

Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society: Special Session on Tibeto-Burman and Southeast Asian Linguistics (2002)

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

PROCEEDINGS OF THE TWENTY-EIGHTH ANNUAL MEETING
OF THE
BERKELEY LINGUISTICS SOCIETY

February 15-18, 2002

SPECIAL SESSION
ON
TIBETO-BURMAN AND SOUTHEAST ASIAN LINGUISTICS

in honor of Prof. James A. Matisoff

edited by:
Patrick Chew
BERKELEY LINGUISTICS SOCIETY

Berkeley Linguistics Society
Department of Linguistics
University of California
Berkeley, California 94720
United States of America

All papers copyright © 2005 by the Berkeley Linguistics Society, Inc.

All rights reserved.

ISSN 0363-2946
LCCN 76-640143

Printed by:
Sheridan Books
100 N. Staebler Road
Ann Arbor, MI 48103

Table of Contents: Special Session	iii
Foreword	iv

Special Session

The Central and Southern Loloish languages of Vietnam JEROLD A. EDMONDSON	1-13
Tone and syllable structure of the Hakha (Lai-Chin) noun LARRY HYMAN and KENNETH VANBIK	15-28
Timing tonogenesis: evidence from borrowing MARTHA RATLIFF	29-41
Different durations of diphthongs in Thai: a new finding RUNGPAT ROENGPITYA	43-54
Nominalization and relativization in Tibeto-Burman SCOTT DELANCEY	55-72
Kinship and spirit terms renewed as classifiers of “animate” nouns and their reduced combining forms in Austroasiatic ANNE DALADIER	73-86
Nominalization, relativization, and genitivization in Thulung Rai AIMÉE LAHAUSSOIS	87-98
On Khumi verbal pronominal morphology DAVID A. PETERSON	99-110
Southeast Asian features in the Munda languages: evidence for the analytic-to-synthetic drift of Munda PATRICIA J. DONEGAN and DAVID STAMPE	111-120
Parallel innovations and coincidence in linguistic areas: on a bi- clausal extent/result construction of mainland Southeast Asia N.J. ENFIELD	121-128
Classifier systems and noun categorization devices in Burmese ALICE VITTRANT	129-148
Contact produced variation in the Tsat of Hainan GRAHAM THURGOOD	149-164

Foreword

The editor would like to first and foremost tender respectful apologies to not only the authors of the articles in this issue, but also to the readership that has probably been thrown off by the delay in publication. The editor assumes full responsibility for any and all typographical errors found in the typesetting of this issue.

This issue of the Proceedings of the Special Session of the 28th Annual Meeting of the Berkeley Linguistics Society finds its roots and inspiration in the pioneering and prolific contributions made by Professor James A. Matisoff, who is now at the time of publication Emeritus faculty. The articles found in this issue present a wide range of topics on the broad languages and linguistics of Mainland Southeast Asia, many of which are of the Tibeto-Burman family.



James A. Matisoff is Professor Emeritus of Linguistics at UC Berkeley. His chief research interests include Southeast Asian languages (especially Tibeto-Burman and Tai), Chinese, Japanese, field linguistics, Yiddish studies, historical semantics, psycho-semantics, language typology, and areal linguistics. He is considered one of the world's foremost authorities on Southeast Asian Linguistics.

After having first taught at Columbia University (1966-69), he joined the Berkeley faculty in 1970. He has conducted fieldwork on Lahu and other Tibeto-Burman languages in Thailand (1965-66, 1970, 1976-77, 1985, 1991) and China (1983, 1984, 1991). He is a former editor of the journal, *Linguistics of the Tibeto-Burman Area*, and is principal investigator of the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) project, which has been supported by NSF and NEH since 1987. He is author of *The Grammar of Lahu*; *Variational Semantics in Tibeto-Burman*; *Blessings, Curses, Hopes, and Fears: Psycho-ostensive Expressions in Yiddish*; *The Dictionary of Lahu*, and *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*.

It is with great honor and pleasure that the editor presents this volume of articles honoring Prof. Matisoff. The editor would also like to include his own personal gratitude for the innumerable advice, guidance, and support Prof. Matisoff has shown over the years as mentor and advisor.

