, 1994: 140). Although this writing system has been troublesome to the Thai speakers trying to read it, some literature has been published such as a dictionary, some folktales, a wordbook, conversation lessons for learning Northern Khmer, and a variety of Christian religious materials (Smalley, 1994: 140-141).

The Differences and similarities between Thai and Northern Khmer writing systems

Compared to Thai, Northern Khmer is non-tonal language with many more consonant and vowel phonemes than Thai language. At the same time, Thai orthography has 2 extra consonant series which are used to distinguish tone series, so these are redundant for writing Northern Khmer phonemes (and there is no attempt to write Northern Khmer etymologically, as is done largely for Standard Khmer). The following tables list the orthographic symbols used in both Thai and Northern Khmer for comparison.

Symbol	Initials: Thai	Initials: N. Khmer	Finals: Thai	Finals: N. Khmer
ก	k	k	k	k
ข	kh			
ค	kh	kh		
ฆ	kh			
ง	ŋ	ŋ	ŋ	ŋ
จ	c	c	t	c
ฉ	ch			
ช	ch	ch		
ซ	s	s		
ฌ	ch			
ญ	j	ɲ		ɲ
ฎ	d			
ฏ	t			
ฐ	th			
ฑ	th			
ฒ	th			
ณ	n			
ด	d	d	t	t
ต	t	t		
ถ	th			
ท	th	th		

Symbol	Initials: Thai	Initials: N. Khmer	Finals: Thai	Finals: N. Khmer
ธ	th			
น	n	n	n	n
บ	b	b	p	p
ป	p	p	p	
ผ	ph			
ฝ	f			
พ	ph	ph		
ฟ	f			
ภ	ph			
ม	m	m	m	m
ย	j	j	j	j
ร	r	r	n	r
ล	l	l	n	l
ว	w	w	w	w
ศ	s			
ษ	s			
ส	s			
ห	h			
ฬ	l			
อ	ʔ	ʔ		

Table 1: Thai orthography symbol used in Thai and Northern Khmer to represent initial and final consonants

Symbol	Thai vowels[78]	N. Khmer vowels[79]
-ะ	a	a
-า	a:	a:
-ิ	i	i
-ี	i:	i:
-ิ◌̣[81]		ɪ
-ี◌̣		ɪ:
-ึ	ɯ	ɯ
-ือ	ɯ:	ɯ:
-ึ		ɤ
-ืออ		ɤ:
-ุ	u	u
-ู	u:	u:
-◌ุ◌		ʊ
-◌ูอ		ʊ:
เ-ะ	e	e
เ-	e:	e:
แ-ะ	ɛ	ɛ

Symbol	Thai vowels	N. Khmer vowels
แ-	ɛ:	ɛ:
โ-ะ / โ-็อ[80]	o	o
โ-ะ	o:	o:
เ-าะ / -็อ	ɔ	ɔ
-อ	ɔ:	ɔ:
เ-◌าะ / -็อ		ɑ
-อ		ɑ:
เ-อะ	ɤ	ə
เ-อ	ɤ:	ə:
เ-อะ /		ʌ
เ-อ		ʌ:
เ - がยะ / เ - がย็		ia
เ - がย	ia	i:a
เ - ぎอะ / เ - ぎอ็		ɯa
เ - ぎอ	ɯa	ɯ:a
-็ว		ua
-ัว	ua	u:a

Table 2: Thai orthography symbol used in Thai and Northern Khmer to represent vowels

The project for development and learning management of Northern Khmer in Surin Province

The project for development and learning management of Northern Khmer in Surin Province was started in last two years for two main reasons: (1) the effective learning in the formal schools of Northern Khmer children was very poor, especially Thai language; (2) the older generations in the community are concerned that the Northern Khmer language and culture were declining. Fortunately, at present language preservation and revitalization are accepted around the world by UNESCO for promotion of language diversity. The view is shared in Thailand and consequently ethnic language minorities have the right to be educated in the language they speak and understand. The bilingual education programs that use mother tongue-based curriculum are being established in the minority groups in Thailand and other countries. For example, in 2003, the Chong people in Chanthaburi province became the first group in Thailand teaching their own language in the formal school for language revitalization. So far, many minority groups are interested in preserving their language as the Chong have. The project for curriculum development and learning management of Northern Khmer was initiated by teachers at Ban Pho Koang School, Surin province. The purposes of this project are the followings:

1. To survey and compile data from the Northern Khmer language and culture from the local scholars' knowledge.

2. To study and investigate the problem of the use of the Thai language among the Northern Khmer students.

---

[78] Thai vowel system is cited from กาญจนา นาคสกุล *ระบบเสียงภาษาไทย* จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพฯ 2545 หน้า 61

[79] Northern Khmer vowel system is cited from William A. Smalley Phonemes and orthography: Language planning in ten minority languages of Thailand 1976 page 57, and Prakorb Pho-ngam A Morphological study of Northern Khmer 1999 page20.

[80] The symbol -็ is represented a short vowel when compared with the same vowel.

[81] A dot under or above the vowel symbol means the tongue position of that vowel is between two vowels in Thai. For example, -◌̣ / ɪ / is realized a mid-high front vowel which is between high front / i / vowel and mid-front vowel / e /.

3. To develop local curriculum and learning management of Northern Khmer in order to improve the learning of Thai at a formal primary school.

Fortunately, a Northern Khmer writing system based on the Thai alphabet had been already devised two decades ago, it only needed to be modified. At this time, teachers in the NK schools want to develop a curriculum for teaching Northern Khmer in the formal school system.

Outcome and output of the project

The output of this project are the local curriculum for a local language subject for one semester, and the teaching materials namely, big books, small books, listening stories, big pictures, playing word cards, etc.

Pictures 1-2: Big books from teaching materials of the Northern Khmer research project (final report submitted to the TRF funding, 2010)

Pictures 3-4: Small books from teaching materials of the Northern Khmer research project (final report submitted to the TRF funding, 2010)

Orthography development

What is an orthography? Why do they use the Thai orthography to develop the Northern Khmer writing system? Susan Malone wrote that:

> An orthography is a system for writing a particular language. Developing an orthography involves selecting and testing the symbols and spelling rules (capital letters, punctuation, hyphens, etc.) to represent the important sounds of the language. The goal in orthography development is that the mother tongue speakers of the language will approve the orthography and use it consistently. Another goal is that orthographies for minority languages will be acceptable to appropriate government departments.
> Malone (2006: 49)

The Northern Khmer orthography using Thai letters has been established since the 1960s, and later revised. This is because native speakers wanted to write Northern Khmer in their own mother-tongue but various attempts were not successful by themselves. It was used in some materials with some adaptation, for many years, but not extensively. They want to be able to use a suitable writing system to record and preserve their culture; for texts such as folktales, proverbs, ceremonial and healthcare publications, including diglot booklets.

The speakers who worked on the Northern Khmer project, decided to use the Thai-based orthography, the following factors are important:

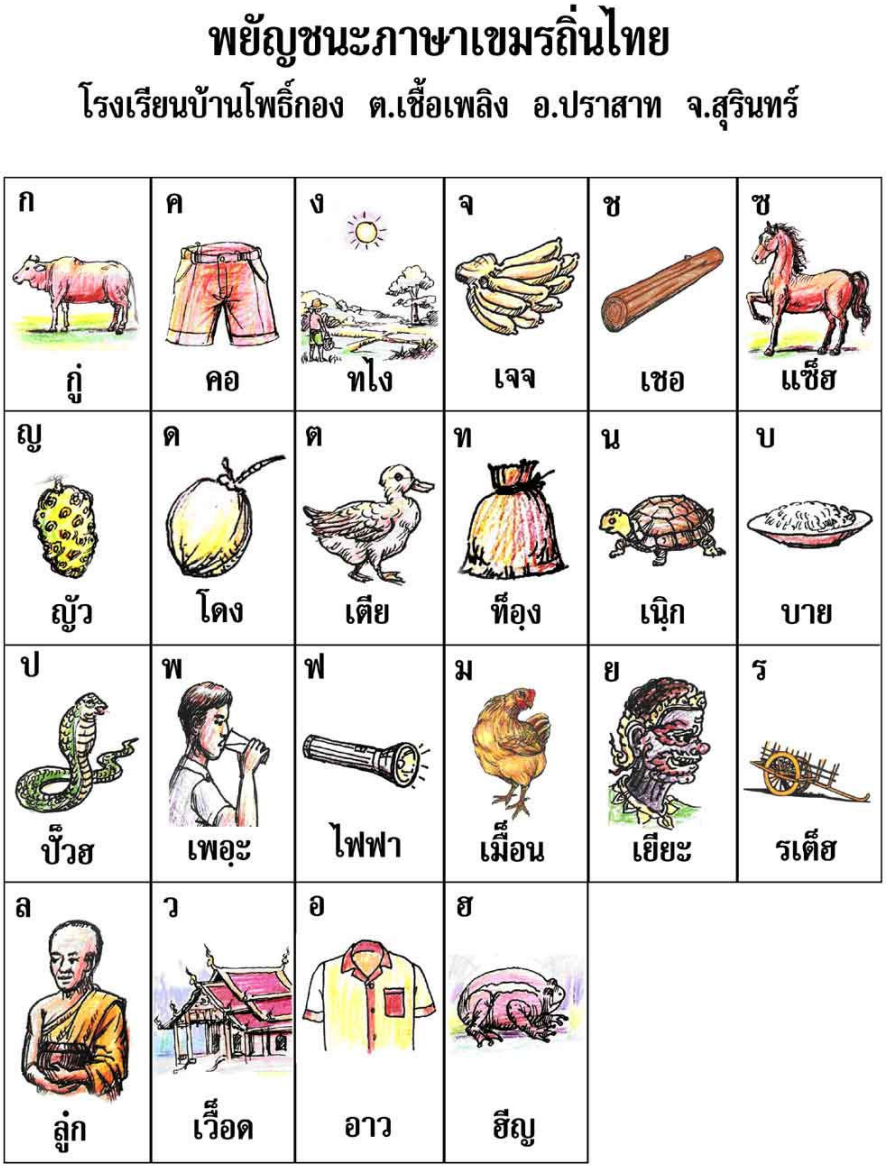
(1) The output, after finishing this project will form the local curriculum, with community knowledge, and will be used to teach Northern Khmer as a subject in the formal school system. If they use the official language characters there will be no problem with official acceptance.

(2) The curriculum will be used to teach grade 4 students who have already learned the Thai alphabet. This means that the students have basic knowledge in the Thai writing system which is useful for learning the Northern Khmer writing system because the mother-tongue language orthography is similar to the majority language orthography. This will make it easier to transfer back and forth between languages.

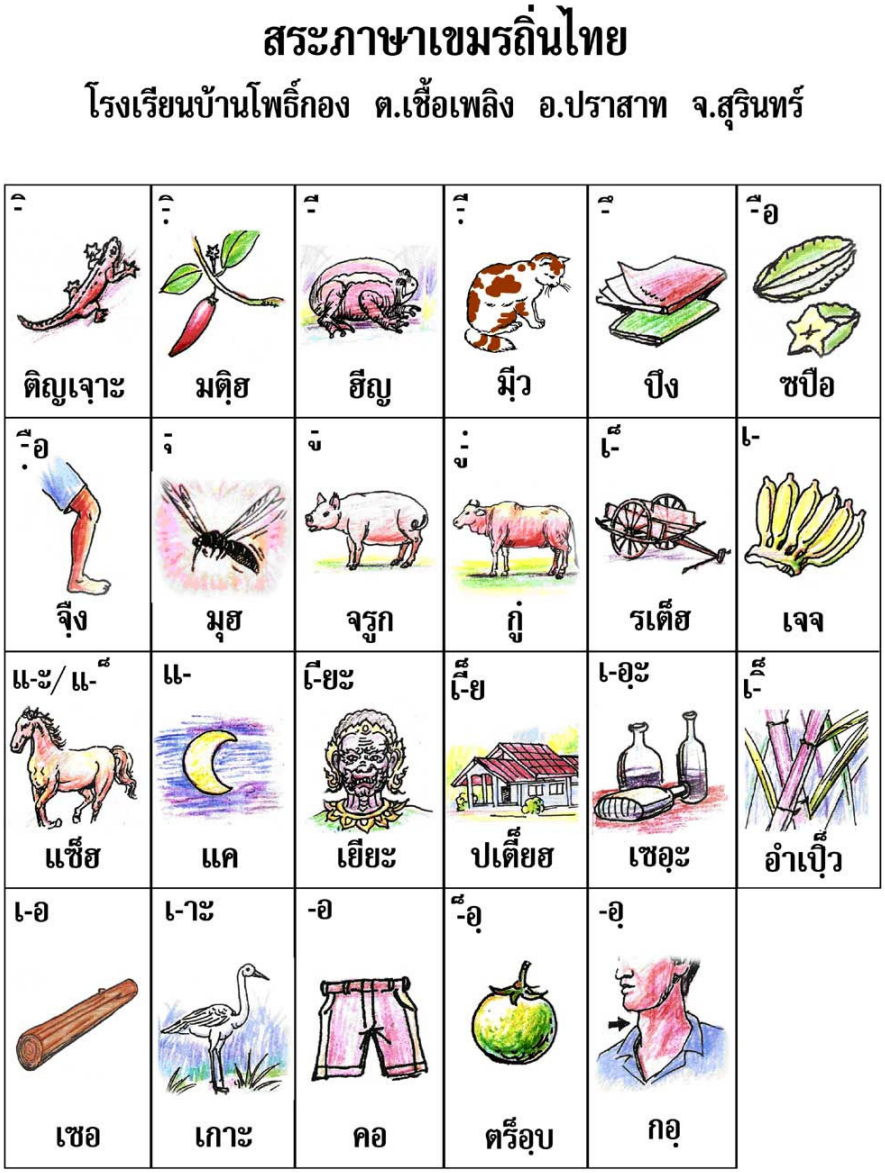
(3) It will help Northern Khmer children identify the differences between the Thai and Northern Khmer languages so that they can use both languages correctly. Moreover, it is expected that by this way the teacher will help students achieve more effective Thai learning.

(4) A Thai-based orthography of Northern Khmer already exists, it only needs to be revised and put into practice.

The Thai-based Northern Khmer orthography is acceptable to Northern Khmer speakers in this project because they participated in the development themselves. Then, they identified key words for each phoneme together with picture drawings. Finally, the phonemes and key words were arranged into alphabet charts as follows:



Picture 5: Example of initial consonant chart of Northern Khmer from teaching materials of the Northern Khmer research project (final report submitted to the TRF funding, 2010)



Picture 6: Example of vowel chart of Northern Khmer from teaching materials of the Northern Khmer research project (final report submitted to the TRF funding, 2010)

The curriculum components and teaching materials

A curriculum is a resource for teachers to help the learners achieve their goals using appropriate teaching materials and learning activities. Why does the community-based education program help children quickly acquire a second language? The reasons can be summarized as follows:

(1) They learn first from basic knowledge in the community and move to more difficult knowledge. In other words, they learn from the communicative language to academic language.

(2) Lessons are taught in the mother-tongue language in which the learners can understand and speak, making them confident in their abilities

(3) When the learners have a strong foundation, they can continue learning in the majority language.

The community- based curriculum components are divided into two tracks, namely, a story track which emphasizes meaning and communication and focuses on whole texts, and a primer track which emphasizes on accuracy and correctness and focuses on the parts of the language. Four elements in language education programs (listening, speaking, reading, and writing) are essential in helping students gain fluency for learning language.

Examples of teaching materials in the story track are shared reading such as big books and small books, listening story, experience story, creative writing, and total physical response (TPR). A primer or an alphabet-teaching book is the material for the primer track.

Making the Northern Khmer primer

A primer is a teaching alphabet book which is used to teach sounds of the language, help new readers get acquainted with the letters, and learn to get meaning from printed texts. The learners will understand that the letters of the text are associated with sounds because letters go together to form words, and each word has meaning (Susan Malone, 2006).

How to develop the Northern Khmer primer?[82]

For the first step, it is important to know that the primer teaches only one new letter at a time. So in this step of creating the Northern Khmer primer, it was necessary to identify from the writing system how many letters are the same as, or diiferent from, the Thai writing system. The sounds and letters that are the same in both languages were selected to be taught first because in this case the Northern Khmer children have already learned the Thai writing system. More than one letter for each lesson can be selected if they have the same sound and same shape in both languages. It is also important to identify which letters are more frequently used so that the keyword of each lesson can be selected. Frequency is determined by counting each letter from the texts or stories which have been written in that minority language, at least 3-4 pages. Then, the letters are listed in order from the most frequently used to the least frequently used.

The second step is thinking of an object, which can be drawn as picture, which is familiar to the learners to be used as the keyword or the main sentence of each lesson. The letters are limited by the number of the lesson. For example, in lesson one more than one letter can be chosen but not too much. In lesson two, when another letter is chosen for creating the keyword, the letters from both lesson one and two can be used. Again, when lesson three is created the letters in lesson one, two, and three can be used, and so on. Only one new letter per lesson is permitted if that letter is different from the national language. This is a challenging task.

For the last step, each lesson must start and end with meaning. This means starting with a picture, then introducing the keyword (the name of object in the picture). After that, the keyword is broken down to the new letters and built up to form the word. Then, the new letter is used to build another new word so that the learner can know other new words from this method. Not only are new words built from the new letter, but new sentences as well. For practice time, the learners use the new letters and words for spelling and hand writing. For the Northern Khmer primer, which is a transfer primer[83] from the Thai language to the Northern Khmer language, it is not necessary to start with one word, but sentences may be used.

---

[82] Adapted from Reading transfer: A Practical Guide to making transition materials, Georgia G. Hunter, SIL International,1994 and Guide to developing a Primer to teach the sounds of the Language, Susan Malone, SIL International, 2006

[83] Transfer primer means a teaching alphabet book which is intended to help a person already literate in one language to become literate in another language that he speaks (Georgia G. Hunter, SIL International , 1994)

Components of a primer

The components of each lesson should be:

(1) new letter or letters.

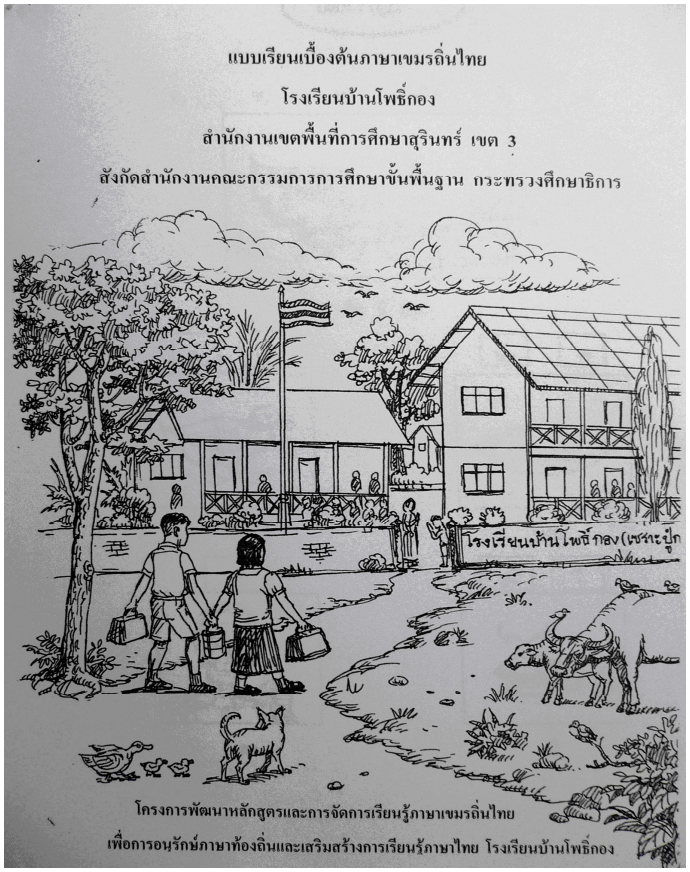
(2) the keyword and its picture or main sentence with a picture.

(3) building new words and making sentences for practice.

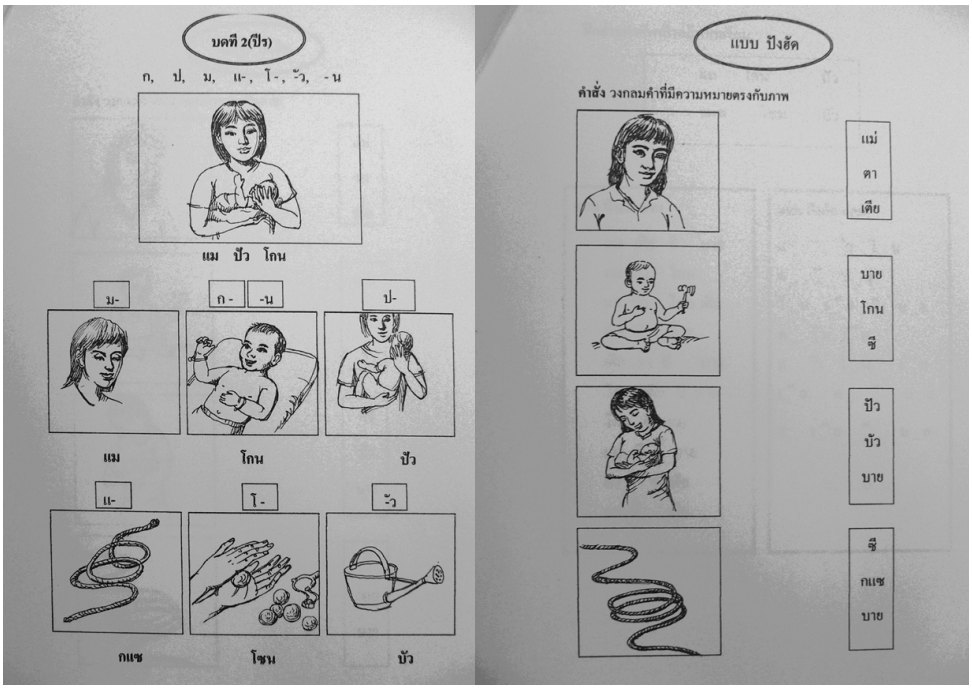
(4) various exercises after the content such as matching word and picture, writing exercise, reading exercise, etc.

A tentative Northern Khmer Primer

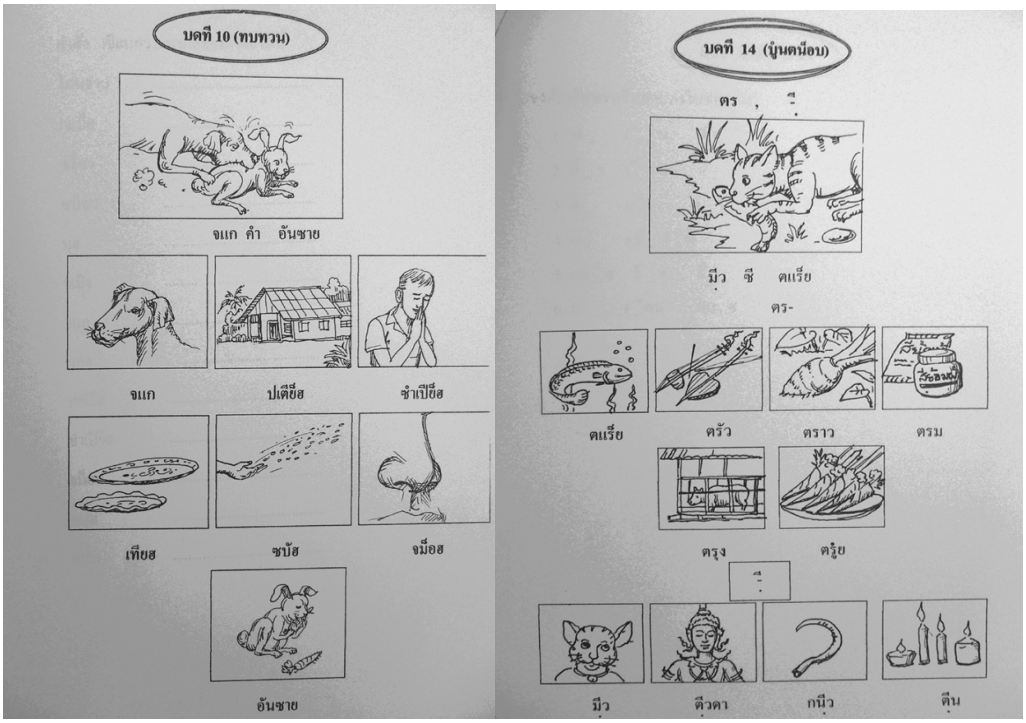
The first tentative Northern Khmer Primer, used to teach for one semester, is composed of 20 lessons.



Picture 7: A tentative Northern Khmer Primer from teaching materials of the Northern Khmer research project (final report submitted to the TRF funding, 2010)



Pictures 8-9: Example of content and exercise in some lessons of the tentative Northern Khmer Primer (from final report of Northern Khmer research project submitted to the TRF funding, 2010)



Pictures 10-11: Example of content in some lessons of the tentative Northern Khmer Primer (from final report of Northern Khmer research project submitted to the TRF funding, 2010)

The opportunities and challenges of creating the Northern Khmer primer based on the Thai alphabet

Northern Khmer phonology is rather complex, especially the vowel system, and quite different from Thai phonology. Although most sounds are the same or similar to Thai, there are a lot that are different., therefore, the Thai alphabet does not have enough symbols to use for Northern Khmer writing without some adaptation. The Northern Khmer speakers solved this problem by using a dot placed under or above the letter, if that vowel is similar to, but differs from the equivalent Thai vowel.

The principle that only one new letter should be taught at a time is not strictly possible, even for the letters that differ from the Thai writing system because sounds always go together with other sounds. For example, the final sounds *-c, -r, -l* and *-h*, which are in the Northern Khmer phonology, always occur with vowels that do not exist in Thai. Thus, at least two new letters need to be introduced in one lesson instead of only one letter. Unlike Chong or Nyahkur, in which the phonological systems are not as complex as Northern Khmer.

Content production of each lesson is very difficult in the Northern Khmer primer. According to the limitation in selecting letters of each lesson, we found that trial and error is needed to determine if the letter selected is suitable for each lesson or not. If not the letter must be changed until the contents of all lessons is complete. Thus, the native speakers who developed the primer had to be much more tolerant and not easily discouraged. The work of the Northern Khmer speakers is very admirable, that they can complete this difficult task even when having to revise the contents over and over without giving up.

Conclusion and Discussion

Developing a primer to teach sounds of the minority languages using a Thai-based orthography is not easy in some ethnic groups such as Northern Khmer. In our experience it depends on the differences between the minority language and the Thai language. It seems that the more complex the phonological system, the more difficulty in creating a primer. The guidelines for developing the primer need to be adapted for each minority language in order to get a suitable primer of each language. From my experience in developing a primer in several minority groups, I think the Northern Khmer language has been the most difficult in the Thailand context. So far, the group has produce only twenty lessons of a primer but there are still a number of letters not used in the lessons. If they can finish, it will be very useful for teaching the Northern Khmer language to the next generations. They will become proud of themselves and this will help in the long-term preservative of their language.

References

David, Thomas and Wanna Tienmee. "Acoustic Study of Northern Khmer Vowels" *Mon-Khmer Studies* 16-17: 99-107.

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In: Sagart, Laurent, Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Routledge/Curzon. 79-82.

Dorothy, Thomas. 1990. "Northern Khmer Orthography Workshops. Surin, 1986-1989" *Mon-Khmer Studies* 16-17:255-265.

Hunter, Georgia G. 1994. Reading transfer: A practical guide to making transition materials. Summer Institute of Linguistics, Inc.

Krashen, Stephen. 1997. *Why Bilingual Education?* http://www.ericdigests.org/1997-3/bilingual.html

Malone, Susan E. 2004. Manual for Developing Literacy and Adult Education Programmes in Minority Language Communities. Bangkok: UNESCO Asia and Pacific Regional Bureau for Education.

Malone, Susan E. 2006. Planning community-based education programs in minority language communities: Resource manual for Mother Tongue speakers of minority languages engaged in planning and implementing education programs in their own communities. Unpublished, used as resource for writing the Manual for Developing Literacy and Adult Education Programmes in Minority Language Communities, published by UNESCO-Bangkok, 2004.

Prakorb, Pho-ngam. 1999. "A Morphological study of Northern Khmer." Ph.D. dissertation, Research Institute for Languages and cultures of Asia, Mahidol University.

Smalley, William A. 1976. "The Problem of Vowels: Northern Khmer" *Phonemes and Orthography: Language Planning in ten minority languages of Thailand.* Canberra: The Australian National University.

Smalley, William A. 1994. *Linguistic Diversity and National Unity*. Chicago: The University of Chicago Press.

Stringer, Mary D. and Nicholas G.Faraclas. 2001. *Working together for Literacy: A Guide book for Local Language Literacy Programmes*. SIL Australia: The Print Connection.

Suwilai, Premsrirat.1997. "Linguistic Contributions to the Study of Northern Khmer Language of Thailand in the last two decades" *Mon-Khmer Studies* 27: 129-136

Suwilai, Premsrirat and other. 2004. *Phɛɛn thii phaasaa khɔɔŋ klum chaattiphan taaŋ taaŋ naj pratheet thai*. Bangkok. (in Thai)

Suwilai Premsrirat. 1997. "Linguistic Contributions To the Study of the Northern Khmer Language of Thailand in the Last Two Decades. " *Mon-Khmer Studies* 27:129-136.

Final report of the project for curriculum development and learning management of Northern Khmer at Ban Pho Koang school, Surin province submitted to Thailand Research Fund (TRF) 2010.

กาญจนา นาคสกุล, 2545. **ระบบเสียงภาษาไทย. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.**

Special Session on Iambic Effects in Austroasiatic Languages

Sesquisyllabicity in Khmer

John Haiman
Macalester Colllege

The well-known Sesquisyllabicity of Khmer, and other Mon-Khmer languages (Henderson 1952, Jenner 1969, Huffman 1972, Matisoff 1973, Thomas 1992) is on the one hand, the outcome of a productive process, the never-ending erosion of the unstressed initial (minor, anacrusic) syllable of bisyllabic words.[84] This process ends, if unopposed, not in iambic words like *ka-pow*, but in the total elimination of the anacrusic syllable (as in *pow*). In that sense, the predominantly iambic word structure of Khmer is only a way-station, and a fortuitous one[85], on the way to the total monosyllabicity of Vietnamese and the Viet-Muong languages in general within Mon-Khmer (Ferlus 1975), the monosyllabicity of Khmer dialects in the Mekong delta that are under the strong influence of neighbouring Vietnamese (Thach 1999), and possibly, the monosyllabicity of Thai within the family of languages including Thai-Kadai and Austronesian, perhaps all under the influence of Chinese (Benedict 1942). Beyond that, not even the monosyllabic word is exempt from further reduction: the end result of erosion is a word whose single syllable has the form CV (Matisoff 1973).

On the other hand, the iamb has become an independent canonical form of the Cambodian word: not only anacrusic syllable reduction, but a number of other processes, both phonological and morphological, both reductive and constructive, seem to conspire to contribute to the creation and maintenance of this favorite or target structure. Polysyllabic words are ground down, but monosyllabic words are also built up, to conform with it.

In both senses, sesquisyllabicity would seem to work against another well-attested canonical tendency in Khmer and other languages of Southeast Asia: the creation and maintenance of symmetrical compounds (Nguyen 1965, Nacaskul 1976, Vongvipanond 1992, Ourn & Haiman 2000, Haiman & Ourn 2009, Haiman 2009, 2010a,b). A sesquisyllabic word is an iamb, while a bisyllabic symmetrical compound is a spondee[86]. Whether these potentially antagonistic tendencies ever get to interact has apparently never been discussed. I will present evidence that they do.

Section 1 of this essay will be devoted to a summary description of the productive processes of initial syllable reduction in Khmer, and says nothing new, cf. in particular Huffman 1972 and the pedagogical grammars of Noss 1966 and Huffman 1970. Section 2 will deal with the iamb as a target structure, and the less widely recognized functional unity of various constructive processes which seem to conspire to build up iambic structures. Finally section 3 will deal with the clash between the drive for iambic structures and the

1. Gregerson 2007 provides a succinct and careful summary of the various reduction effects of the anacrusic syllable in Mon-Khmer, including total loss, and (presumably en route to this) the transition from strong to weak stress, with the concomitant loss of both vowel and consonant inventories (paradigmatic reduction) and vowel and consonant sequences (syntagmatic reduction) and prosodic independence.

[85] Ken Gregerson suggests that this "fortuitous way station" may be compared to a "support level" in the stock market -- a kind of barrier, which seems to resist erosion, but, once broken, entails panic and collapse. No one can doubt the psychological reality of such support levels: everybody noticed when the DJ index sank beneath 10000 after the crash of 2008.

[86] A spondee is a metrical foot consisting of two long syllables (by weight or by stress) as in James Shirley's lines "your heads must come/ to the cold tomb".

Haiman, John. 2011. "Sesquisyllabicity in Khmer." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.147-161.

drive for parallelism between adjacent words. The conclusion is that the processes described in sections 1 and 2 are subject to the overarching requirement to maintain symmetry in decorative compounds.[87] [88]

1. Reduction

1.1. The assimilation of borrowings

Khmer has borrowed heavily from Indic languages and some borrowings are instantly recognizable. Many, however, have been shaped to conform to the iambic canon. Long initial pretonic syllables become short or disappear. The leftmost forms below occur in careful reading pronunciations only. In the starred examples, the spoken reduction of the initial vowel is reflected in alternative spellings in the conventional orthography as well as in the pronunciation:

1.	*aːcaː* > *acaː*	'priest'	(<P. *aacariya*)
	aːkaːh > *akaːh*	'air'	(<P. *aakaasa*)
	aːkia > *akia**	'building'	(<Skt. *aagaara*)
	aːknee: > *aknee:**	'Southeast'	
	aːrawm > *arawm*	'feeling'	(< Skt. *aarammaŋa*)
	aːtet > *atet*	'sun'	(< Skt. *aadicca*)
	aːvut > *avut*	'weapon'	(< P. *aavudha*)
	baːrej > *prej*	'cigarette'	
	boːraːn > *boraːn*	'ancient'	(< P. *puraaŋa*)
	piːbaːk > *pibaːk*	'difficult'	
	piːntu > *pintu*'	'grade'	
	saːlaː> *sala*:	'hall'	

Short initial vowels are replaced by schwa or disappear. In the following words, there is no careful vs. casual alternation possible.

2.	P. *kapala* > *kbaːl*	'head'
	P. *pati* > *pdej*	'husband'
	Skt. *same* > *smaeu*	'same'

[87] An initial draft of this paper was written, and delivered as a talk, during a sabbatical fellowship at the Max-Planck Institute for Evolutionary Anthropology in Leipzig in 2003-4. I am grateful to Bernard Comrie and to the staff at the institute for their generous hospitality. Thanks to Noeurng Ourn and Veasna Keat for all the hours they have spent teaching me Khmer since 1996. Deep thanks also to Ken Gregerson and Jim Matisoff for their practical and insightful suggestions and gentle corrections of some blunders in the version of this paper which was delivered in Bangkok in 2009. The practical orthography employed below was developed by Stephanie Farmer and myself, working with Mr. Keat, on a Macalester Keck grant in 2007, and will, I hope, reduce my own risk of apopleptic seizures in future dealings with copy-editors and proofreaders. I apologize to purists who may risk some apoplectic seizures of their own with this orthography, but readers who know Khmer will easily recognize {aw} and {au} as the first and second register default vowels, {e} as both epsilon and a neutralized vowel which occurs before syllable-final palatals and [v] (e.g. {tev} 'go', {cenj} 'exit', {tec} 'a bit', {bej} 'three'), {ee} as [e]. Schwa as the offglide in other diphthongs (after nuclear u, i, w, o, and e) and in most unstressed syllables is rendered by {a}. For expository purposes, in this article the phonetic version of schwa in anacrusic syllables is represented by [a] (as opposed to its written form {aw} or {au}), while epenthetic schwa which breaks up word-initial consonant clusters is here represented by a space (as opposed to its written form, which is nul). Both of these would surface in our new transliteration as {a}, that is, like schwa in unstressed syllables in general.

[88] Editors' note: regarding the author's use of his "practical orthography" in this paper instead of the International Phonetic Alphabet (IPA), which was otherwise required of all contributors for this and its companion volume: this was not agreed to lightly. Various other contributors went to great lengths to comply with our editorial preference, we believe that those efforts greatly contributed to the consistency and usefulness of these volumes, and for this we express our gratitude.

1.2. Reduction of initial word in two-word compounds

This reduction affects the initial element of not only compounds, but phrases. As Khmer is a head-initial language, modifiers are stressed and heads are destressed. Thus in NP like *tngaj nih* 'today', the nuclear stress falls on *nih* 'this', and the head noun *tngaj* 'day' is destressed.

Long syllables become short:

3.	*a:* + *ckuat* → *ackuat*	'blockhead'
	a: + *kaw:* → *akaw:*	'swindle'
	kha:ng+ceu:ng → *khang ceu:ng*	'North'
	ta:j+haong → *ta(:)j haong**	'unnatural death' (Thai borrowing)
(?)	*ba:n#cia* → *bawnjcia*	'command'[89]

The rhyme potions of short syllables are replaced by schwa or disappear:

4.	*muaj+ daw:ng* → *mdaw:ng* 'once'	
	muaj + *rau:j* → *mhau:j*	'one hundred'[90] (the r~h alternation is a lenition)
	muaj + *tngaj* → *mngaj*	'one day'
	mwn + *dael* →*mdael*	'never'
	pon +*ma:n* → *pma:n*	'how much, many'
	pon+tae → *ptae*	'but'
	pram+bej → *pmbej*	'eight'
	preah + *ʔawng* → *pʔawng*	'lord'
	tev lee:ng → *tlee:ng*	'go visit'

1.3. The reduction of initial syllable of native disyllabic words

To begin with, the inventory of possible syllable types in the anacrusic syllable is already severely reduced relative to what is possible in the main syllable (Huffman 1972: 55, Haiman 1998:612):

- The only possible syllable onsets are Stop + (r).
- The only possible syllabic nucleus is the default vowel {*aw*} ~ {*au*} (pronounced as schwa, that is [a] in the present transliteration, except before nasals in second register words, where it surfaces as [u]).
- The only possible coda is a nasal.

In allegro pronunciations, and in the invariable practice of younger speakers and less literate ones, this structure is further reduced (Noss & Proum 1966 passim, Huffman 1970 passim, Huffman 1972:59-61, Haiman & Ourn 2003a:157). Among the recurrent processes that can be identified in initial syllable reduction are the following:

Allegro Rule #1: Glottal Stop + V + Nasal → (Syllabic Nasal)

The glottal stop is a phoneme in Khmer, but is lenited in word-initial position to near-silence. Perhaps for this reason the reduction of unstressed initial syllables with glottal stop onsets proceeds differently from that of initial syllables that begin with other stops.