Patrick Chew, vol. editor

The Central and Southern Loloish Languages of Vietnam

JEROLD A. EDMONDSON

University of Texas at Arlington

1. Introduction.

In *Three Tibeto-Burman languages of Vietnam* (2004) I outlined the Vietnam locations and situations of three Northern Loloish languages—Phu Kha (Phù Lã), Xá Phó, and Lôlô.¹ In this paper I present data and analysis on the remaining TB groups of that country—the Côông, the Sila, the Lahu, and the Hani, all found in Lai Châu Province in the far northwest and all belonging to the Central and Southern sub-groupings of the Loloish language. Like the three Northern Loloish language, all these are found very near the border with China and all—except possibly Sila—are presumed to have ultimately come from the north. However, we are only beginning to understand the obviously complex language history that has led to many linguistic groups living in close proximity and the sequencing of migration and conflict that are woven into the intricate tapestry of Mường Te District.² Indeed, until now there has been very little known in general about these four languages aside from basic information about their home territory, numbers, and some cultural features. That is not to say that all these languages have been points of utter darkness. The Lahu and Hani languages of Thailand and China, for example, have been described and analyzed in great depth. The work of Matisoff 1973, 1978 is especially notable for Lahu, and Hansson 1989 and Li and Wang 1981 have published much on Hani. But information about the other two languages—the smaller groups, Côông and Sila—has been brief and incomplete. These places do not allow of a full statement about any of these languages, but I hope, nevertheless, to provide more details about all these languages and how they compare to language forms outside Vietnam, cf. also my website³ for a tabulation of about 500 items taken from my field study of language of this area.

In the following, I will first discuss Côông and Sila and then go on to Lahu and

¹ The research reported on here has been sponsored by a 1995 grant NEH RT-21754-95 from the National Endowment for the Humanities and by the grants SBR 9511285 and SBR9729043 from the National Science Foundation to the author and Dr. Kenneth J. Gregerson all entitled “Languages of the Vietnam-China Borderlands”. I wish also to acknowledge the assistance of Profs. Nguyễn Văn Lợi, Hoàng Văn Ma, To Văn Thang, who arranged and accompanied me on the field trips that led to the data and analysis here. Many thanks as well are due Pete Unseth, who spent many hours digitizing the data from my original tape recordings, and Trần Thuần for help with some of the Vietnamese data. Most of all I wish to thank Graham Thurgood who was able to unlock the system of tonal development in all of these languages.

² Lai Châu province has the most complex linguistic situation of any place in Vietnam and much of that complexity is due to the number of groups in Mường Te. In addition to the Tibeto-Burman groups, one finds there White Thái farmers and the little studies Mon-Khmer grouping – Máng.

³ <http://ling.uta.edu/~jerry/>.

Hani. All these groups live in Lai Châu Province, Mường Te District on the China and Lao borders, regarded as territory among the most difficult of access in all Vietnam. To reach our field location one full day of bone shattering jeep travel was necessary from Điện Biên Phủ, passing the then flooded provincial capital at Lai Châu, then tracing the course of the Black River to road's end at Mường Te District capital. This remote site has allowed retention of a traditional lifestyle and the development of a linguistic microcosm of unexpected diversity and vitality at the end of a long and hard road.

2. The Công.

The Công people of Mường Te District live in five villages: Bo Lếch (Can Hồ Commune), Nậm Khao, Nậm Pục (Nậm Khao Commune), Tác Ngá (Mường Mò Commune), and Nậm Kè (Mường Tong Commune). Their population was given as 1261 in the last official census 1989. The population is estimated to have reached 1560 by the year 2000 PV (1998:21). It is said that their ancestors originally came from China, but our informant, Mr. Lý Văn Lăng, about 55 years of age in June 1999, had no information about the time or source of this migration. Bradley (1977:68) states that the Công probably fled China as a consequence of the Moslem uprisings in Yunnan Province during the 19th century and the first decades of the 20th century and then were resettled during the wars between the Burmese and Vietnamese into NW Vietnam.

The Công autonym is also a puzzlement. In EMPV (363) it states that the most widely used name is a toponym from one of their villages, Bo Lếch, a White Thái designation meaning 'iron mine'. Thus the Công refer to themselves in their own language as [sam³³ khoŋ³³ (tshaŋ³³ a³¹)] 'iron mine people'. The [tshaŋ³³ a³¹] is used to designate 'people, group, ethnicity', such as [a³¹ kha³³ tshaŋ³³ a³¹] 'Hani' and [za³³ ze³³ tshaŋ³³ a³¹] 'Yao'. At Nậm Khao and Nậm Pục the autonym [phuɪ³³ a³¹] is known but little used. It also resembles the name the Công use for the Lahu [kha⁵⁵ phuɪ³³]. A number of people have also suggested the name *Công Lô Ma*, which is said to refer to a place in China where they once lived.

Bradley regards Công to be a language closely related to Phunoi (1977:68, 1979, 1997), "In Vietnam, the Phunoi are called Công, and speak a slightly different dialect..." Công was first recorded by LeFèvre-Pontalis 1892, which we have not consulted. We, however, have been able to examine Bradley's word list.⁴ In his description of Phunoi Bradley (1979:45-7) notes the existence of minor syllables, as the *ja* in *ja -ba³³* 'elephant', initial voiceless nasals /hm hn hŋ hmj/, a voiceless lateral /hl/, and a voiceless palatal glide /hj/. Phunoi, moreover, has final consonants /-p -t -m -n/ and four tones described as high level, mid level, low level and low rising. The vowel nuclei are /i u ɛ ə o ai a au/. Of the minor syllables, Bradley says (47) that the word for hand *là* also appears as a minor syllable [lə] in some compounds.

2.1. Distinctive features of Công.

The Công of Vietnam has a high level tone (55), a mid-falling tone (31), and a

⁴ I was also able to listen to data recorded in the 70's in Vientiane, Laos by Jimmy G. Harris. There were about 1000 items in that list. Harris later trained this speaker how to write his language in a romanized script and how to organize a dictionary.

mid-level tone (33) corresponding to the tones of the parent language *1, *2, and *3. There appear also to be some cases of a rising tone. Initials are:

p	ph	b	m	w
pj	phj	bj	mj	
t	th	d	n	s
ts	tsh	dz	ɲ	ɕ j
k	kh	g	ŋ	

Rhymes include /i u u im in iŋ um un uŋ ø e o ɔ a am an aŋ/ in our list of about 500 words.

2.2. Comparative comments.

We are now able to state that Côông appears to be a language of significantly different properties from Phunoi, as the comparisons below show.

<u>Phunoi</u>	<u>Công</u>
a. minor syllables take the tone of the main syllable but has no independent tone	first syllable is often unstressed but has independent tone
b. voiceless nasals, laterals, palatals	no voiceless nasals, laterals or palatals
c. final consonants /-p -t -k/	no final oral stop consonants, only /-m -n -ŋ/

Công also demonstrates several innovations not found in the Phunoi data at our disposal. Bradley's 1977 Phunoi material shows the voiced stops /b d/ corresponding to Công /m/ and /l/ or /n/. This feature is very important as many of these examples stem from Proto-Loloish vocabulary with *C-nasal initials, which in Bisoid languages (the sub-branch to which Bradley assigns Phunoi and presumably Công) regularly develop into voiced stop initials. Công fails to reflect voiced stops and instead exhibits nasals. Perhaps Công has undergone a recent mutation changing voiced stops to nasals or perhaps it fails to have this landmark features but is still a Bisoid language. Below are some comparisons that show the voiced stops-nasal alternation.

Gloss	Phunoi	Công
woman	khǎbja ³¹	taŋ ³¹ kha ³¹ ma ³¹
daughter	je ¹³ bi ³³	za ³¹ mi ³¹
fire	bi ³¹	mi ³¹
dream	jup ³¹ ba ³³ ba ³³	zu ³¹ ma ³³
black	ʔā ⁵⁵ da ³³	na ³³ la ³³
eye	ʔā ⁵⁵ bia ³³	me ³³ nu ³³
elephant	jəba ³³	za ³¹ ma ³³
a fly	ma ³¹ ba ³³	nuu ⁵⁵ ma ³³
river	lā ⁵⁵ ba ³³	u ³¹ ma ³³
near	ʔā ⁵⁵ di ³¹ a ³³	ni ³¹ ni ³¹ a ³¹

Phunoi nasal finals have often disappeared where they are preserved in Côông:

Gloss	Phunoi	Công
sun	<i>mo³¹ ni³³ si³¹</i>	<i>muŋ³¹ nuŋ⁵⁵</i>
mushroom	<i>hmu⁵⁵</i>	<i>muŋ⁵⁵</i>

Voiceless onset nasals, palatal continuants, and laterals can begin a syllable, but Côông cognate vocabulary evidences only nasals, palatals, and laterals that are voiced or sometimes /h/.