---

[89] Argued in Haiman 1999: *ba:n cia* 'cause to be' is a phonetically impeccable, and semantically plausible source, for *bawnjcia* 'command'. (More generally, *ba:n* 'get' may be the lexical source for the causative prefix *bVN-*. For a skeptical rejoinder, cf. Enfield 2001.)

[90] This alternation occurs more regularly in Kiengiang Khmer, with concomitant tonogenesis: *rian* > *hian* (low tone) 'study, learn' (Thach 1999:87). Elsewhere in Khmer, {', h,r, v} constitute a natural class of "velars", /N/ becoming [ng] before each of them.

5. *ʔawmbeul → mbeul* ‘salt’
ʔawmpeu: → mpeu: ‘action’
ʔawnsaw:m → nsaw:m ‘steamed cake’
ʔawngka:l → (ng)ka:l ‘when’
*ʔ*awngkaw: → ngkaw: ‘rice’

In a number of common words with initial stops other than the glottal stop, it is again the initial stop, rather than the syllable rhyme, which is deleted: this happens regularly when the initial consonant of the unstressed syllable is identical with the initial consonant of the main syllable (Jacob 1990:25), and sporadically elsewhere: *bawntec → ntec* “a bit”, *tunsa:j → nsa:j* “rabbit”, *runjcuaj (dej) → njcuaj (dej)* “(earth)quake”, *kawngkaep → ngkaep* “frog” (avoiding **kakaep*), *pnom penj → mpenj* “Phnom Penh” (avoiding **papenj*). [91]

Allegro Rule #2: Initial Stop (+ r) + vowel → initial stop + (schwa)[92]

6. *krawma: → kama:* ‘scarf’
prawhael → pahael ‘approximately’
praulwm → palwm ‘dawn’
trawlawp → talawp ‘return’
siavphev → saphev ‘book’

(Note that Allegro Rule #2 , which is exceptionless, deletes the post-initial string /rV/ from the anacrusic syllable.)

Allegro Rule #3: Initial Stop + V (+ Nasal) → voiceless initial stop + (schwa) (+Nasal)

7. *bawmra:m → pra:m* ‘warning’
bawntup → patup ‘room’
bawngha:nj → paha:nj ‘show’
bawngʔaem → pangʔaem ‘sweet, dessert’
dawdael → tadael ‘same’
kawnda:l → kada:l ‘middle’
kawnlaeng → kalaeng ‘place’

Note that Allegro Rule #3 deletes the post-initial string /VN/ in the anacrusic syllable. Most likely, the causative prefix *bawN-* is reduced to *p-* (via an intermediate *pauN-*) by this same rule.

8. bawng+kaeut → pakaeut ‘create’

Both [bangkaeut] and [pakaeut] are acceptable pronunciations of ‘create, originate’, spelled {bawng + kaeut}. Virtually every causative spelled with initial {*bawN-*} exhibits this alternation. For a relatively small number of causatives, such as

9. {*paunjaul*} ‘explain’
{*paunleut*} ‘extinguish’
{*paungri:k*} ‘develop’

only the {*pauN-*} spelling is acceptable[93], and for yet others such as {*p coap*} ‘join’, only the {*p-*} spelling is

[91] Again, this kind of reduction occurs regularly in the Kiengiang variety: *kawmphwh > mphwh* ‘shrimp’, *cawngkra:n > ngkra:n* ‘kitchen’ , possibly followed by the total disappearance of the nasal, especially in the speech of younger and less literate speakers (Thach ibid.83-4).

[92] In Kiengiang, this cluster simplification occurs regularly in main stressed syllables, with concomitant tonogenesis: *srae > se:* (low tone) ‘paddy’ (Thach ibid. 87).

[93] The full causative prefix is {bawN} (first register), while the reduced but still syllabic prefix is written as {pauN} (second register). I have no explanation for this. The distinction collapses, of course, in fully reduced pronunciations of the prefix.

acceptable. There are a handful of what seem to be purely orthographic doublets: there seems for example to be no difference in meaning between the spellings

10. {bawn +dual} and {p+dual} both meaning 'cause to fall, knock down'
{bawn+sawm} and {p+sawm} both meaning 'unite, join'.

In a smaller handful of cases, the contrast between the two pronunciations has become lexicalized:

11. {bawng+rian} means "teach", but only {p+rian} can mean 'edify'.

It is hard to know what to make of apparent cases of near-synonyms such as in the statement (from a grade school primer):

12.

Lbah	*psawm*	*kaeut*	*laeung*	*daoj sa:*	*bawnsawm*	*lbah*
sentence	join	arise	go=up	because	join	sentence

'Conjoined sentences arise through sentence conjunction.'

Are *psawm* and *bawnsawm* synonyms here? Common sense would suggest that they are. There are three different paths for the lenition of syllable-initial /r/. In the dialect of Phnom Penh (Noss 1966), in stressed syllables, initial /r/ becomes [h] (*roam → [hoam]* 'dance'), while /r/ as the second consonant in a cluster usually disappears (*craeun → caeun* 'lots'), although [h] also shows up in a small number of words (*trej → thej* 'fish'). In initial unstressed syllables, syllable-initial /r/ for speakers of all dialects undergoes allegro rule #4.

Allegro Rule #4: Initial /r/ + V (+ Nasal) → /l/ + schwa → (glottal stop) + schwa

13.

raubam → labam	'dance [N]'
raubeh → labeh	'rubble'
raubiang → labiang	'alley'
raumhaeuj ([rumhaeuj]) → lahaeuj	'cool'

Allegro Rule #5: Initial Stop (+ schwa) → zero (total loss of the anacrusic syllable):

14.

bawbaw: → baw:	'rice gruel'
bawntaw: → taw:	'continue'
knjom → njom	'I'
pteah → tea(h)	'house'

Allegro Rule #6: Initial Nasal + schwa → zero before a homorganic nasal (again, total loss)

15. *maumi:ng maumeang → mi:ng meang* 'dazed on waking up'

2. The iamb as target structure

2.1. The assimilation of borrowings

To conform with the iambic target, pretonic syllables are reduced, as noted. But this is not all. As polysyllabic suffixing languages, Pali and Sanskrit have many words with post-tonic syllables and these syllables are dropped in assimilated loans (Gorgoniev 1966:36-43, Jacob 1990:48).

16.

jana > cian	'nation(ality)'
jiivita > civeut	'life'
vajra > pec	'diamond'
vaktra > pheak	'face'
jaya > cej	'victory'
bhuumi > phu:m	'village'
purusa > burauh	'man'
karana > ka:	'deed'

kamma > kam	'action'
khatta > khet	'province'
krodha > kraot	'angry'
ksanta > ksan	'serene'
ksata > ksawt	'poor'
lesa > leh	'pretext'
laabha > lo:p	'greedy'
sattva > sat	'animal'
dravya >troap	'goods'
dvaara > tvia	'door'
yanta > jon	'machine'
yodheti > jut	'fight'
uttaraa > utdaw:	'North'
vipatti > vi'bat	'disadvantage'

(Note that this process of adaptation does not so much create iambs as it does eliminate trochees (metrical feet consisting of a strong syllable followed by a weak one). It creates iambs, but it also creates monosyllabics.)

2.2. An epenthetic schwa (represented here by a space) can be inserted between the two consonants of a word-initial cluster:

17.	*tnaot → t naot*	'sugar palm'
	lngung → l ngung	'stupid'
	lhong → l hong	'papaya'
	lngiac → l ngiac	'evening'
	lpev → l pev	'pumpkin'
	pteah → p teah	'house'

Gorgoniev (1966:32) calls this insertion of epenthetic schwa (which could be either voiced or voiceless, appearing as [h]) following voiceless initial consonants), one of the most "essential" characteristics of Khmer. Virtually any combination of consonants is possible word-initially, frequently as a result of the allegro rules, and schwa insertion occurs between C1 and C2 unless C1 is /s/, or C2 is a liquid or /h/ following a stop (Martini 1942-5:125). This rule alone, which creates a reduced initial syllable, is responsible for converting an enormous number of "phonologically" monosyllabic words into phonetically iambic words in Khmer. (The scare quotes around "phonologically" reflect the underlying uncertainty in speakers' minds about the "true" cognitive status of word-initial consonant clusters. This section is largely about that uncertainty.)

Sometimes a more substantial insertion is possible. Gorgoniev also notes (ibid. 48) that the (meaningless and according to him often interchangeable) initial consonant "prefixes" (that is, simply the letters *k-,c-,s-, t-, m-* occurring before other consonants) may be "reenforced by nasalization", that is, by the insertion of a meaningless string /VN/ after that initial consonant (cf. Haiman & Ourn 2003 for meaningless {-Vm(n)-} insertion), or (presumably) "reenforced by rhotacization", that is, by the insertion of the string /rV/ in the same position. (Gorgoniev provides examples of the phenomenon for both changes, but uses the word "reenforcement" only in the first case.) Some examples of meaningless /rV/ insertion are:

18.	*kngaeng → krawngaeng*	'bent, deformed (arm)' ([krangaeng])
	kngawng → krawngawng	'bent, curving' ([krangawng])
	ksae → krawsae	'string, line' (the latter considered a more elegant form)

Prakorb 1992 reports that in the dialect of Northern Khmer, of which he is a native speaker, the meaningless string / rV/ can be optionally inserted between virtually any two word- initial consonants. The

implications of this stunning assertion should be pondered by all students of Khmer; possibly it should be repeated as a mantra several times daily.[94]

It is noteworthy that the optionally inserted strings /VN/, /Vm/, and /rV/ described by Gorgoniev, Prakorb, and Haiman & Ourn for Central and Northern Khmer are exactly the strings that are most regularly eliminated by anacrusic syllable reduction. In those cases where the inserted string is etymologically motivated, its pronunciation can be regarded as simply a careful restoration of an inherited underlying structure. In other cases of analogical extension, the inserted string has a purely decorative function. This seems to be the case in pairs of words like:

19. *c(awm)ʔeung* 'gaunt, bony'
c(aum)peuːh 'crooked'
k(awm)hoːc 'wily, naughty, depraved'
k(awm)sawt 'poor, wretched'
s(awm)kaum 'gaunt'

Again, it is difficult to know how to interpret near-synonyms in passages such as

20.a) *via daeu dawmrawng teuk*
it go directly water
'It heads directly for the water.'

20.b) *nej cam (nej trawng)*
meaning main meaning direct
'Its main (that is, literal) meaning..'

In English, "direct" and "literal" are quite different words. But it is not clear that in Khmer *dawmrawng* and *trawng* are thought of as different at all. At the very least, it can be asserted that the infix has some "stylistic" but no syntactic or derivational function.

In Haiman 1998, 2003, it was proposed that the nearly productive and generally meaningful derivational infix {-*Vm*(*n*)-} may owe its origin to exactly this insertion process. An initially meaningless decorative restoration has been morphologized to create in some cases meaningful infixes denoting nominalization (in hundreds of words like *k-awmn-aeut* 'birth'), causation (as in dozens of words like *s-awm-lap* 'kill')[95], and finally, apparently meaningless infixes in a number of other words that should be nominalizations or causatives by the look of things but aren't (an ostensibly derivational procedure which results in no derivational change of state). This is the state of affairs we called 'syntactic backsliding' in Haiman & Ourn 2003. In other words, if this hypothesis about the origin of -*Vm(n)*- is correct, then the most productive infixation process in Khmer is itself an iambic or sesquisyllabic phenomenon. Like epenthetic schwa insertion, it converts a monosyllabic word into a disyllabic one[96].

Prakorb 1992 confirms that in Northern Khmer, a comparable morphologization may have affected even the epenthetic schwa inserted between word-initial contiguous consonants. Thus schwa in N. Khmer is not just a vowel that breaks up consonant clusters, as it seems to be in the examples of (17) in Central Khmer: it is a nominalizing infix in words like *t-a-hom* 'size' (compare Central Khmer *t-um-hom*) , from *thom* 'big'.

---

[94] Ken Gregerson has informed me that meaningless [r] insertion in this position is also observed in many West Bahnaric languages of NE Cambodia and SE Laos. In the Kavet dialect (known to speakers of neighbouring dialects as Kravet), Skt. Samudera ocean (=Khmer sa'mot) becomes sramot.

[95] These reinterpretations may be favored, first, by the independent existence of {'*VN*-} and {*bVN*-} as nominalizing and causativizing prefixes, and more generally, by the iconic assumption that "more (form) is more (semantic content)".

[96] In proposing the morphologization of a hypercorrect or decorative string as the possible origin for the infix -Vm(n)- in Khmer, I do not claim that there is any basis for such an account of other infixes with constant meaning, such as the instrumental -n-, which are found over a wide area within MK. The infix -Vm(n)- however, seems to be largely home-grown. This and the variety of meanings (including no meaning) with which it is associated make it a plausible contender for a morphologization account.

(It is also possible that in N. Khmer, the schwa is diachronically just what is left after prior erosion of what -- in the Central Khmer dialect -- is the nominalizing infix *-Vm-*.[97])

In all the cases of infixation dealt with so far, the infix is a analyzed as a (possibly mistaken) restoration of elided material within the root. But some of the less productive infixes may have originated as prefixes. Haiman 1998 notes that agentive {-m} occurs as both a prefix (*m-cah* 'master' < *cah* 'old') and an infix (*l-m-eu:h* 'criminal' < *leu:h* 'transgress'). Farmer 2009 suggests that the infixation of {b}, which occurs in about 70 mainly liquid-initial words like {*l-b-aeng*} "game" (< *lee:ng* 'play') (Pou 2004) , may also be the result of metathesis, derived from an earlier conjectured form **p-lee:ng*. The plausibility of this speculation is enhanced by pairs of extant words related by exactly this kind of obstruent+liquid ~ liquid + obstruent metathesis, among the *sra:j* 'untie, solve' and *r sa:j* 'come loose'. What could have motivated such a change? Farmer observes that words beginning with the cluster {Stop + Liquid} are monosyllabic (recall that epenthetic schwa is not inserted when C2 is a liquid), but the result of the metathesis does provide an input string for schwa insertion, and thus the creation of a phonetically sesquisyllabic word *l baeng* [labaeng].

The general observation holds that there is a continuing back-and-forth alternation between monosyllabic words with complex initial onsets, and disyllabic words with reduced or reducable anacrusic syllables. The space between the first and second consonants of monosyllabic words with complex onsets is both phonologically and morphologically ambiguous, the locus of an ongoing "tug-of-war" (Farmer 2009). Phonologically, the phonetic material that appears there may be the remains of a syllable coda that has been eroded by an allegro rule (in which case I have represented it here as [a]), OR it may be the result of an infixation process (represented here as [ ] an empty space). There is no warrant, in standard phonetic practice, for representing the same sound in different ways. I am doing so here to make the point that this sound has different possible interpretations. Morphologically, if the material is infixed, it may be meaningless (purely decorative), OR it may be semanticized as a derivational affix. That is, restitution may become morphologization in Khmer via a three-stage process: first a phonetic string is just that, and associated with a careful pronunciation of a word that contains it; then, as the reduced pronunciation comes to be the standard one, the added string is made into a decorative morpheme; finally, that morpheme acquires a meaning. While some students, like Pinnow 1957, Jacob 1960, and Lewitz 1968 have demonstrated the diachronic priority of disyllabic structures in a number of instances (e.g. Old Khmer *kanmee:ng* > Modern Khmer *kmee:ng* 'youth' [k mee:ng]), there is no solid basis for doing the same synchronically. Both erosion and restitution seem to be productive ongoing processes.

3. The Drive for Parallelism in Khmer

Like Malay, Thai, Burmese, Lao, Vietnamese and many other languages of the region, Khmer evidences a love of purely decorative or non-referential symmetry[98]. Words of all kinds (nouns, verbs, prepositions, conjunctions), and short phrases, are coupled with:

- (near-) synonyms including doublets from other languages (like *aches and pains*);
- meaningless decorative echo forms ("servant words"[99] like *jibber* in *jibber jabber*);

[97] Noss & Proum 1966 insist that this is also the case in standard Khmer. They emphasize that there is a phonetic, as well as a morphological, difference between *cnga:j* 'far' and allegro *c nga:j* 'distance' (op.cit 272, 322). That is, the epenthetic schwa is a nominalizer, as is /awm/ in the careful pronunciation /c-awm-nga:j/. Huffman 1972:62 makes the same case for the distinction between slaw: 'cook' and s law: 'a stew', the latter a reduction of *s-awm-law:*. The ambiguity of schwa is reflected in the statement of the allegro rules, where a full vowel is replaced by a schwa. Given the independent necessity for a rule of epenthetic schwa insertion, the allegro rules could have been given as simple deletions, with a subsequent rule of epenthetic schwa insertion separating consonant clusters whether they originated as such or arose as a result of allegro elision. The point of section 2, however, is that speakers are forever treating this sound as ambiguous.

[98] Referential symmetry (as in correlative constructions and counterfactual conditionals) is iconic: symmetry of form reflects symmetry in the meaning (Haiman 1985, Haiman & Kuteva 2002). Decorative symmetry is meaningless. The contrast between referential and decorative symmetry is parallel to the contrast between representational and decorative art.

- synonyms which have been tricked out with etymologically unmotivated
- "Procrustean"affixes[100] which make them alliterate with the main word (like *kit,cat, and ca-boodle*);
- non-synonyms which happen to alliterate (rather than rhyme) with the main word, and are "conscripted" to accompany the main word for reasons of euphony alone (perhaps like our rhyming *loose-y goose-y, true blue).*
  (Nguyen 1965, Nacaskul 1976, Vongvipanond 1992, Ourn & Haiman 2000, Haiman & Ourn 2009, Haiman 2010a,b).

This reduplication frequently has none of the familiar functions of reduplication (cf. Moravcsik 1978): it does not convey plurality, intensification, or iteration (it is not iconic), nor disdain (like the pejorative plural-like Yiddish, Turkish, or Hindi varieties of schmo-reduplication, cf. Stolz 2008), nor does it convey irony (*yeah, yeah*), nor the playful histrionics of ideophones (Haiman, to appear). Its primary or sole function is to be symmetrical. In any event, decorative compounding creates spondees, whenever the reduplicand is monosyllabic.

In principle, it seems that there is an imminent clash between sesquisyllabicity and the drive for parallelism. Are the varied decorative compounds of Khmer themselves subject to the processes which create iambic structures or does the drive for parallelism override the drive for initial syllable reduction?

One possible outcome, one might suspect, is that parallelism is exclusively a feature of formal registers, while sesquisyllabicity arises in conversational styles only. The two tendencies might thus be effectively quarantined from each other. In fact, however, the use of decorative symmetry is a feature of both spoken and written Khmer, at every level of formality, although it is undoubtedly hypertrophied in writing and formal speech[101]. The two drives thus do get to meet in casual allegro speech.

One structural feature which suggests that sesquisyllabicity is a "respecter of parallelism" is the behavior of one variety of decorative parallel structures which to my knowledge has not been noted in the traditional literature. This is reduplication via the "Adam's rib" effect: a word is decoratively coupled not with a morphologically distinct near-synonym, nor with a meaningless echo-word, but with its own cognate accusative construction (Haiman & Ourn 2009).

kwt	'think'	is coupled with *baek kumnwt*	'turn thought'
ceh	'know'	is coupled with *mian cawmneh*	'have knowledge'
chian	'pace,step, stride'	is coupled with *bawh cumhian*	'throw a step'
saeuc	'laugh'	is coupled with *awh sawmnaeuc*	'exhaust laughter'
tlaj	'valuable'	is coupled with *mian dawmlaj*	'have worth'
ʔuat	'boast'	is coupled with *bawnjcenj ʔawmnuat*	'emit boasts'

and so forth. What is curious is the order in which these conjuncts appear. Contrary to the nearly universal law of lengthening constituents (Behaghel 1932, Malkiel 1959, Hawkins 1994), it is the cognate accusative which almost invariably comes first in pairs of this sort:

---

[99] These are known as *bo'ri'va: sap* 'servant words'. They were first described in Maspero 1915, and have been extensively catalogued by native grammarians, among them Sisovat 1972, Chiang Caun 2002, and Chun Leuh 2007. There is a running debate, first, on whether they are nonce formations, and second on whether they are meaningful, with eminent authorities taking opposing views. Sisovat asserts that they are meaningful, Chiang that they are not. Maspero 1915:226, Gorgoniev 1966:73, and Chiang 2002:10 claim that they can be made up on the spot, while both Mr. Ourn and Mr. Keat maintain that they are learned like other words. For an initial critical assessment of some positions, and an attempt to characterize the form of *bo'ri'va: sap* according to generative principles, see Farmer 2008.

[100] Native grammarians like Ieuv-Koeus 1945 and the monolingual Dictionnaire cambodgien of the Institut Bouddhique recognized meaningless affixes (like Gorgoniev's "reenforcing" {r} and {N} infixes, and what I call "Procrustean" adaptations in some "servant word" compounds, cf. Haiman 2010b) as *akum* -- 'magic, incantatory' (< P. *aagamana* 'scripture') letters.

[101] This may be most compactly illustrated with a single example. The symmetrical compound *creah croam* "chow down" includes a servant word as its second member, and is at the same time extremely vulgar.

21. a) *bawh cumhian chian* 'step' (and not **chian bawh cumhian*)
    b) *awh sawmnaeuc saeuc* 'laugh' (and not **saeuc awh sawmnaeuc*)

Whatever the reasons for this may be, it is notable that the starred order (a light constituent followed by a heavy one) is "already" an asymmetrical structure of the sort favored by the pressure for sesquisyllabicity. The phonetic processes of allegro speech would make short work of the short initial constituent. The typologically deviant existing order (a heavy constituent flowed by a light one), on the other hand, offers some "padding" against the phonetic reduction of the initial element, and helps to maintain a balance between conjuncts.

Another striking symmetry-preserving feature is the behaviour of disyllabic reduplicants. The prevailing pattern seems to be that reduction of the unstressed initial syllable occurs either "across the board", that is to both conjuncts, or to neither one. For example, we encounter

22. a) *{kawnteunj kawntonj}* (both conjuncts full, represented as written) and
    b) [*kateunj katonj*] (both conjuncts reduced, represented as spoken), but neither
    c) **kawnteunj katonj* (only the second reduced) nor
    d) **kateunj kawntonj* (only the first reduced) for 'short and squat'[102];

23.a) *krawnge:k krawngak* (both conjuncts full) and
    b) *kange:k kangak* (both reduced), but neither
    c) **krawnge:k kangok* (second reduced) nor
    d) *kange:k krawngok* (first reduced) for 'zigzag';

24.a) *bawbaw: bawbiak* (both conjuncts full) and
    b) *pabaw: pabiak* (both reduced), but neither
    c) **bawbaw: pabiak* (second reduced) nor
    d) **pabaw: bawbiak* (first reduced) for 'rice gruel';

25.a) *crawpi:h crawpeu:h* (both conjuncts full) and
    b) *capi:h capeuh* (both reduced), but neither
    c) **crawpi:h capeu:h* (second reduced) nor
    d) **capi:h crawpeu:h* (first reduced) 'crooked';

26.a) *cawmpi:h cawmpeu:h* (both full) and
    b) *capi:h capeu:h* (both reduced), but neither
    c) **cawmpi:h capeu:h* (second reduced) nor
    d) **capi:h cawmpeu:h* (first reduced) for 'crooked' again.

(Note in (25) and (26) how the same root *cpeu:h* 'crooked' exhibits what I would call decorative infixation of /rV/ and /Vm/ with no noticeable meaning alternation.)

This enforced symmetry of reduction or non-reduction is an extremely common pattern. Other paired sesquisyllabic words which behave in a similar fashion include:

---

[102] These judgments were not, unfortunately, collected on the fly from spontaneous utterances. Instead they were elicited in the following manner. First I copied about 300 utterances from a variety of written sources (folk tales, novels, magazine articles) in which these symmetrical compounds had occurred. The consultants in separate interviews spaced out over several sessions, then looked over these, and eliminated those which included any vocabulary or constructions that were not in their active vocabulary (a loss of between 10 and 20 utterances from the data base). Then I asked the consultants to read each passage twice: the first time slowly, the second time quickly. After reading through the first two examples, the consultants volunteered that speaking style was different from even rapid reading style, and happily agreed to produce speaking versions instead of simply allegro reading versions of the texts. This resulted in three versions of each pair. Every production in all styles was symmetrical. I then asked whether they would accept the constructed asymmetrical results of (22)-(26) in which one or another of the conjuncts was reduced (but not both). Their smiles of bewildered and condescending derision have been translated into the asterisks you see.

27. *dawmrae dawmrev ~ tamhae tamhev* 'straighten out, correct'
dawmnam dawmnaoc ~ tamnam tamnaoc 'plants'
crawbo:t crawbawl ~ cabo:t cabawl 'complicated, confused, corrupted'
cunghi:k cungho:k ~ cahi:k caho:k 'bumpy, pitted, with potholes'
cawmki: cawmkuat ~ caki: cakuat 'crazy'
cawmnae cawmnam ~ canae canam 'recognize, remember'
cawmteut cawmto:ng ~ cateut cato:ng 'hind end uppermost'
cawmʔeh cawmʔah ~ caʔeh caʔah 'insult, cuss at'
cawmʔeung awmʔaeng ~ caʔeung caʔaeng 'emaciated'
trawmawng trawmaoc ~ tamawng tamaoc 'solitary, lonely'
akrawk akrej ~ akvawk akej 'bad, nasty" (irregular r→v lenition)
bawmpia bawmpian ~ pampia pampian 'force, violate'
cawcee:c cawca:c ~ cacee:c caca:c 'chatter, gossip'
kawmpeuk kawmpok ~ kapeuk kapok 'small, trivial'
kawndaeng kawndaoc ~ kadaeng kadaoc 'lonely'
kroam kria ~ koam kia 'internal wounds, depression'
krawhee:m krawhaw:m ~ kahee:m kahaw:m 'red'
krawha:j krawhawl ~ kaha:j kahawl 'feel hunger pangs'
krawhee:ngrawho:ng ~ kahee:ng kaho:ng 'full of holes'
ʔawnteah ʔawnteang ~ nteah nteang 'restless'
ʔawndaet ʔawndo:ng ~ ndaet ndo:ng 'reverie'

In fact, since reduction is a matter of degree rather than an all-or-nothing phenomenon, we can go further and see that both conjuncts must be reduced to exactly the same degree. *Bawbaw:* "rice gruel" can be partially reduced, to *pabaw:,* as illustrated above. But it can be further reduced, by the total loss of the anacrusic syllable, to *baw:*. The accompanying decorative echo-word can be reduced in the same way, from *bawbiak* to *pabiak* or to *biak.* But both conjuncts must agree in the degree of reduction. Other symmetrical compounds like {*bawbaw: bawbiak*} are:

28. *dawmnae dawmneung* '*information*' (decorative)[103] → *tamnae tamnawng* → *tanae tanawng*

dawmrae dawmrawng '*straighten*' (decorative) → *tamvae tamvawng* → *tamae tamawng*

dawmʔae dawmʔawk '*loiter*' (decorative) → *tamʔae tamʔawk* → *taʔae taʔawk*

bawmpheut bawmphej '*scare*' (decorative) → *pampheut pamphej* → *papheut paphej*

bawmplec bawmpla:nj '*destroy*' (decorative) → *pamplec pampla:nj* → *paplec papla:nj*

bawngkhia bawngkheang '*prohibit*' → *pakhia pakheang* → *khia kheang*

In addition to the allegro rules already described in section 1, there is another final one which seems limited to paired sesquisyllables. We have already seen that /rau/ lenites to [la] in anacrusic syllables. In pairs, further lenition of [la] ~ [l](that is, from the phonetic sequence [l] + schwa)] whether from /rau/ via erosion or from original /l/ via epenthesis) to [a] is frequent:

[103] It is notable that decorative forms sometimes precede and sometimes follow the meaningful words with which they are coupled. All of the decorative forms in (28) happen to precede.

29. *l vaut l vaun → avaut a vaun* ‘supple, sensuous’
l veung l veu:j → aveung aveu:j ‘vast’
rauli:m rauliam → ali:m aliam ‘mischievous’
raunji: raunjoa → anji: anjoa ‘tremble’
raungi: raungeu: → angi: angeu: ‘dazed’

Again, either both conjuncts are reduced or neither one is. That is, like movement out of a coordinate structure in syntax (Ross 1967, Haiman 1985), reduction of the anacrusic syllable via any kind of erosion cannot disturb the symmetry of a coordination. Note that not only rules of reduction but rules of construction are limited in the same way. We have seen that epenthetic schwa can be inserted between elements of an initial consonant cluster. Thus *lngi: lngeu: → l ngi: l ngeu:* ‘stupid’. The initial /l/ can be elided, yielding [angi: angeu:]. The initial schwa is due here to epenthesis, not to erosion. No matter: either both conjuncts have it, or neither one does.

Another feature which suggests, however, that sesquisyllabicity is “no respecter of parallelism” is the behavior of distributive or repetitive total reduplications like *psee:ng psee:ng* ‘various’. Huffman (1970:186) notes that “repetitive compounds are typically pronounced (especially in rapid speech) with reduced stress on the first element and full stress on the second element” (although with no apparent stress-driven segmental modifications). Such reduplications are iconic, rather than decorative, however, and it is notable that iconic reduplication need not be symmetrical within Khmer. In fact, frequentative reduplication is eminently sesquisyllabic:

ka:j ‘dig up, scratch’ yields the frequentative *kaw-ka:j* ‘scrabble’, etc.

The crucial remaining instances are those where both conjuncts are monosyllabic and occur in decorative symmetrical compounds. These are not so numerous as paired sesquisyllables, but there are hundreds of such compounds, like *klia klej* ‘sentence’ (whose second element is an alliterating reduplicant), and *ruac haeuj* ‘escape finish’ (both elements meaningful and nearly synonymous) . What happens to the initial element in such compounds? (Reduction of the onset cluster in such cases goes beyond sesquisyllabicity, but the same processes which erode the anacrusic syllable also affect the initial clusters of monosyllables. Recall that result of erosion is no just monosyllables, but monosyllables of form CV). Remarkably, we encounter the same pattern. Either both conjuncts are reduced, mostly by the changes limited to the Phnom Penh dialect r → h, Cr → C:

30. *cro:m cre:ng → co:m ce:ng* ‘prop up from all around’
crul cruah → cul cuah ‘slip out of bounds’
criav cruanj → ciav cuanj ‘wrinkled’
crial cree: → cial cee: ‘start to set’ (sun)
kria kraw: → kia kaw: ‘poor’
kriam kroam → kiam koam ‘parched’
kriat kreu:m → kiat keu:m ‘rough, hasty’
ree: ria → hee: hia ‘back and forth’
rwang ra:v → hwang ha:v ‘affair’

Or neither one is:

31. c*ha:v chee:v* (no reduction occurs) ‘impatient’
ciat cev (no reduction) ‘tasty’
ckoam ckau:ng (no reduction) ‘out of joint’[104]
cliav cla:t (no reduction occurs) ‘intelligent, cunning’
dac daoc (no reduction occurs) ‘ragged’

[104] Given the rule of epenthetic schwa insertion, the conjoined elements here are phonologically monosyllabic, but phonetically bisyllabic. In allegro speech, this epenthetic schwa may be subject to deletion, but in the examples I have heard, the schwa is present in both conjuncts. (I have been able to identify no allegro rule which reduces C+schwa to C.)

kliav klawm	'flawless'
knan knoan	'offended'
khwl kho:c	'depraved'
kho: khev	'violent'
lhae lhaeuj	'alleviate'

The only sporadic exceptions to this pattern, where one of the conjuncts exhibits some reduction, while the other does not, occurred when it was the second element that was reduced:

32.	*cwn prwn*	→ *cwn pwn*	'Chinese' (total loss of [r])
	khaeng re:ng	→ *khaeng haeng*	'haughty' (lenition of [r] to [h])

However these spotty examples are to be accounted for, they clearly do not bear witness to the power of iambicity.

4. Conclusion

It seems that Khmer exhibits a system of formal grammatical agreement between symmetrical conjuncts. Not even the pressure to destress and reduce the first syllable of a disyllabic word can violate this propensity for agreement. To the contrary: reduction is just one more of the features that symmetrically conjoined constituents share. In fact, it is more than that: if decorative compounds alliterate in the first place, the examples of (22)- (31) demonstrate that they must agree in (non-)reduction of their initial syllables (if disyllabic) or of their initial consonant clusters (if they are monosyllabic).

The phenomenon is strikingly similar to that of incipient number agreement in Puerto Rican Spanish (described in Poplack 1980, and reviewed in Haiman & Ourn 2009). In this as in other New World dialects of Spanish, where the presence of the plural marker {-s} is a sociolinguistic variable, there is a tendency for "concord" or agreement: either all elements in plural noun phrases like

32.	*la-s*	*chica-s*	*bonita-s*
	the-plural	girl-plural	pretty-plural
	'the pretty girls'		

will bear the affix, or none of them will. This statistically robust result comports with neither random variation nor with any version of functionalism. By random variation, any one of the eight logically possible outcomes (-s-,-s-,-s ….. -∅, ∅, ∅) would occur with equal probability. The principle of least effort would consistently favor marking no element (leading to a loss of the grammatical category), while the countervailing principle of maximum clarity would no less consistently favor marking all of them. A compromise between them would favor marking a single element, but no more. The tendency to concord is, however, compatible with the daring hypothesis of Ferguson and Barlow (1988:17) that familiar grammatical agreement may have been motivated not only by cognitive factors, but by an aesthetic drive for something like rhyme.

And the facts in Khmer also offer strong support for such a position. If the balance between symmetrical conjuncts has become a grammatical constraint, it is clear that none of the conventionally cited cognitive motivations for grammatical agreement (referential tracking through the marking of number, gender, and case, for example) are relevant, since the consistently (non-)reduced syllables are associated with no meaning. Rather, it is symmetry for the sake of symmetry alone which trumps all other considerations. In a language like Khmer which so strongly favors the creation of rhythmically lopsided words, this drive for symmetry is even more remarkable.

Like all changes which effect reduction, the allegro rules of section 1, which lead to the erosion of the unstressed syllable in Khmer, are not simply the blind result articulatory laziness, or of the unconstrained principle of least effort. Rather, this principle has been channeled and conventionalized in Khmer, as in other languages, to apply exactly where and how it does. This limitation on unrestricted reduction is already patent from a consideration of the fact that iambic words are not only the result of erosion, but also of the (quasi-) restoration of eroded material: they are built up to conform to a prevailing pattern, as discussed in section 2.

The interplay between sesquisyllabicity and symmetry discussed in section 3 raises questions about further shaping and conventions to which reduction and other processes may be subject. It suggests that speakers are mindful of aesthetic motivations even as they mumble.

References

Behaghel, O. 1932. *Deutsche Syntax, IV*. Heidelberg: Carl Winter.

Benedict, P. 1942. Thai, Kadai, and Indonesian: A new alignment in Southeast Asia. *American Anthropologist* 44:576-601.

Chiang Caun. 2002. *Vee:jjakaw: kmae (Khmer grammar)*. Pnom Penh: Privately printed.

Cun-Leuh. 2007. *Ve;jjakaw: kmae*. NP: Pon Caun & Tiang Ciat

Enfield, N. 2001. Remarks on John Haiman 1999. *Studies in Language* 25/1: 115-24.

Farmer, S. 2008. *The origins of nonsense: Bo'ri'va: sap in Khmer*. Honors thesis, Macalester College.

Farmer, S. 2009. On the possibility of metathesis as a source for infixation in Khmer. *Southeast Asian Linguistic Society Papers 14*, Paris.

Ferguson, C. & M. Barlow. 1988. Introduction. Agreement in natural languages. Chicago: *Center for the Study of Languasge and Information*

Ferlus, M. 1975. Vietnamien et proto-Viet-Muong. *Asie sud-est et monde insulindien* 6/4:21-55

Gorgoniev, Y. 1966. *Khmer grammar*. Moscow: Nauka.

Gregerson, K. 2007. *Mon-Khmer reduction effects*. Unpublished letter to T. Ealing.

Haiman, J. 1985. Symmetry. Haiman, J. (ed.) *Iconicity in syntax*, 85-105. Amsterdam: Benjamins.

Haiman, J. 1998. Possible origins of infixation in Khmer. *Studies in language* 22:597-617.

Haiman, J. 1999. Auxiliation in Khmer. *Studies in Language* 23/1: 149-72.

Haiman, J. 2003. Explaining infixation. Polinsky, M. & J. Moore (eds.) *Explanation in linguistic theory*, 105-20. Palo Alto: CSLI

Haiman, J. 2010.a. Competing motivations. Song, J. (ed.) *Oxford handbook of typology*. Oxford University Press.

Haiman, J. 2010b. The origin of new words. *Linguistics*, to appear.

Haiman, J. to appear. Why do ideophones reduplicate?

Haiman, J. & T. Kuteva.2002. The symmetry of counterfactuals. Bybee, J. & M. Noonan, Eds. *Clause combining in grammar and discourse: A Festschrift for Sandra A. Thompson*. Amsterdam: Benjamins.

Haiman, J. & N. Ourn. 2003. Nouns, verbs, and syntactic backsliding in Khmer. *Studies in Language*

Haiman, J. & N. Ourn. 2003a. *Creative forces in Khmer*. Iwasaki, S. et al. (eds.) SEALS XIII, 157-64. Canberra: RSPacS, Australian National University.

Haiman, J. & N. Ourn. 2009. Decorative symmetry in ritual (and everyday) language.Corrigan, R. et al. eds.) *Formulaic language*, Volume 2: 567-87. Amsterdam: Benjamins.

Hawkins, J. 1994. *A performance theory of word order and constituency*. Cambridge University Press.

Henderson, E. 1952. The main features of Cambodian pronunciation. Bulletin of the School of Oriental and African Studies.

Huffman, F. 1972. The boundary between the monosyllable and the disyllable in Cambodian. *Lingua* 29: 54-66.

Huffman, F. , Charan Promchan, Chhom-Rak Thong Lambert.1970. *Modern spoken Cambodian*. Yale U. P.

Ieuv-Koeus. 1945. *Vee:jjakaw: kmae*. Phnom Penh: Institut Bouddhique.

Jacob, J. 1960. The structure of the word in Old Khmer. Bulletin of the School of Oriental and African Studies 23/2: 359.

Jacob, J. 1990. Introduction to Cambodian. Oxford University Press.

Jenner, P. 1969. *Affixation in Khmer*. Unpublished PhD dissertation University of Hawaii.

Lewitz, S. 1968. L'accentuation syllabique en cambodgien. *Papers of the CIC Far Eastern Language Institute*, 155-67.

Malkiel, Y. 1959. Studies in irreversible binomials. *Lingua* 8: 113-60.