Gloss	Phunoi	Công
forget	<i>hmin³³ la³¹ tse³³</i>	<i>min³³ ka³¹ li³⁵</i>
buy	<i>hne³³ ce³³</i>	<i>hu³³ e³³</i>
red	<i>ʔā⁵⁵ hne⁵⁵</i>	<i>ne⁵⁵ le⁵⁵</i>
today	<i>hna³³ mi⁵⁵ ni³³</i>	<i>nam³³ loŋ³¹ koŋ³³</i>
mushroom	<i>hmu⁵⁵</i>	<i>muŋ⁵⁵</i>
beard	<i>ban³¹ hmot</i>	<i>man³¹ mu³¹</i>
hot	<i>ʔā⁵⁵ hlā⁵⁵</i>	<i>loŋ⁵⁵ e⁵⁵</i>

Phunoi final voiceless stops are lost in Côông:

Gloss	Phunoi	Công
lightning	<i>mo³¹ biap³¹</i>	<i>muŋ³¹ mian³¹</i>
kill	<i>sat³¹ ce³³</i>	<i>se³¹ i⁵⁵</i>
vomit	<i>pha³¹ ce³³</i>	<i>phe³¹ hai³⁵</i>
shirt	<i>hlat³³</i>	<i>a⁵⁵ kha³³</i>

Phunoi rime -au corresponds to Côông -ø:

Gloss	Phunoi	Công
bone	<i>ʔā⁵⁵ jau³¹</i>	<i>ʔaŋ³¹ jø³¹</i>
horn	<i>ʔā⁵⁵ chau⁵⁵</i>	<i>ʔaŋ³¹ khø⁵⁵</i>
sweet	<i>ʔā⁵⁵ chau⁵⁵</i>	<i>təhø⁵⁵ lø⁵⁵</i>
nine	<i>cau³¹</i>	<i>kø³¹</i>
sugarcane	<i>pon³¹ chau⁵⁵</i>	<i>pon³¹ təhø⁵⁵</i>
shout	<i>hau⁵⁵ ce³³</i>	<i>hø⁵⁵ e⁵⁵</i>
steal	<i>kha³¹ ce³³</i>	<i>khø³¹ e³³</i>
widow	<i>bəchau³¹</i>	<i>mā³³ te³³ təhø³¹</i>

Phunoi minor syllables possess a limited inventory of initials, schwa vowels, and a tone the same as the following main syllable. Côông word structure has some features of a weak first syllable; it is, for example, iambic with a somewhat shortened first vowel for many words. This vowel is frequently not schwa and the

first syllable can have its own fully independent tone. On the other hand, the tone categories of the two languages also match closely in value and distribution.

Gloss	Phunoi	Công
comb	<i>təcha</i> ³³	<i>tu</i> ³¹ <i>kha</i> ³³
navel	<i>məcha</i> ^{31?}	<i>pe</i> ⁵⁵ <i>to</i> ^{33?}

There are also a great many lexical differences between Phunoi and Công. Our vocabulary is at present limited, so caution is in order. Phunoi and Công are, of course, in contact with Lao and Vietnamese respectively and loanwords may be the source of these differences (e.g. for ‘year’ and ‘four’). It also appears that that the palatal glide *j* has changed to *z* under Vietnamese influence.