Martini, F. 1942-5. Aperçu phonologique du cambodgien. *Bullertin de la Societe lingyuistique de Paris* 42:112-31.

Maspero, G. 1915. *Grammaire de la langue Khmere*. Paris: Imprimerie Nationale.

Matisoff, J. 1973. Tonogenesis in Southeast Asia. Hyman, L. (ed.) *Occasional Papers in Linguistics* 1:72-95. Los Angeles: University of Southern California.

Moravcsik, E. 1978. Reduplicative constructions. Greenberg, J. (ed.) *Universals of human language, volume* 3: 297-334. Stanford U.P.

Nacaskul, K. 1976. Types of elaboration in some Southeast Asian languages. P. Jenner et al (eds.) *Austroasiatic Studies*, part 2: 873-90. Honolulu: U. Hawaii Press.

Nguyen 1965. Parallel constructions in Vietnamese. *Lingua* 15: 125-39.

Noss, P. 1966. The treatment of r in the modern Khmer dialects. Zide, N. (ed*.) Studies in comparative Austroasiatic linguistics*, 89-95. The Hague: Mouton

Noss, P. & I. Proum. 1966. *Cambodian basic course, volume 1*. Washington: FSI.

Ourn, N. & J. Haiman. 2000. Symmetrical compounds in Khmer. *Studies in language* 24:483-514.

Pinnow, H-J. 1980 [1958]. Remarks on the structure of the Khmer syllable and word. *Mon-Khmer Studies* .8:131-7.

Poplack, S. 1980. Deletion and disambiguation in Puerto Rican Spanish. Language 56/2: 371-85.

Pou, S. 2004. Selected papers on Khmerology. Phnom Penh: Reyum Publishing.

Prakorb, P.-N. 1992. The problem of aspirates in Central Khmer and Northern Khmer. *MKS* 22: 252-6.

Ross, J. 1967. *Constraints on variables in syntax.* PhD dissertation, MIT.

Sisovat, P. 1972. *Bo'ri'va: sap knong phiasa: kmae (Servant words in Khmer).* NP: Privately printed.

Stolz, T. 2008. Total reduplication vs. echo-word formation in language contact situations. Siemund, P. & N. Kintana (eds.) *Language contact and contact languages*, 107-32. Amsterdam: Benjamins.

Thach, N.M. 1999. Monosyllabization in Kiengiang Khmer. *Mon-Khmer Studies* 29:81-95.

Thomas, D. 1992. On sesquisyllabic structure. *Monh-Khmer Studies* 21: 206-10.

Vongvipanond, P. 1992. Lexicological significance of semantic doublets inThai. Compton, C. & J. Hartman (eds.) *Papers on Tai languages, linguistics, and literature in honor of William Gedney on his 77th birthday*, 141-53. University of Michigan monograph series on Southeast Asia, volume 16

Merge Right: Iambic Effects, Grammaticalization and Cliticization in Brao Grammar

Charles Keller
ICC & SIL International

Jacqueline Jordi
ICC & SIL International

Kenneth Gregerson
SIL International

1. Introduction

This paper observes some reductions of surface form in Brao[105] (See Keller, Jordi, Gregerson and Baird, 2008),[106] a West Bahnaric (Mon-Khmer) language in Ratanakiri Province, Cambodia. We present certain cases of grammaticalization and cliticization and recognize some principles of iambic process which appear to underlie these historical effects (See Donegan and Stampe, 1983, for the larger metrical framework in Austroasiatic). One special effect is the merging of pre-verbal pronouns (and tense) with the root, which has triggered a syntactic respecification of subject pronouns, creating thereby a hitherto non-existent grammatical agreement system. The 'driving' empirical principle throughout this study is a simple one, though its effects are not, *viz.,* MERGE RIGHT:

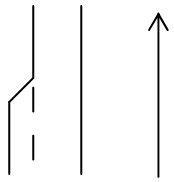


Figure 1: MERGE RIGHT

2. Phrasal conflation

The forces of iambic restructuring have had their effects in Brao at the phrase level as evidenced in certain pronouns.

2.1 Pronouns

Two Brao Krung pronouns seem to reflect a historical "repackaging" of original syntactic strings of the following type: PRO + NUM + N. Consider the pronoun array below:

[105] Brao Krung data are due to Charles Keller; Brao Ombaa, Kavet and Lun material is from the research of Jacqueline Jordi.

[106] This recent survey of Brao varieties by Keller et al., identified in preliminary fashion the following Brao varieties, Northern Brao: Jrii, Kavet, Hamong, Ombaa, Kanying Southern Brao: Krung, Lowland Brao (Brao Tanaap), Lun.

Keller, Charles, Jacqueline Jordi & Kenneth Gregerson. 2011. "Merge Right: Iambic Effects, Grammaticalization and Cliticization in Brao Grammar." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.162-177.

	Singular	Dual	Plural
1	*ʔaj*	*jɨɨ*	*ŋaaj* (Inc.) *ʔjaa* (Excl.)
2	*saj*	***səbraa***	*vɛɛ*
3	*hɔɔ*	***ṃbraa***	*mɛɛ*

Figure 2. Brao Krung Free Pronouns

Krung pronouns are generally monosyllabic in form, as seen above, with the exception of the dual forms for second and third persons, *viz*., *səbraa* and *ṃbraa*, respectively. Although they are nowadays perceived and treated as monomorphemic, they seem straightforwardly to be constituted of the following elements:

(1) s*əbraa*: *saj* + *baar* + *raa*
'2sg' 'two' person'
'you two persons'

(2) *ṃbraa*: mɛɛ + *baar* + *raa*
'3pl' 'two' 'person'
'they two persons'

In survey notes on Brao Lun at Taveng, the 3rd dual cognate of (2) is, in fact, pronounced in its full original form as *mɛɛ baar raa*, rather than being reduced to a single (sesquisyllabic) phonological word.

2.2 The metrical template

The prototypical full phonological word (PW) in many Mon-Khmer languages famously takes a *sesquisyllabic* ('syllable and a half') form (Matisoff, 1973:86). Metrically, such MK PWs constitute an *iambic* construction of the type:

PW = Reduced Syllable + Full Syllable
W(eak) Stress S(trong) Stress

It is this irrepressible---and much observed---metrical template that reduces elements on the left, drawing them into a right-headed orbit and remodeling their constituent structures to conform to a prototypical iambic PW pattern. Predictably enough, elements to the right of the metrical head escape this process relatively unscathed. Benedict (1993:121) characterized this process 'canonical reduction on the left'.

In respect to the two Krung complex dual pronouns above (1 and 2), something like the following (3) historical assimilation into the body of the sequisyllabic word is reconstructed here as having taken place, with *səbraa* '2nd dl' plotted in the left and *mbraa* '3rd dl' in the right columns below:

(3) First Cycle: *baar* reduction

	***səbraa* '2nd dual'**	***ʔmbraa* '3rd dual'**
Stage 1 Free form *baar* :	(*saj*) +(***baar***)+ (*raa*) '2sg' 'two' 'person' (PW1)+(PW2)+(PW3)	(*mɛɛ*)+(***baar***)+(*raa*) '3rd pl' 'two' 'person' Idem

PWords 1, 2, 3 are independent, each with S(trong) stress, relatively.

Stage 2 Phrase Reduction:	(*saj*) + ((***bàar***)-*raa*)) (PW1)+((PW2)-PW3)	(*mɛɛ*)+((***bàar***)-*raa*)

PW2 as the adjacent constituent to the left of PW3 takes a weakened phrase stress *vis-à-vis* PW3 as it gets metrically grouped with PW3.

Stage 3 Word reduction: (*saj*) + (***bə̀*** - *raa*) | (*mɛɛ*) + (***bə̀*** - *raa*)
PW1 + PW2-3

PW2 gets fully merged into PW3 as a presyllable in which baar > bə̀.

Stage 4 Syllable reduction: (*saj*) + (*b-raa*) | (*mɛɛ*) + (*b-raa*)
PW1 + PW2-3

Weak syllable *bə̀* reduced to cluster initial *b-*.

Briefly put, then, *baar* is thus reduced by stages to *b-* as summarized below:

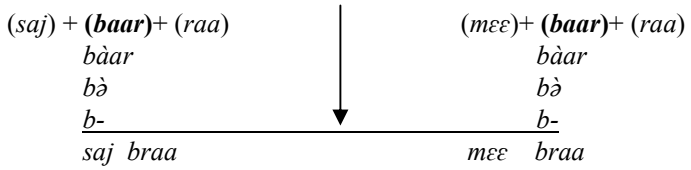


Figure 3. Reduction of *baar* > *b-*

Next, with *baar* now reduced to a single segment within a cluster, the constituent *saj* and *mɛɛ* find themselves the *new adjacent elements on the left* and hence within the immediate 'scope' of metrical (iambic) forces 'pulling' to the right. The fate of *saj* and *mɛɛ* are plotted in the Second Cycle below, where we use the briefer format to summarize successive stages of *saj* and *mɛɛ* reduction:

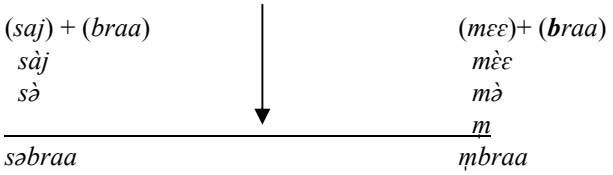


Figure 4. Reduction of *saj* and *mɛɛ*

Thus, a syntactic phrase is drawn inexorably through successive reductions into a prototypical target PWord configuration by iambic forces at work in Brao Krung. The stages above reconstruct a process that acts first on elements immediately to the left (Proximal), leaving non-immediate (Distal) constituents "beyond reach" as in Stages 2 and 3 above. Then, these stages having been achieved, the formerly Distal constituent to the left gets delivered next to a position of metrical adjacency to the iambic head and becomes the "new Proximate" and susceptible, therefore, of attraction to the right. In this way *saj - braa* constitutes a kind of phrasal sequisyllabic 'way station' *en route* to its complete remodeling as a Krung PW, *səbraa*. That is, in *saj - braa* the syllable weakens to *sàj* as a reduced dependent word preceding *braa* as a fully stressed one. However, in order for assimilation to be fully realized, *saj* has to be reduced to a permitted first (pre)syllable shape, since *CVC- weak presyllables are not optimally allowed in Krung. Thus, *saj* reduces to *sə̀*, a prototypically perfect weak syllable shape in a Krung PW.

A parallel process can be seen above to have effected the final outcome producing *m̩braa*, with the addition of a Stage 4 in which *mə-* receives a further reduction to a syllabic continuant *m̩* . In both of these pronoun forms a trajectory of reduction of the following sort can be observed:

Free Words in a Phrase >
Iambic Phrase Reduction >
Iambic Word Reduction >
Iambic Syllable Reduction

These processes appear to suggest some general principles of Iambic Reduction:

i) Iambic Domain Principle: The Iambic Stress Wave (Weak + Strong) ranges across not only the familiar Mon-Khmer Phonological Word, but also across Phonological Phrases with an Iambic Head-----and even, we note in passing, within a monosyllable, where vocalic prominence can shift from left to right ('elephant' **ruus* > *rúih* in Rengao but has shifted to *rəwéh* in Tampuan, Bunong. That is, the iambic stress contour may occur across various "size levels" and may be represented as follows:

	WEAK	STRONG
PPHRASE	PWORD	PWORD
PWORD	SYLL	SYLL
SYLLABLE	VOCALIC	VOCALIC

Figure 4. Iambic Domains

ii) Adjacency Principle: Iambic Reduction operates on a principle that an *immediately* adjacent constituent to the left of an Iambic Head gets *reduced first*, as represented below for various levels of structure:

	Immediately Adjacent Constituent		Iambic Head Constituent
PPHRASE:	PWORD	>	PWORD
PWORD:	SYLL	>	SYLL
SYLLABLE:	VOCALIC	>	VOCALIC

Figure 5. Scope of immediacy of iambic reduction on the left

Corollary: After forms are reconfigured by principles i) and ii), cyclical diachronic iteration of i) may apply again if the conditions of ii) exist (see Cycles 1 and 2 above).

Harris and Campbell (1995:343-4) discuss the notion of 'cyclical,' 'spiraling' and 'renewable resources' as forces at work in the directionality of grammaticalization (e.g., words > clitics > affixes). Clearly, iambic stress in Mon-Khmer languages constitutes such a premier renewable resource that keeps the wheels of change turning.

iii) Sesquisyllabic Target Principle: Iambic patterns across PHRASES tend to reduce and remodel constituents in the direction of sesquisyllabic WORDS as 'ideal' Mon-Khmer targets.

Corollary: Mon-Khmer disyllabic forms, being already metrically 'well-formed' evince fewer iambic attractions from the left than do monosyllabic forms.

Comment: The historical fact that Vietnamese, as a Mon-Khmer language, has reduced original disyllabic forms absolutely to monosyllables, of course, reflects iambic reduction to the ultimate degree. The point of iii) here, however, is that an unconstrained iambic *reduction to monosyllabicity is not inevitable* across Mon-Khmer. Sesquisyllabicity remains a very viable metrical target for reductive processes.

3. Clausal Conflation (Verb-headed constructions)

Just as Noun-headed constructions can yield metrically reduced iambic forms, so also can Verb-headed strings. The Brao Krung intransitive clause, for example, may take the following structure:

SUBJECT TENSE VERB

The PWs in (5) and (6) below appear to reflect precisely this grammatical ordering of constituents, but *as reductions* of the clausal schema above:

(4) *sŋrɛʔ* (< saj maa rɛʔ)
2sg-FUT-go
'You will go'

(5) *hn̥rɛʔ* (< hɔɔ maa rɛʔ)
3sg -FUT- go
'He will go'

In parallel fashion with NPs in 2.1 above, the iambic processes that produced PW forms (4) *sɛnrɛʔ* ~ sn̥rɛʔ and (5) *hn̥rɛʔ* from clause constituent strings are plotted in the left and right columns respectively below:

(6) First Cycle: *ma* reduction

	***sn̥rɛʔ* 2nd sg-FUT-go**	***hn̥rɛʔ* 3sg-FUT go**
Stage 1 Free forms:	(saj) + (maa) + (rɛʔ) 2nd sg- FUT- go	(hɔɔ) + (maa) + (rɛʔ) 3sg- FUT- go
(Equal stress, relatively)	(PW1)+ (PW2) +(PW3)	Idem

PWs 1, 2, 3 are independent, each with S(trong) stress, relatively.

Stage 2 Phrase Reduction:	(saj)+ ((mà) - (rɛʔ)) (PW1)+((PW2+(PW3))	(hɔɔ) + ((mà) - (rɛʔ)) Idem

PW2 as the proximate constituent to the left of PW3 takes a weakened *phrase* stress (mà) and gets metrically grouped with PW3 as ((mà) - (rɛʔ)).

Stage 3 Word reduction:	(saj) + (mərɛʔ) (PW1) + (PW 2-3)	(hɔɔ) + (mərɛʔ) (PW1) + (PW2-3)

PW2 gets fully merged into PW3 becoming a presyllable, i.e., *ma* > *mə*.

Stage 4 Syllable reduction:	(saj) + (m̥rɛʔ) (PW1)+ (PW2-3)	(hɔɔ) + (m̥rɛʔ) (PW1) + (PW2-3)

Weak syllable *mə* is reduced to syllabic segment *m̥*-.

Stage 5 Segment reduction:	(saj) + (n̥rɛʔ) PW1 PW2	(hɔɔ) + (n̥rɛʔ) PW1 PW2

Assimilation [m̥] > [n̥] / __r-

Next, with *ma* FUT already reduced to a single segment [n̥], the constituents *saj* and *hɔɔ* (right column above), become the *new adjacent elements* on the left and hence subject to forces pulling towards the iambic head on the right. Its reduction is tracked in the Second Cycle, as plotted in the more compact representation below:

(7) Second Cycle *saj* '1sg' and *hɔɔ* '3sg' reduction

(saj) + (n̥rɛʔ)	(hɔɔ) + (n̥rɛʔ)
(PW1)+(PW2-3)	(PW1)+(PW2-3)
S W-S	S W-S
sàj	*hɔ̀ɔ*
sə	*hə*
s-	*h-*
***sn̥rɛʔ* '2sg-FUT-go'**	***hn̥rɛʔ* '3sg-FUT go'**

Thus, *snrɛʔ* "you will go" and *hnrɛʔ* "He will go" result from nearly parallel processes that conflate three original PWs as shown in the successive stages above. In this way, over time, original grammatically independent constituents on the left are weakened (W) and brought under the sway of the all-powerful (S) iambic head, whether nominal or verbal, such that the sesquisyllabic "cookie cutter" recreates PHRASES as Mon-Khmer WORDS.

4. Grammaticalization, transparency and respecification

The metrical effects discussed in sections 1 – 3 above have served in Brao to reduce[107], delete, conflate and generally remodel a number of words, phrases and clauses in their surface *phonological* forms. These surface effects have also resulted in the *grammaticalization* of certain words as well. Consider the pronouns below, which were given in (1) and (2):

(8) *səbraa*[108] *saj + baar + raa*
'2sg' 'two' person'
'you two persons'

(9) *m̩braa*: *mɛɛ + baar + raa*
'3pl' 'two' 'person'
'they two persons'

This pair of dual pronouns appear to qualify as instances of grammaticalization in the sense that they:

i) Are derived from earlier free lexical items such as prounouns *saj* '2sg' and *mɛɛ* '3rd pl'; number *b* < *baar* 'two', and *raa* 'person'

ii) The forms are reduced in overt shape m̩ , sə, b, a common feature of grammaticalization.

iii) The forms are 'bleached', i.e., generalized and stripped of features, semantically: *saj* '2sg' reduces to only a person marker '2nd' and *nɔɔʔ* 3sg" to '3rd person', Number *b* < *baar* 'two', and *raa* 'person' take on more 'abstract' grammatical functions.

iv) Both *səbraa* and *m̩braa* are treated as unanalyzed pronouns on a par with *aj* 'I' or *ŋaaj* 'we exclusive', the very paradigm itself being presumably complicit in this reinterpretation

Similarly, as detailed above, the following sequence of PRONOUN + FUT + VERB have also coalesced under iambic stress into a single phonological word:

(10) *sɛnrɛʔ* < *saj maa rɛʔ*
2sg-FUT-go
'You will go'

hn̩rɛʔ < *hɔɔ maa rɛʔ*
3sg-FUT-go
'He will go'

Again, these grammaticalized cases are unanalyzed forms that are used perfectly appropriately as to sense, but without apparent recourse to the opaque elements of which they are composed.

A striking case of grammaticalization is found in clauses in the Brao Ombaa dialect:

(11) *aj* *ə-dɔk*
1sg 1sg-go
'I am going'

That is, a free 1st Person sg subject pronoun *aj* is followed by the verb *dɔk* 'go' to which is obligatorily attached a prefixal morpheme *ə-* '1st person sg., as in the tree below:

[107] SECTION 4 Reductions of vowels and minor syllables in this fashion is observed by Donegan (1993:23) to be a predictable concomitant of the stress-timing which characterizes Mon-Khmer.

[108] Jacq and Sidwell 2000 list only 2nd person dual as **ɟaar*: Lvn *ɟəər*, Nha *ɟaan*, Sap *caar* 'they dual', Chg *cəən*'?'. These forms may be reductions from a phrase that included *baar* 'two' but just how or whether are not clear.

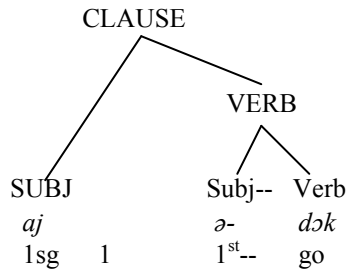


Figure 7. Subject-Verb agreement in Brao Ombaa

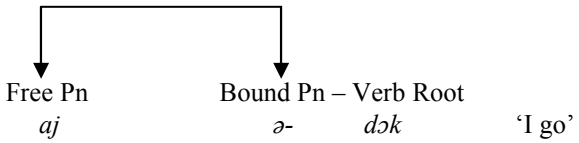
This configuration builds on the Brao Ombaa set of free pronoun as shown in Figure 8:

	Singular	Dual	Plural
1	*ʔaj*	*baa* (incl.)	*ŋaaj* (incl.)
		ɲɨɨ (excl.)	*ɲaa* (excl.)
2	*haj*	*səbraa*	*vɛɛ*
3	*nɔɔʔ*	*m̩braa*	*mɛɛ*

Figure 8. Brao Ombaa Pronouns

In Brao Ombaa, all verbs studied so far take prefixed pronominal forms that agree as to Person and Number with the Subject. For example, with Pronoun subject *aj* 'I' the agreeing prefix is *ə-,* as exemplified below:

(12) Person-Number Agreement



Similarly,

aj ə-caa	I eat
aj ə-məət	I want
aj ə-həriəŋ	I desire

It is clear that the prefixal forms are REDUCED versions of the free pronoun which have merged with the verb root under iambic stress conditions, as illustrated in (13) with the verb *dɔk* 'go':

(13)

ʔaj ə-dɔk	(1p sing.)	'I am going'
haj hə-dɔk	(2p sing.)	'You (sg) are going'
nɔɔʔ lə-dɔk	(3p sing.)	'He is going'
ɲɨɨ ɲə-dɔk	(1p dual excl.)	'We two (incl) are going'
ɲaa ɲə-dɔk	(1p plu excl.)	'We all (excl) are going'
baa bə-dɔk	(1p dual incl.)	'We two (excl) are going'
ŋaaj ŋə-dɔk	(1p pl.incl.)	'We (incl) are going'
vɛɛ və-dɔk	(2p pl.)	'You (pl) are going'
mɛɛ mə-dɔk	(2p pl.)	'You (pl) are going'

The above forms are transparent reductions of the free pronoun; only in the instance of *nɔɔ lə+dɔk* '3sg go' does a variation occur, with initial *n-* ~ *l-* in its affixal counterpart. Thus, in Brao Ombaa an iambically driven conflation of verb constituents has apparently rendered original pronouns, which have become phonologically attached to the verb head, morphologically opaque to speakers with the result that in

clauses *person reference gets respecified as independent syntactic pronouns*---just as the reduced pronouns originally were!

On a comparative note, subject-verb agreement as to person and number is not a widespread grammatical feature among Mon-Khmer languages. Such agreement is, however, reported for Khasi (See Bedell, , and others cited in his References), where they are regarded as clitics, which itself implies that they may have been delivered historically to their dependent status by means of metrical effects on the cohesion of constituents. Also, in Temiar (Aslian) Benjamin (To appear), page 17 details the function of pronominal agreement clitics, which relate in his analysis "topic" and "comment" co-reference

As an interesting aside, Jordi notes that with Serial Verb constructions in Brao the prefixal agreement form *lə-* is typically attached only to the FIRST VERB in a serialized construction, as in the sentence below:

(13) *nɔɔʔ lədɔk daŋ caa* 'he he-walk look eat'
he go look-for eat
'He is going to look for food to eat.'

That is, the 3sg agreement prefix *lə-* is attached to *dɔk* 'go' but not also to *daŋ* 'look for' and *caa* 'eat,' reinforcing the claim that serialized verbs constitute in some sense a functional unity

However, in one case of serial verbs from a text the agreement prefixation (**həN-** '2sg + N-Fut') is also REPEATED on the last verb of the series---a kind of bracketing effect:

(14)

həŋ*hərooc tii* *hərooc jiŋ* *hərooc tuuh* ***hən****caa*
2sg+fut+tear off hand tear off leg tear off head 2sg+fut+eat

'You will tear off his hand, tear off his leg, tear off his head and you will eat them'

Finally, though we cannot treat of all the cases of surface reduction and grammaticalization in the present paper, on-going research by present authors, Jordi and Keller continue to turn up interesting findings. Jordi notes recently, for example, the following sentence:

(15) *ʔaaw ʔaj ləə tərdaak*
shirt 1S 3S wet
'my shirt is wet'

In which the verb tərdaak 'get wet' 3sg + Verb is to be understood in terms of the following underlying sequence:

(16) *ʔaaw ʔaj ləə trɔɔ daak*
shirt 1S 3S got wet

That is, trɔɔ, which has a variety of senses, here has the meaning of 'undergo, must, etc. and gets REDUCED and MERGED with the verb root: *trɔɔ daak* > *tərdaak*. This is reminiscent of Khmer tɨtɨk 'be wet', which raises the question whether it is also a reduction of trəw tɨk 'suffer being wet' Similarly, Brao Ombaa *tərtih* 'be big' (referring to pigs) apparently also derives from *trɔɔ tih*, though the sense is a bit elusive.

A grammaticalization of *am* 'to give' is reflected in the following reductions:

(17) *bəəm ŋ = kɛɛt* < *bəəm ʔam kɛɛt* 'make to die'
bəəm n = doow < *bəəm ʔam doow* 'cause to get'

Again, close-knit preverbal morphemes get reduced and merged with the verbal head

5. Iambic effects and the status of reduced forms

The Brao forms we have discussed so far all have in common that they are reductions of independent words under conditions of iambic *stress*. The status of such elements have been characterized as stress-related *cohesion* by Harris and Campbell (1995:63-64) in the following terms (italics added):

> "By *cohesion* we mean the status of a linguistic sequence as a fully *independent word, a clitic, an affix or an unanalyzable part of a larger unit*. In many [historical] changes, the four statuses form a continuum; an element that is at one time a fully independent word may become a clitic and then an affix, ending up as an unanalyzable part of another word, no longer having the status of a morpheme...Cohesion, in this sense, is also *related to surface phenomena, including stress*.....

That is, the forms cited by Harris and Campbell constitute a cline from independent to fully merged elements as follows:

Word > Clitic > Affix > Unanalyzable part of larger unit

We examine next the Brao examples in terms of one or another of the characterizations above. For expository ease, it is perhaps useful to consider them in reverse order as:

Unanalyzable part of a larger unit (5.1)
Affix (5.2)
Clitic (5.3)

5.1 Iambic reductions as unanalyzed elements of a larger form.

Perhaps the most plausible case of absolute merger under iambic stress conditions is that of the already discussed Brao dual pronouns:

(18) s*əbraa*: < *saj* + *baar* + *raa*
2sg two person
'you two persons'

m̩braa: < *mɛɛ* + *baar* + *raa*
3pl two person
'they two persons'

These two dual pronouns in the array below have arisen over time from phrases that have reduced to the point of fusion as a single sesquisyllabic referring form. The iambic pressures of reduction from many to one is emblematically reflected by their place in a field of otherwise monomorphemic pronouns (See Brao Ombaa Free Pronoun) chart:

	Singular	Dual	Plural
1	*ʔaj*	*baa* (incl.) *ɲii* (excl.)	*ɲaa* (incl.) *ŋaaj* (excl.)
2	*haj*	*ɛəbraa*	*vɛɛ*
3	*nɔɔʔ*	*m̩braa*	*mɛɛ*

Figure 9. Fused Brao Ombaa Dual Pronouns

Arguing in favor of a monomorphemic (unanalyzable) interpretation of the two duals above are the following:

i) Historical change has obscured on the surface of it the relation between the modern 2nd person sg *haj* and the earlier **ɛaj* uniquely preserved here in *ɛə-*.
ii) The originally motivated *baar* 'two' for dual has only the residual *b* as part of a cluster and fused

with *raa* 'person' as *braa*, which has no independent meaning.

iii) There is idiosyncratic behavior also in the employment for dual person marking, in which the 2nd person *ɛaj* is used for *singular,* but the 3rd person *mɛɛ* turns up for *plural*.

All these factors would seem to reflect an unanalyzed whole status for these dual forms.

5.2 Iambic reduced preverbal forms as inflectional prefixes: grammatogenesis:

Clearly, Brao Ombaa clauses like the following involve what looks like grammatical agreement between the syntactic pronominal Subject:

(19)

vɛɛ və-dɔk

'You (pl) are going'

That is, Subject: *vɛɛ* takes a reduced version of itself (*və-*) as an agreement marker attached phonologically to the verb root. See section 4 above for the whole set of free pronouns and their counterpart preverbal reduced forms.

Significantly, the Brao data reflect a case in which PHONOLOGICAL (iambic) processes bring into being a GRAMMATICAL agreement system. The merger of originally free pronouns with verb roots recalls, of course, the long history in linguistics of attempts to explain personal inflections (especially in Indo-European) as "simply personal pronouns that have lost their independence" (Sweet 1900, 53, ll3 as cited in Harris and Campbell 1995, 20). Without rearguing nineteenth century debates on "Agglutination theory" (Franz Bopp), "Glottogonic" growth (August Schleicher) as over against the pronouncements of the "Neogrammarians" (Karl Brugmann), one can reaffirm that empirical data like that of Brao does, indeed, reflect the grammaticalization and reinterpretation of free forms as grammatically functional bound forms. And lying behind these changes are not semantic, syntactic, or morphological motivations, but in the Brao case the 'blind forces' of prosody.

5.3 Iambic effects as cliticization

We have referred so far to bound pronouns of agreement in Brao as 'prefixes.' However, a plausible interpretation of some of the Brao grammatical forms is that they are, more accurately, *clitics*, which have been characterized, among other things, as follows:

- 'Simple clitics' have been described in the literature (Zwicky, 1977) as bound forms which often have free counterpart morphemes with the same meaning. Pronouns often provide classic instances of this category of clitics.
- Wakernagel (1892) viewed clitics as fundamentally 'little words without their own accent'. In more recent parlance clitics have been said to be 'prosodically deficient' and 'stray' phonological elements that get 'adjoined' to or integrated into full-fledged prosodic words structures (Anderson, 2005:13).

After a wide-ranging review of various treatments of clitic-hood, Anderson boils things down to the following:

> We….. have a reasonable understanding of what clitics are: they are linguistic elements that display prosodically deficient phonology, anomalous morpho-syntax, or both.
> Anderson (2005:33)

On the question of Subject Agreement in dialects other than Brao Ombaa, Jordi provides the following notes:

In Brao texts from subvarieties such as Kavet and Lun Tengaaj 3rd person sg. is also marked on the verb, for example in the Kavet sentence:

(20) *muuj raa* *hnɔɔʔ* *naʔ* *hŋiw*
one person 3sg still 3sg+alive
'one person he is still alive'

This construction is quiet regular but in Kavet, in this case having a free NP Subject *muuj ra* 'one person' as well as (redundantly) a free PN *hnɔɔʔ* '3sg" ----and a reduced preposed h- '3sg' on the verb as *hŋiw*. However, Kavet Subject Agreement happens:

i) Only with the 3rd person and
ii) Only in the *present tense*, never in the future tense.

These grammatical restrictions on the 'prosodically deficient' form *h* suggest that in Kavet we have CLITIC agreement, the verb above to be represented as: *h=ŋiw* 'be alive'

Likewise, from data at hand, Brao Lun has Subject Agreement *ONLY* on the third person and is quite regular for non-future tense, e.g.,:

bɛɛʔ *nɔɔʔ* *nəkɨn ...*
father 3sg 3sg+think
'father (he) thinks (something)'

This kind of anomolous and restricted pattern of occurrence, again, suggests a CLITIC interpretation of the verb as *nə=kɨn* 'he thinks.'

For the future tense (irrealis mode?), however, Lun follows the Brao Ombaa Subject Agreement patterns quite regularly, i.e. involving ALL persons, rather than restricted to 3rd person, for example:

(21) *ʔaj* *ʔə=ŋ=kuu*
1sg 1sg=fut=stay
'I will stay'

(22) *paʔ* *hə=ŋ=kacɛɛt* *moʔ qaj*
if 2pr sg+fut+kill sister 1sg
'if you kill my sister....'

In the present tense, Brao Lun follows the Brao Ombaa mostly in *direct* speech as in forms below (whereas in *indirect* speech their isn't usually a marking on the verb) e.g.:

(23) *ə=mɛɛt* 1sg + want
hə=bəəm 2sg + do
mə=rɛʔ 3rd pl + go

Clearly, across the Brao dialects in Southern Laos and Northeastern Cambodia there remains a great deal yet to be learned about agreement patterns and constraints, on which the present paper can only offer initial hints.

Conclusion

Tracing the prosodic, specifically iambic, effects on the various dialects of Brao yields insights on more than one front.

First, the phonological effects of iambic stress in Brao include the powerful role of the Weak + Strong (rising) contour and its dominance not only over Phonological Words but also over Phonologtical Phrases and even within the Syllable. This template exerts its influence over synchronic constraints and focuses the direction of diachronic change as well.

Secondly, from a grammatical perspective, Brao reflects how the workings of prosodic forces of reduction and merging of phonological 'stuff' delivers at the same time reinterpretations of grammar,

resulting in the rise of agreement systems implicating subject and verb relations. It turns out, thus, that grammar can arise, not for grammatical or semantic reasons, but for purely phonological (prosodic) ones

Finally, the *sprachgefuhl* of native Brao speakers reflects a kind of 'amnesia' of perception as to the analysis of certain linguistic forms that leads to a re-specification and addition of grammatical information (via new free subject pronouns) required for the successful speech act.

References

Anderson, Stephen. 2005. *Aspects of the Theory of Clitics.* Oxford University Press.

Bedell, George. Bedell, George. 2011. "The Syntax of Khasi Adverbial Clauses." In Sophana Srichampa and Paul Sidwell (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 2*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.1-10.

Benedict, Paul K. 1993. Tibeto-Burman "split cognates." *Linguistics of the Tibeto-Burman Area* Volume 16:2:121-122.

Benjamin, Geoffrey. To appear (2009?). Temiar morphology: a view from the field. *Aslian Languages*, edited by Nicole Kruspe & Niclas Burenhult. Canberra: Pacific Linguistics

Donegan, Patricia. 1993. Rhythm and vocalic drift in Munda and Mon-Khmer. *Linguistics of the Tibeto-Burman Area* 16.1: 1 – 43.

Harris, Alice C. and Lyle Campbell. 1995. *Historical cross-linguistic perspectives*. Cambridge Studies in Linguistics 74. Cambridge University Press.

Jacq, Pascale and Paul Sidwell. 2000. A comparative West Bahnaric dictionary. *Languages of the world/dictionaries 21*. Muenchen, Lincom Europa.

Keller, Charles, Jacqueline Jordi, Kenneth Gregerson, Ian G. Baird. 2008. Brao dialects: lexical and phonological variations. *Revue de l'Institute de la Langue Nationale de l'Académie Royale du Cambodge*, pp. 87-152.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman, ed. *Consonant Types and Tone*, pp. 71-95. *Southern California Occasional papers in Linguistics,* No. 1. Los Angeles: UCLA.

Wakernagel, Jacob. 1892. Uber ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1:333-436.

Zwicky, Arnold M. 1977. On clitics. Bloomington: Indiana University Linguistics Club.

Appendix

1. Brao Ombaa and Brao Krung Pronouns and Proclitics

BRAO OMBAA			KRUNG	
PRESENT TENSE (Reduced Pn as proclitic)			Emphatic	Neutral
ʔaj ə-dɔk	(1p sing.)	'I am going'	*ʔaj ə-rɛʔ*	*ʔaj rɛʔ*
haj hə-dɔk	(2p sing.)	'You (sg) are going'	*saj sə- rɛʔ*	*saj rɛʔ*
nɔɔʔ lə-dɔk	(3p sing.)	'He is going'	*hɔɔ hə- rɛʔ*	*hɔɔ rɛʔ*
ɲɨɨ ɲə-dɔk	(1p dual excl.)	'We two (incl) are..'	*jɨɨ jə- rɛʔ*	*jɨɨ rɛʔ*
ɲaa ŋə-dɔk	(1p plu excl.)	'We all (excl) are ..'	*ʔjaa ʔjə- rɛʔ*	*ʔjaa rɛʔ*
baa bə-dɔk	(1p dual incl.)	'We two (excl) are..'	*baa bə- rɛʔ*	*baa rɛʔ*
ŋaaj ŋə-dɔk	(1p pl.-incl.)	'We (incl) are..'	*ŋaaj ŋə- rɛʔ*	*ŋaaj rɛʔ*
vɛɛ və-dɔk	(2p pl.)	'You (pl) are going'	*vɛɛ və- rɛʔ*	*vɛɛ rɛʔ*
mɛɛ mə-dɔk	(3p pl.)	'They are going'	*mɛɛ mə- rɛʔ*	*mɛɛ rɛʔ*

Brao Ombaa has in the Present Tense a free subject pronoun with its reduced proclitic counterpart on the verb root. Brao Krung treats that sequence as "emphatic" and the uncliticized set as "neutral."