Gloss	Phunoi	Công
daughter	<i>ja</i> ¹³ <i>bi</i> ³³	<i>za</i> ³¹ <i>mi</i> ³¹
dream	<i>jup</i> ³¹ <i>ba</i> ³³ <i>ba</i> ³³	<i>zu</i> ³¹ <i>ma</i> ³³
fish	<i>ta</i> ³³ <i>te</i> ³³	<i>loy</i> ³¹ <i>te</i> ⁵⁵
crab	<i>wəcha</i> ³¹	<i>lay</i> ⁵⁵ <i>to</i> ³¹
eye	<i>ʔa</i> ⁵⁵ <i>bja</i> ³³	<i>me</i> ³³ <i>nu</i> ³³
year	<i>hi</i> ³³ <i>pi</i> ³³	<i>xo</i> ³³ <i>la</i> ³¹
four	<i>si</i> ³¹	<i>un</i> ³¹
nine	<i>te</i> ³¹	<i>kə</i> ³¹
teeth	<i>she phe</i> ⁵⁵	<i>ʔay</i> ³³ <i>so</i> ³¹
otter	<i>ʔu</i> ⁵⁵ <i>bo</i> ³¹	<i>lay</i> ⁵⁵ <i>sam</i> ⁵⁵
star	<i>bəku</i> ⁵⁵ <i>si</i> ³¹	<i>pi</i> ³¹ <i>ku</i> ⁵⁵

Phunoi of Laos has been thought to be a rather diverse group, and the recent survey by Shintani et al 1999 has confirmed that suspicion. However, none of the studied Phongsaly locations yet studied seems strongly to resemble Công of Vietnam. While much more fieldwork study of both Phunoi and Công is in order, it appears that Công and the Phunoi recorded by Bradley are no longer very close.

3. Sila.

The Sila people of Vietnam believe their ancestors came from Laos. According to their old people, they once lived in the high mountainous areas near Mường U and Mường Lá of Phongsaly Province, Laos. Owing to harsh conditions and exploitation, seven families left about 175 years ago for Vietnam under the leadership of Hủ Chả Hoa. They first settled in Mường Tùng and then moved again several times until they reached their current homelands. A part of this story can be confirmed, as a quick comparison of Sila data from Vietnam with the Sila of Laos (Shintani et al 1999) shows—contra the results with Công—a very strong resemblance. Today the Sila—numbering about 700 people (589 in the 1989 census)—are found in three villages: Seo Hay and Xi Theo Chai (Can Hồ Commune) and Nậm Sín (Mường Nhé). According to Chazée (1999:166), the Lao Sila population call themselves *Sila* or *Sida*, our informant, Ms. Vàng Kó Um of Can Hồ Commune, about 35 years of age, informed us that she uses the Sila autonym [ko³⁵ za³³ su³¹] ‘Sila male’ or [ko³⁵ za³³ ma³¹] ‘Sila female’, though even the old

people are uncertain of its meaning, EMPV (1977:369).

3.1. Distinctive features of Sila.

The initials of Sila show five places of articulation, labial, dental, palatal, velar, and glottal with manners of articulation unaspirated, aspirated, and voiced (/d/ is not represented in our corpus).

p	pj	t	k	ʔ
ph		th	kh	
		tɕ		
		tɕh		
b	bj		g	
f	fj	ɕ	x	h
v		ʐ		
m		n	ɲ	ŋ
w		l		

Rimes in Sila are always simple, that is to say there are no syllable codas, only the nuclear vowels

i	y	u	u
e		ə	o
ɛ			ɔ
		a	

Sila has tone contour 35 corresponding to *1 of the parent language;⁵ tone contour 31 corresponding to *2, and tone contour 33 corresponding to *3. Closed-syllable word shapes CVC occur in pitch contour 31, 33, or 35 all accompanied with creaky glottal constriction at the end of the syllable, which largely disappears when followed closely by a second syllable in a compound. We indicate it with a tilde under the last digit of the tone, e.g. *va*³¹ ‘pig’.

3.2. Comparative comments.

Bradley (1997:45) assigns this language a position in Southern Loloish quite close to Hani/Akha and our data support this decision. The tone shapes of Sila are identical to those of Viet Hani. Sila, like Hani, also has strong decay of original syllable codas. There are today no nasals or stops word finally.

Like Cđông, Sila has experienced some influenced from Vietnamese. For example, Sila /ph-/ has mutated to /f-/ , as in *fe*³¹ ‘rooster’ (<*po¹), *fo*³⁵ ‘silver’ (<*plu¹), *fi*³¹ ‘day after tomorrow’ (<*prak¹), *fa*³¹ ‘leaf’ (*C-pak¹), *fi*³⁵ ‘chili pepper’ (<*C-pat¹), *fe*³¹ ‘frog’ (<*k-/pa²).

Sila is also quite a lot like Hani lexically. The similarity in the construction of compounds is especially evident in these pairs of Sila and Hani items: *phi*³⁵ *khə*³¹ *la*³¹ *lu*³¹ vs. *phi*³¹ *xə*³³ *lha*³³ *ya*³¹ ‘clothing’; *ze*³³ *ju*³⁵ *ze*³¹ vs. *za*³¹ *xjo*³³⁷ *za*³¹ ‘son’; *ze*³³ *mi*³⁵ *ze*³¹ vs.

⁵ Although the pitch trajectory rises from mid to high in this tone category, it may be because the high or 55 tone shape must begin at mid-level at syllable onset.

za³¹ mi³¹ za³¹ ‘daughter’. Although the position of the Sila language appears quite close to Hani, the people have clothing, customs, and practices that confirm a separate ethnic identity.

4. Lahu.

There are three kinds of Lahu spoken in Vietnam: Yellow Lahu, Black Lahu, and White Lahu. We were able to study only the Black Lahu of this area. The total Lahu population in 1989 was 5,319 and estimated by PV to have reached 6,600 by 2000. The Lahu have many names in Mường Te. The Black Lahu group often refer to themselves as *Khucong* or [khu³³ tsho³³]. They look down on their Yellow Lahu neighbors, calling them contemptuously [ne⁵³ tu³³] ‘Jungle Spirits’.⁶ According to the EMPV (354) the local White Thai majority term all the Lahu in Mường Te *Xá Toong Lương*, White Thai for ‘Spirits of the Yellow Banana Leaves’. Other scornful exonyms are *Xá Quỷ* meaning ‘Devil Savages’. In addition to these names the Lahu have distinctive monikers for each of their subgroups: (1) *La Hủ Sủ* or Yellow Lahu (living in the two communes Pa Vệ Sủ and Pa Ủ as well as in the villages of Là Pé, Nhu Tè, and Hóm Bô of the Ca Lăng Commune), (2) *La Hủ Nà* or Black Lahu (living in the village of Nậm Phìn, as well as Nậm Khao, Nậm Cầu, Phìn Hồ, Nậm Xá of Ca Lăng commune), and (3) *La Hủ Phung* or White Lahu (living often together with the Yellow Lahu in the villages of Xà Hồ, Ủ Ma, Pha Bù, Pa Ủ and Khô Ma of Pa Ủ commune as well as Hà Xe of Ca Lăng commune). It is reported that the Lahu originally came from the Jīnpíng area of Yúnnán Province, China.

4.1. Distinctive features of Lahu.

Since Lahu has been so exhaustively described in Matisoff 1973 and 1988 and Bradley 1978, I will dispense with sketching its features and simply note that it has the following inventory of initial consonants /p t tɕ k q ph th tɕh kh qh b dz g m n ŋ f ɕ h v ɣ l/ and vowels /i u u e ə o ɛ a ɔ/. The seven tones for Vietnam Black Lahu are 33, 35 53, 31 212, 53? and 31?.

4.2. Comparative comments.

The Lahu of Mường Te speak a language that differs some from the Black Lahu recorded in Matisoff 1988 and the *Zàngmiǎnyǔ yǔyīn hé cìhuì* 1991 in many respects. These differences seem focused mostly in the lexical domain.

One major difference is the variation of velar and uvular stops.

Gloss	Matisoff 1988	Viet Lahu	Semao Lanchang
mountain	qho ³³	qho ³³	
excrement	qhe ⁵³	khe ⁵³	
return	qo ^{31?}	qo ^{31?}	
ashes	qho ^{31?}	qha ^{31?}	
road	qo ³³	ya ^{31?} qo ³³	
village	qha ^{53?}	qha ^{53?}	
ditch	ɣu ³¹ qha ⁵³	u ³¹ kha ⁵³	ɣu ³¹ qha ⁵³
year	qho ^{31?}	qho ^{31?}	qho ^{31?}

⁶ The form *ne⁵³* in Matisoff 1988 is defined as ‘spirit, demon’.

There are also some differences in regard to voicing.