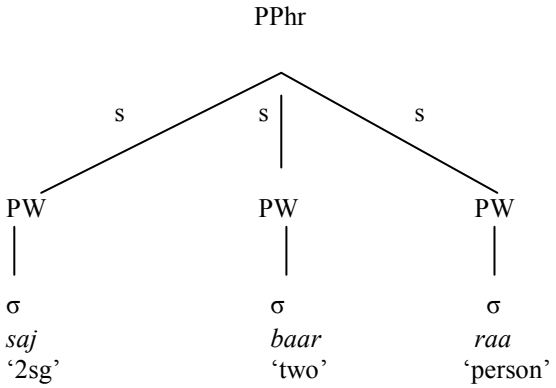
BRAO OMBAA FUTURE TENSE (Reduced Pn + FUT + verb root)			BRAO KRUNG Emphatic	 Neutral
ʔaj ə-n-dɔk [ʔn̩dɔk]	(1pers sing)	'I will go'	*aj ə-n-rɛʔ*	*aj maa rɛʔ*
haj hə-n-dɔk	(2persy sing)	'You (sg) will go'	*saj sə-n--rɛʔ*	*saj maa rɛʔ*
nɔɔʔ lə-n-dɔk	(3prs sing)	ETC.	*hɔɔ hə-n-rɛʔ*	*hɔɔ maa rɛʔ*
ɲɨɨ ɲə-n-dɔk	(1prs dl excl)	↓	*jɨɨ jə-n-rɛʔ*	*jɨɨ maa rɛʔ*
ɲaa ɲə-n-dɔk	(1prs pl excl)		*ɲaa ɲə-n-rɛʔ*	*ɲaa maa rɛʔ*
baa bə-n-dɔk	(1prs dl incl)		*baa bə-n-rɛʔ*	*baa maa rɛʔ*
ŋaaj ŋə-n-dɔk	(1prs pl-incl)		*ŋaaj ŋə-n-rɛʔ*	*ŋaaj maa rɛʔ*
vɛɛ və-n-dɔk	(2prs pl)		*vɛɛ və-n-rɛʔ*	*vɛɛ maa rɛʔ*
mɛɛ mə-n-dɔk	(3rd pl)		*mɛɛ mə-n-rɛʔ*	*mɛɛ maa rɛʔ*

In Brao Krung "emphatic" *saj sə-n--rɛʔ* (2sg) the pronoun clitic + Nasal FUT marker actually occur phonetically with a syllabic nasal: *saj sn̩--rɛʔ* and so on for the other persons. Note that Krung "Neutral" in Future Tense has no reduced constituents

2. Iambic Meter and Stray Adjunction: reduction of phrase and word structure as tree representations:

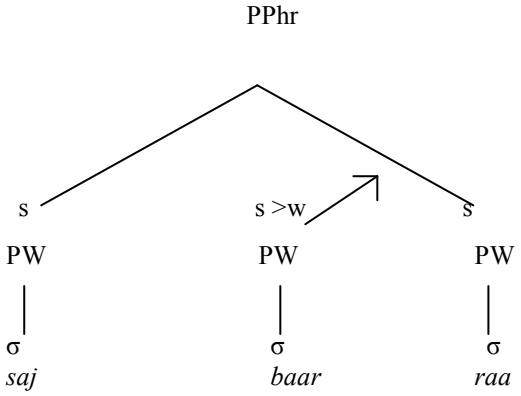
The Brao Krung form *səbraa* '2nd person dual' is assumed to have undergone the following stages to reach its present shape. Using Anderson's (2005) terminology, the trees below represent successive stages of adjunction or merger of 'stray' (metrically unattached) syllables into a more cohesive phrasal construction and ultimately into a prototypical iambic sesquisyllabic Phonological Word:

1) A Phonological Phrase (PPhr) is composed of Phonological Words (PW) each of which has its own Strong (S) stress on a monosyllable (σ):



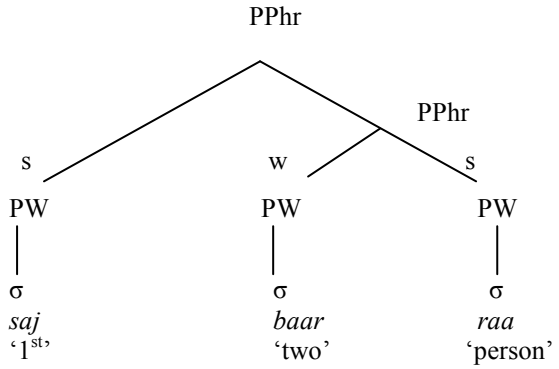
2) The PW (baar 'two') directly adjacent to the rightmost iambic head PW (raa 'person') weakens (2. a) to form a W(eak) + S(trong) Phrasal contour (2.b)

a)



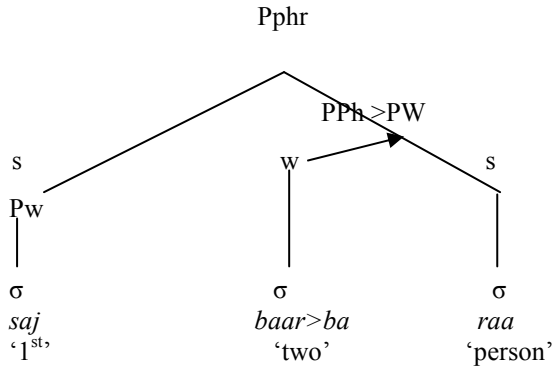
‘1st’ ‘two’ ‘person’

b)



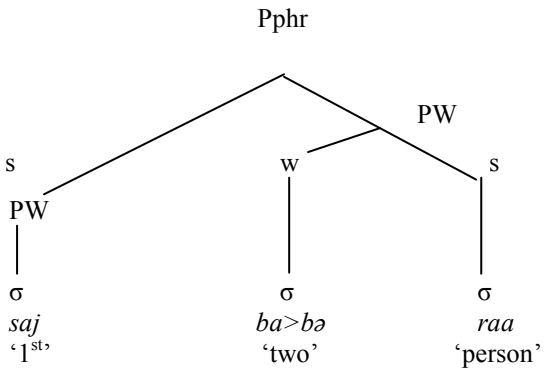
Thus, we recognize a consolidation into a Weak + Strong iambic *Phrase* as the opening gambit in constituent reductions of this kind in Brao.

3) The weakened adjacent phrasal form *baar* ‘two’ is reduced further (>*ba*), which allows its attachment as the sesquisyllabic weak syllable at a lower (PW) node:

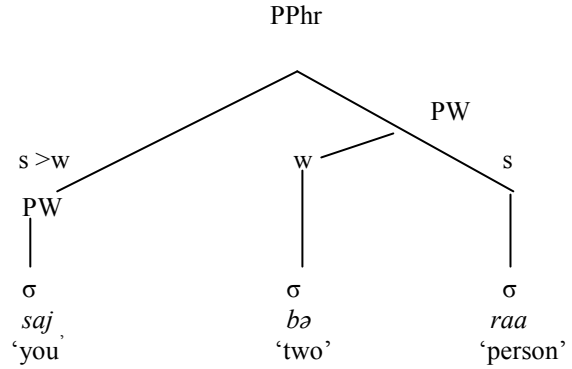


The transition *baar >ba* represents in segmental terms the assumption of an intermediary way station between a full phrasal status for *baar* and its ultimate PW status in the form of *bə* .

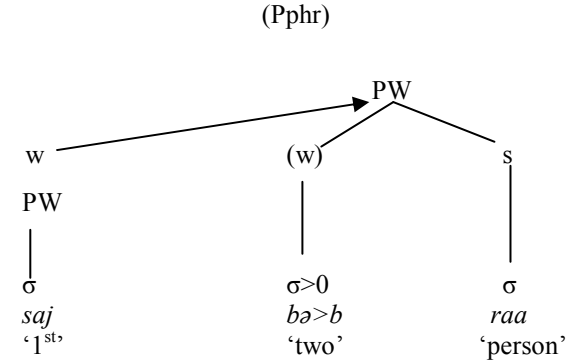
4) The consolidation of the weak syllable ba > bə consumates the adjunction of an original free syntactic constituent baar ‘two’ into a prototypical Mon-Khmer presyllable *bə* within an iambic target sesquisyllabic PW:



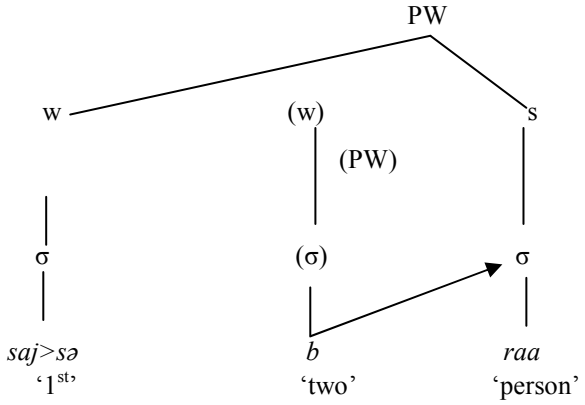
5) Iambic forces next weaken (s>w) the ‘new’ adjacent leftmost PW in PPhr, attracting its monosyllable *saj* ‘you’ into the orbit of the rightmost PW and its strong (s) iambic head *raa* ‘person’:



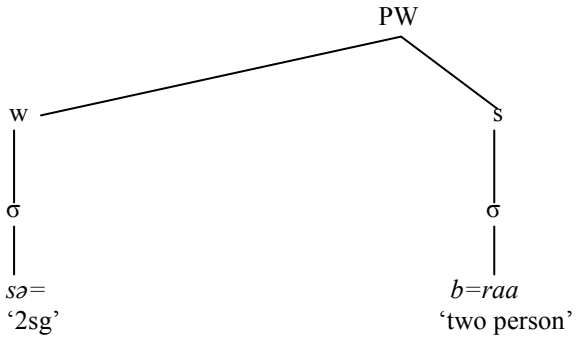
6) The weakened leftward PW *saj* now becomes 'neither fish nor fowl' as a 'prosodically deficient' element and gets 'demoted', becoming a 'stray adjunction' that attaches to the rightward PW as shown below:



As part of this cohesion process, the presyllable is further reduced (*bə>b*) to be reinterpreted as part of a *br*-cluster in the main strong syllable, thus, *braa*.



7) At this stage there is one disyllabic PW sajbraa. But since this fails as a well-formed sesquisyllable in Brao, one more reduction is required, viz., *saj>sə* to form the perfect Mon-Khmer presyllable, as shown below:



Thus, the contemporary form *səbraa* (sə=b=raa) 'you dual' is achieved---via the magic of iambic attraction to the right, leaving in its wake various reductions on the left.

Retention and Reduction in Reduplicants of Semai

Timothy C. Phillips
Summer Institute of Linguistics

Abstract

One dialect of Semai has two homophonous verb forms *ɟɨk*, one of which means "to stand" and the other "to marry". When these two verbs are reduplicated, however, the former surfaces as *ɟɨŋɟɨk* and the latter as *ɟəkɟɨk*, the difference being explainable historically by a segment retained in the reduplicant which has been lost in the base. This paper explores the wide variety of reduplication, including this one, found across various Semai dialects, then proposes a series of iambic reductions that serve to explain these phenomena, replacing in the process the notion of 'incopyfixation.' Examples of this same general pattern of reduplication and reduction are also provided from several other Aslian languages.

1. Background

The Semai language belongs to the Aslian branch of Mon-Khmer. There are reportedly 43,892 Semai people,[109] who live mostly in remote areas of the Malaysian peninsula in the states of Perak and Pahang. Conservative dialects of Semai exhibit up to 45 phonemes:[110] 19 consonants, and 14 oral vowels and 12 nasal vowels (Phillips 2007). Proto-Semai appears to have had an additional 3 oral vowels, including a diphthong, and additional nasal vowels as well (Diffloth 1977).

1.1. Word

Most Semai roots, consistent with broader Mon-Khmer typology, fit the following syllable template:

$$\underbrace{(C_3\ V_2\ (C_4)\ )}_{\text{Minor}} \quad \underbrace{C_1\ V_1\ C_2}_{\text{Major}}$$

The final syllable is regarded as the **major** syllable and the penultimate syllable, if present, is regarded as the **minor** syllable. Semai roots are usually monosyllabic or disyllabic. Multisyllabic words are generally limited to zoonyms and phytonyms, expressives, and affixed forms.

The minor vowel V_2 is usually a very short, non-phonemic, excrescent[111] [ə], Its occurrence in any given word is often optional if C_4 is not present and the two consonants C_3 and C_1 are easily pronounced without the excrescence, particularly in fast speech. For this reason iambic disyllabic roots with no coda and reduced minor syllables are sometimes called "sesquisyllabic," meaning one and one-half syllables (Matisoff 1973).

[109] According to "Orang Asli Population Breakdown" on the Center for Orang Asli Concerns website, for 2003.

[110] In this paper the short vowels will be represented with a single letter and the long vowels with a double letter. The latter is a departure from standard IPA. Further, 'r' is used to symbolize the flap.

[111] This V_2 vowel is viewed as "epenthetic" in most analyses, but the term "excrescent" (Levin 1974) is perhaps the better term. Excrescent vowels in Aslian are an epiphenomenal articulation uttered as the oral articulators are repositioned between consonants. See also Yap (2009).

Phillips, Timothy C. 2011. "Retention and Reduction in Reduplication of Semai." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.178-197.

There are a number of words that have minor syllables with minor vowel segments (V_2) other than [ə], namely [a], [i], and [u]; these are pronounced with slightly greater length than when V_2 is [ə]. In this paper these words are considered true disyllabic words, even though the minor syllable is reduced in terms of its vowel inventory.

1.2. Stress

In all Aslian languages words stress is on the ultimate (major) syllable. In sesquisyllabic and disyllabic words the minor syllable does not carry the same weight, phonetically or phonemically, as the major syllable. Consequently, such words have an iambic contour, i.e., weak stress followed by strong stress.

Aslian languages strongly prefer iambic forms, tending to squeeze words into a canonical iambic shape. For example, the following shows the development of several Semai forms from proto-Semai:[112,113]

	Proto form[114]	Final form	Gloss	
(1)	**sᵊ.mᵘ. ˈwaʔ*	*sᵊm. ˈwaʔ*	'all'	(Malay borrowing)
(2)	**kᵊ.rᵊk. ˈbaak*	*kᵊr. ˈbaak*	'butterfly'	(some dialects)
(3)	**kᵊ.nᵘ. ˈwaan*	*kᵊn. ˈwaan*	'child'	(southern dialect)
(4)	**tᵊ.hᵘ. ˈwaal*	*tu. ˈhaal*	'to blow'	(southern dialect)

1.3. Nasal Preplosion

Nasal preplosion, i.e., the insertion of a homorganic plosive before a nasal, occurs when a syllable-final nasal is preceded by an oral vowel. Most Aslian languages exhibit some form of nasal preplosion (Phillips 2006). In most Aslian languages this phenomenon is only found word-finally (i.e. on major syllable codas); however, Bishop (1996) has reported preplosion in minor syllables in Kensiw, and non-final preplosion is found in one type of reduplication in Semai (see §2.2 below).

Blust (1997:161) sees nasal preplosion as part of a progression of denasalization. Several points along this putative trajectory are represented across the many dialects of Semai.

Proto-MK		SE Semai		Kuala Kenip Semai		Perak Semai
*-m	→	-ᵇm	→	-pᵐ	→	-p

Thus, a reduction of final nasals to voiceless plosives results in a merger of *-m with *-p, and as will be shown, this has significant implications in Semai reduplication.

2. Reduplication patterns in Semai

Proto-Semai and contemporary conservative dialects of Semai have several types of reduplication. This section will lay out the basic reduplication patterns of verbal and expressive reduplication.

2.1. Verbal reduplication - incopyfixation

Semai verbs indicate the progressive or indeterminate aspect by reduplication of the root. The form of reduplication depends on the form of the root. These processes are summarized below:

---

[112] Semai reconstructions in this paper are from Diffloth (1977) and Phillips (2005).

[113] Another Aslian language, Mah Meri, has many forms originating from Malay that demonstrate iambification, for example [bəŋguk] 't.o. fish' (Malay *bengkarung*); [benti] 'to stop' (Malay *berhenti*); [məlak] 'rotten (of eggs)' (Malay *tembelang*). (Data from Kruspe 2010.)

[114] The excrescent vowel will typically be superscripted as a reminder that the vowel is not phonemic, but rather more of an articulatory artifact.

	Root	Progressive
Monosyllabic	C_1VC_2	$\mathbf{C_1əC_2}.C_1VC_2$
(5)	ciip ‘walked’	cəpciip ‘is walking’
Sesquisyllabic	$C_3əC_1VC_2$	$C_3\mathbf{əC_2}.C_1VC_2$
(6)	təhool ‘to blow’	təlhool ‘is blowing’

In (5) above, monosyllabic roots are reduplicated by preposing copies of the root onset and coda plus an excrescent vowel between the two. Sesquisyllabic roots (of the form $C_3əC_1VC_2$ in (6) above) are reduplicated by what has been termed ‘incopyfixation’ (Matisoff 2003), also known as coda-copy (e.g. Kruspe 2004), wherein a copy of the major (ultimate) syllable coda appears to be infixed into the minor (penultimate) syllable. However, incopyfixation, while an elegant and splendidly iconic term for describing the surface form, will be shown to be somewhat misleading historically speaking. Despite this, I will continue to use the term for this type of reduplication.

The penultimate vowel bearing the minor syllable weak stress of the reduplicant is usually a schwa. However, when the coda of the reduplicant is a glottal (that is, /h/ or /ʔ/), the penultimate vowel is /a/. When the coda is a palatal (/j, c, ɲ, s/), the penultimate vowel is /i/. For symmetry reasons it is expected that when the coda is the bilabial /w/ that the penultimate vowel will be /u/, but verification of this has proved elusive.

Full disyllabic roots (that is, roots that have penultimate syllables with codas and/or non-excrescent vowels) do not exhibit incopyfixation, but rather have a **bə-** prefix, apparently borrowed from Malay **bər-**.

2.2. Verbal reduplication – full root

There is a second type of verb reduplication found in many dialects of Semai wherein the full root is repeated multiple times. This form of reduplication iconically indicates continuous, intensive, multiple or repetitive action.

	Root	Progressive
Monosyllabic	C_1VC_2	$\mathbf{C_1a}.C_1VC_2.\mathbf{C_1VC_2(.C_1VC_2...)}$
(7)	*bɛɛt* ‘to sleep’	*(bi.)ba.ˈbɛɛt.ˈbɛɛt.ˈbɛɛt* ‘(he) keeps sleeping’[115]
Sesquisyllabic	$C_3əC_1VC_2$	$\mathbf{C_3a.}C_3ə.C_1VC_2\mathbf{.C_3ə.C_1VC_2\ (.C_3ə.C_1VC_2...)}$
(8)	*təhool* ‘to blow’	*(bi.)ta.təˀ.ˈhool.təˀ.ˈhool.təˀ.ˈhool* ‘(it’s) blowing non-stop’
Disyllabic	$C_3vC_1VC_2$	$\mathbf{C_3a.}C_3v.C_1VC_2\mathbf{.C_3v.C_1VC_2\ (.C_3v.C_1VC_2...)}$
(9)	*cagoh* ‘to pray’	*(bi.)ca.ca.ˈgoh.ca.ˈgoh.ca.ˈgoh* ‘(he’s) praying constantly’

With this form of reduplication each repetition of the full root is pronounced as a prosodic word with ultimate stress. Note that for roots ending with a preploded nasal, the nasal is preploded with each repetition. This is a departure from most nasals, including incopyfixed nasals, which are pronounced as simple nasals (that is, non-preploded) in the non-final position. This non-final preplosion, along with the stressed syllables throughout the word, argues that each repetition behaves as if it constitutes its own prosodic word.

[115] The ***bi-*** prefix means 3rd-person-singular, and is limited to certain Pahang dialects. In other Pahang dialects and in Perak, the prefix ***ki-*** is used instead.

2.3. Expressive reduplication

The Semai language is rich in expressives, a word class found in many Mon-Khmer languages.[116] Expressives in Semai exhibit their own patterns of reduplication. Diffloth describes several patterns of reduplication used in expressives, two of which are related here.

Type I: Major reduplication

In major reduplication, the major (ultimate) syllable is repeated two or three times. The meaning conveyed is an iterative one; that is, of 'repetition at intervals of time.'

	Root	Expressive	
Monosyllabic	C_1VC_2	$C_1VC_2.C_1VC_2$ ($.C_1VC_2$)	
(10)	*tus*	*tus.tus.tus*	'repeated sounds of running fast'
Sesquisyllabic	$C_3əC_1VC_2$	$C_3ə.C_1VC_2.C_1VC_2$ ($.C_1VC_2$)	
(11)	*dᵊ.jɔ̃ɔ̃l*	*dᵊ.jɔ̃ɔ̃l.jɔ̃ɔ̃l*	'the appearance of an object floating down a river and getting stuck here and there'

Type II: Minor reduplication

In minor reduplication, the reduplication is a prefixed CəC syllable, as in (12) and (13) below:

	Root	Expressive	
Monosyllabic	C_1VC_2	$C_1ər.C_1VC_2$	
(12)	*hããc*	*hᵊr.hããc*	'the sound of dragging something'
Sesquisyllabic	$C_3əC_1VC_2$	$C_3əC_2.C_1VC_2$	
(13)	*dᵊ.jɔ̃ɔ̃l*	*dᵊl.jɔ̃ɔ̃l*	'the appearance of an object which goes on floating down a river'

The onset of the reduplicant is a copy of the first consonant of the root. The coda of the reduplicant is /r/ if the root is a monosyllable, or a copy of the final consonant of the root if the root is more complex. The meaning conveyed is that of "prolongation or continuous repetition in time" (Diffloth 1976b).

Diffloth also provides other expressive forms, such as /cw-cruhaaw/ 'the sound of water falling', derived from the root /*cruhaaw/. This form is of particular importance because the leftmost copy of the root has been reduced to its 'shell'; that is, only its first and last segments. Later in this paper we will see this same phenomenon occurring in verbal reduplication.

3. Varieties of verbal reduplication in Semai

This paper will be limited primarily to looking at incopyfixation, the verbal reduplication outlined in §2.1. Across the numerous Semai dialects there is a wide variety of manifestations of this reduplication, from conservative dialects which preserve the full coda in the reduplicant to dialects that no longer have productive reduplication.

[116] All of the information in this section on expressive reduplication is from Diffloth's paper on Semai expressives (1976b).

3.1. Conservative or "Classic" Aslian Incopyfixation: Betau dialect (full coda copy)

Sesquisyllabic Roots

When a verb undergoes incopyfixation, a copy of the final consonant C_2 is infixed between the first two consonants, C_3 and C_1. The excrescent vowel is placed between the resulting first two consonants, yielding the following CvC-CVC pattern:

C_3 v C_1 V C_2 → C_3 v **C_2** C_1 V C_2

(14)	*tᵊhɔr*	*tᵊrhɔr*	'to incant'
(15)	*kᵊwaac*	*kⁱcwaac*	'to scratch (chicken)'
(16)	*bᵊlɨh*	*bᵃhlɨh*	'to see'
(17)	*bᵊdil*	*bᵊldil*	'to shoot'
(18)	*hɛɛᵍŋ*	*hᵊŋhɛɛᵍŋ*	'to fly'

Monosyllabic Roots

A typical Semai monosyllabic root follows the same incopyfixation pattern as sesquisyllabic roots, making up for the missing C_3 by first making a copy of the first consonant C_1. Diffloth describes it as a 2-step process: (1) initial reduplication and (2) incopyfixation of the final. The resultant form is:

C_1 v **C_2** C_1 V C_2

(19)	*ɟar*	*ɟᵊrɟar*	'to run'[117]
(20)	*bɨɨc*	*bⁱcbɨɨc*	'to lick'
(21)	*lɛɛp*	*lᵊplɛɛp*	'to hunt'
(22)	*nɔɔm*	*nᵊmnɔɔm*	'to urinate'
(23)	*koh*	*kᵃhkoh*	'to cut'

3.2. Retention of nasals in reduplicant: Pahlawan dialect

All Semai dialects in the state of Perak have undergone a reduction of preploded final nasals to simple voiceless plosives. In many of these dialects, however, the historical final nasal is still retained, but only in the reduplicant, as in (24) and (26) below.[118]

(24)	*ɟɨk*	*ɟⁱɲɟɨk*	'to stand'	(proto-Semai: **ɟɨᵍŋ*)
(25)	*ɟɨk*	*ɟᵊkɟɨk*	'to marry'	(proto-Semai: **ɟɨk*)
(26)	*hɛɛk*	*hᵊŋhɛɛk*	'to fly'	(proto-Semai: **hɛɛᵍŋ*)

3.3. Reduction of plosives in reduplicant: Kampar dialect

In several Perak dialects of Semai, final plosives in the root are reduced to glottal stops in the reduplicant.

(27)	*ciip*	*cᵊʔciip*	'to walk'
(28)	*ɟɔt*	*ɟᵊʔɟɔt*	'to sway'
(29)	*pec*	*pⁱʔpec*	'to throw away'
(30)	*kɨɨʔ*	*kᵃʔkɨɨʔ*	'to vomit'

Interestingly, the vowel of the penultimate syllable still reflects the original consonant even though it has

[117] These examples are from the Terisu dialect.

[118] Note that there is a practical issue here. If there are two homophonous verbs (e.g. /ɟɨk/ - one means 'to stand' and the other means 'to marry' in some Perak dialects) that have no difference on the surface of the base form and yet are reduplicated in two different ways, then obviously no theory of reduplication can possibly predict how all words will reduplicate based solely on their root form. Dictionaries must therefore specify the reduplicated form for affected verbs or else provide the historic origin of such words just so that the reduplicated form can be determined.

subsequently been reduced to a glottal stop. For instance, the /i/ in (29) reflects the final consonant /c/, and the /ə/ of (27) and (28) reflects a non-glottal final consonant.

3.4. Reduction of fricatives in reduplicant: Tapah dialect

In a few Perak dialects of Semai, the final fricative /-s/ in the root is reduced to the glottal fricative /-h/ in the reduplicant.

(31)	*pɨɨs*	*p^{i}hpɨɨs*	'to sweep'	(proto-Semai: **pɤɤs*)
(32)	*wees*	*w^{i}hwees*	'to leave behind'	(proto-Semai: **wees*)

Note that the palatal nature of the /s/ is reflected in the /i/ excrescent vowel, which only occurs before palatal codas in the minor syllable. Normally the vowel /a/ is expected before /h/ in the reduplicant.

3.5. Reduction of all C in reduplicant: Bota dialect

In a few Perak dialects of Semai, all final consonants except for nasals and approximants in the base are reduced to glottal stops in the reduplicant.

(33)	*b^{ə}dil*	*b^{ə}ʔdil*	'to shoot'	
(34)	*pɨɨs*	*p^{ə}ʔpɨɨs*	'to sweep'	
(35)	*puut*	*p^{ə}ʔpuut*	'to blow'	
(36)	*hɔ̃ɔ̃ŋ*	*h^{ə}ŋhɔ̃ɔ̃ŋ*	'to whistle'	
(37)	*ceek*	*c^{i}ɲceek*	'to sew'	(proto-Semai: **ciə^{g}ŋ*)
(38)	*gɨɨj*	*g^{i}(j)gɨɨj*	'to sit'	

3.6. Reduction of nasals to plosives in reduplicant: Terisu dialect

In a few Perak dialects of Semai (at the northern reaches of the Semai territory), final nasals in the root are reduced to simple voiceless plosives in the reduplicant as well as the root.

(39)	*coop*	*c^{ə}pcoop*	'to dig'	(Proto-Semai: **coo^{b}m*)
(40)	*lɛɛp*	*l^{ə}plɛɛp*	'to hunt'	(Proto-Semai: **lɛɛp*)
(41)	*liip*	*l^{ə}pliip*	'to swallow'	(Proto-Semai: **lii^{b}m*)
(42)	*t^{ə}hool*	*t^{ə}lhool*	'to blow'	(Proto-Semai: **təhool*)

3.7. Loss of all final -C in reduplicant: Tanjung Malim dialect

In the Tanjung Malim area in southern Perak, incopyfixation is effectively no longer productive. Most verbs cannot be reduplicated, although some frozen reduplicated forms do exist. There are very few verbs where both the non-reduplicated form and the reduplicated form were both acceptable, and even then the acceptability was deemed marginal by native speakers. It appears each verb was frozen either in its base form (most commonly) or in its reduplicated form (rarely). Nevertheless, where the reduplicated form does exist, yet another manifestation of reduplication is evident: all final consonants (except nasals) have been deleted from the reduplicant.

(43)	*sɛɛʔ*	*s^{a}sɛɛʔ*	'to be wrong'
(44)	*tɔɔh*	*t^{a}tɔɔh*	'to spit'
(45)	*gɨl*	*g^{ə}gɨl*	'to flow'
(46)	**ɟɨk*	*ɟ^{i}ɲɟɨk*	'to stand'

However, once again the V in the reduplicant reflects its historical roots: the /a/ in the penultimate in (43) and (44) reflects a glottal final consonant (-ʔ/-h).

4. Analysis: the reason for the rhyme

Reduplication is extremely common in the world's languages. Reduplication is generally defined as the repeating of some Base string within a word [Wilbur 1973, Bauer 1988, Matthews 1991]. It is quite common for the repeated part of the word, the Reduplicant, to be morphologically reduced. What is rare is for the Reduplicant to contain parts of both ends of the Base, as is seen in Aslian incopyfixation. Rarer still is for the Reduplicant to retain segments that apparently are part of the root (at least historically), yet not present in the Base. This retention is found in Semai and a few other Aslian languages. The following sections contain a very brief summary of some recent theories, as well as a new angle for analyzing Aslian reduplication from the perspective of iambic constraints.

4.1. Previous analyses

This type of reduplication, which is prevalent in several Aslian languages, has been called "coda copy" (e.g. Kruspe 2004) or "incopyfixation" (Matisoff 2003) because of the apparent infixing of a copy of the coda.

(47) *t ə h ɔ* ***r*** → *t ə* ***-r-*** *h ɔ r* 'to incant'

BASE REDUPLICATION

However, this characterization has a weakness in that for various dialects of Semai the coda of the penultimate is not always a copy of the base, at least not of the contemporary word, as in the following example:

(48) *g ə s aa* ***k*** → *g ə* ***-n-*** *s aa k* 'to dance' (proto-Semai: *******gəsaa*^{g}*ŋ*)

Incopyfixation has also been noticed by a few researchers who have been scouring the available literature looking for exotic reduplication forms against which they have tested their theories (see for instance Gafos 1998a&b; Hendricks 2001). In a nutshell, Aslian incopyfixation severely complicates most theories because it requires "long-distance" copying of segments. Skipping segments and/or the crossing of association lines is not allowed by most reduplication theories. Furthermore, most reduplication theories assume a "Base-Reduplicant" scheme wherein part of the word can be identified as the Base, all or part of which is copied in the Reduplicant. Semai incopyfixation tends to confound these theories because the two different copies of the root preserve different segments of the *original* (historic) base. To wit, if we call the penultimate (pre-)syllable the reduplicant (because of its generally reduced nature), it becomes difficult to explain why there are segments in the reduplicant that cannot be considered copies of the base because the base itself (the ultimate syllable) may no longer preserve the original (historical) segment, as in (48) above.

4.2. Iambic reductions

It has been proposed that all reduplications originate from full copies of the root which are then reduced. For instance, Berry (1998:162) states, "Full-copy analysis ... avoids the problems of association to segmental templates, but also lacks an explanatory account of the reduplication patterns. In full-copy analysis, the full root is copied and then reduced, by rule, to meet template requirements." Not surprisingly, the Aslian languages also have their own "template requirements." Of particular interest to this discussion is the iambic shape of most roots and reduplicated forms. As mentioned in §1.2 above, there appears most crucially to be prosodic pressure to squeeze words into a canonical iambic shape.

Iterative Iambification

In this scenario, the iambification process is seen as being applied iteratively and at different structural levels. The following example uses the root /təhɔr/ "to incant" from the conservative Betau dialect.

(49)

a) *təhɔr*	$C_3{}^əC_1VC_2$ **W S**	Base The root word has an iambic contour of Weak plus Strong syllables.
b) *təhɔr – təhɔr*	$C_3{}^əC_1VC_2 – C_3{}^əC_1VC_2$ PW_1(**W**) PW_2(**S**)	Full Copy Replication: PW_1-PW_2 Initially, two full copies of the root word are posited. The resulting PPh also has an iambic contour: PW1 is Weak and PW2 is Strong.
c) *tər – təhɔr*	$C_3{}^ə(C_1V)C_2 – C_3{}^əC_1VC_2$	Evisceration of PW_1 As the weaker PW, PW1 is 'gutted of its innards' leaving the shell of its first and last segments.
d) *tər – hɔr*	$C_3{}^əC_2 – (C_3{}^ə)C_1VC_2$	Loss of PW_2 presyllable Although PW2 is strong, its presyllable is weaker than PW1 and thus is lost.
e) *tərhɔr*	$C_3{}^əC_2C_1VC_2$	Final Iambic Form

To recapitulate the process in (49) above, two full copies of the root word are first posited. Because of the iambic contour of the prosodic phrase (PPh), the first prosodic word (PW_1) is weak while the second prosodic word (PW_2) is strong. The first application of iambification is at the PPh level: as the weaker of the two, PW_1 undergoes the first reduction and is 'gutted of its innards' leaving only the 'outer shell' of its initial and final consonants. Then, another iteration of the iambic process operates on the second PW (PW_2), causing it to drop its penultimate syllable altogether in order to maintain an overall iambic contour as the two prosodic words merge into one aggregate iambic prosodic word. In other words, the minor syllable of PW_2 is weaker than both the (now gutted) PW_1 and the major syllable of PW_2, hence it is deleted resulting in the final iambic form. Another way to view the motivation for the second process is that preservation of some portion of PW_1 is of a higher rank than preservation of the full form of PW_2. Indeed, were it not so, PW_1 would be lost entirely and with it any trace of the reduplication.

Iterative Iambification can also be interpreted as the transmogrification of the PPhrase into a PWord by limitation of the number of syllabic peaks. The core of syllabicity (that is, the vocalic peak) of PW_1 is reduced, then the remaining consonants are re-aligned according to the remaining syllabic peaks. This can produce some unwieldy sequences, which over time are ameliorated by further assimilation or deletion.

Note that this series of reductions also applies to monosyllabic roots, only that the final step (d) is unnecessary because the PW_2 does not have a penultimate syllable.

Elision Plus Iambification

Another scenario involves a phonotactic constraint requiring the elision of the middle consonant in a medial cluster of three consonants. This analysis unfolds as follows:

(50)

a) *təhɔr*	$C_3{}^əC_1VC_2$ **W S**	Base The root word has an iambic contour of Weak plus Strong syllables.
b) *təhɔr - təhɔr*	$C_3{}^əC_1V\mathbf{C_2} – \mathbf{C_3{}^əC_1}VC_2$ PW_1(**W**) PW_2(**S**)	Full Copy Replication: PW_1-PW_2 Initially, two full copies of the root word are posited. The resulting PPh also has an iambic contour: PW_1 is Weak and PW_2 is Strong.

c) *təhɔr - hɔr*	$C_3{}^əC_1V\mathbf{C_2} - (\mathbf{C_3}{}^ə)\mathbf{C_1}VC_2$	Middle Consonant Elision: C_3 of $C_2C_3C_1$ As the two PWs merge into one PPh, a dispreferred sequence of 3 consonants is generated and the weakest of the 3, the presyllable of PW_2, is elided.
d) *tər - hɔr*	$C_3{}^ə(C_1V)C_2\,C_1VC_2$	Evisceration of PW_1 As the weaker PW, PW_1 is 'gutted of its innards' leaving the shell of its first and last segments.
e) *tərhɔr*	$C_3{}^əC_2C_1VC_2$	Final Iambic Form

To recapitulate the process in (50) above, two full copies of the root word are first posited. The resulting phonological phrase over time undergoes pressure to become one phonological word. As this happens, a sequence of three consonants is created, consisting of the final consonant of PW_1 and the first two consonants of PW_2 (that is, $C_2C_3C_1$). The medial consonant of the three, C_3, is elided because it is the weakest of the three, belonging to the weak penultimate syllable of PW_2. Finally, as a means of creating an allowable penultimate syllable, PW_1 is gutted of its innards, leaving only its initial and final consonants intact.

In summary, whether by Iterative Iambification or by Elision Plus Iambification, the end result is the same. The final form preserves the ultimate (and more salient) syllable of the root, and it conforms to the iambicity that Semai prefers. Although reductions have been made, segments from both copies of the root are preserved. The reduplicated forms found in other dialects can be explained by adding further reductions to the aforementioned process. Note that strictly speaking, only reductions have actually taken place, not infixation; hence the term "incopyfixation" is more of a synchronic label arising from the *apparent* infixing of the final consonant.

4.3. Incarnations of incopyfixation in Semai

As demonstrated in §3 above, Semai has a wide variety of incopyfixation-like reduplication. This section will demonstrate how each manifestation of Semai incopyfixation can be viewed as a consequence of full-copy reduplication followed by various reductions resulting in an iambic form.

Langkap dialect – preservation of nasals in the penult, loss in the ultima

As mentioned earlier, Perak dialects have reduced historical final nasals – if not following a nasal vowel – to a homorganic voiceless plosive. However, the original nasal is preserved in the reduplicant in many of the Perak dialects. For example, the historical form /ɟɨᵍŋ/ 'to stand' is still used in many of the dialects of Pahang. In the Langkap (Perak) dialect the word is /ɟɨk/, as witnessed in the imperative.

(51) ɟɨk! 'Stand up!'

Although the final nasal has been reduced to a voiceless plosive, the nasal still surfaces in the reduplicated form: /ɟɨɲɟɨk/.

(52) 'to stand'	*ɟɨᵍŋ*	root (historic form)
	ɟɨᵍŋ – ɟɨᵍŋ	full copy replication
	ɟəŋ – ɟɨᵍŋ	reduction of internal segments of 1st PW
	ɟəŋ – ɟɨᵍŋ	[loss of penult of 2nd PW – no effect]
	ɟəŋ – ɟɨk	reduction of final nasal to homorganic plosive
	ɟiɲ – ɟɨk	assimilation of penultimate nasal coda and vowel
	ɟʲɲɟɨk	final iambic form

Thus, familiar iambic processes of weakening on the left in a reduplicated PPhrase can be seen in (52), where it is clear that such forms reflect:

a) Full reduplication as a starting point.

b) More than one stage of weakening/assimilation in the leftward reduplicant PW.

c) Final devoicing (-ŋ > -k) has affected all free forms of this morpheme whether in simple or reduplicated items. Internal cases of original -ŋ in the reduplicated PPhrase are protected from the application of this historical rule of change.

Kampar dialect – reduction of plosive coda in the penult

It is not the case that the coda of the reduplicant just so happens to be a conservative environment in general. The historical changes illustrated in (52) affect constituents at their final margins and therefore do not change internal codas within a reduplicative phrase. Such processes are to be fundamentally distinguished from the internal processes of reduction as iambic effects that take place. In several Perak dialects all final plosives in the root are reduced to a glottal plosive in the reduplicant as a result of an internal process of reduction due to iambic effect. Such reductions are found in the Kampar dialect.

(53)	'to walk'	*ciip*	root
		ciip – ciip	full copy replication
		cəp – ciip	reduction of internal segments of 1st PW
		cəʔ – ciip	reduction of reduplicant's final plosive to glottal
		cᵊʔciip	final iambic form

As mentioned in §1.1, if the coda of the root is a palatal (that is, /c/, /ɲ/, /j/, or /s/), the reduplicant (left constituent) vowel is /i/ rather than /ə/, as illustrated in (54):

(54)	'to throw'	*pec*	root
		pec – pec	full replication
		pic – pec	reduction of internal segments of 1st PW
		piʔ – pec	reduction of coda plosive
		pⁱʔpec	final iambic form

Remarkably, this vowel (/i/ in (54) is preserved even when the coda is subsequently reduced to a glottal stop; that is, the vowel in the penult reflects assimilative effects of the original palatal /c/, not the synchronic glottal /ʔ/.

If the coda is glottal, the reduced reduplicant vowel is /a/ rather than /ə/, as in (55):

(55)	'to vomit'	*kɨiʔ*	root
		kɨiʔ – kɨiʔ	full copy replication
		kaʔ – kɨiʔ	reduction of internal segments of 1st PW
		kaʔ – kɨiʔ	[reduction of coda plosive – no effect]
		kᵃʔkɨiʔ	final iambic form

Tapah dialect – reduction of fricative coda in the penult

In the Tapah dialect, a final /s/ in the root is reduced to the glottal fricative /h/ in the penultimate (reduplicant) copy.

(56)	'to sweep'	*pɨis*	root
		pɨis – pɨis	full copy replication
		pis – pɨis	reduction of internal segments of 1st PW
		pih – pɨis	reduction of coda of 1st PW
		pⁱhpɨis	final iambic form

Moreover, since /s/ patterns as a palatal in Aslian languages, the reduced reduplicant vowel is /i/. This assimilation happens before /s/ is reduced to /h/. Note that any proposal that calls for copying the final coda of the base onto the reduplicant fails because of historical changes to codas in either reduplicant or base.

Batuuq Enam-belaih dialect – reduction of /s/ in the ultima

In Batuuq Enam-belaih, which is geographically and linguistically very close to Tapah, the final /s/ appears to have undergone changes in both the reduplicant penult and in the ultima, as shown in (57):

(57)	'to sweep'	**pɨɨs*	root (historic; modern root is pɨɨy)
		pɨɨs – pɨɨs	full copy replication
		pis – pɨɨs	reduction of internal segments of 1st PW
		pih – pɨɨs	reduction of coda plosive of 1st PW
		pih – pɨɨj	reduction of coda plosive of 2nd PW
		p^{i}hpɨɨj	final iambic form

The final /s/ is becoming /ih/, and in many words the sound change has progressed further and dropped the fricative altogether, resulting in simply /i/ (or rather /j/, since this sound now fills the coda consonant position). The result is that neither the reduplicant nor the base now contain the original final /s/. Note, however, that the change in the reduplicant penult of /s/ to /h/ is an iambic reduction, whereas the change of the final /s/ to /j/ is a "generic" sound change having nothing to do with iambification.

Terisu dialect – reduction of nasal in both copies

Although Cameron Highlands is on the Pahang side of the central mountain range, Semai speakers in this area speak more like their Perak counterparts in that final nasals have been reduced to homorganic voiceless plosives. However, for speakers of the Terisu dialect the final nasal has been reduced in both the copies of root, not just the final copy as with the majority of Perak dialects.

(58)	'to swallow'	*liip*	root (modern) (Proto-Semai: **lii^{b}m*)
		liip – liip	full copy replication (of modern root)
		ləp – liip	reduction of internal segments of 1st PW
		l^{ə}pliip	final iambic form

This case may seem less interesting because it looks more like standard reduplicant-base reduplication. However, in light of the fact that most of the other dialects retain the final nasal of the root in the first copy of the reduplication, it may prove difficult to explain why this one dialect reduces both copies to plosives. Incopyfixation is an ancient Aslian (not just Semai) process and reduction of nasals to homorganic plosives is relatively recent and only applies to final nasals.[119] Why, then, does the Terisu dialect reduce medial nasal codas to homorganic plosives in reduplicants, not just in the base?

Tanjung Malim dialect – loss of non-nasal codas

In the Tanjung Malim area, the final consonant (if non-nasal) is lost in the reduplicant. However, in certain cases a trace has been left behind when the penultimate (reduplicant) vowel still reflects the nature of the coda before its loss, as seen in (59).

(59)	'to be wrong'	*sɛɛʔ*	root
		sɛɛʔ – sɛɛʔ	full copy replication
		saʔ – sɛɛʔ	reduction of internal segments of 1st PW
		sa – sɛɛʔ	loss of coda of 1st PW
		s^{a}sɛɛʔ	final iambic form

[119] Forms such as *k^{ə}ntɔp* "right (side)" (Proto-Semai *k^{ə}ntɔ^{b}m) and *k^{ə}ŋjɛɛk* "wing" (Proto-Semai *k^{ə}njɛɛ^{g}ŋ) prove that Terisu reduces final nasals, but not medial nasals, to plosives.

The vowel /a/ in the penult (reduplicant) would not be expected unless followed by a laryngeal segment (/ʔ/ or /h/). Historically there was the glottal /ʔ/ in the coda position of the penult, but this has been lost over time. Loss of coda in the penult is consistent with iambic reduction of the left reduplicant copy.

Summary of Incopyfixes of Semai dialects

The following table summarizes the rhymes of the reduplicant for each final consonant (in its proto-form). Blank cells represent non-existent or insufficient data.

Table 1. Rhyme of reduplicated presyllable, by dialect

*–C	Betau	Terisu	Tapah	Langkap	Kampar	Bota	T. Malim
m	*əm*	*əp*	*əm*	*əm*	*əm / əʔ*	*əm / əʔ*	
n	*ən*	*ət*	*ən*	*ən*	*ən / əʔ*	*ən / əʔ*	
ɲ	*iɲ*	*ic*	*iɲ*	*iɲ*	*iɲ / iʔ*	*iɲ / əʔ*	
ŋ	*əŋ*	*ək*	*əŋ*	*əŋ*	*əŋ / əʔ*	*əŋ / əʔ*	*əŋ*
p	*əp*	*əp*	*əʔ*	*əʔ*	*əʔ*	*əʔ*	*ə*
t	*ət*	*ət*	*əʔ*	*əʔ*	*əʔ*	*əʔ*	*ə*
c	*ic*	*ic*	*iʔ*	*əʔ*	*iʔ*	*əʔ*	
k	*ək*	*ək*	*əʔ*	*əʔ*	*əʔ*	*əʔ*	*ə*
ʔ	*aʔ*	*aʔ*	*aʔ*	*aʔ*	*aʔ*	*aʔ / əʔ*	*a*
s	*is*	*is*	*ih*	*əs*	*iʔ*	*əʔ*	*ə*
h	*ah*	*ah*	*ah*	*aʔ*	*aʔ*	*əʔ*	*a*
r	*ər*	*ər*	*ər*	*ər*	*əʔ*	*əʔ*	*ə*
l	*əl*	*əl*	*əl*	*əl*	*əʔ*	*əʔ*	*ə*
j	*ij*	*ij*	*ij*	*ij*	*iʔ*	*ij/əʔ*	*i*
w	*əw*	*əw*		*əw*	*əʔ*	*əʔ*	

The diachronic picture

The reduction of reduplicants in Semai dialects is continuing to undergo changes. As indicated in Table 1 above, the Kampar and Bota dialects have some verbs that historically had nasal endings where now the reduplicant shows either the original nasal or sometimes a plosive. Apparently the original nasal ending was lost in the base and is now slowly being lost in the reduplicant. Older language speakers preserve more of the nasals. Furthermore, several speakers commented that both forms (with the nasal preserved or lost) could be heard, but that the form with the nasal preserved recalled to mind how their grandparents once spoke.

The data from many of the dialects tended to have a few inconsistencies and/or multiple forms that were deemed acceptable. In virtually all cases, however, one of the forms fits an older, more conservative pattern of reduplication (represented in Table 1) and the other form represents a newer, regularized pattern where more information is lost regarding the historic root. A few examples are given here.

	Root	Conservative	Regularized	Gloss	
(60)	*huuc*	*h^{i}ʔhuuc*	*h^{ə}ʔhuuc*	‘to climb’	(Kampar)
(61)	*hoʔ*	*h^{a}ʔhoʔ*	*h^{ə}ʔhoʔ*	‘to love’	(Bota)
(62)	*kə^{b}m*	*k^{ə}mnəm*	*k^{ə}pnəp*	‘to get’	(Terisu, nominal form)

5. Incopyfixation in other languages

5.1. Other Aslian

Incopyfixation is an ancient process in Aslian, and has survived in all three major divisions – North Aslian (NA), Central Aslian (CA), and Southern Aslian (SA). This section provides a summary of incopyfixation that has been found to date, including the differences from the phenomena found in Semai.

Semnam (CA)

Semnam has preploded nasals in words with long vowels, but final nasals are reduced to plosives in words with short vowels in the major syllable (Burenhult & Wegener 2009).

	Root	Progressive	Phonetic	Gloss
(63)	*kɔɲ*	*kɲkɔc*	[kəɲkɔc]	'to sit'
(64)	**h^{u}ooc*	*hch^{u}ooc*	[hich^{u}ooc]	'to whistle'
(65)	**blaaj*	*bjlaaj*	[bilaaj]	'to be high'
(66)	*h^{u}oom*	*hmh^{u}oom*	[həmh^{u}oo^{b}m]	'to hug'
(67)	**ktɔm*	*kmtɔp*	[kəmtɔp]	'to spit'

The excrescent vowel appears to be [i] before /c/ or /j/ but [ə] before /ɲ/ or /s/. In Semnam, when the vowel of the major syllable is non-nasal, final nasal consonants are preploded after long vowels, as in (66), and have been reduced to homorganic voiceless stops when the vowel is short, as in (67). Note that while (66) evidences just one iambic effect (the reduction of the vowel in the left reduplicant copy), example (67) shows two iterations of iambification (the reduction of the innards of the left reduplicant copy and the loss of the penultimate syllable of the right copy). Also note that (67) has parallels to Semai example (48) in that both involve reduplications where the left and right copies preserve different segments of the historic root.

Jahai (NA)

The Jahai imperfective form of verbs (Burenhult 2005) follows the "classic" Aslian incopyfixation pattern that looks precisely like that of conservative Semai dialects.

	Root	Progressive	Phonetic	Gloss
(68)	*kwẽs*	*kswẽs*	[kiswẽs]	'to sweep'
(69)	*kɟeŋ*	*kŋɟeŋ*	[kəŋɟe^{g}ŋ]	'to listen'
(70)	*gulɛm*	*gumlɛm*	[gumlɛ^{b}m]	'to carry'
(71)	*tigil*	*tilgil*	[tilgil]	'to go around'
(72)	*bakɛs*	*bskɛs (*baskɛs)*	[biskɛs]	'to grow up'

The excrescent vowel is /i/ before palatals /c, ɲ, s, j/, /a/ before glottals /ʔ, h/, and /ə/ elsewhere. Final nasals are preploded after non-nasal vowels, but are realized as simple nasals when reduplicated in the penultimate syllable.

A notable difference in Jahai is that the incopyfixation of the final consonant can occur on disyllabic roots where the vowel in the penultimate syllable is specified; that is, something other than the excrescent schwa. For these words the final consonant appears to be infixed as before (as the coda of the penultimate), but the penultimate vowel remains as it was in the original root. The exception to this rule for disyllabic roots is when the penultimate vowel is /a/, in which case the penultimate /a/ is replaced with the standard excrescent vowel, phonetically assimilated to the coda consonant (that is, /s/).

Jahai also displays a number of other types of verbal affixation that include incopyfixation as part of its morphology; those data will not be covered in this paper.

Kensiw (NA)

Kensiw has incopyfixation, but it has not been fully characterized yet (Peterson, p.c.).

	Hypothetical Root	Reduplicated	Gloss
(73)	*ŋap*	*ŋ^{ə}p.ŋap*	'to chew'
(74)	*hipɯh*	*hih.pɯh*	'to moan'
(75)	*plit*	*p^{ə}t.plit*	'to spit'
(76)	*ʔsek*	*ʔek.sek*	'to cough'
(77)	*ʔipa^{d}n*	*ʔi^{d}npa^{d}n*	'to hurry'

In some cases the pattern is classic Aslian incopyfixation, but there are also some remarkable variations such as (a) the nasal in the reduplicant (penultimate syllable) is often preploded, as in (77) 'to hurry'; (b) the vowel of reduplicant is sometimes a copy of the base, as in (76); (c) the vowel of the reduplicant is unpredictable, as in (74); and (d) in some cases there is a consonant cluster in the base, but not in the reduplicant, as in (75).

Jah Hut (position uncertain)

Diffloth (1976c) spells out a curious variation in Aslian reduplication for process verbs in the Jah Hut language.

	Root	Progressive	Phonetic	Gloss
(78)	*klɯŋ*	*kŋʔlɯŋ*	[kəŋʔlɯŋ]	'to speak'
(79)	*rɛɲ*	*rŋʔrɛɲ*	[rəŋʔrɛɲ]	'to gnaw'
(80)	*tin*	*tʔtin*	[təʔtin]	'to sharpen'
(81)	*cjɛk*	*cʔcjɛk*	[ciʔcjɛk]	'to sleep'
(82)	*wec*	*wʔwec*	[wuʔwec]	'to climb'

The onset of the reduplicant (penultimate) is always a copy of the initial segment of the base. For roots that have a non-plosive initial and a nasal final segment, the reduplicant rhyme is /əŋʔ/; elsewhere, the reduplicant rhyme is generally /əʔ/. In the elsewhere case there are also some 'anticipatory' phonetic changes where the schwa in the reduplicant is actually [u] if the initial segment of the root is bilabial, and [i] if the initial segment is palatal.

The elsewhere case, where nearly every non-nasal final is reduplicated as a glottal stop, looks similar to reduplication seen in Bota Semai. Where nasals are preserved they are rendered as /ŋʔ/ in the reduplicant. The glottal stop is a bit of mystery. However, Jah Hut (Diffloth 1976c: 110) as well as other Aslian languages[120] appear to have co-articulated glottal closures with many of their final segments. This could easily account for the glottal stop.

Temiar (CA)

The situation with Temiar is complicated by dialectal variations.[121] For example, where words in the northern dialect end with a voiced stop, the southern dialect has a voiceless stop; conversely, where the southern dialect has a voiced final stop, the northern dialect has a voiceless stop. By comparing Temiar to other Aslian languages, it appears that the northern dialect's final voiceless plosives are a reflex of a proto final nasal when the nasal is not following a nasal vowel.[122]

[120] See, for instance, Kruspe's dictionary of Mah Meri (Kruspe 2010).

[121] The examples provided in this section are from Northern Temiar. All of the data come from Benjamin (1976a).

[122] This is a tentative analysis. It helps explain why there is a nasal at the end of some words with nasal infixes; namely, the nasal infix nasalizes the vowel of the major syllable which in turn prevents the final nasal from being reduced to a plosive; e.g. /*kəm/ 'to know' is reduced to [kəp], but /kɛpnəm/ 'knowledge' preserves its final nasal.

	Gloss	Proto	N. Temiar	Other Aslian
(83)	'to cry'	**jaam*	*jaap*	*jaa^{b}m* [Semnam]; *ja^{b}m* [Semaq Beri]
(84)	'to bury'	**kəm*	*kəp*	*kɨ^{b}m* [Jahai]; *kə^{b}m* [Semelai]

As such Northern Temiar patterns much as the Tapah and Pahlawan Semai dialects do for roots that historically had a nasal ending: final nasals are preserved in the reduplicant, but reduced to a voiceless plosive at the end of the word.

	Root	Progressive	Gloss
(85)	*jaam*	*j^{ɛ}mjaap*	'to cry'
(86)	*s^{ə}lɔɔŋ*	*s^{ɛ}ŋlɔɔk*	'to hunt successfully'
(87)	*s^{ə}lɔg*	*s^{ɛ}glɔg*	'to lie down'
(88)	*s^{ə}maaɲ*	*s^{ɛ}ɲmaaɲ*	'to ask'
(89)	*t^{ə}lɛŋ*	*t^{ɛ}ŋlɛk*	'to teach'

Southern Temiar has voiced plosives as the reflex of proto final nasals. When reduplicated, the synchronic plosive shows up in the reduplicant regardless of whether it is voiced or voiceless. The historic final nasal is essentially lost. Southern Temiar therefore patterns very much like Terisu Semai in that the synchronic final segment always shows up in the reduplicant. It is probably not a coincidence that Terisu is at the very northern reaches of the Semai territory and indeed overlaps with the southernmost Temiar territory. For both Southern Temiar and for Terisu Semai a simple explanation is that this type of reduplication is highly productive and operates freely on contemporary forms rather than being tied to historic roots, hence when a root/base reduces a final nasal to a homorganic voiceless stop, the same voiceless stop is then reflected in the reduplicant.

The bases and reduplicants are summarized in the following table:

	Northern Temiar		Southern Temiar	
ProtoAslian	Reduplicant	Base	Reduplicant	Base
**-m*	*-m*	*-p*	*-b*	*-b*
**-p*	*-b*	*-b*	*-p*	*-p*

Other Aslian

My brief visits to the Mah Meri (SA) speakers did not reveal any verbal reduplication. Work done by Asmah (2006) and Kruspe (2010) appear to confirm that Mah Meri does not have verbal reduplication.

Verbal reduplication did appear in the rest of the wordlists that I took or to which I had access at other locations. However, generally no attempt was made in these lists to determine if the apparent root was attested in its non-reduplicated form, or to find out what meaning reduplication added to the root. Therefore, the following items give evidence of reduplication but must be investigated further.[123]

	Root	Reduplicated	Gloss	Language
(90)	**l^{ə}puɯs*	*l^{a}spuɯs*	'fragrant'	Tonga/Mos (NA)
(91)	**n^{ə}hawf*	*n^{ə}fhawf*	'to breathe'	Mintil (NA)
(92)	**p^{ə}lɔɲ*	*p^{ə}ɲlɔɲ*	'to sing'	
(93)	*k^{ə}lɨgŋ*	*k^{ə}ŋlɨgŋ*	'to speak'	Chewong (NA)
(94)	*(n^{ə})lɨk*	*n^{ə}glɨk*	'to laugh'	
(95)	*(n^{ə})hagŋ*	*n^{ə}ŋhagŋ*	'to dance'	

[123] The Tonga/Mos data come from unpublished data provided by Mary M. Peterson. The Mintil data are from Benjamin 1976b. The Semaq Beri data are from personal field notes plus unpublished SIL lists.

(96)	**cᵊdɨk*	*cᵊkdɨk*	'to hiccup'	Mendriq (NA)
(97)	**hᵊɟaŋ*	*hᵊŋɟaŋ*	'to stand'	
(98)	**kᵊdɨh*	*kʲhdɨh*	'to talk'	
(99)	*kuj*	*kʲjkuj*	'to be heading'	Batek (NA)
(100)	**hɨj*	*hⁱhɨj*	'to yawn'	
(101)	**kᵊseŋ*	*kᵊŋseŋ*	'to dance'	
(102)	*ɟɨk*	*ɟᵊkɟɨk*	'to breathe'	Kintak (NA)[124]
(103)	*cᵊkoᵍŋ*	*cᵊŋkoᵍŋ*	'to carry'	
(104)	**kᵊwac*	*kʲcwac*	'to swim'	
(105)	**ʔahɔj*	*ʔajhɔj*	'to yawn'	Semaq Beri (SA)
(106)	**cədɨk*	*cɪkdɨk*	'to hiccup'	
(107)	*goŋ*	*gŋgoŋ*	'to carry'	Semelai (SA)[125]
(108)	*smaɲ*	*sɲmaɲ*	'to ask for (s.t.)'	
(109)	*cŋɛw*	*cwŋɛw*	'to look down'	
(110)	*ʔɔh*	*ʔhʔɔh*	'to blowpipe'	

5.2. Regional varieties of Malay

There are three regional varieties of Malay that have been documented to have reduplication patterns which are distinctly odd for Malay, but analogous to the Aslian reduplication discussed in this paper.[126] Namely, in each of these three languages there are examples of reduplicants with codas that reflect the final consonant. These reduplicants bear distinct resemblance to the incopyfixation found in Aslian in that the reduplicants appear to be reduced copies of the entire root, preserving both the initial segment as well as features of the final segment.

In the Perak dialect of Malay, two reduplication phenomena noted by Zaharani (1991:112) are noteworthy.

(111)	*siket*	'a little'	*səʔsiket*	'very little'
(112)	*gəlap*	'dark'	*gəʔgəlap*	'very dark'
(113)	*ɟaman*	'time'	*ɟəŋɟaman*	'for a long time'
(114)	*kəʁen*	'dry'	*kəŋkəʁen*	'very dry'

When the root ends in a plosive, the coda of the reduplicant is the glottal /ʔ/. When the root ends in a nasal, the coda of the reduplicant is the nasal /ŋ/, which assimilates its place of articulation to the next segment. Zaharani comments that "phonologically this is quite an odd rule. It is hard to determine what is the phonetic motivation for it." However, the homeland of the Perak dialect is claimed to be the Bota region, which overlaps the western reaches of the Semai territory; therefore, it seems entirely plausible that Perak Malay has been influenced by Semai. Note, however, that this reduplication process in Perak Malay involves intensification of nouns and adjectives, not verbs.

Ulu Muar Malay is documented by Hendon (1966) as having words formed with reduplicants of a similar nature, replacing final plosives with /ʔ/ and final nasals with a placeless nasal.

---

[124] Asmah (1976) only mentions CV- and CVC- reduplication of verbs in Kintak, where it carries the meaning of continuative aspect. Data in her thesis (Asmah 1963), however, contain examples of incopyfixation in reduplicated verbs. My research confirmed the existence of such forms.

[125] Data from Kruspe (2004). Semelai exhibits classic Aslian incopyfixation, which she refers to as "coda copy."

[126] Further research needs to be done in order to determine if these varieties of Malay have syllable template constraints similar to Aslian's iambic constraints, or if the pattern of reduplication is simply borrowed from Aslian neighbors.

(115)	*laŋit*	‘sky’	*laʔ+laŋit*	‘palate’
(116)	*galap*	‘dark’	*gaʔ+galap*	‘is repeatedly dark’
(117)	*siaŋ*	‘daytime’	*siɲ+siaŋ*	‘during the daytime’
(118)	*paraŋ*	‘sword’	*pam+paraŋ*	‘sword-like decoration’

The differences here are that the vowel in the reduplicant remains unreduced, and there is no consistent semantic meaning associated with the reduplication process. However, the form of the reduplicant is again reminiscent of Aslian incopyfixation. Geographically Ulu Muar borders the Semelai territory.

For Johor Malay, Onn (1976) gives examples of reduplication which denotes intensification of the root. In most cases the reduplicant is simply Cə-, but when the root end in a plosive, the reduplicant is Cəʔ-.

	Root	Intensified	Gloss
(119)	*tutup*	*təʔtutup*	‘to close’
(120)	*tembak*	*təʔtembak*	‘to shoot’

5.3. Other Mon-Khmer

Nancowry, a Nicobarese language documented by Radhakrishnan (1970:49ff) has reduplication that has parallels to incopyfixation.

	Root	Reduplicated	Gloss
(121)	*kap*	*ʔukap*	‘to bite’
(122)	*tin*	*ʔintin*	‘to push’
(123)	*cim*	*ʔumcim*	‘to mourn’
(124)	*cat*	*ʔitcat*	‘to jump’

For Kammu, a language spoken in northern Laos, Svantesson (1983) provides the following forms:

	Root	Reduplicated	Gloss
(125)	*l̀mààc*	*lc̀mààc*	‘to be struck’
(126)	*stééñ*	*sń̄tèèñ*	‘small steady still light’
(127)	*kʔááw*	*kẃʔààw*	‘open’
(128)	*r̀ŋɨ̀p*	*rp̀ŋɨ̀p*	‘many people lie down’

Perhaps it is also significant that Shorto (2006) invokes reduplication to explain the forms of various Proto Mon-Khmer reconstructions:

	Proposed source	Proto-MK	Gloss
(129)	**ʔaak-ʔaak →*	**kʔaak*	‘crow’
(130)	**lɔk-lɔk →*	**klɔk*	‘container’
(131)	**leb-leb →*	**blep*	‘to immerse’
(132)	**hiil-hiil →*	**lhiil*	‘bare, smooth, worn’
(133)	**ʔuh-ʔuh →*	**huh*	‘to blow’

6. Conclusion

Generally languages have reduplication in some form or another. However, languages do not always reduplicate in the same way, and as this paper has shown, even dialects within a given language (such as Semai) can exhibit a great deal of variation in reduplication. Semai verbal reduplication is remarkable for other reasons as well. For one, Semai has apparent discontinuous reduplication, where the reduplicant contains segments from both the initial and the final segments of the base. Secondly, in Semai verbal reduplication the reduplicant (penultimate syllable) often preserves segments from the historic root that are lost in the base (ultimate syllable). This fact presents a special challenge for Base-Reduplicant theories since there is no mechanism for the reduplicant to copy information that has been lost in the synchronic form of the base. This also means that dictionaries will need to specify reduplicated forms, at least for certain verbs.

This paper has proposed that Semai verbal reduplication can be analyzed rather straightforwardly as a result of full-copy reduplication with prosodic (specifically iambic) pressures then molding the resulting form into a canonical iambic shape. The term "incopyfixation," while eloquently capturing the synchronic image of the final coda being infixed into the penultimate syllable, is alas misleading – no infixation has taken place, in the view of this paper. That is, no copy has been infixed, but rather a residue of the original full copy survives in these forms.

The approach taken here demonstrates the importance of taking into account the prosodic and metrical dimensions of phonological and grammatical elements in Semai and beyond. It turns out that long-standing problems become more tractable as certain well-known factors are considered. Specifically, this paper suggests that the effects of iambic processes of reduction on the left and merging toward the right can yield some valuable new insights.

Finally, this paper has shown that verbal reduplication, including incopyfixation, can be found in all three major branches of Aslian and therefore must have been a part of proto-Aslian. Data from other languages, both from nearby but unrelated languages as well as geographically distant but historically related languages, exhibit similar patterns of reduplication. Clearly Aslian reduplication has a rich history that we have only just begun to discover.

References

Asmah Haji Omar. 1963. Bahasa Semang – Dialek Kentakbong. Jabatan Pengajian Melayu, Universiti Malaya. Kuala Lumpur.

_____. 1976. "The verb in Kentakbong." In Austroasiatic Studies, eds. P. N. Jenner et al., Part II:951-70. Honolulu: The University Press of Hawaii.

Asmah Haji Omar, (ed.). 2006. Bahasa Mah Meri. Kuala Lumpur: Universiti Malaya.

Bauer, Laurie. 1988. *Introducing Linguistic Morphology*, Edinburgh: Edinburgh University Press.

Benjamin, Geoffrey. 1976a. "An outline of Temiar grammar." In *Austroasiatic studies, Part I*, eds. Philip N. Jenner, Laurence C. Thompson and Stanley Starosta. Honolulu: University Press of Hawaii, pp. 129-187.

_____. 1976b. "Austroasiatic Subgroupings and Prehistory in the Malay Peninsula." In *Austroasiatic Studies, Part I*, eds. Philip N. Jenner, Laurence C. Thompson and Stanley Starosta. pp. 37-128.

_____. 1989. "Achievements and Gaps in Orang Asli Research." *Akademika* 35: pp. 7-46. University Kebangsaan Malaysia.

_____. 2009. "The current situation of the Aslian languages." Revised and updated version. Paper presented at the *Symposium on Language and Diversity*. Division of Linguistics and Multilingual Studies, Nanyang Technological University, 5–6 March 2009. To appear in a *Pacific linguistics* volume on Aslian languages, edited by Nicole Kruspe & Niclas Burenhult.

Berry, Lynn. 1998. Alignment and Adjacency in Optimality Theory: Evidence from Warlpiri and Arrernte. PhD thesis. University of Sydney.

Bishop, Nancy. 1996. "A preliminary description of Kensiw (Maniq) phonology." *Mon-Khmer Studies* 25: pp. 227-253.

Blust, Robert. 1997. "Nasals and nasalization in Borneo." *Oceanic Linguistics 36* (1): 149-179.

Burenhult, Niclas. 2005. *A grammar of Jahai.* Canberra : Pacific Linguistics Research School of Pacific and Asian Studies.

Burenhult, Niclas & Claudia Wegener. 2009. Preliminary Notes On The Phonology, Orthography And Vocabulary Of Semnam (Austroasiatic, Malay Peninsula). *Journal of the Southeast Asian Linguistics Society* 1:283-312.

Carey, Iskandar. 1976. Orang Asli: The Aboriginal Tribes of Peninsular Malaysia. Oxford University Press.

De Lacy, Paul. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge Studies in Linguistics #112. Cambridge, UK; New York, Cambridge University Press.

Diffloth, Gérard. 1968. "Proto-Semai Phonology." *Federation Museums Journal*, Vol. XII, pp. 65-74. Kuala Lumpur.

_____. 1976a. "Minor-syllable vocalism in Senoic languages." In *Austroasiatic studies, Part I*, eds. Philip N. Jenner, Laurence C. Thompson and Stanley Starosta. Honolulu: University Press of Hawaii, pp. 229-247.

_____. 1976b. "Expressives in Semai." In *Austroasiatic studies, Part I*, eds. Philip N. Jenner, Laurence C. Thompson and Stanley Starosta. Honolulu: University Press of Hawaii, pp. 249-264.
_____. 1976c. "Jah Hut, an Austroasiatic language of Malaysia." In *Southeast Asian Linguistic Studies, II.* , ed: Liem, Nguyen Dang. Canberra: Pacific Linguistics C-42, pp. 73-118.
_____. 1977. "Towards a history of Mon-Khmer: proto-Semai vowels." Kyoto: Tonan Azia Kenkyu XIV.4, pp. 463-495.
Gafos, Adamantios I. 1998a. "A-templatic reduplication", Linguistic Inquiry, 29:515-27.
_____. 1998b. "Eliminating long-distance consonantal spreading", NLLT 16:223-78.
Fix, Alan G. 1977. *The Demography of the Semai Senoi*. Anthropological Papers. Museum of Anthropology, University of Michigan. No. 62.
Hendricks, Sean. 1999. "Reduplication without template constraints: a study in bare-consonant reduplication", Ph.D. dissertation, University of Arizona.
_____. 2001. "Bare-consonant reduplication without prosodic templates: expressive reduplication in Semai", Journal of East Asian Linguistics, 10: 287-306.
Hendon, Rufus S. 1966. *The Phonology and Morphology of Ulu Muar Malay*. Yale University Publications in Anthropology number 70. New Haven, Dept. of Anthropology, Yale University.
Inkelas, Sharon and Zoll, Cheryl. 2005. Reduplication: Doubling in Morphology. Cambridge: Cambridge University Press.
Kroeger, Paul R. 1986. "Discontinuous reduplication in vernacular Malay." Proceedings of the 15th annual meeting of the Berkeley Linguistics Society. 193-202.
Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
_____. 2010. A dictionary of Mah Meri as spoken at Bukit Bangkong. University of Hawai'i Press.
Levin, Juliette (1987). Between epenthesis and excrescence: what happens after redundancy rules. *WCCFL 6*: 187–201.
Lye Tuck-Po. 2004. *Changing pathways: Forest degradation and the Batek of Pahang, Malaysia.* Lexington Books. Malaysian edition ©2005 published by Strategic Information Research Development.
Marantz, Alec. 1982. "Re reduplication." Linguistic Inquiry 13:483-545.
Matisoff, James A. 1973. "Tonogenesis in Southeast Asia." In Larry M. Hyman, ed., Consonant Types and Tone, pp. 71-95. Southern California Occasional papers in Linguistics, No. 1. Los Angeles: UCLA.
_____. 2003. "Aslian: Mon-Khmer of the Malay peninsula." *Mon-Khmer Studies* (Bangkok) 33:1-57.
Matthews P. 1991. *Morphology*, 2nd edn. Cambridge University Press, Cambridge, UK.
McCarthy, John and Alan Prince. 1986. Prosodic morphology. University of Massachusetts and Brandeis University, ms.
Onn, Mohamed Farid. 1976. Aspects of Malay Phonology and Morphology. PhD dissertation, University of Illinois.
"Orang Asli Population Breakdown." Center for Orang Asli Concerns. Retrieved 14 Oct 2009. <http://www.coac.org.my/codenavia/portals/coacv2/code/main/main_art.php?parentID=11374494101180&artID=11600257781189 >
Peterson, Mary M. 1997. Kensiw Grammar: Basic Declarative, Interrogative and Imperative Propositions. Thammasat University.
Phillips, Timothy C. 2005. *Linguistic Comparison Of Semai Dialects*. Unpublished manuscript. Economic Planning Unit, Prime Minister's Department, Malaysia.
Phillips, Timothy C. 2006. "A Survey of Nasal Preplosion in Aslian Languages." Paper presented at the International Conference on Indigenous People, Kuala Lumpur, 5 July 2006.
_____. 2007. "Brief Phonology of Semai, Betau dialect." In *Reflections in Southeast Asian Seas: Essays in Honour of Professor James T. Collins (Book II)*, eds. Chong Shin, Karim Harun, Yabit Alas. Pontianak: STAIN Pontianak Press.
"Population of Orang Asli Sub-Groups, 1999." *Orang Asli Archives.* 21 Nov 2002. Keene State University. Retrieved 14 Oct 2009. <http://www.keene.edu/library/OrangAsli/pop2000.doc>.
Radhakrishnan, R. 1970. A preliminary descriptive analysis of Nancowry. Ph.D. dissertation. University of Chicago.
Shaw, Patricia A. 1993. "The prosodic constituency of minor syllables", WCCFL 11:117-32.
Shorto, Harry L. 2006. *A Mon-Khmer comparative dictionary*. Edited by Sidwell, Paul; Cooper, Doug and Bauer, Christian. Canberra: Australian National University. Pacific Linguistics.
Steriade, Donca. 1988. Reduplication and Syllable Transfer in Sanskrit and Elsewhere. Phonology 5:73-155.
Svantesson, Jan-Olof. 1983. Kammu phonology and morphology. Lund, Sweden, Liber, Forlug.

Wilbur, Ronnie. 1973. *The phonology of reduplication*. PhD dissertation, University of Illinois at Urbana-Champaign.

Yap Ngee Thai. 2009. Nonexhaustive syllabification in Temiar. *Journal of the Southeast Asian Linguistics Society* 1:267-281.

Zaharani b. Ahmad. 1991. *The Phonology and Morphology of the Perak Dialek.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against "Sinosphere" and "Indosphere"

Mark W. Post
The Cairns Institute, James Cook University

1. Introduction

In both the Austroasiatic (AA) and Tibeto-Burman (TB) language families, we find a rough overall cline in typological organization. In some languages, we find the following set of features: a complex morphological word, finiteness asymmetries, extensive suffixing, polysyllabic prosody, mostly simple onsets, and mostly monophthongal vocalism. In other languages, we find a relatively simpler morphological word, verb serialization, prefixation, syllabic prosody, occasional onset clustering (or sesquisyllabism), and complex diphthongs. Examples of the first type of language in TB would include Garo, Newar and Kiranti languages, among others, while in AA this typology is found mainly in Mundan. Examples of the second type of TB language include Mizo, Lahu and Lisu, among many others, while in AA this typology is found in Khmer and, especially, in Vietnamese. Seemingly, then, this cline in typological organization is further correlated with geographical location. Languages of the first type are found mainly in the Subcontinent, while languages of the second type are more common to mainland Southeast Asia.

Probably because of this geographical correlation, the within-family typological contrasts outlined above have sometimes been explained in terms of *contact influence*, namely from Indic languages in the West and Sinitic languages in the East. The labels *Indosphere* and *Sinosphere*, coined by Matisoff (1991b, if not sooner) and further popularized by Bradley, LaPolla et al. (2003), encapsulate this view, and are by now so widely-used that they hardly seem open to question.

While certainly diagnostically useful, and geographically sound, the problem as I see it is that in particular cases, clear evidence of pre-modern contact with either Indic or Sinitic languages is lacking. In the cases of many languages of the Eastern Himalayan region, it is all but certain that there were no such early contacts at all. And yet, as Matisoff (1991b: 485) correctly points out, Eastern Himalayan languages are "firmly...Indospheric" in terms of their broad typological characteristics.

In this paper, I will argue that a concept of "contact influence", depending as it would on the actual historical occurrence of population contacts (e.g. Indic > TB or Sinitic > TB), is insufficient to explain the broad typologies of TB languages of the Eastern Himalaya, at a minimum. Instead, adapting the model of prosodic shaping of language typology developed by Donegan and Stampe (1983; 2004) with primary reference to AA, I will suggest that the effects of *rhythmic prosodies* – particularly, the development of a trochaic rhythm in the "Indosphere" – provides a more plausible motivation for the development of most of the typological features which have been attributed to spheres of contact influence, in TB languages just as in AA. Although ultimate Indic and Sinitic sources for differing rhythmic profiles in TB and AA languages cannot be ruled out, they should neither be assumed as a proximate cause, nor (and far less) to reflect the historically dominant/subordinate population relationships that seem to be implied by the use of these labels.

Post, Mark W. 2011. "Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against "Sinosphere" and "Indosphere"." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.198-221.

2. Typological convergence in greater mainland Southeast Asia

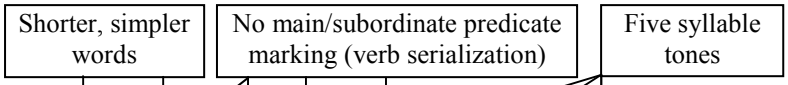
Many of the languages of greater mainland South-East Asia[127] – especially, those of AA and TB stock – can be characterized in terms of a typological cline which spans several linguistic dimensions. To the North, East, and Southeast, we often find…

- rich tone systems (often, syllable tone systems with large inventories)
- diphthongal vocalism
- simple onsets
- basic monosyllabism (syllable = morpheme = word) or sesquisyllabism ("one and a half" syllables per word, with a "minor" syllable initial which is often a prefix or earlier prefix)
- compounding rather than derivation (or lightly-grammaticalized derivations)
- relatively isolating morphology (rare or marginal affixation)
- productive or semi-productive prefixes
- no finiteness asymmetry (verb serialization)
- "SVO" word order (not always, but often)

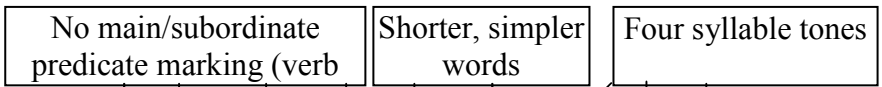
To the South, West, and Southwest, we often find…

- few or no tones (often, word tone systems with small basic inventories)
- monophthongal vocalism
- onset clusters
- basic polysyllabism (words may be two, three or more syllables)
- extensive derivational affixation, with or without compounding
- relatively synthetic, agglutinative morphology (extensive affixation)
- mainly suffixes
- robust finite asymmetries (clause chaining, subordinate clause embedding)
- uniform SOV word order

While not every feature value is of course found in every language in these regions, at all or in the same degree, as rough overall tendencies (or sets of potentially related tendencies) they hold well enough to at least attract attention. An example of the first type of language from the TB family would be Lahu (Lolo-Burmese, Burma/Thailand/Southwest China), a sentence from which is first compared with a similarly-structured Mandarin Chinese sentence in (1)–(2).



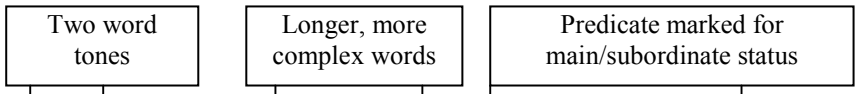
1) *lâ* *pɔ̀ʔ* *chèʔ* *câ* *pə* *šē* *ve* *cê.*
tiger jump bite eat finish ADVS NZR QUOT
'The tiger jumped (out) and bit (into them) and ate (them) all up!' **Lahu, Eastern TB** (Matisoff 1991a:411)



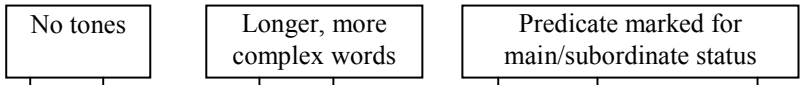
2) *tā* *jiù* *jì-xù* *pá* *dào* *shù* *shàng* *qù* *zhǎi* *lí.*
3 SFOC continue climb arrive tree top go pick pear
'So he continued climbing up the tree to pick pears.' **Mandarin Chinese, Sinitic**

[127] By "greater mainland Southeast Asia" I mean most of modern-day China and Tibet, Southeast Asia up to the Thai-Malay border, most of North-East India and the Northern/Eastern periphery of Bangladesh, and most of the Himalayan region.