Gloss	Matisoff 1988	Viet Lahu	Semao Lanchang
walk	<i>dzu</i> ⁵³	<i>tsu</i> ⁵³	<i>dzu</i> ^{53/}
pull	<i>ɣɔ</i> ³¹	<i>gɔ</i> ³¹	<i>ɣɔ</i> ³¹
crow chicken		<i>bɔ</i> ³¹	<i>bu</i> ³¹
stand	<i>vu</i> ³⁵	<i>xu</i> ³⁵	
scrape, skin	<i>qɛ</i> ^{31/}	<i>khe</i> ^{31/}	<i>qɛ</i> ^{31/}
wash clothes	<i>tshu</i> ⁵³	<i>tshu</i> ⁵³	<i>tshɿ</i> ^{53/}
earrings	<i>na</i> ¹¹ <i>ɣɔ</i> ⁵³ <i>dzu</i> ³¹	<i>na</i> ²¹² <i>pɔ</i> ⁵³ <i>tɛu</i> ³¹	<i>na</i> ³¹ <i>ɣɔ</i> ⁵³
pillow	<i>u</i> ³⁵ <i>gɛ</i> ⁵³	<i>u</i> ³⁵ <i>ko</i> ⁵³	<i>u</i> ³⁵ <i>gɛ</i> ⁵³

Some voiced stops and affricates have become pre-nasalized, while others have devoiced entirely. Velar fricative *x* is *h*. It is probable that Viet Lahu has undergone some influence through language contact. For example, Black Lahu *p̄h̄ lâ* (“dog-tiger”) kind of small wildcat that eats pigs or dogs in Vietnam Black Lahu it is *fu*⁵³ *la*⁵³ with the meaning *chó sói nhỏ* ‘small wolf’. The change *ph* → *f* is typical kind of change effected in other minority language from Vietnamese.

Sun 1992 mentions several ways in which Kucong of Yunnan with the same autonym as the Viet Lahu) shows special features not found more widely in Lahu. These differences were not sufficient to prevent a decision on August 9, 1987 to incorporate the Kucong of Yunnan (30,000 population) into the much larger Lahu nationality. Some items of difference concern nasal codas that have developed in harmony with nasal initials at syllable onset. The Kucong data we gathered from Vietnam, however, shows mostly the Lahu proper forms and not China Kucong forms, cf. China vs. Vietnam Kucong: ‘spider’ *a*⁵⁵ *kaŋ*³³ *laŋ*³³ vs. *a(ŋ)*⁵⁵ *ka*³³ *la*³³, ‘sweet’ *t̚shaŋ*³³ vs. *t̚sho*³³, ‘sell’ *xan*³¹ vs. *ho*³¹. Also some Kucong forms in China show vowel raising, *a* → *ɤ*, e.g. China vs. Vietnam Kucong ‘bamboo’ *vɤ*⁵³ vs. *va*⁵³, ‘cloth’ *phɤ*³³ vs. *pha*³³, ‘good’ *nɤ*³³ vs. *na*³³. In all these cases, however, Viet Kucong agrees with Lahu generally and not with China Kucong. But for the other cases of vowel raising *ɔ* → *o*, then Viet Kucong agrees with China Kucong, ‘high’ *mo*³³ for both, not **mɔ*³³, ‘horse, sky, old’ *mo*³¹ not **mɔ*³¹ as well as people *t̚sho*³³ not **t̚shɔ*³³.

By and large Viet Lahu/Kucong differs only very minimally from the larger concentrations in China, Thailand, and Laos.

5. Hani (Hahni).

The Hani people of Vietnam are found in two provinces, Lai Châu at Mường Te and Phóng Thổ Districts as well as in Lào Cai Province at Bát Xát District. All the Hani people of Vietnam are thought to have migrated from Jīnpíng and Lǚchūn Counties of Yunnan Province, China. The earliest pioneering families – perhaps five or six in number – came to Vietnam about 325 years ago and settled in Lai Châu Province at Mường Te. The Phóng Thổ and Bát Xát Hani entered Vietnam much later, 175 years ago. Despite geographic distance, bad weather, and bad roads, the Hani of Vietnam say they are able to communicate with each other in the Hani language. The villages where they live are: Sín Thầu, Chung Chải, Mù Cà, Ca Lăng, and Thu Lũm (Mường Te) and Y Tí and A Lù (Bát Xát). In Mường Te (cf. map), the Hani settlement areas are separated by more than 50 km with no land or