An example of the second type of language would be Galo (TB > Tani, North East India), which is here compared with the neighbouring and typologically similar Eastern Indo-Aryan language Assamese (3)–(4).



3) *ʔihìi ʔoó má-làa rì-gərə́-làa ʔarúm=àm áa-rə́-kú*
wood vegetable search.for-NF do-ACNC-NF evening=ACC come-IRR-CMPL
'After searching for firewood and vegetables (they'll) return in the evening.' **Galo, Western TB**



4) *ta-i ula-i dza-i ula-i go-i pela-i bhab-e.*
3-F emerge-NF go-NF emerge-NF go-NF throw-NF think-3.SUB
'She goes ahead on out and thinks.' **Assamese, Eastern Indo-Aryan** (Data from Joana Jansen, glossing adjusted by this author)

Although this article's interest will mainly lie in the typological characterization of TB languages, the above generalizations in fact seem to be largely independent of the genetic origins of the languages involved. Vietnamese and Lahu resemble one another typologically to a greater extent than either resembles Khasi or Boro, despite that Vietnamese and Khasi are both AA languages and Lahu and Boro are both TB. Sora and Belhare resemble one another typologically to a greater extent than either does Khmuʔ or Karen – again, despite that Sora and Khmuʔ are both AA and Belhare and Karen are both TB. And Tai languages, of course, resemble Sinitic languages to such an extent that most scholars until quite recently assumed that they were probably genetically related (Grierson 2005 [1928]).

To the extent that these generalizations are true, and given the geographical correlations observed above, many linguists would suspect that we are dealing here with *areal typologies*, which have most likely come about as a result of prolonged population contacts resulting in population interactions, bilingualism, cultural exchange, and a consequent structural convergence (Thomason and Kaufman 1988; Thomason 2001). And indeed, this is probably the prevailing view, among Tibeto-Burmanists at least:

3. One popular view: "Sinosphere" and "Indosphere"

Primarily due to the pioneering work of Matisoff (1978; 1982; 1990; 1991a; 1999, etc.), the importance of language contact to an understanding of areal language structures in greater mainland Southeast Asia and their historical development has long been recognized. It is quite clear both that language contact exists in greater mainland Southeast Asia, and that it has played a role in numerous cases of what we might call cultural-linguistic convergence.[128] In addition, it is sometimes further argued that there are broader patterns of contact and areal convergence which in a sense shape the overall picture of cultural-linguistic convergence in the Tibeto-Burman region at least, and which ultimately explain the existence of typological prototypes such as those outlined in §0. As Matisoff writes,

> [I]t is convenient to refer to the **Chinese and Indian spheres of influence** as the 'Sinosphere' and the 'Indosphere'…. Some languages are firmly in one or the other…the Munda and Khasi branches of Austroasiatic and the Kamarupan branch of TB are Indospheric; while…the Loloish branch of TB and the Viet-Muong branch of Mon-Khmer are Sinospheric…. Whatever their genetic affiliations, the **languages of the ST area have**

[128] Prominent recent studies of the role of language contact in shaping various aspects of area language grammars and phonologies from different genetic stocks include Enfield (2001), LaPolla (2001) and Thurgood (2010).

> **undergone massive convergence in all areas of their structure** – phonological, grammatical, and semantic…. Hundreds of words have crossed over genetic boundaries in the course of **millenia of intense language contact**, so that it is often exceedingly difficult to distinguish ancient loans from genuine cognates.
> (Matisoff 1991b:485-486, emphasis mine)

As labels, *Sinosphere* and *Indosphere* accomplish two jobs simultaneously. First, they identify and label two broad subsets of greater mainland Southeast Asian languages, in which distinct sets of typological convergence phenomena are observed. Second, the labels explicitly attribute this convergence to massive, direct and sustained influence from two implicitly dominant cultural-linguistic forces, Sinitic and Indic, over the many smaller and geographically less-widely-represented populations of the area.

The concept of a Sinosphere/Indosphere divide has gained a fair amount of traction among Tibeto-Burmanists, at least; a recent and well-known collection of papers on variation and change in TB languages, for example, seems to assume it as a starting point for inquiry (Bradley, LaPolla et al. 2003), and references in the literature on TB languages are common. But how plausible is the logic behind their use? Another way of phrasing this question: given the observed, geographically-correlated typological cline which has been observed to hold (albeit roughly) among AA and TB languages alike, how plausible is it that contact with, and influence from, Indic and Sinitic languages will be able to exhaustively explain this situation?

4. Newar and Galo: Two case studies of Indospheric TB languages

To conclude that a term like "Indospheric" can be plausibly applied in a particular case, we would need to demonstrate at least two things: one, that a candidate language indeed exhibits typological features which are established as characteristic of Indospheric languages, and two, that a contact history with Indic languages can be independently demonstrated.[129, 130] In this section, we will look briefly at two case studies of candidate Indospheric TB languages: Newar and Galo.

Newar is spoken in the Western Himalaya, primarily in the Kathmandu Valley in modern-day Nepal. Galo is spoken in the Eastern Himalaya, primarily in mid-central Arunachal Pradesh state in modern-day North East India. Figure 1 shows the approximate geographical locations of Newar and Galo, and includes an extremely rough and haphazard geographical bifurcation of the greater mainland Southeast Asian region into an "Indosphere" and a "Sinosphere" (Figure 1).[131] The environments in which Newar and Galo speakers live are similar – in the Himalayan foothills – and both are, at least in modern times, at the geographical doorstep of the great Indo-Aryan *Sprachbund* of northern South Asia.

---

[129] By "plausibly" here, I do not mean "accurately". That is to say, demonstration of both (a) typological convergence and (b) contact in a given case does not necessarily indicate a causal link (a) > (b). However, the correlation, if repeatedly observed, would suggest a causal link to be at least *plausible*. On the other hand, if no contact can be shown to have occurred, or if contact can be shown *not* to have occurred, then a causal link between (a) and (b) would be viewed as less plausible or implausible, respectively.

[130] Since I am primarily interested in structural convergence in this paper, "independent" evidence would in this case include such things as loanwords (since borrowing of lexical items does not necessarily correlate with structural convergence toward the donor language). Of course, if *language*-independent evidence such as sharing of cultural features or written records attesting to historical population contacts can also be brought to bear, this would be even more welcome.

[131] This bifurcation is provided for heuristic purposes only, and should not be taken to represent a real border between the two "spheres" in question which has been arrived-at through systematic consideration of the typological profiles of all TB languages. Indeed, such a neat, precise borderline cannot possibly exist as such, and I doubt very much that even the most ardent advocate of an Indosphere/Sinosphere divide would claim that it does. Nevertheless, it may help to visualize the relative geographical locations of the languages under discussion vis-à-vis the rough geographical positions of these two "spheres".

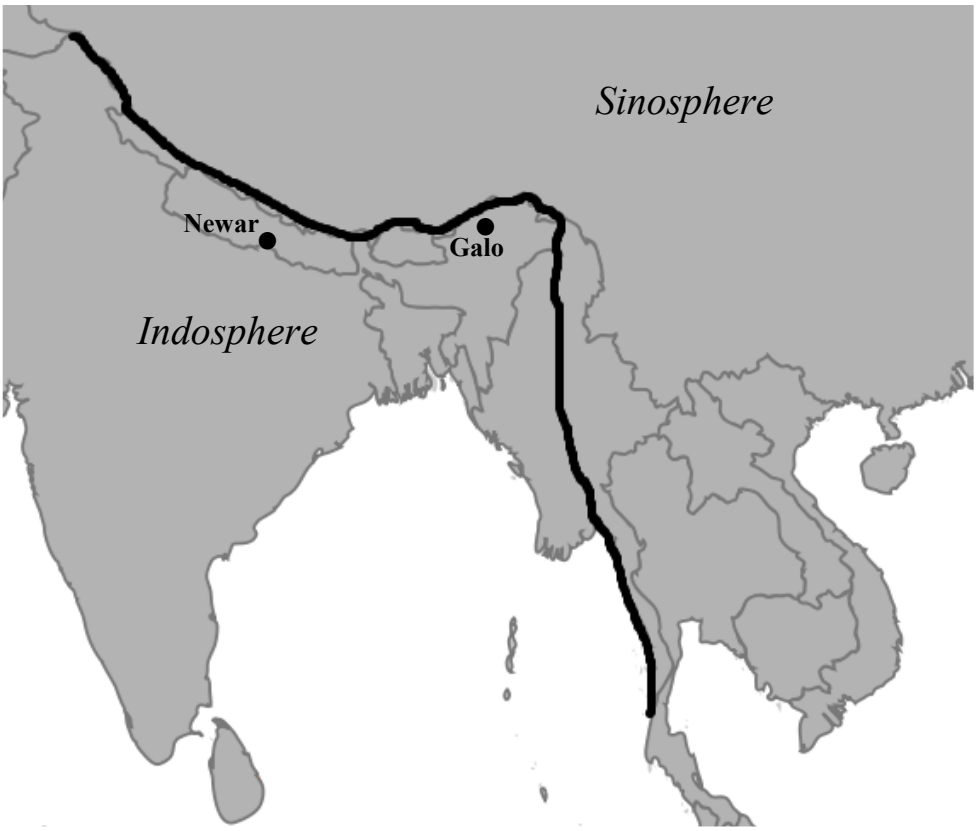


Figure 1: Newar and Galo: two "Indospheric" TB languages?

Dolokha Newar (Genetti 2007) and Lare Galo (Post 2007) have both been described as having characteristics which are typical of Indospheric languages. Aligning Newar explicitly with the Indospheric type, Genetti writes,

> In [Indospheric] languages there is often considerable inflectional morphology, from fully developed case-marking systems to extensive pronominal morphology found on the verb. These languages generally mark a number of types of inter-clausal relationships and have distinct constructions involving verbal auxiliaries [as opposed to serialization]. Newar dialects are clearly representative of the [Indospheric] type. (Genetti 2007:3, bracketed insertions mine)

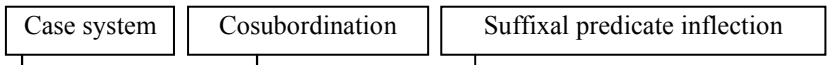
A similar characterization can be applied to Galo, briefly summarizing the larger discussion of Galo's typological profile in Post (2007:§1.2.10):

> Galo has extensive inflectional morphology, with a fully developed case-marking system and numerous obligatory TAM suffixes found on the predicate. Galo marks a number of types of inter-clausal relationship and has distinct constructions involving several types of non-finite predicate forms. Galo is thus clearly representative of the Indospheric type.

Examples (5)–(6) roughly exemplify some of these features for Dolokha Newar and Lare Galo respectively:

	Case system		Cosubordination		Suffixal predicate inflection	
5)	*āmun*	*jā*	*na-en*	*ye-eu*	*ka.*	
	3.SG.ERG	rice	eat-NF	come-3.FUT	ASS	

'He will eat his meal and come back.' **Dolokha Newar** (Genetti 2007:430)



6) *bɨɨ* *ʔacín* *dó-làa* *áa-rə́-kú* *dáa.*
3.SG.NOM rice eat-NF come-IRR-CMPL ASS
'He will eat his meal and come back.' **Lare Galo**

Thus, it would seem that Dolokha Newar and Lare Galo each satisfy the first criterion of Indospheric status, in that they each exhibit a similar suite of characteristically Indospheric typological features. But how about the second? Can a history of contact with Indic language speakers be independently demonstrated in each case?

In the case of Dolokha Newar, the answer is an unequivocal "yes". As Genetti writes,

> One fact that…Dolokha history makes clear is that the Dolokhae people have not been isolated from other linguistic groups…. Lexical borrowings from Indo-Aryan languages (especially Nepali, Sanskrit and Hindi, but possibly Maithili and other languages) [abound, and] structural borrowings are also in evidence…. Almost all Dolokha Newars are bilingual in Nepali and Dolokhae. Nepali is the lingua franca of the local area, as it is throughout Nepal. (Genetti 2007:22)

And furthermore,

> The extent of borrowing from Indo-Aryan may be attested by the large number of Indo-Aryan nouns found in my lexical database, composed primarily of words found in narrative, but also additionally containing elicited words. Of these, about **39% were Indo-Aryan loans** and about 59% were either native Dolakhae lexemes or **nativized Indo-Aryan borrowings**… Borrowed nouns are treated just like native nouns morphologically, and thus undergo the same suffixation and cliticization processes. (Genetti 2007:111)

According to Genetti, then, Dolokha Newar exhibits extensive evidence of contact with and influence from Indic languages. Bilingualism in the local Indo-Aryan lingua franca is the overwhelming norm, and this and other forms of exposure to Indo-Aryan languages has led to massive borrowing, with a large percentage of vocabulary items in Genetti's corpus being of Indo-Aryan origin. To these linguistic observations, we can add certain cultural observations. Newar traditionally practice Hinduism and Buddhism – both being of Indo-Aryan origin – and have adopted a caste system, an unmistakable signature of influence from Indo-Aryan groups. The material culture of Newar groups is also highly developed, with traditional woodworking and metalworking practices as well as sculpture and painting typically being oriented toward or inspired by Buddhist and/or Hindu mythology. Finally, Newar literature dates from the 12th century, when Sanskrit texts begin to be translated into Newar language (Genetti 2007:§3.2). In short, there can be little argument against viewing Newar as an Indospheric TB language in terms of exhibiting the requisite evidence of formative Indo-Aryan influence, over a great number of centuries if not indeed millennia. In turn, one might easily suppose that the relatively Indospheric typological profile of Newar is, in turn, part of this same nexus of contact-derived cultural and linguistic features. Now, how about Galo?

As was briefly mentioned above, Galo is spoken in mid-central Arunachal Pradesh state, North East India, in the mid-Eastern Himalaya, an area which in many ways (topography, climate, flora and fauna…) resembles mid-central Nepal (Figure 2). It was also mentioned that Galo, like Newar, is in close proximity to the Indo-Aryan *Sprachbund* at its southern border. But this is where I will argue that the comparison ends:

Figure 2: The Tani cultural-linguistic area in rough outline,
showing the approximate concentration of Galo speakers between the Subansiri and Siang rivers

Very little information of any reliability has historically been available regarding the geography, peoples and languages of the mid-Eastern Himalaya, up to and including the present day. Although "annexed" by British India in the late 19th/early 20th century, very little of the region was in fact governed, and virtually no regional integration was attempted (Mackenzie 1884). A definitive account of British expeditions into the region produced in 1912 included a highly revealing map insert, in which the extent of detailed British geographical knowledge of the mid-Eastern Himalaya is shown to extend roughly 40 kilometers up the Siang River valley (source of the Brahmaputra) – covering perhaps 4-5% of the "annexed" terrain (Hamilton 1912).[132] The de facto boundary with Tibet was in essence imagined; it had never been properly surveyed.

Since Indian independence, this information gap has in many ways remained unchanged. Restrictions on the publication of accurate maps (even locally) are in force, internal travel remains difficult and sometimes dangerous due to frequent landslides, and a highly restrictive permissions regime effectively reduces research in the region to a tiny trickle. Many of what one might suppose to be definitive fieldwork-based ethnographic surveys of the region – such as the *Anthropological Survey of India* – appear upon scrutiny to be at least in part derived from decades-old British survey records. Many of these were, in turn, based on secondhand accounts provided by individuals from distinct tribes with which British administrators (not anthropologists or linguists) happened to be in contact. Unsurprisingly, such works tend to be unreliable.

In the absence of facts, it is easy to make assumptions. It is easy, for example, to assume that Arunachal Pradesh, resembling Nepal in terms of overall setting, must resemble it culturally and linguistically as well; indeed, in *very* early times, it may well have. But in terms of one key factor, namely evidence of pre-modern Indo-Aryan (IA) contact with TB cultures – the comparison breaks down. There is very little evidence of extensive and enduring contact between IA and Eastern Himalayan populations until relatively late in the 20th century.

[132] These expeditions were mainly punitive, and designed to dissuade various hill tribes (especially the "Abor", i.e. Adi) from carrying out slave raids or extortion campaigns in British-administered Upper Assam.

Some relevant facts regarding the Tani cultures, the majority group of central Arunachal Pradesh, will be of help here. Tani cultures of Arunachal Pradesh,[133] of which the Galo are one, are in essence hill tribes, with a material culture closely resembling that of TB and Tai groups of modern-day Northern Burma. They traditionally build raised bamboo-and-wood houses, and primarily practice shifting or rotating hillside cultivation. Wet rice agriculture is practiced extensively by the plateau-dwelling Apatani, into which pisciculture has been more recently integrated (von Fürer-Haimendorf 1955). However, the terraced paddy agriculture so common to the Western Himalaya has been introduced only very recently in the Tani region and remains relatively little-practiced. Spiritually, Tani cultures are traditionally animist/shamanist (Blackburn 2010). Very, very few Tani peoples are practicing Buddhists. Hinduism is currently making strong inroads in the Tani area; however, it is doing so, in a sense, covertly, by means of a very recent Sanskritization of existing animist traditions (rather than by outright conversion). Christianity is rapidly on the rise in many areas, and there are apparently some converts to Islam among lowland Nyishi communities as well (Yankee Modi, personal communication). The point to underscore here, however, is that all of these shifts have happened within the most recent handful of decades; many residents of Arunachal Pradesh who are alive today are able to remember a time when Tani cultures were uniformly animist/shamanist, and when no "religion" existed among them that could be named. Finally, we should note here that there is little or no evidence of a caste system among Tani groups, such as one finds among the Newar of Nepal or in the modern-day Indian state of Manipur (Sen 1992).[134] Similarly, there is no evidence at all of the present or past existence of kings, an aristocracy, or even a village chief among Tani cultures, nor is there even a native vocabulary capable of reference to such institutions. In Eastern and Central Tani cultures, at least, the non-hereditary *kəba(ŋ)* 'village council' was the sole arbiter of disputes.

These statements apply to Tani cultures of Arunachal Pradesh only. To the East of the Tani area, one finds "Mishmi" groups whose material and spiritual culture appears to resemble that of the Tani to a large extent, but which have yet to undergo any comprehensive ethnographic or linguistic investigation. Closer to the Bhutan and Tibet borders, one finds Tibetic groups (Monpa, Menba) whose spread within the region is clearly relatively recent. Between the Tibetic groups and the Tani, one finds a diverse group of speculatively TB languages and cultures (Sulung/Puroik, Bugun, Bangru, Sherdukpen, Hruso Aka, Koro Aka, Miji…), whose traditional animism is in many cases robust, and who in some cases preserve a material culture which very likely predates that of most Tani groups in several respects. For example, sago palm harvesting is traditionally practiced in preference to dry rice cultivation among the Sulung/Puroik (Riba undated). Along the modern-day Burma border, one finds TB groups with close relations to Northern Burma hill tribes, and which may be considered part of the same overall cultural-linguistic sphere.

To summarize, then, in the Tani area, and in most other regions of Arunachal Pradesh excepting certain Tibet frontier areas, one finds a characteristically Southeast Asian rather than South Asian material culture (see also Burling 1965), little or no evidence of pre-modern organized religion and little or no evidence of craftwork, excepting certain non-representational canework effigies which are, however, not generally designed to outlast a particular ritual. There is no true caste system, and no system of permanent political hierarchies ("priests", "kings" or "chiefs" versus "common" people). To this one might add that there are no well-documented archeological sites in the region which can be clearly associated with earlier IA or Ahom administration of the hill region,[135] and – up to and including the present day – there is not one

[133] I here use the terms "Tani (languages, cultures…) of Arunachal Pradesh" or "…of the Eastern Himalaya" rather than simply "Tani" to avoid reference to Mising, being the only Tani tribe inhabiting the Assam plains area. It is seemingly the case that Mising speakers have migrated to the Assam plains from the East Siang region (where mutually-intelligible languages are spoken) in relatively recent times, where a very different set of contextual circumstances may be found. Consequently, the Mising have undergone a number of salient cultural and linguistic changes which render their situation relatively anomalous in the overall Tani context.

[134] Clan-based or, much more rarely, tribe-based dominant-subordinate relationships may be found here and there in the Tani and surrounding areas, but generally relate to the past practice of slave-taking rather than to an enduring clan hierarchy of any sort (von Fürer-Haimendorf 1955). That is to say, one could become a slave by virtue of being captured or bartered as a slave; however, this misfortune would not necessarily be inherited by a slave's children, at all or in the same way.

[135] Administrative or religious structures often highlighted as "indigenous to Arunachal Pradesh" (such as Malininthan and Ita-fort) are generally located in foothill or plains areas adjacent to the modern-day border with Assam.

permanent settlement (i.e., village or town) of IA language speakers in the entire Tani region within Arunachal Pradesh (Figure 3).[136]

Figure 3: (a) Milang (Tani) elder *Aamín Moodə̀* (b) *Moobùk* (Milang) village (c) Galo canework effigy

Unlike the case of Newar, then, there is very little evidence from culture, or from what little we know about the history of the region, of IA contact with Tani groups of Arunachal Pradesh, at least. But what about language?

Tani languages are of course spoken in different areas, with different degrees of proximity to the modern-day border with Assam, where the Eastern IA language Assamese is spoken both as a lingua franca and as the native language of tens of millions of people. Being an Assam-bordering language, Galo would be expected to be among those Tani languages exhibiting strong evidence of Indic language contact. However, in my Galo corpus of 5,049 base lemma (that is, not including derived words), the full set of nativized and/or unrecoverable loanwords can be exhaustively presented in a small table (Table 2).

The increasing prevalence of a semi-creolized form of Hindi, used as a lingua franca in modern-day Arunachal Pradesh, makes it easy for visitors to some areas to come away with the impression that Indic language contact is high, and may have always been. In urban areas in which people from numerous tribes congregate, it is sometimes difficult to find people under the age of 30 who speak anything other than "Arunachali Hindi" (Modi 2005). However, Galo texts collected from expert speakers above the age of 35 who are village inhabitants reveal a paucity of Indic loanword use, even in the ostensibly high Assamese contact region of Lower West Siang District, Daring Circle (approximately 20 km from the Assam border, along the West Siang trunk road). For example, a personal narrative text spoken by a 35 ~ 40-year-old woman of four minutes in length entitled 'Pig Spirit' contained only two loanwords, both of Assamese origin: *dukàn* 'shop' and *kulí* 'open (for business)'. A folktale spoken by a 75 ~ 80-year-old man of twenty-six minutes in length entitled 'The Wives of Abo Tani' contained only three Assamese loanwords *besi* 'more', *baki* 'remainder', and *dakon* 'basket'. Other texts exhibit similar ratios.

While these relatively inextensive structures do fall politically within modern-day Arunachal Pradesh, they are architecturally more plausibly associable to plains-based populations.

[136] This may seem surprising in view of the region's political inclusion in modern-day India, ensuing political and cultural integration, and the large population of Eastern IA-speaking groups at Arunachal Pradesh's southern doorstep who might well be expected to have an interest in pressing north. These facts conceded, permanent settlement of non-indigenous inhabitants in Arunachal Pradesh remains proscribed by a post-independence law designed by ex-British anthropologist-cum-missionary Verrier Elwin, who later became an Indian citizen, advocate of tribal affairs and adviser to then-Prime Minister Nehru (Elwin 2005 [1957]). Illegal non-indigenous settlers are routinely removed from their encampments and transported to the border.

Galo	Gloss	Source	Form	Gloss
ʔazár	'thousand'	Asm	*hezar*	'thousand'
pohàa	'money'	Asm	*poisa*	'money'
dukàn	'shop'	Asm	*dukan*	'shop'
bozár	'market'	Asm	*bozar*	'market'
kulí	'open (for business)'	Asm	*kuli*	'open (in general)'
bónd	'close(d) (for business); strike'	Asm	*bɔndɦ*	'close (in general); strike'
gám	'village headman'	Asm	*gaõ*	'village'
nahór	'Ceylon ironwood tree'	Asm	*naɦor*	'Ceylon ironwood tree'
untəráa	'orange (citrus fruit)'	Asm	*sumtra*	'orange'
rəbáp	'pomelo (grapefruit)'	Asm	*rɔbɔp*	'pomelo'
umbitáa	'papaya'	Asm	*umbita*	'papaya'
kurìi	'cat'	Asm	*mekuri*	'cat'
tamúr	'betelnut'	Asm	*tamul*	'betelnut'
tikə́	'contract'	Asm	*tika*	'contract'
potáa	'paper; letter'	Asm	*pɔtrɔ*	'paper'
potə̀	'license'	Asm	*pɔtrɔ*	'paper'
gurée	'horse'	Asm (via ET)	*gɦora*	'horse'
hàa	'tea'	Asm	*sa*	'tea'
balìi	'sand'	Asm	*balu*	'sand'
kɨrkíi	'window'	Asm	*kirki*	'window'
lagí	'want/need'	Asm	*lag-* + *-i*	'want/need; attach + NF'
porì	'study; read'	Asm	*porh-* + *-i*	'study; read + NF'

Table 2: Well-assimilated Indic loanwords in Lare Galo (Asm = Assamese)

To reiterate, Galo is in this case an example of a Tani language of Arunachal Pradesh in which Indic linguistic influence would be expected to be *high*, due to its relative proximity to the Assam border. In other Tani languages, evidence of Indic influence is even lower. For example, in *Dohɨŋ*, a Minyong village along the East Siang trunk road, most women above the age of 60 appeared to be unable to converse in any Indic language, and were aware of very few Indic words. In *Moobùk* (Peki-Modi), a Milang village along the Upper Siang/Dibang Valley district border, only a small number of Indic words were known by the majority of residents; children seemed unable to converse in any Indic language at all. Although certain Indic loanwords were in evidence in these villages, they were quite clearly recent loans resulting from the inevitable concomitants of modern-day administration (i.e., such words as *gam* 'village headman' (cf. Table 2), something which every village is required by law to have). These, in turn, were apparently borrowed from the neighbouring Padam language, together with characteristically Padam phonological adaptations.

To summarize, then, there is little if any evidence of extensive pre-modern Indic-Tani population contacts (again see also Post (in press-a)). There are few if any clear examples of extensive cultural exchange, and few well-assimilated loanwords. Furthermore, loanwords which do exist typically represent novel items or concepts which, more likely than not, would have arrived at some point during or after the onset of British Indian administration of the region in the late 19th-early 20th centuries (Mackenzie 1884). Even following that point, however, integration with larger India proceeded at a snail's pace, in part borne of a deliberate effort to "protect" the cultures of the Eastern Himalayan region from foreign influence (Elwin 2005 [1957]). This policy effectively remained in place until the "Chinese Aggression" of 1962, after which immigration of Hindi-speaking schoolteachers and establishment of Hindu mission schools increased dramatically (Bose 1997). Galo, and most if not all other Tani languages of the Eastern Himalaya, thus fails to fulfill the second criterion of Indospheric identity.

So if pre-modern contact with IA language speakers seems unlikely, what *can* explain the "firmly Indospheric" typological profile of Galo and many other Eastern Himalyan languages?

We might ask another question at this point. Up to now, we have been discussing relatively abstract typological features, and "cherry-picking" sentence examples which are explicitly designed to illustrate an idealized typological convergence. But how closely do "Indospheric" TB or AA and IA languages *really*

resemble one another, if their typological features are considered independently, and with no background assumption of convergence?

In the case of Tani languages, at least, the answer is "relatively little". That is to say, although it is possible to pick out certain IA-like sentences or structural features, it is just as possible to focus on features which are anything but IA-like. Despite lacking prototypically IA-like characteristics, many such features seem to be diachronically secondary, in the sense that they appear to post-date Proto-Tibeto-Burman. Accordingly, the presence of "Indic influence" during these formative periods would appear doubtful.

For example, one might consider the complex set of predicate formatives found in the Tani languages and, indeed in most of the Indospheric TB languages of North East India.[137] Consider, for example, the Minyong sentence in (7), which consists of a verb root followed by five derivational predicate formatives, an inflection, and a clitic sentential particle.

7) *amí* *ə́=kòm* *gók-tà-kí-rám-hí-káa-tó=î.*
person IDEN=ADD call-INCP-TENT-FRUS-REFL-RES.FOC-PFV=QTAG
'The guy also tried in vain to have a go at calling, eh.' (Upper Belt Minyong)

It would be difficult indeed to relate such complex, single-verb predicate structures to the simpler and often multi-verb structures of IA languages (see again example (4); also see virtually any chapter from Cardona and Jain (2003), especially Goswami and Tamuli (2003)).[138] Although both IA and many Indospheric TB languages are *relatively* synthetic and agglutinating by comparison with, say, Chinese, the precise characters of their morphological profiles are in fact quite different – sometimes to the extent that, if de-contextualized from any supposed "sphere of influence"-like framework, convergence would hardly be suspected.

To summarize: purported contact with IA languages is insufficient to explain the observed typological profiles of certain Indospheric TB languages, at a minimum, inasmuch as contact with IA language-speaking populations is in at least some cases highly unlikely to have occurred. Furthermore, although some features of the typology of Indospheric TB languages indeed do resemble the parallel features in IA languages, many others do not. In short, the "Indosphere/Sinosphere" model inadequately explains the observed typological divide among Subcontinental and Mainland Southeast Asian languages on two levels, at least as far as some of the relevant languages are concerned. Accordingly, it is worth wondering whether an alternative account is available which might better explain the situation.

5. An alternative account

Over a series of papers, Donegan and Stampe (1983; 2004; 2009) have argued that shifts in rhythmic alignment not only *sufficiently* explain cases of deep structural change in languages, they may even be *necessary* to their explanation – whether or not the quite separate effects of language contact (borrowing, calquing) might also be involved. The following sets of roughly opposite typological characteristics are observed by Donegan and Stampe to correlate with a prevailing falling or rising rhythm in a variety of (mostly AA) languages, and are hypothesized to be motivated by the prosodic outcomes of opposite rhythmic preferences (Table 3).

[137] For an overall description, supported by a discussion regarding the diachronic recency of the relevant construction, see Post (2010) for Tani languages and DeLancey (2010) for Bodo-Garo.

[138] Similar observations are made by Donegan and Stampe (1983), arguing against the reality of Mundan structural convergence to the IA type.

Falling rhythm	**Rising rhythm**
Trochaic word accent	Iambic word accent
Enclisis and suffixation	Proclisis and prefixation
CVC	Sesquisyllabism, onset clusters
Monophthongs, vowel harmony	Diphthongs, vowel reduction
Simple, register tones	Complex, contour tones
OV constituent order	VO constituent order
MOD-H modification	H-MOD modification
Case (usually suffixes)	No true case system
Synthetic, agglutinative morphology	Isolating, analytical morphology

Table 3: Typological outcomes of prevailing falling and rising rhythms

Figure 4 reproduces Donegan and Stampe's (1983: 11) exemplification of the effects of opposite rhythmic preferences on syllable canons, which provides a succinct illustration of the potential power of rhythmic preference over the diachronic shaping of linguistic form. Beginning with a reconstructed AA form consisting of a sesquisyllabic initial followed by a stressed, bimoraic final – hypothetically reflecting a rising rhythmic preference – we find a shift to falling rhythmic preference in Proto-Mundan giving rise to a trochaic meter. Loss of accentual prominence in the final syllable next motivates a final rhyme truncation, while a gain in accentual prominence in the initial syllable motivates progressive vowel harmonization in the initial syllable; while this may appear counter-intuitive, the reason is because harmonization offers a means of assigning the rhyme a relatively polar, thus more salient, vowel quality. "Progressive" Mundan languages then develop further in this direction: the final vowel is eventually lost, while its associated mora is transferred to a consonant coda; this further strengthens the erstwhile initial, stressed syllable.

The opposite effects are found in Mon-Khmer (MK): "progressive" MK languages exhibit collapse of the de-accentuated initial syllable into the initial of an onset cluster, and accentual prominence of the erstwhile final syllable rhyme motivates diphthongization.

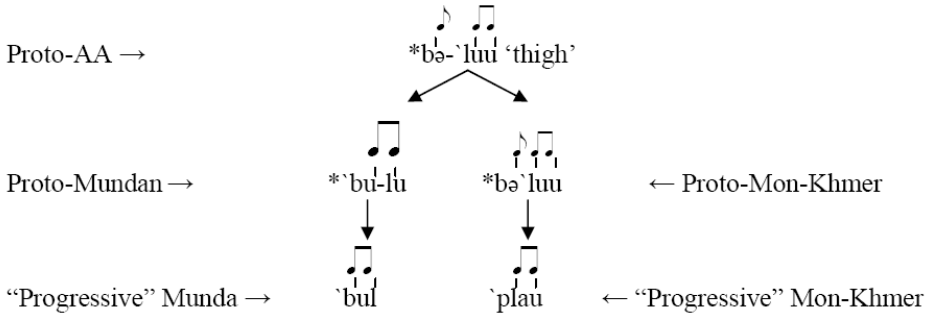


Figure 4: Rhythm and opposite prosodic drifts in Austroasiatic
(adapted from Donegan and Stampe (1983: 11))

Examples (8) and (9) contrast a Sora (Mundan) sentence with its translation equivalent in Khmer (MK), and illustrate some of the typological outcomes which are hypothesized by Donegan and Stampe to result from opposite rhythmic preferences. Sora shows a preference for simple onsets, monopthongal vocalism and CVC syllables, while cluster initials and diphthongs are found in Khmer. While Sora is relatively synthetic and agglutinating, Khmer morphology is relatively isolating and analytical. While Sora has extensive suffixation, Khmer has prefixes if it has affixes at all. In addition, the appearance of cluster initials in Khmer suggests the earlier occurrence of prefixes.

8) *anin* *dɔŋ-ɲɛn* *darəj-ən* *ə-tiy-ben* *idsim-tɛ* *ted.*
3SG OBJ-1SG rice-ART INF-give-INF want-3PR NEG
'S/he doesn't want to give me the rice.'
(Sora (Mundan), Donegan and Stampe 2004: 3) (Falling, trochaic)

9) *kŏət ʔət caŋ ʔaoy baay kŋom.*
3SG NEG want give rice 1SG(ACC?)
'S/he doesn't want to give me the rice.'
(Khmer (M-K), Donegan and Stampe 2004: 3) (Rising, iambic)

The above examples illustrate a small subset of the typological features which are observed by Donegan and Stampe to correlate with opposite rhythmic profiles. A fuller set of features, together with a more detailed explanation of the mechanisms which are hypothesized to account for their diachronic evolution, may be found in Donegan and Stampe (1983; 2004; 2009); I will thus refer the reader to these papers for a complete exposition of their theory.

Our goal in the present context, however, will be to determine whether Donegan and Stampe's model can provide a more plausible account of the shaping of "Indospheric" typologies than the "contact influence" subtext of the "Indospheric" label appears to be able to, in the case of Tani languages such as Galo, at least. In order to evaluate the case in terms of Galo, we'll need to understand at least four things. First, we'll need to understand something about the typological features of Proto-Tani, Galo's earliest reconstructible post-PTB ancestor. Next, we need to understand something about the rhythmic preference of Galo, both in the modern language and in its ancestral stages. Thirdly, we need to look for evidence of rhythmic effects in the re-shaping of Galo's broad typology since the Proto-Tani stage. Finally, as has already been discussed above, we would need to evaluate whether Galo's evolved typological profile brings it in line with the "Indospheric" typology characteristic of high-IA-contact TB languages.

6. Rhythm and the synthetic drift of Tani languages

6. 1. A brief, partial reconstruction of Proto-Tani

Post (2006; 2007: §2.2), building on the seminal insights of Sun (1993), has argued that Proto-Tani probably had the following characteristics:

1. lexemes primarily monosyllabic ~ sesquisyllabic (**mí* 'person', **tá-bɨ́* 'snake'…)
2. two tones at the level of the monosyllabic morpheme (**dú* 'forearm/elbow', **dù* 'dig with tool'…)
3. maximal CGVX[139] syllable structure (**bráŋ* 'roast', **pjóŋ* 'steal'…)
4. phonologically reduced and semantically general CV- prefixes (**ʔa-*, **ta-*, **ja-*, **sya-*...)
5. morphologically isolating otherwise (few grammaticalized predicate inflections...)
6. head-initial noun modification (**mík-máŋ* 'eye-lack' 'blind (person)')
7. compounding of simplex monosyllables (see again 6)
8. a mixture of pre-head and (mostly) post-head operators

Modern Tani languages generally have the following characteristics, albeit to different degrees and with slight regional differences:

1. lexemes primarily disyllabic (Minyong *ami* 'person', *tabɨ* 'snake'…)
2. tonal at the level of the phonological word (Galo *lagdú* 'forearm', *dunàm* 'to dig with a tool'…)
3. maximal CVX syllable structure in many languages (Galo *báa-* 'roast', *cóo-* 'steal'…)
4. prefixal collapse and root-harmonization (PT **ʔá-pòŋ* 'NPFX-liquor' > Galo *opòo* 'liquor')
5. morphologically synthetic/agglutinating, many predicate derivations and inflections (some reflecting fusion of earlier sequences)
6. head-final noun modification (Galo *ʔaɲík-talúu* 'eye-plank' 'spectacles')
7. compounding of etymologically complex, disyllabic words (see again 6)
8. all productive operators post-head

The following examples roughly schematize this set of developments from a reconstructed Proto-Tani stage (10) to modern-day Lare Galo (11).

[139] G = glide/liquid; X = C or V.

10) **ʔá-mróʔ* **hì=gu=m̀* **ŋó* **mám* **lìŋ* **máŋ=*dùŋ.*
NPFX-aconite PRX=IND=ACC 1.SG feel.with.hands want not=sit/exist
≅ 'I don't want to touch this-one arrow poison.' (Proto-Tani reconstruction)

11) *ʔomoʰ* *hɨgì-m̀* *ŋó* *mám-lìi-máa-dùu.*
arrowhead PRX.IND-ACC 1.SG feel.with.hands-DESD-NEG-IPFV
'I don't want to touch this here arrowhead.' (Modern Lare Galo)

To the extent that we can currently discern, then, Proto-Tani exhibited typological characteristics which align it closer to the "Sinospheric" type, with "Indospheric" characteristics emerging in later stages of the Tani languages' development. Although it is not currently possible to reconstruct most aspects of Proto-Tani prosody due to ongoing lack of adequate comparative data, in terms of Donegan and Stampe's modelled alignment of typological features with rhythmic preference, we would predict that Proto-Tani (or a near ancestor) would most likely have exhibited a rising rhythm. By contrast, most modern Tani languages, being typologically relatively "Indospheric", would be expected to exhibit a falling rhythm.

While we continue to lack adequate data for the majority of Tani languages, we can make a fair number of observations concerning prosody and its effects on various aspects of the structure of Galo:

6.2. Rhythmic effects in the shaping of modern Galo

Modern Galo exhibits a strongly falling rhythmic preference, leading to a robustly trochaic (strong-weak) meter at the level of the phonological word. This has motivated a series of strengthening and weakening effects in initial and final syllables, respectively. We review some of these in the following sections §6.2.1 and §6.2.2.

6.2.1. Strengthening effects

Neither monosyllabic words nor initial syllables of disyllabic words exhibit reduction or weakening phenomena in Galo. Instead, the full complement of reconstructed PT vowels is represented in modern Galo monosyllables and disyllable initials alike, in both short and long forms: **a/aa, *i/ii, *u/uu, *e/ee, *o/oo, *ə/əə, *ɨ/ɨɨ*. In addition, the following strengthening effects are observed:

6.2.1.1. Root-nuclear harmonization

Rhymes of certain weak (CV- or V-) word-initial syllables have been fortified in Galo by means of progressive harmonization with final syllable nuclei:

(a) ****ʔá-pòŋ*** 'NPFX-liquor' > Lare Galo ***ʔopòo*** 'liquor'
(b) ****tá-pá*** 'MDIM-maize' > Lare Galo ***təpə̀*** 'maize' (follows Word-final weakening, see §6.2.2.2)
(c) ****mə̀-kɨ̀*** 'fire-smoke' > Lare Galo ***mɨkə̀*** 'smoke' (precedes Word-final weakening)
(d) ****ʔá-jù*** 'NPFX-spirit' > Lare Galo ***ʔuì*** 'spirit' (precedes Word-final weakening and Inter-vocalic glide deletion, see Post (2007: §2.4.4.6))

This process is irregularly distributed in the modern Galo lexicon; for example, while an initial prefix **ʔá-* progressively harmonizes in (a) and (d), and in many other cases, it is also possible to find segmentally comparable words in which harmony is not observed. For example, *ʔalóo* 'bone' (< PT **ʔá-lóŋ* 'NPFX-bone', compare with (a)) and *ʔaú* 'fat/grease' (< PT **ʔá-fù* 'NPFX-fat/grease', compare with (d)). It is possible that these represent the different outcomes of prefixal lexicalization at different historical stages, with consequently different inputs into a regularly-applied phonological rule (as was suggested in Post (2007: §2.4.3.1)). Looking at other Tani languages, however, in which distributions of progressive root-nuclear harmonization are often similarly irregular (and often affect forms which are unaffected in Galo, and vice versa), it seems perhaps more likely that this is simply an irregularly-diffusing process in each case.

In addition to being a light (CV- or V-) initial, in the majority of cases, the affected syllable is also an etymological prefix. However, it is not the case that all and only prefixes have progressively harmonized; (c) is an example of a lexical root which has irregularly harmonized, and *tá-* 'MDIM' is an example of a Galo

prefix which harmonizes but rarely, as in (b). Irregularity in distribution notwithstanding, the most likely motivation for this change is prosodic: metrically prominent syllables whose rhymes are not high in salience have increased their salience by copying the vowel of a following syllable.

6.2.1.2. Initial coda fortification

Again with seemingly irregular application in the lexicon, a number of etymologically light word-initial syllables are fortified by means of copying a following syllable-initial onset consonant for use as an initial coda. Perhaps unsurprisingly, this process is very often found among words with an intrinsically emphatic semantic value:

(a) Lare Galo *ʔa**tt**ór* 'firm; hard; strong' < PT **ʔá-**t**ór* 'NPFX-strong (of a current state)'
(b) Lare Galo *ʔa**dd**ɨ̀ɨ̀* 'strong; resilient; durable' < PT **ʔá-**d**ɨ̀ɨ̀* 'NPFX-strong (of an enduring quality)'

Or, it may be found in certain pragmatically emphatic variant forms of lexemes:

(c) Lare Galo *ʔa**ll**ɨ̂ɨ̂* 'very good' < *ʔa**l**ə́* 'good'
(d) Lare Galo *ʔa**ll**ôo* 'way over there' < *ʔa**l**ò* 'over there'

Otherwise, it may be found with seemingly irregular distribution among certain lexemes, including some – such as numerals – which do not obviously seem to carry an emphatic semantic value or lend themselves easily to emphatic pragmatic use.

(e) Lare Galo *ho**tt**úm* 'bear' < PT **s(y)á-**t**úm* 'PFX:ANIM-bear'[140]
(f) Lare Galo *ʔa**ŋŋ**ó* 'five' < PT **ʔá-**ŋ**ó* 'NPFX-five'

The above examples represent cases in which initial coda fortification is irregularly lexicalized in a particular set of forms. There is also a synchronically-operating morphophonological process in Galo with approximately the same outcome and presumed functional motivation, called *Triggered foot-strengthening* in Post (2007: §4.1.4.6). In this process, concatenation of a vowel-initial enclitic to a bimoraic phonological word triggers gemination of the word-final syllable onset. In practice, this most often applies to words with underlying V.CV or CV.CV structures, although CVV words are also affected. The resulting output of (C)VC.CV effectively fortifies the initial syllable by generating a consonantal coda, much as in the irregular process described earlier in this section.

(g) Lare Galo *ta**b**á* 'snake' + *=ə́* 'TOP' > *ta**bb**ə́* 'the snake…'
(h) Lare Galo *ʔa**y**ò* 'night' + *=ə̀m* 'ACC' > *ʔa**yy**òm* 'at night'

6.2.2. Weakening effects

6.2.2.1. Syncope and apocope

Syncope and apocope (reduction or deletion of a vowel from within and at the end of a word, respectively) are both common processes in Galo phonology, with a common prosodic motivation. Since Galo is a weight/quantity-sensitive language (i.e., it treats "heavy" and "light" syllables differently), the precise outcomes of these two processes are different depending on the structure of a targeted syllable, as well its position in the phonological word and phrase. Importantly, however, only the *final* syllables of a disyllabic phonological word are affected; initial syllables of disyllabic phonological words, whether heavy or light, receive a stress accent and do not undergo any restructuring processes. Some of the segmental outcomes of this process are as follows:

(a) word-final vowel length is neutralized in phrase-final contexts

[140] Contrast Upper Belt Minyong *hitum* 'bear' and Pasighat Adi *situm* 'bear', in which gemination is not observed.

Morphemes with an underlying CVV structure which occur in phonologically phrase-final contexts are realized with a rhyme which is not phonetically longer than an underlying CV syllable in the same position. The following examples illustrate this principle. First note that the underlying rhymes of the second syllables in *ʔikìi* 'dog' and *ʔabó* 'father' are long (VV) and short (V) respectively; however, when pronounced in a phonologically phrase-final context as in (12) and (13), the final vowels are phonetically identical in length. That *ʔikìi* 'dog' and *ʔabó* 'father' are indeed underlyingly contrastive in length can be shown by placing them in a phonologically phrase-medial context (12)–(15).[141]

12) *ŋôk ʔik**ì***
ŋó-kə ʔik**ìi**
1SG-GEN dog
'my dog'

13) *ŋôk ʔab**ó***
ŋó-kə ʔab**ó**
1SG-GEN father
'my father'

14) *tan**íi** gó*
tan**íi** gó
person IND
'a person'

15) *ʔáb^{ə} gó*
ʔab**ó** gó
father IND
'a father'

(b) word-final short vowels are reduced in phrase-medial contexts

As will have been seen in (15), underlyingly short vowels are reduced in phonologically phrase-medial contexts, often to the point of surfacing only as a release of the preceding consonant. In this paper, such vowels are conventionally transcribed via a superscripted variant of the underlying vowel.[142]

(c) nuclei of word-final -CVC syllables are reduced in phrase-medial contexts

Again in phonologically phrase-medial contexts, the nucleus of a word-final syllable of CVC structure is reduced to schwa or consonantal release (16).

16) *kâak^{ə}nə́!*
káa-k**ə̀**n=ə́
look-good/easy=cop
'How beautiful!'

6.2.2.2. Word-final weakening

Similar to condition (b) in §6.2.2.1, although applying diachronically at the level of lexical representations rather than synchronically at the level of the phonological phrase, certain PT word-final short vowels are reduced in innovative dialects of Galo such as Lare (though not in some other Galo dialects such as Northwestern). Most straightforward among these are PT **-a*, **-i*, and **-ə*, which all merge to *-ə* in word-final environments. The proto-vocalism can be ascertained via by examining the same etymological morpheme in a word-initial environment – something that, due to the prevalence of compounding in the Tani lexicon, is often easy to do. For example:

[141] In some Galo dialects, the rhymes in (12) and (13) are distinguished by a voiceless, aspirated release of the underlying short vowel *ʔabó^{h}*, contrasted with a clear-voice release of the underlying long vowel *ʔikì*. In other dialects, no phonetic reflex of the underlying length distinction has been found in phrase-final contexts. It is worth further noting that, since utterance of individual words qualifies as a phonologically phrase-final context, this means that underlying vowel length in final open syllables cannot be discovered by a traditional "wordlist"-style lexical elicitation. Accordingly, contrastive vowel length in these contexts is not represented in most of the published sources on Tani languages.

[142] In Post (2007), such vowels were generally transcribed via schwa, reflecting their typical phonetic value if at all realized. It was later found that this transcription was inadequate: since schwa is phonemic in Galo, transcribing syncopated vowels via schwa made it impossible to distinguish between a full and a reduced schwa, when schwa was the underlying vowel in question. Hence, the present somewhat non-standard convention was adopted.

(a) PT **s(y)á-rá* 'PFX:ANIM-boar' > Lare Galo *horə́* 'boar' but *ranə̀* 'wild sow'
(b) PT **tá-bí* 'MDIM-snake' > Lare Galo *tabə́* 'snake' but *birə́m* 'python'
(c) PT **ʔá-ɦə̀* 'bamboo' > Lare Galo *ə̀ə* 'bamboo', *əlùu* 'bamboo grove'

Among the remaining PT final short vowels **-i*, **-e*, **-o*, and **-u*, **-e* and **-u* usually merge to *-i* in palatal environments, while **-u* merges to **-o* in non-palatal environments word-finally[143]:

(d) PT **ʔá-mé* 'NPFX-elder sister' > Lare Galo *ʔaɲí* 'elder sister'
(e) PT **ʔá-ɲì* 'NPFX-two' > Lare Galo *ʔaɲì* 'two'
(f) PT **ʔá-jù* 'NPFX-spirit' > Lare Galo *ʔuì* 'spirit'

(g) PT **ʔá-kù* 'NPFX-old' > Lare Galo *ʔakò* 'old'
(h) PT **ʔá-ɦò* 'NPFX-child' > Lare Galo *ʔaò* 'son'

While the shifts among these particular qualities are not obviously motivated by their relative lack of prominence (unlike the shift of **-a* and **-ɨ* to schwa), the aggregate effect of these changes is to severely reduce the inventory of short vowels which are available in word-final positions, to *-ə*, *-o* and *-i*. By contrast, the full complement of seven PT vowels remains available in the rhymes of word-initial syllables (cf. §6.2.1).

6.3. Discussion

The preceding subsections have reviewed a number of cases in which initial and final syllables of disyllabic words have been restructured, with initial syllables tending to strengthen and final syllables tending to weaken. Without committing to any particular theory of metrical phonology, we can nonetheless see fairly clearly that such phenomena are indicative of a trochaic stress pattern, in which initial syllables are fortified due to stronger metrical prominence, and final syllables are reduced due to lower metrical prominence.[144] Accordingly, it would appear that we have ample evidence that modern Galo exhibits signature features of the falling rhythmic preference which Donegan and Stampe associate with the "Indospheric" set of typological features which were previously identified in Galo.

Furthermore, to the extent that we can discern, falling rhythm appears most likely to be a secondary feature of Galo, i.e., to have arisen in of one of its post-PT ancestral stages. This is because we have seen that the effects of falling rhythm, such as progressive nuclear harmonization, applied to structures, such as prefixes, which are correlated with the preferentially rising rhythmic alignment which we earlier argued to be the most likely profile associated to the PT stage. By contrast, no signature features of rising rhythm, such as prefixal weakening or the rise of new prefixes, have been detected at any reconstructible stage in Galo's post-PT history.

In sum, we have been able to demonstrate that Galo's prosodic profile is indeed in line with the expectations of Donegan and Stampe's model, inasmuch as the presence of a falling rhythmic preference correlates with the expected set of typological features. Although somewhat less concretely, we have also seen that the bulk of available evidence suggests that both falling rhythm and the correlated set of typological features have arisen relatively recently, restructuring a typology whose most salient features are in line with Donegan and Stampe's rising rhythmic type. This view can be very roughly schematized as in Figure 5.

---

[143] Due to a separate and possibly earlier change, short **-e* and **-i* lengthened in non-palatal environments (in all qualifying morphemes, regardless of position in the world), removing this available set of short final vowels.

[144] Elsewhere, I have argued that the diachronic rise of a trochaic metrical foot in Galo and in other Tani languages has motivated coalescence of a prototypically disyllabic phonological word, whose prosodic salience has in turn led to a number of restructuring effects in the lexicon and grammar alike. Some of these restructuring effects include the overall shift in the Tani lexicon to a prototypically disyllabic, etymologically bimorphemic lexical word (Post 2006), the fusion of erstwhile sequences of monosyllabic functors, often suffixes or postpositions, into unanalyzable disyllabic units (Post 2009: §7.1) and the nascent development of verbal auxiliaries from erstwhile disyllabic sequences of grammatical suffixes (Post 2009: §7.4).

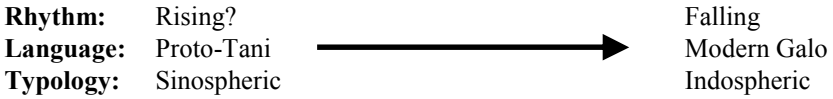


Figure 5: Correlating rhythmic and typological drift in Galo

To offer a preliminary summary, the evidence reviewed above suggests that Donegan and Stampe's model may indeed offer a more consistent explanation for the typological divide discussed in §2 than the "Indosphere/Sinosphere" contact-with-influence model was able to, at least where Galo is concerned. This was the primary goal of the article, and I will have more to say about this in the summary conclusion in §8 below. First, however, I would like to adopt a broader view, in an attempt to locate the provenance of the rhythmic effects we have observed in Galo. Before doing so, I will straightforwardly warn the reader that much of what follows is, of necessity, highly speculative, discussing as it does several languages whose prosodic profiles and historical circumstances are but little known. Readers who lack the taste for speculation at this level can, therefore, skip over the following section should they choose to, at little cost to the article's main aims.

7. A broader view: The status and provenance of rhythmic effects beyond Galo

Looking beyond modern Galo, we might wonder: if it is the case that Galo has undergone a rhythmic and, consequentially, typological shift, how might it have come about? Or, put somewhat more concretely, what are the contextual circumstances of Galo, in terms of genetic origins, geographical location and language contact conditions, and how well can they be correlated with the diachronic account sketched-out in Figure 5?

As discussed above, modern-day Tani languages are primarily spoken in the Eastern Himalaya, in the central region of the modern-day Indian state of Arunachal Pradesh. The ultimate origin of Tani peoples and languages is at present unknown, but is maintained by Tani oral traditions to lay somewhere outside of the present area of concentration (Post in press-a). As was briefly discussed in §4, Tani culture in general more closely resembles the hill tribal type common to the Mainland Southeast Asian region (Burling 1965); there is accordingly ample reason to suspect an origin in this area. However, linguistic evidence in general points to a "northern origin" theory, the view most commonly found in Tani oral traditions as well (Blackburn 2003/2004; Geiyi undated). Specifically, since the two primary Tani subgroups – Western Tani and Eastern Tani (Figure 6) – are concentrated in the north-south flowing Subansiri and Siang rivers respectively, and since the highest intra-Tani genetic diversity is, in general, found in the northern reaches of these two rivers, it would seem most likely that Proto-Tani is traceable to somewhere in this northern region (Figure 7); see Post (in press-a) for additional discussion.

The Brahmaputra Valley is an area about whose history a little more can be said:

Prior to 1000 BC, it is difficult to conjecture about the cultural-linguistic composition of the area, although there is at least a possibility of Austroasiatic predominance (Kakati 1995; Diffloth 2005). From 1000 BC to 400 AD we find the South-westward spread of Bodo-Garo, most likely from an initial position in the Northern Burmese/North-East Indian hill regions, where "Sal" languages such as Tangsa are spoken in great variety to this day (DeLancey 2012). From 400 AD to the present, we find the North-eastward spread of the Eastern Indo-Aryan languages Bengali and Assamese (Baruah 1960 [1933]). From 1200 AD, we find the arrival and subsequent decline of Ahom (Tai) from the Northern Burmese Shan states, plus small communities of later Tai arrivals from the same area such as Khamti, Aiton, Phake and Khamyang (Morey 2005). Sometime in or before the 19th century, we find the South-westward spread of Mising (Eastern Tani), apparently following the course of the Siang River, which in Assam becomes the Brahmaputra (Post in press-a) (Figure 8).

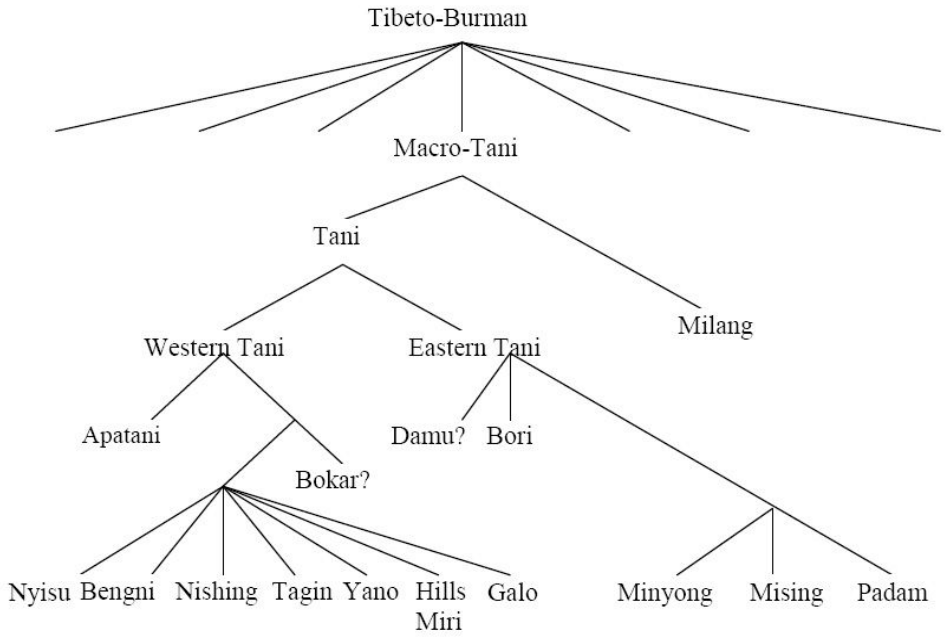


Figure 6: Tani family tree, following Post and Modi's (in press) minor revision of Sun (1993)

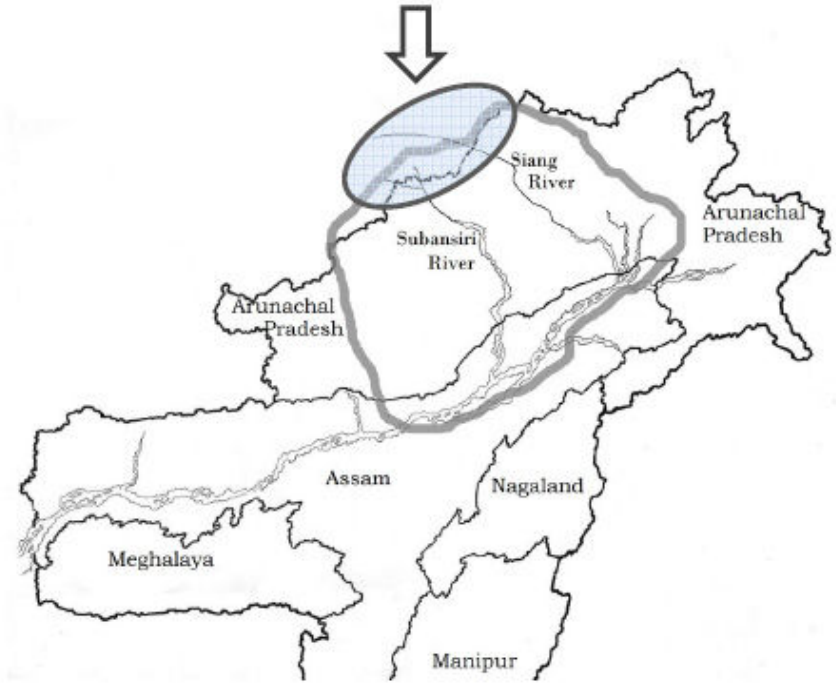


Figure 7: A possible homeland of the Tani cultures and/or languages

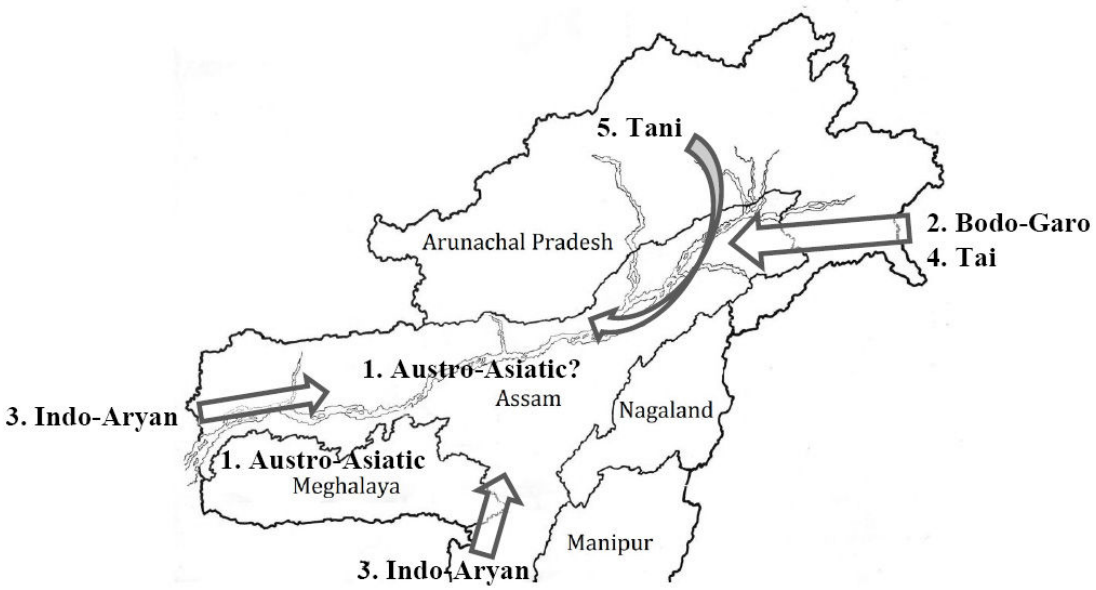


Figure 8: The peopling of the Brahmaputra Valley

Little has been written about the prosody of any of these languages, with the exception of the iambic Tai languages whose spread and influence in the region has, however, been comparatively slight (Morey 2005). Most if not all modern-day Bodo-Garo languages appear to be trochaic, although certain aspects of their grammatical typology, such as a prevalence of fossilized prefixes (Burling 2004; Joseph 2008), suggest an earlier iambic profile. Eastern IA languages, meanwhile, are uniformly trochaic (author's field notes), as indeed are most IA languages. Accordingly, it is at least possible that the entire region was characterized by a rising rhythmic profile – possibly associated with a more widespread early distribution of Austroasiatic speakers and substrates – but that in many languages, shifts to a falling rhythm have come about. Potentially, these shifts could be associable to the spread of Eastern IA languages in the region.

This is one possibility. But looking again at the Tani languages: although all known Tani languages appear to exhibit a falling rhythm and agglutinating morphological profile, there is an uneven distribution of trochaic effects. In the Eastern Tani area, clustering around the Siang River valley, we find evidence of trochaic effects to be relatively weak. Although certain initial syllable strengthening effects are observed, few or no final weakening or reduction effects have been found – in fact, there is evidence of relatively recent iambic effects in some of these languages, such as weakening of the initial **ʔá-* prefix to *ə-*. Relatively more trochaic effects are found in Galo, a "Transitional" Tani language,[145] as was discussed in §6.2. The greatest evidence for trochaic effects is found in Upper Belt Nyishi, a Western Tani language spoken in a geographically north-western section of the Tani area. This differential distribution of trochaic effects is schematized in Table 4.

Why should trochaic effects be stronger in the west? Well, the only honest answer is that we don't know. But, if one is to speculate: it is in and to the west of the Tani area that we find some of the highest genetic diversity in the Tibeto-Burman region. Almost completely undescribed, the languages spoken in this region are speculatively considered to be Tibeto-Burman but contain, at the same time, both cultural and linguistic characteristics which are distinctively non-Tibeto-Burman; these languages and the populations which speak them could well harbour remnants of a much more diverse pre-Tibeto-Burman past (Blench and Post in press). Could these languages, which almost certainly would have had a wider distribution in pre-Proto-Tani times, be the proximal source of falling rhythm, in Tani – and potentially in other area languages as well? Unfortunately, since we continue to lack even a single comprehensive, reliable description of any of these languages, there is nothing to do but wonder.

[145] "Transitional" here is shorthand for "genetically Western, with contact-induced convergence to Eastern" (Post in press-b).

Word	PT	Western Tani	Transitional	Eastern Tani
		Nyishi	Galo	Minyong
'dog'	**ʔá-k(w)ì*	*ʔikj*	*(ʔ)ikìi*	*əkì*
'four'	**ʔá-prí*	*ʔapj*	*(ʔ)appíi*	*appí*
'men's sitting area'	**bráŋ-gò*	*baagw*	*baagò*	*baŋgo*
'seven'	**kV-nɨ̀t*	*kan*	*kanə̀*	*kənɨt*
'eight'	**prí-ɲì*	*piin*	*piinə̀*	*piiɲɨ*
'snake'	**tá-bí*	*tab*	*tabə́*	*tabɨ*
'leg hair'	**lə̀-mít*	*lim*	*ləmə̀*	*ləmɨ̀t*
'come'	**váŋ-*	*ɦaa-*	*(ɦ)áa-*	*áŋ-*

Table 4 – Differential distribution of rhythmic effects in Tani

8. Conclusion

This article's primary goals have been, first, to suggest that while the basic typological observations underlying the "Indosphere/Sinosphere" model are accurate in a broad sense, and do roughly coincide with the geographical locations of Indic and Sinitic cultures, the "contact-influence" subtext of this model cannot be sustained in at least some cases. Specifically, there are at least some "Indospheric" languages which, though they exhibit the requisite set of typological characteristics, yet do not exhibit evidence of pre-modern contact with Indic languages. Furthermore, while these "Indospheric" languages may resemble Indic languages in a broad sense, such as being relatively agglutinating, or exhibiting finiteness asymmetries, the finer details of their typological features are not so Indic-like as to sustain a view that they have probably come about via borrowing or calquing. Second, adopting a model developed by Donegan and Stampe (1983; 2004), in which rhythmic alignment is viewed as the primary linguistic engine of typological drift, I suggested that the synchronic and diachronic facts of Galo and other Tani languages instead support a view in which a relatively "Sinospheric" language, presumably correlated to a rising rhythm, developed into modern-day languages with relatively "Indospheric" typological profiles, in concert with a hypothetical shift to today's observed falling rhythmic alignment. Accordingly, my argument is that Donegan and Stampe's model can provide a more consistent explanation for the observed typological divide in Greater Mainland South East Asian languages than does the "contact-influence" subtext of the "Indosphere/Sinosphere" model.

I then proceeded with an admittedly speculative inquiry regarding the possible provenance of falling rhythm in the Tani area. While little can be said with confidence regarding the pre-history of the Eastern Himalayan region, the bulk of evidence adduced to date suggests that 1) the Tani languages spread in the area relatively recently, from a more-or-less northern position 2) a group of genetically diverse languages probably pre-dated Tani languages in this area, and could be retained in the Tani area as one or more substrates 3) the area to the south of the Tani was perhaps initially dominated by Austroasiatic, followed by Bodo-Garo, and lastly by Indo-Aryan speakers, up to and including the present day. Accordingly, the diachronic rise of a falling rhythm in the region, with attendant changes in the grammatical typologies of regional languages, could potentially be attributable to pre-existing substrate languages, or else could be associated with the spread of Indo-Aryan languages in the modern-day Indian Northeast.

So does this lead us back to where we started? If the spread of falling rhythm in the region coincides with the spread of Indo-Aryan languages, doesn't this mean that we are dealing, after all, with an "Indosphere"? Well, no, or at least, not in the sense originally implied. As virtually any detailed study of language contact phenomena will amply demonstrate, the structural features of a language do not spread like a disease. Diffusion of structural features requires more than simply contact: it requires learning and understanding: bilingualism and interaction (Thomason and Kaufman 1988; Aikhenvald 2007, among many others). By contrast, prosodic diffusion requires little more than contact; contact, that is, followed by the almost inevitable convergence upon the behavioural characteristics of others which results from the fundamental learning and adaptive processes which characterize human beings and all other mammals: it requires only imitation; not understanding (Epps 2007). Through imitation of the observable behaviour of others, prosodic features can, from a particular area of concentration, spread over vast geographical distances, bringing languages into close alignment with respect to some aspects of their linguistic profiles, despite their speakers never in fact having come into contact with one another.

The conclusion to be drawn, then, is: whether or not the ultimate source of falling rhythm in the "Indospheric" region of Greater Mainland South East Asia can be traced to Indo-Aryan languages, it is rhythm – not language contact, bilingualism and population interchange – that most likely provides us with the proximal cause of "Indospheric" language typology, at a minimum. Accordingly, it is worth dispensing with the labels "Indosphere" and "Sinosphere", not only because of the possibly incorrect characterization of the proximal cause of typological alignment that they provide, but because of the pre-historical dominant/subordinate population relationships that they imply, for which – in several cases at least – no evidence whatsoever is available.

The alternative "rhythmic alignment" account of Donegan and Stampe, I have argued, is consistent with the Tani data, but remains to be tested on a wider set of languages, TB, AA and other. It is hoped that a qualitative and quantitative improvement upon the current level of prosodic description in grammars and associated descriptive works treating regional languages will one day make this possible. But in the meantime, studies which indicate the presence or absence of correlations between the rhythmic profiles and typological characteristics of individual languages or subgroups, and which further address their diachronic development, would make welcome contributions.

Abbreviations:

AA	Austroasiatic	MDIM	Masculine diminutive
ACC	Accusitive	MK	Mon-Khmer
ACNC	Additive concessive	MOD	Modifier
ANIM	Animate	NEG	Negative
ADD	Additive	NF	Non-final
ART	Article	NOM	Nominative
Asm	Assamese	NPFX	Noun prefix
ASS	Assertive	NZR	Nominalizer
C	Consonant	O	Transitive object
CMPL	Completive	OBJ	Object
DESD	Desiderative	PFX	Prefix
ERG	Ergative	PRX	Proximate
ET	Eastern Tani	PT	Proto-Tani
F	Feminine	QUOT	Quotative
FRUS	Frustrative	RES.FOC	Result focus
FUT	Future	SFOC	Sequential focus
G	Glide (liquid)	SG	Singular
GEN	Genitive	SUB	Subject
H	Head	TAM	Tense, aspect and modality
IA	Indo-Aryan	TB	Tibeto-Burman
IDEN	Identifiable	TENT	Tentative
INCP	Incipient	TOP	Topic
IND	Individuator	V	Vowel
INF	Infinitive	V	Predicate
IPFV	Imperfective	X	Segment (consonant or vowel)
IRR	Irrealis		

References:

Aikhenvald, A. Y. (2007). 'Grammars in contact: A cross-linguistic perspective.' In A. Y. Aikhenvald and R. M. W. Dixon, Eds. *Grammars in contact: A cross-linguistic typology*. Oxford, Oxford University Press: 1-66.

Baruah, K. L. (1960 [1933]). *Early history of Kāmarupa, from the earliest times to the end of the Sixteenth Century*. Guwahati, Assam, Lawyer's Book Stall.

Blackburn, S. (2003/2004). 'Memories of migration: Notes on legends and beads in Arunachal Pradesh, India.' *European Bulletin of Himalayan Research* 25/26: 15-60.

--- (2010). *The Sun Rises: A Shaman's Chant, Ritual Exchange and Fertility in the Apatani Valley*. Leiden/Boston, Brill.

Blench, R. and M. W. Post (in press). 'Re-thinking Sino-Tibetan phylogeny from the perspective of North East Indian languages.' In N. Hill, Ed. *TBA [Proceedings of the 16th Annual Himalayan Languages Symposium]*. London, School of Oriental and African Studies.

Bose, M. L. (1997). *History of Arunachal Pradesh*. New Delhi, Concept.

Bradley, D., R. J. LaPolla, B. Michailovsky and G. Thurgood, Eds. (2003). *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*. Canberra, Pacific Linguistics.

Burling, R. (1965). *Hill Farms and Padi Fields: Life in Mainland Southeast Asia*. Englewood Cliffs, NJ, Prentice-Hall.

--- (2004). *The Language of the Modhipur Mandi (Garo). Volume I: Grammar*. New Delhi, Bibliophile South Asia.

Cardona, G. and D. Jain, Eds. (2003). *The Indo-Aryan Languages*. London, Routledge.

DeLancey, S. (2010). 'Bodo adverbial suffixes and the genesis of grammar.' *Paper presented at the 5th International Conference of the North East Indian Linguistics Society*. Gauhati University, Assam, India, February 12-14.

--- (2012). 'On the origins of Bodo-Garo.' In G. Hyslop, S. Morey and M. Post, Eds. *North East Indian Linguistics Volume 4*. New Delhi, Cambridge University Press India: 3-20.

Diffloth, G. (2005). 'The contribution of paleoanthropology to the homeland of Austro-Asiatic.' In L. Sagart, R. Blench and A. Sanchez-Mazas, Eds. *The Peopling of East Asia: Putting Together Archaeology, Linguistics, and Genetics*. London, Routledge-Curzon: 77-80.

Donegan, P. J. and D. Stampe (1983). 'Rhythm and the holistic organization of linguistic structure.' *Chicago Linguistic Society* 19(2: Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax): 337-353.

--- (2004). 'Rhythm and the synthetic drift of Munda.' In R. Singh, Ed. *The Yearbook of South Asian Languages and Linguistics*. Berlin/New York, Mouton de Gruyter: 3-36.

--- (2009). 'Iambic vs. trochaic effects in Austroasiatic.' *Paper presented at the 4th International Conference on Austroasiatic Linguistics*. Mahidol University, Bangkok, Thailand, October 29-30.

Elwin, V. (2005 [1957]). *A Philosophy for NEFA*. Itanagar, Arunachal Pradesh, India, Directorate of Research, Government of Arunachal Pradesh.

Enfield, N. J. (2001). 'On genetic and areal linguistics in Mainland Southeast Asia: Parallel polyfunctionality of 'acquire'.' In A. Y. Aikhenvald and R. M. W. Dixon, Eds. *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*. Oxford, Oxford University Press: 255-290.

Epps, P. (2007). 'Tucanoan influence on Hup.' In A. Y. Aikhenvald and R. M. W. Dixon, Eds. *Grammars in Contact: A Cross-Linguistic Typology*. Oxford, Oxford University Press: 267-289.

Geiyi, N., Ed. (undated). *Mirror of K.A.W.A. [Descendants of Karko Welfare Association]*. Naharlagun, Memkee Publishers.

Genetti, C. (2007). *A Grammar of Dolokha Newar*. Berlin, Mouton de Gruyter.

Goswami, G. and J. Tamuli (2003). 'Assamese.' In G. Cardona and D. Jain, Eds. *The Indo-Aryan Languages*. London/New York, Routledge: 391-443.

Grierson, G. A., Ed. (2005 [1928]). *Linguistic Survey of India, Volume 2: Mōn-Khmēr and Siamese-Chinese Families (Including Khassi and Tai)*. New Delhi, Low Price Publications.

Hamilton, A. (1912). *In Abor Jungles: Being an Account of the Abor Expedition The Mishmi Mission and the Miri Mission*. London, Eveleigh Nash.

Joseph, U. V. (2008). 'Causative prefixes in four Bodo-Garo languages.' In S. Morey and M. W. Post, Eds. *North East Indian Linguistics*. New Delhi, Cambridge University Press India.

Kakati, B. (1995). *Assamese, its formation and development: A scientific treatise on the history and philology of the Assamese language*. Gauhati, Assam, India, Lawyer's Book Stall.

LaPolla, R. J. (2001). 'The role of migration and language contact in the development of the Sino-Tibetan family.' In R. M. W. Dixon and A. Y. Aikhenvald, Eds. *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*. Oxford, Oxford University Press: 225-254.

Mackenzie, A. (1884). *History of Relations of the Government with the Hill Tribes of the North East Frontier of Bengal*. Calcutta, Royal Asiatic Society.

Matisoff, J. A. (1978). *Variational Semantics in Tibeto-Burman: The "Organic" Approach to Linguistic Comparison*. Philadelphia, Institute for the Study of Human Issues.

--- (1982). 'Conjugal Bliss: An Indo-Aryan Word-Family 'pair'/'yoke'/'join' in Tibeto-Burman.' *South Asian Review* 6(3): 42-50.
--- (1990). 'On Megalocomparison.' *Language* 66(1): 106-120.
--- (1991a). 'Areal and Universal Dimensions of Grammatization in Lahu.' In E. C. Traugott and B. Heine, Eds. *Approaches to Grammaticalization, I: Focus on Theoretical and Methodological Issues*. Amsterdam, John Benjamins: 383-453.
--- (1991b). 'Sino-Tibetan linguistics: Present state and future prospects.' *Annual Review of Anthropology* 20: 469-504.
--- (1999). 'Tibeto-Burman tonology in an areal context.' In S. Kaji, Ed. *Proceedings of the Symposium 'Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics'*. Tokyo, Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies: 3-32.
Modi, Y. (2005). *Emergence and impact of Hindi as a lingua franca in Arunachal Pradesh* M. Phil Thesis, Jawaharlal Nehru University.
Morey, S. (2005). *The Tai Languages of Assam: A Grammar and Texts*. Canberra, Pacific Linguistics.
Post, M. W. (2006). 'Compounding and the structure of the Tani lexicon.' *Linguistics of the Tibeto-Burman Area* 29(1): 41-60.
--- (2007). *A Grammar of Galo*. PhD Thesis, La Trobe University.
--- (2009). 'The phonology and grammar of Galo "words": A case study in benign disunity.' *Studies in Language* 34(4): 931-971.
--- (2010). 'Predicate derivations in the Tani languages: Root, suffix, both or neither?' In S. Morey and M. Post, Eds. *North East Indian Linguistics Volume 2*. New Delhi, Cambridge University Press India: 175-197.
--- (in press-a). 'The language, culture, environment and origins of Proto-Tani speakers: What is knowable, and what is not (yet).' In S. Blackburn and T. Huber, Eds. *Origins and Migrations among Tibeto-Burman Speakers of the Extended Eastern Himalaya*. Leiden, Brill.
--- (in press-b). 'The Siyom River Valley: An essay on intra-subgroup convergence in Tibeto-Burman.' In G. Hyslop, S. Morey and M. W. Post, Eds. *North East Indian Linguistics Volume 5*. New Delhi, Cambridge University Press India.
Post, M. W. and Y. Modi (in press). 'Language contact and the genetic position of Milang in Tibeto-Burman.' *Anthropological Linguistics*.
Riba, M. (undated). Like Butterflies in the Mud [Documentary film]. Naharlagun, Arunachal Pradesh, India, Centre for Cultural Research and Documentation (CCRD).
Sen, S. (1992). *Tribes and Castes of Manipur: Description and Select Bibliography*. New Delhi, Mittal.
Sun, T.-S. J. (1993). *A Historical-Comparative Study of the Tani (Mirish) Branch of Tibeto-Burman* PhD Dissertation, University of California.
Thomason, S. G. (2001). *Language Contact: An Introduction*. Edinburgh, Edinburgh University Press.
Thomason, S. G. and T. Kaufman (1988). *Language contact, creolization and genetic linguistics*. Berkeley, University of California Press.
Thurgood, G. (2010). 'Hainan Cham, Anong, and Eastern Cham: Three languages, three social contexts, three patterns of change.' *Journal of Language Contact - Varia* 3: 39-65.
von Fürer-Haimendorf, C. (1955). *Himalayan Barbary*. London, John Murray.

A Case for Clitics in Pacoh

Richard Watson[146]
SIL International

Introduction.

When my wife and I began studying Pacoh in 1961, the term "clitic" had not yet entered our vocabulary. They all looked like prefixes to us. Although I was told by at least one colleague that Mon-Khmer languages only had derivational prefixes, such as pa- 'causative', not inflectional ones, I thought Pacoh must be an exception because it had subject markers on verbs and dative and genitive markers on pronouns. In 2000, delving back into *Mon-Khmer Studies* after many years away, I read an article by Solntseva (1995) describing the subject markers of Ta-oaih as clitics[147] and I found myself in agreement and eager to see if more Pacoh inflectional presyllables might be clitics rather than prefixes.

Anderson (2005:20) says, "Simple clitics are unaccented variants of free morphemes, which may be phonologically reduced and subordinated to a neighboring word. In terms of their syntax, though they appear in the same position as one that can be occupied by the corresponding free word." With this definition in mind we could treat the Pacoh 3rd person marking system as showing forms which are "subordinated to a neighboring word".

The system is quite simple: the Pacoh pronoun *ɗɔ:* '3rdSg' is the free form, yet it is restricted in its distribution according to the phonology of the verb. When the verb is monosyllabic (as opposed to sesquisyllabic, this will be discussed below) *ɗɔ:* is not used, instead 3rd person subjects are indexed by the attachment of unaccented *ʔu=* to the verb, in the manner that we had earlier thought of as a prefix. Similarly the unspecified 3rd person form *ŋəh* alternates with *ʔi=*. The exceptional context is with negatives, in which the free forms are ordinarily used, e.g.

(1)	***lɤjʔ***	*ʔu=pok*	**but**	*ɗɔ:*	***lɤjʔ***	*pok*
	not	3rdSg.go		3rdSg	**not**	go
	'He didn't go.'			'He didn't go.'		

Although free subjects do occasionally follow negatives as well, e.g.

(2)	***lɤjʔ***	*ŋa:j*	*dyo:n*	*ʔi.ko:ɲ*	*ʔi.duh*	*ca:*
	not	they	let	poss.male	forementioned	eat
	They didn't let their father eat it.					

Mark Alves (2002) makes a good case for the grammaticalization of prefixes in Pacoh, but I contend that in several cases the process has stopped at the clitic stage. As long as they only occur as replacements of independent words when attached to monosyllabic roots, they are morpho-syntactic units subject to

[146] I cannot take 100% credit for the text of the present paper, since it has benefitted considerably from extensive suggestions and corrections made by Paul Sidwell, and anonymous reviewer remarks which were taken into account in reworking some parts of the text.

[147] In this paper Pacoh clitics are proclitics and are considered to be syntactic elements, only appearing to be prefixes because of phonological patterning.

Watson, Richard. 2011. "A Case for Clitics in Pacoh" In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.222-232.

phonological constraints. Most of the variants I describe as clitics are, I assert in this study, related to unbound words, in clear contrast to unambiguous prefixes, such as *ta-* 'involuntary action', *tər-* 'reciprocal', *pər-* 'causative-reciprocal', etc.[148]

In my experience, and subjective judgment, Pacoh speakers favor iambic word structure, and this suggests an explanation whereby the full unbound forms of pronouns may be phonologically reduced to create the clitics discussed here. It has also been suggested (by Paul Sidwell in personal communication) that the clitic may have origins in other pronoun forms that survive as free forms in other Mon-Khmer languages, but remain in Pacoh only in these bound forms. These ideas will be discussed below, and I will leave it to the wider community of scholars to consider these ideas and draw their conclusions about which may be more correct. In the first section I limit the description to the two subject pronoun clitics for simplicity, then discuss evidence for clitic status in section 2 before going on to the description of other proposed clitics in section 3.

1. Subject pronoun clitics

1.1 Third person singular

As stated above, the third person singular pronoun *ɗɔ:* usually alternates with the clitic form *ʔu=* before one-syllable verbs. However, there is a pragmatic complication as *ʔu=* occurs in two usages:

a. non-derogatorily when there is no issue of status. In some texts *ʔu=* is not used at all, as names and other references are used instead. A character who is first referred to by name and other references before being referred to by *ʔu=* is probably in the non-derogatory usage. This is also the case in everyday speech where *ʔu=* is the regular form before monosyllabic verbs.

b. derogatorily when animals, spirits, children, enemies, or despised persons are subject of a one syllable verb. In this case it may be singular or plural, without distinction, and there is seldom any previous level of reference, unless needed for identification. If a two-syllable verb is used, *ɗɔ:* '3rdSg' is required and can also be plural in this usage; however, in this usage a speaker is far more likely to only use one-syllable verbs with *ʔu=*, and disdain may be audible. This usage is similar to Vietnamese *nó* '3rdSg'.

The distinction is not always clear, such as in the 'underdog' situation when the narrator must indicate that the hero is despised by one group but esteemed by others. In folktales it is common for the orphan or number ten son, for example, to be belittled by use of ʔu=, although he turns out to be the hero, helped by the spirits.

1.2 'Unspecified subject'

ŋəh 'unspecified subject' ordinarily yields to the clitic *ʔi=* before one-syllable verbs. Grammatically, 'unspecified subject' (abbreviated 'unspec') occurs in three constructions:

a. in a main clause when an unspecified third person is subject, e.g.

(3) *ʔi=hom ŋɛʔ, ʔi=ho:m ma: ʔi=ho:m tu.man, ʔɤn*
unspec=see all, unspec=see but unspec=see close REL
jo:ŋ ma: ʔi=ho:m tu.man toʔ təl.lɤŋ
far but unspec=see close to mirror
'One sees everything; one sees but one sees close; what is far one sees near in a crystal ball.'

In the following example the auxillary verb *ɗɔ:ʔ* 'PRF'takes *ʔi=*, but *ʔi.ɟɯ:* 'to remember' being a two syllable verb, is preceded by *ŋɤh*, e.g.

[148] see S.Watson 1966.

(4) *maː ʔən ʔi=ɗɔːq taq thɛt maː lâyq ŋəh ʔi.juː*
but *REL* unspec=PRF do wrong but not unspec. remember
'But what one has done wrong one doesn't remember.'

b. in an embedded purpose clause modifying a noun, e.g.

(5) *koh ʔu=poːk toʔ ʔn.nɔːŋ ʔi=taʔ piɗaj*
then 3rdSg=go to forest-plot unspec=do field
'Then he went to a plot for making a field.'

(6) *kər.cɔː lɤjʔ cɔːm ʔa.mɤh ʔi=ʔiɲ*
kər.cɔ not know whatever unspec=want
'Carcho didn't know what to want (what one should want).'

In English we could assume that this *ʔi=* refers to the same subject as the matrix clause, as in an English infinitive clause. However, this subject is not specific and could as easily be translated 'people in general'.

c. when a 2nd person is subject of a command. The 'unspecified subject' is less direct and so perhaps less blunt than using the 2nd person pronoun, e.g.

(7) *ʔa.kɤp ʔi=poːk!* **versus** *ʔa.kɤp maj poːk!*
'Don't anyone=go!' **versus** ' Don't you go!'

The free pronoun *ŋəh* is rarely found with a one-syllable verb, but can be. When it is, it usually indicates reduplicative style,[149] as in the two examples below:

(8) *ʔi.koh maː ʔa.ʔɔː ŋɤh ɗoːŋ ŋɤh cị̈am ʔa.ɲiː*
thus but so-good unspec carry unspec care-for uncle
'But carry and care for uncle well.'

(9) *ɓɯːs yaːŋ tah ŋɤh ta.piəŋ ŋəh klɥ̈ːn nam ɓɯːs ʔi=lɥ̈ːs.*
not spirit trash *unspec.* date unspec. play if not unspec=wrong
'A spirit won't trash *you* for courting and playing if you don't wrong anyone.'

As with *ʔu=, ʔi=* only attaches to verbs, e.g.

(10) *lɯː lɤjʔ ʔiɓoːn hoːm ɗuŋ vị̈ːl ʔn.naŋ*
Really not get.to see home community any.longer.

Example (4) provides an unusual example of *ŋɤh* preceding a verb while following a negative.

2. Pacoh word structure and phoneme distribution

The explanation that I find most satisfactory for the occurrence of clitics in Pacoh is tied to iambic word structure, on the principal that the internal forms of clitics are tied to the phonological restrictions in the presyllable (2.2).

2.1. Iambic Word structure

Pacoh words can be monosyllabic or disyllabic. Disyllabic words are described as sesquisyllabic[150] and metrically iambic; their first syllable is always unstressed and more limited in phoneme inventory, while the second, i.e. main syllable, is more strongly stressed. Disyllabic words may have presyllables that are simply part of a single morpheme or they may have a prefix or clitic followed by a monosyllabic root. With

[149] For more on reduplication in Pacoh see R.Watson 1966a and b.

[150] For more on this term see Matisoff 1973 or Thomas 1992. For more on Pacoh phonology see R.Watson 1964 & 1995 or Alves 2006.

regard to iambic effects, there is no perceived difference between presyllables, whether non-morphemic, prefix or clitic. My thesis is that speakers' preference for iambic word structure is such that various morphemes preceding monosyllabic verbs, already somewhat unstressed because of their grammatical status, have been reduced phonetically to the same status as prefixes when preceding monosyllabic roots.

Keller and Gregerson (n.d.) use examples from Krung, also Mon-Khmer, to suggest three principles of iambic reduction:

i) Iambic Domain: The Iambic Stress Wave (Weak + Strong) ranges across not only the Mon-Khmer phonological word, but also across phrases with an Iambic Head.

ii) Adjacency Principle: Iambic Reduction to the Right operates on a principle dictating that an immediately adjacent constituent to the left of an Iambic Head gets reduced.

iii) Sesquisyllabic Principle: Iambic patterns across PHRASES tend to reduce and remodel constituents towards sesquisyllabic WORDS as targets.

2.2 Phoneme distribution

Pacoh phonology is marked by strong restrictions on the distributions of phonemes within word structure. These must be understood if we are to explain proposed phonetic reductions.

a. Main syllables

In Pacoh the phonological template for main syllables is Cm (CL)Vm (Cf), as in *po:k* 'go' or *pla:ŋ* 'thatch'. Following is the full phoneme distribution for main syllables:

Cm: any of the 20 consonants, except /, jʔ, wʔ/, i.e: /p,t,c,k,ʔ,ɓ,ɗ,ʃ,m,n, ɲ,ŋ,w,j,l,r,h, ʃ/.

CL: liquids or h can occur only when following /p,t,k/, but not /tl/, i.e: /ph,pl,pr,th,tr,kh,kl,kr/. (There are a few examples of /bl, br,cr/ in Pahi.)

Vm: can be any of the 30 vowels in a closed syllable, but none of 12 short vowels in open syllables).[151]

Cf: any consonant other than the three voiced implosives (ɓ,ɗ,ʃ).

b. Presyllables

Any main syllable can potentially be preceded by a presyllable composed of a first consonant, first vowel, and, optionally, a second consonant, i.e. C1V1 or C1VəC2, such as /pa/ or /pər/. Presyllables are considerable reduced in phoneme inventory, as seen in the following list.

C1: can be any of the five voiceless stops. Other consonants can occur in C1 only as reduplications of Cm. (Any exceptions are very rare and assumed to have been borrowed.)

V1: can be only /i, a, u/.

C2: only nasals and liquids. Nasals always assimilate to Cm.

Və: only schwa in closed syllables. I assume that schwa merges with C2 when following glottal.

2.3 Presyllable distribution relevant to clitics

a. Initial consonants of clitics

The question I wish to test is whether it is viable to propose an origin in Pacoh of the suface forms *ʔu=* and *ʔi=* as originally coming from phonetically reduced *ɗɔ:* and *ŋɤh* respectively. But why would /ɗ/ and

[151] see R.Watson 1964, 1996

/ŋ/ both reduce to a glottal stop? As noted above, only voicless stops can occur in C1, unless C1 is a reduplication of Cm. Here and in section 3 we find that all clitics have initial glottal stops. This makes sense since clitics have no reduplicative connection with Cm although some prefixes do. Also, some presyllable consonants also arise from original syllable initials, such as the nominalization of *ca:* 'to eat' *tən.na* 'food' (where /c/ > /t/ since /c/ is not allowed in presyllables when it does not reduplicate Cm).

b. Clitic vowels

How could /u/ and /i/ be related to /ɔ/ and /ɤh/ respectively? I suppose that the /u/ of *ʔu=* derived from /ɔ/ of *ɗɔ:* '3rdSg' as the only back rounded vowel available in presyllables.

Secondly, I suppose that the /i/ of *ʔi=* did not derive directly from *ŋɤh* 'unspecified subject' but rather from the high front coda of *ŋa:j* '3rdPl', since I also hypothesize that *ŋɤh* actually grammaticalized from *ŋa:j*, since in Pacoh grammaticalization to /ɤh/ forms is common with pronouns. In addition to *ŋɤh* for 'unspecified subject', there is *ʔɤh* '3rdSg inalienable poss.', *mɤh* '2ndSg poss.', *ʔa.ɲɤh* '3rdDual poss.' and *ʔa.pɤh* '3rdPl poss.' *(ʔa)ɗɤh* 'Dative' is a grammaticalization of *ʔa.ɗɔ:* Dative'.

Krung also does not allow 'weak' CVC presyllables,[152] but its solution is to keep the /a/ and drop the /j/, whereas Pacoh apparently dropped the /a/ and kept /j/ as /i/. (I wonder about a similar connection between the final /j/ of *maj* '2ndSg' the *ʔi-* prefix of *ʔi.ɲa:* '2ndDual' and *ʔi.pɛ:* '2ndPl'.)[153]

2.4. Other explanations for ʔi= *and* ʔu*:*

As pointed out by Paul Sidwell, forms potentially related to Pacoh clitics *ʔi=* and *ʔu=* do occur in related languages as pronouns. The following are indicative examples taken directly from the Mon-Khmer Languages Project online database (sealang.net/monkher):

Form	Gloss	Language	Index
		Aslian	
ʔuʔ	'it'	Kensiu	Pha2006:C:336-1
ʔi-	'his, her, its'	Semai	Mea1987:C:2552
		Bahanric	
ʔi:	'it'	Alak	Huf1971:C:3217-432-7
ʔi:	'she'	Sapuan	Jac1999:C:133
		Khasic	
ʔu:	'he, it, an, the (masculine)'	Khasi	Sin1906:C:6986
ʔu	'he (3p masc. pronoun)'	War [Amwi]	Wei1975:C:1140
ʔi	'he, she (honorific marker)'	Pnar [Jowai]	Bar2010:C:2417-1
ʔi	'relative pronoun'	War [Amwi]	Wei1975:C:2273
		Mangic	
ʔi ˥˥	'he/she/it'	Bolyu	Edm1995:C:1139
		Palaungic	
ʔu	'one'	Palaung	Mil1931:C:2513
ʔi:	'person'	Rumai	New1994:C:247
ʔí	'people, others'	U	Sva1988:C:269
		Nicobaric	
ʔu	'demonstrative pronoun'	Car	Whi1925:C:6284

The difficulty I have with this data is that I don't find any evidence of free forms *ʔi* and *ʔu* being used in the 3rd person singular and unspecified meanings in the closest languages to Pacoh, such as Kadô, which you might expect in that case. If such forms had existed in Pacoh, why did *ɗɔ:* and *ŋɤh* even arise at all? It makes more sense to me to suggest that *ɗɔ:* and *ŋɤh* were free forms that deduced phonetically to that *ʔu=*

[152] That is, presyllables with semivowels in final position.

[153] Another connection might be made between the nɛh/nɤh 'this/here' of Pacoh and the naj/nɤj 'this/here' of Vietnamese.

and *ʔi=* according to an extension of the normal processes that we observe in respect of Pacoh iambic sesquisyllables.

2.5. Summary of evidence and arguments:

a. Pacoh subject clitics *ʔi=* and *ʔu=* substitute for free words *dɔ:* and *ŋɤh* only when followed by monosyllabic verb roots. This is evidence that *ʔi=* and *ʔu=,* although phonologically attached to verbs, function syntactically on the phrase level like all other subject pronouns rather than on the word level.

b. *ʔi=* and *ʔu=* occur in the same position as their free counterparts preceding verbs, except in negated clauses when they attach to the verb rather than to the less stressed negatives that free subjects usually, but not always, precede.

c. *ʔi=* and *ʔu=,* along with their free counterparts, can occur with any verb, without paradigmatic gaps as affixes often have.[154]

d. Typologically, neither head nor dependent-marking is expected in Eastern Mon-Khmer languages.

e. Reduction to clitic forms can be explained by a single or limited set of phonological rules or contexts, that is, iambic effects create 'collapse on the left' and hence reduction and coalescence. (There is never collapse on the right.) In this case a phonological word is composed of two syntactic units.

f. An etymologically motivated explanation can be suggested that involes grammaticalization of Mon-Khmer pronouns.

3. Additional proposed clitics

3.1 The Dative clitic* ʔa= + *PRONOUN

Synchronically, in Pacoh the dative preposition *ʔa.dɔ:* occurs with nouns and two-syllable pronouns, but it **appears** to be reduced to the clitic *ʔa=* on one syllable pronouns. Below, a different history is proposed.

(11)	*kɯ:*	'1stSg'	>	*ʔa=kɯ:*	'to me'
	maj	'to you'	>	*ʔa=maj*	'to you'
	dɔ:	'3rdSg'	>	ʔa=dɔ:	'to him/her/it'

Other combinations are *ʔa=ɲaŋ* 'to us-dual', *ʔa=hɛ:* 'to us-Pl', *ʔa=ŋa:j* 'to them-general'. Curiously, *dɔ:* '3rdSg' can be preceded by either the clitic or the full form. We find both *ʔa.dɔ: dɔ: nɛh* and *ʔa=dɔ: nɛh* 'to him here', as well as both *ʔa.dɔ: dɔ: koh* and *ʔa=dɔ: koh* 'to him there'. For example:

(12) *ku.mi̯ʔ kɯ dɔ:ŋ ʔa.cu̯: ʔa.dɔ: dɔ: koh ɓu̯aiʔ*
a.little.while I carry return to 3rdSg that fish
'In a little while I'll bring him fish.'

(13) *ʔiɲ cɔ:m ʔi.mɔ: lɯ: ʔa.ʔaj ʔa.dɔ: koh*
want know why really sick to-3rdSg that
'(We) want to really know why he is sick.'

Diachronically there is good evidence that *ʔa:* was the dative preposition in all environments, and it became attached to one syllable pronouns as a clitic by the same iambic effects as described in section 2 above. Then *ʔa=dɔ:* '3rdSg Dative' was likely combined into *ʔa.dɔ:* in order to keep the unstressed *ʔa=* from merging with a following presyllable, such as the first syllable of *ʔa.ɲa:* 'they-Dual' and *ʔa.pɛ:* 'they-Pl'?.

[154] Anderson 22005:33 "Affixed words are more likely to have idiosyncratic shapes than host-clitic combinations."

Evidence for this is found in the closely related Kadô language,[155] where *ʔa:* 'Dative' is the free variant rather than *ʔa.ɗɔ:*. This free *ʔa:* in Kadô is given full stress to avoid merging with a following *ʔa* presyllable; but it loses length and stress to attach to monosyllabic pronouns as a clitic, e.g., *ʔa:* + *ʔan* '3rdSg' yields *ʔa=ʔan* '3rdSg Dative'.

3.2. Nominalizer clitic ʔN=

ʔɤn 'relativizer' (REL) functions more broadly as a nominalizer.[156] *ʔɤn* frequently follows a noun, pronoun or proper noun as a relativizer introducing a modifying clause. However, such a 'relative clause' does not need a noun head--it can stand alone as a full noun phrase. A missing head noun might point to something known from the context, however long distant, or something commonly known, such as, *parnai* 'matter' in:

(14) *hɛ:* *ɗɔ:ʔ* *ʔi.juɯ:* *ʔɤn* *ʔa.ʔi:* *ʔa.ʔam* *ɗɔ:ʔ* *to:ŋ* *pa.tam*
we PRF remember REL mo. fa. PRF say advise
'We have remembered (matter) that parents have advised us.'

However, in that case or the following it might be that the nominalized phrase or clause simply functions as a noun phrase argument, for example:

(15) *keh* *ʔɤn* *Phɯang* *ʔa.dah muəŋ* *hɛ:* *tɯ:* *ʔɤn* *hɛ:* *po:k* *taʔ* ʔŋkeh
so that Phɯaŋfear region us from that we go do that
'So the Phương fear our region from that (time when) we went and did that.'

Stative verbs can modify nouns with or without *ʔɤn*, e.g., *ʔacɔ: pɯ:t* 'dog big' vs. *ʔacɔ: ʔɤn pɯ:t* 'dog that.is big'. Without *ʔɤn* 'that', the sense is more descriptive, but with *ʔɤn* it is more identificational. When 'relativizing' or 'nominalizing a clause or stative verb, *ʔɤn* is never reduced to a clitic.

On the other hand, whenever *ʔɤn* nominalizes pronouns, demonstratives, and certain stative verbs; it is reduced by iambic pressure to the clitic *ʔN=*[157], assimilating to the point of articulation of the initial consonant of the following monosyllabic word:

a. *ʔN=* + DEMONSTRATIVE derives pronouns. For example:

(16)	*nɛh*	'this'	>	*ʔn.nɛh*	'this one'
	koh	'that'	>	*ʔŋ.koh*	'that one'
	tih	'yonder'	>	*ʔn.tih*	'that yonder one'
	mɔ:	'which/where'	>	*ʔm.mɔ:*	'which-one' 'which-place'

When *ʔnnɛh* 'this' or *ʔŋ.koh* 'that' modify a noun, stated or elided, they appear to function as anaphoric pronouns rather than simply demonstratives, for example *ɗɔ: koh* '3rdSg there' versus (*ɗɔ:*) *ʔŋ.koh* '(3rd) that one there'; *toʔ koh* 'to there' versus *toʔ ʔŋ.koh* 'to that place'.

b. *ʔN=* + PRONOUN derives possessive pronouns. For example:

---

[155] It's unclear whether Kadô should still be considered as a dialect of Pacoh or a separate language. The basic word list indicates only a dialect difference. More extensive materials indicate a much greater difference. ?an is '3rdSg' as in Pahi and Bru.

[156] The nominalizer for verbs is an infix –N/L- with several nasal and liquid allomorphs, e.g., *taʔ* 'do/make' > tən.taq 'doing/thing-made', *po:k* 'go' > *pəm.po:k* 'going' *ca:* 'eat' > *tanna* 'food', *kla: ŋ* 'to pipe' > *kalla:ŋ* 'bamboo pipe', *krup* 'to cover' > *karrup* 'a cover'.

[157] Phonologically I would write *ʔəN*, but the schwa is imperceptible between glottal stop and nasal or liquid in presyllables.

(17)	*kɯ*	'1stSg'	>	*ʔŋ=kɯ:*	'mine'
	maj	'2ndSg'	>	*ʔm=maj*	'yours'
	dɔ:	'3rdSg'	>	*ʔn=dɔ:*	'his/hers/its'
	ɲaŋ	'1stDual'	>	*ʔŋ=ɲaŋ*	'ours'
	hɛ:	'1stPl'	>	*ʔŋ=hɛ:*	'ours'
	ŋa:j	'3rdPl'	>	*ʔŋ=ŋa:j*	'theirs-general'

Since the longer variant *ʔn.dɔ:* 'belongings' is required for two-syllable pronouns, e.g., *ʔndɔ: ʔiɲa:* 'yours-dual', *ʔn.dɔ: ʔa.ɲa:* 'theirs-dual', *ʔn.dɔ ʔi.pɛ*, one might assume that the clitic *ʔN=* was derived from *ʔn.dɔ:* as the longer variant. But following the example of *ʔa:* vs. *ʔa.dɔ:* 'Dative', it is more likely that *ʔn.dɔ:* is composed of *ʔɤn+dɔ:*.

ʔɤn bears a grammatical resemblance to Chinese *de*, also labeled as relativizer and nominalizer, but labeled 'GEN' when marking a pronoun or noun as possessor, e.g. *wǒ de péngyǒu* 'my friend' (Li & Thompson 1981:126)[158]. In Pacoh *ʔɤn* meets all of those labels (also called an 'associative marker'). Following a noun either the plain pronoun or the nominalized variant can occur, e.g., *praʔ kɯ:* or *praʔ ʔŋ=kɯ:* 'my money'. The nominalized variant is used for contrast. For example:

(18)	*ʔaʔam ɲɥŋ,*	*ʔih*	*pɯ:t*	*ʔndo*	*pɛdaj*
	father look,	is.not	big	belonging	us?

'Look father, isn't ours bigger?' (contrast with that of #10 son)

c. *ʔN=* + STATIVE VERBS derive a semantic class

The relativizer clitic does not attach to stative verbs in general, but only a few like the ones below.

(18)	*kɥ:ɲ*	'male	>	*ʔŋ.kɥ:ɲ*	'the males'
	kan	'female	>	*ʔŋ.kan*	'the females'
	ʔɔ:	'good/pretty'	>	*ʔŋ.ʔɔ:*	'the good ones'
	pɯ:t	'big'	>	*(ʔa.pɛ:) ʔm.pɯ:t*	(they) who are big=the important ones

3.3. The adverbializer clitic ʔa=

Pacoh has an adverbializer *ʔɔ:* 'resultative' that is usually reduced to *ʔa=* before one-syllable verbs (since /ɔ/ cannot occur in a presyllable). This is almost always the case when it precedes the stative verb *ʔɔ:* 'good, well', which is the common farewell:

(20)	*ʔa=ʔɔ:*	*ʔi=po:k*	*ʔaw*
	so=good	unspec=go	imperative

'Go well!'

And the departing persons say:

(21)	*ʔa=ʔɔ:*	*ʔi=ʔat*	*ʔaw*
	so=good	unspec=stay	imperative

'Stay well!'

In written form there is a recent tendency to spell out the adverbializer, i.e., **o o** 'so good', probably indicating more concern with meaning than with phonics. There are some entries in our lexicon that might indicate 'frozen' or fully lexicalized combinations, such as, *ʔa.wa:s* 'to create', *ʔa.ɓɯ:s* 'to avoid', and *ʔa.liəh* 'again'. However, we find that these are productive forms that can also occur as two words, e.g., *ɲɥ:n ʔɔ: wa:s* 'cause so-as-to appear' *ʔɔ: ɓɯ:s* 'so-as-to not', and *ʔɔ: liəh* 'so-as-to come-back'.

Curiously, Pacoh also has a number of negative verbs that begin with *ʔa-*, such as, *ʔa.ɓiɲ* 'not want', *ʔa.ʔiɲ* 'hate' vs. *ʔiɲ* 'want', and *ʔa.vi* 'not have' vs. *vi* 'have/exist'. And the Pahi dialect has a homophonous word *ʔa.wa:s* 'not', as well as *ʔa.wa:s* 'to create', but I cannot suggest a source for this *ʔa* within Pacoh.

[158] See Matisoff 1972 for same phenomenon in many SEAsian languages.

3.4. The ʔN/L= 'single' clitic?

Some classifiers and quantifiers can take the *ʔN/L=* clitic, indicating a single one or small amount. It seems that nominalizing the classifier emphasizes the individualizing function of the classifier, e.g. instead of *mɤh lam ʔacɔ:* 'one class. dog', one might say, *mɤh ʔl.lam* 'one single one' or *ʔl.lam ʔacɔ:* 'a single dog'. Another theory is that this clitic is a reduction of *mɤh* 'one', which already appears to be a grammaticalization of *mu̯:j* 'one'. However, *mɤh* can be followed by both *ʔN/L* and a classifier, e.g., *mɤh ʔl.lam* 'one single thing'. Other examples are *ʔn=naʔ* 'one thing', *ʔn=to:m* 'one stem/trunk', *ʔm=ɓeaq* 'a few', *ʔn.ɟiaʔ* 'a few'.

In the Pahi dialect the Flood text refers to *ʔn=lam ʔa.cɔ: ʔn=lam ti.kuəj* 'one dog (and) one person', in which /ʔn/ doesn't assimilate, lending, more evidence to the *ʔɤn* hypothesis. If it seems strange that this morpheme can be realized as nasal or liquid, note the similarity with English *in-* 'not', in *impossible*, *ineligible*, *illegible, irregular,* etc. Another hypothesis is that this originated as a reduplication of the classifier *lam*, giving *ʔm=*, which assimilates in the same way that *ʔn=* would.

3.5. Demonstrative clitics or contractions

Since all Pacoh clitics are considered to be contractions based on iambic effect, 'clitic vs. contraction' might seem to be an unnecessary contrast. However, I believe that the examples below show that a distinction is necessary, even though somewhat fuzzy.

ʔi=nɛh 'like this' and *ʔi=koh* 'like that' appear to be reductions of *ʔi:n .nɛh* 'like this' and *ʔi:n .koh* 'like that thing', in which the nasal is elided in addition to dropping length and stress. To have adapted to the CVC presyllable template would have made it homophonous with *ʔn.nɛh* 'this one' and *ʔŋ.koh* 'that one' of section 3. The interrogative demonstrative *ʔi=mɔ:* 'like how/why' also appears to be a reduction of *ʔi:n mɔ:* 'like which/where/how/why'. The free word *ʔi:n* never occurs separately with the demonstratives, but rather with nouns and pronouns, e.g., like *ʔi:n dɔ: koh* 'like 3rdSg that', and often in the synonymous compound *ʔi:n ʔarɤʔ* 'like like', as in *ʔi:n ʔarɤʔ dɔ: koh* 'just like 3rdSg that.

In addition to *ʔi=mɔ:* above, the interrogative demonstrative *mɔ:* 'which/where/how/why' occurs in four other contracted variants:

ʔm.mɔ: was discussed in section 3 above, e.g. *toʔ mɔ:* 'to where' vs. *toʔ ʔm.mɔ:* 'to what place'.

toʔ mɔ: 'to where?' reduced to *tu=mɔ:*, especially in the Pahi and Kadô dialects.

The example most suspicious of being only a contraction rather than a clitic is *dɔ: mɔ:* '3rdSg which/whoever' reduces to *ɗa=mɔ:, which* can also be reduplicated in *ɗa.mo: ɗa.mo:* 'everyone' Two counts against identifying it as a clitic are the fact that *dɔ:* should not cliticize to *ɗa* since the *ɗ* does not reduplicate Cm, the main syllable initial consonant (see 2.3a above). Secondly, there seems to be a growing tendency for writers to write it out fully as either ***do mo*** or ***domo***, even though it was only found in its full form once in a collection of early texts.

We would expect to find 'from where' spoken as **tɛ: mɔ:* or reduced to **ta.mɔ:;* but we don't find either one. Instead, whenever *tɛ:* 'from' is followed by a demonstrative, it requires *ta-*to be added, e.g., *tɛ: ta.mɔ:* 'from where'*, tɛ: ta.nɛh* 'from here', *tɛ: ta.koh* 'from there', *tɛ: ta.tih* 'from up there' and *tɛ: ta.tɔh* 'from down there'. It appears as if *ta=* arose as a clitic of *tɛ:* but was not deemed adequate, so both *tɛ:* and *ta=* are maintained.

3.6. Kin terms

Kin terms, like most fauna and many flora names, begin with *ʔa-*, which appears to be a class marker. The *ʔa-* on kin terms can be replaced by *ʔi-* to mark inalienable possession. For example, *ʔa.miəŋ* means older brother of a female', but a girl cannot say, *ʔa.miaŋ ku:* 'my older-brother', but rather *ʔi-miəŋ ku:*. Some roots also change to possessed forms, e.g., *ʔa.ca:j ʔa.ʔɛ:m* 'older and younger siblings', become *ʔi-sa:j ʔi-si̯am*. In the case of *ʔa.ʔam ʔa.ʔi:* 'father and mother', the High Pacoh attach *ʔi-* to the nouns for

'man' and 'woman', i.e. *ʔi-kɥ̯ːɲ ʔi-kan* when referring to 'someone's father and mother'. It is not clear to me whether this *ʔi-* should be considered a prefix or a clitic, although it may ultimately have the same origin as the *ʔi=* '3rdSg' marker.

However, an argument can be made concerning the source for the *ʔa=* and *ku=* markers on Pacoh boys' nicknames. In the areas that use *ʔa.ʔam* for 'father', boys' nicknames add *ʔa=* to the first word of the rhyme created for the nickname; whereas in the dialect areas that use *kɥ̯ːɲ* for 'father', boys' nicknames usually add *ku=* instead.[159] *ʔaː* is found as a third person pronoun in other Katuic languages (e.g. Ngeq *ʔaː* 'he' [Huf1971:C:2827-380-16], Souei *ʔaː* 'it' [Huf1971:C:3216-432-6]) it is likely that the *ʔa=* on nicknames is a phonological reduction and attachment consistent with the preference for sesquisyllables. This may have provided an analogical model for the use of *ku=,* since in the minds of Pacoh speakers *ʔa.ʔam* and *kɥ̯ːɲ* provide the sources for contrast between *ʔa=* and *ku=*.

Girls nicknames start with *kan* 'female' which is the same as *kan* in the mother title, except that in the nickname it precedes a meaningful word rather than a meaningless child's name. There are also various village names that begin with /kən/, such as *kənthoŋ*. I suspect that these may be based on women's names, but investigation is needed to determine if they really are women's names and if *kan* was reduced to *kən=* in this environment.

The grandfather title *vɥ̯ʔ* or grandmother title *kaʔ* can be replaced by *ʔu=*as a title of respect, e.g. *ʔu=nɔːn* 'highly respected Non'. In discussion, I think I hear *ʔu* as both a stressed free form and as a clitic attached to the following name. Since Pacoh names (not nicknames) are normally only one syllable, I have no clear contrasts with *ʔu* before disyllabic words. Since *ʔu=* '3rdSg' is often used derogatorily, it might seems odd that it is also a term of respect, but the one is a pronoun only attached to verbs while the other is a title only attached to names.

References

Alves, Mark J. 2003. Pacoh Pronouns and Grammaticalization Clines. In Shoichi Iwasaki, Andrew Simpson, Karen Adams & Paul Sidwell, eds. SEALSXIII: papers from the 13th meeting of the Southeast Asian Linguistics Society (2003). Canberra, *Pacific Linguistics*, 2007, pp. 1-12.

Anderson, Stephen. 1995. Rules and Constraints in Describing the Morphology of Phrases. In *Chicago Linguistic Society 31 Vol. 2: The Parasession on Clitics*, pp. 15-31.

_____. 2005. *Aspects of the Theory of Clitics. OSTL 11*. Oxford University Press.

Jenny, Mathias. 2003. New Infixes in Spoken Mon. *MKS* 33:183-194.

Keller, Charles, Jacqueline Jordi & Kenneth Gregerson. 2011. "Merge Right: Iambic Effects, Grammaticalization and Cliticization in Brao Grammar." In Sophana Srichampa, Paul Sidwell and Kenneth Gregerson (eds.) *Austroasiatic Studies: papers from ICAAL4. Mon-Khmer Studies Journal Special Issue No. 3*. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. pp.152-167.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Matisoff, James A. 1972. Lahu nominalization, relativization, and genitivization, in J. Kimball, ed., *Syntax and Semantics*, Vol. I:237-257.

_____. 1973. Tonogenesis in Southeast Asia. In L. Hyman (ed.), *Consonant types and tone*, 71-95. Los Angeles: University of Southern California.

Mon-Khmer Languages Project database: http://sealang.net/monkhmer/dictionary/

Nichols, Johanna. 1986. Head-Marking and Dependent-Marking Grammar. *Lg*. 62:56-119.

Solntseva, V. Nina. 1996a. Genitive case in Altaic languages and in some languages of Southeast Asia. *MKS* 25:255-262.

Solntseva, V. Nina. 1996b. Case-marked Pronouns in the Taoih Language. *MKS* 26:33-36.

Thomas, David D. 1992. On Sesquisyllabic structure. *MKS* 21:206-210.

Watson, Richard L. l964. Pacoh phonemes. *MKS* l:l35-48. Saigon: Linguistic Circle of Saigon and SIL.

_____. l966a. "Clause to sentence gradations in Pacoh." *Lingua* l6:l66-89.

[159] See Watson (1969).

_____. l966b. *Reduplication in Pacoh. (Hartford studies in linguistics* 21, xi, 138 pp.) Hartford: Hartford Seminary Foundation.
_____. l969. "Pacoh names." *MKS* 3:77-89.
_____. 1996. Why Three Phonologies of Pacoh? *MKS* 26:197-205.
Watson, Saundra K. 1964. Personal Pronouns in Pacoh. *MKS* 1:81-97.
_____. 1966. Verbal Affixation in Pacóh. *MKS* 2:15-30.
_____; and Cubuat. l979. *Nóh Pacóh-Yoan-Anh (Pacoh-Vietnamese-English dictionary)*. Huntington Beach, CA: SIL. xi, 445 pp.

This is the second 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 9780858836426