river links and found on both banks of the Black River (Sông Đa). There are also some villages very near the China border in the far northwest of Mường Te District. Those in the Phóng Thổ District of Lai Châu Province and Bát Xát District of Lào Cai are the Black Hani or *Hà Nhì Đen*, which we were not able to study. We did, however, interview two speakers from two differing Mường Te locations, Ms. Pơ Go Sur of Bản Mù Cả of Mù Cả, age 50, member of the *Hà Nhì Cồ Chồ* clan, and Ms. Lò Mi Sô of Bản Chang Chau Pa of Xạ Hua Bun, age 42, member of the *Hà Nhì La Mí* clan. Their forms of the Hani language were quite similar but not identical. Mostly they differed lexically. The system of tones, initials, and finals was effectively the same. Both speakers agreed on the autonym [ha³³ ni⁵⁵ za³¹]. With regard to tones, their speech had three tones in smooth syllables: 55 (corresponding to *1 of the parent), 31 (corresponding to *2 of the parent), and 33 (corresponding to *3 of the parent). There were also two syllables shapes that ended in strong glottal constriction, one transcribed as [31ʔ] and the other as [33ʔ].

The initials of Vietnam Hani are:

p	pj	t	tj	k
ph	phj	th	thj	kh
b	bj	d	dj	g
		ts		tɕ
		tsh		tɕh
		dz		dʒ
m	mj	n	ɲ	ŋ
		l		
		lh		
f		s	ɕ	x
			j	

According to *Zàngmiǎnyǔ yǔyīn hé cíhuī* 1991 there are three kinds of Hani attested in China.

5.2. Comparative notes.

The Hani of Vietnam differ from the Akha described by Hansson (1989:55-89) in several important ways. The first difference involves voicing of the syllable initial. Many items of Vietnam Hani have lost voicing of the initial consonant that is found in Akha of Thailand.

Gloss	Viet Hani	Hani (Luchun)	Akha (Thai)
thunder	tɕi ³¹	dʒi ³³	dʒe ³³
copper	ku ³¹	gu ³¹	gu ³¹
star	ku ⁵⁵	gu ⁵⁵	gu ⁵⁵
fire	mi ³¹ tsa ³¹	mi ³¹	mi ³¹ dza ³¹
ear	na ³¹ pɔ ⁵⁵	bɔ ⁵⁵	na ³¹ bɔ ⁵⁵

The second difference concerns the loss of final nasal codas. In Vietnam all final velar nasals have disappeared, whereas in Akha (Thailand) one finds -aŋ and

-m.⁷

Gloss	Viet Hani	Hani (Luchun)	Akha (Thai)
hair	<i>tshe⁵⁵ kho⁵⁵</i>	<i>tshe⁵⁵ khay⁵⁵</i>	<i>tshe⁵⁵ khay⁵⁵</i>
thread	<i>sa³¹ khu⁵⁵ ɔ⁵⁵</i>	<i>sha³¹ khay⁵⁵</i>	<i>sha³¹ khay⁵⁵</i>
heavy	<i>i³³ kho³³</i>	<i>jo³³ khay³³</i>	<i>khay³³</i>
day before yesterday	<i>xu³¹ nɔ³³</i>	<i>fu31 naŋ³³</i>	<i>hu31 naŋ³³</i>
name	<i>tshɔ⁵⁵ mi⁵⁵ ɔ⁵⁵</i>	<i>tshɔ⁵⁵ mjaŋ⁵⁵</i>	<i>tscho⁵⁵ mjaŋ⁵⁵</i>

The third difference is the presence of the partially voiced or breathy lateral initial *lh-*.

Gloss	Viet Hani	Hani (Luchun)	Akha (Thai)
wind	<i>u³³ hle⁵⁵</i>	<i>li⁵⁵</i>	<i>le⁵⁵</i>
tongue	<i>hla⁵⁵ ma³³</i>	<i>hla⁵⁵</i>	<i>la⁵⁵</i>
hot	<i>hlɔ⁵⁵</i>	<i>lɔ⁵⁵</i>	<i>laŋ⁵⁵</i>
boat	<i>hlɔ³¹</i>	<i>hlu³¹</i>	
river	<i>lɔ⁵⁵ ba³¹</i>	<i>lɔ⁵⁵</i>	<i>la⁵⁵</i>

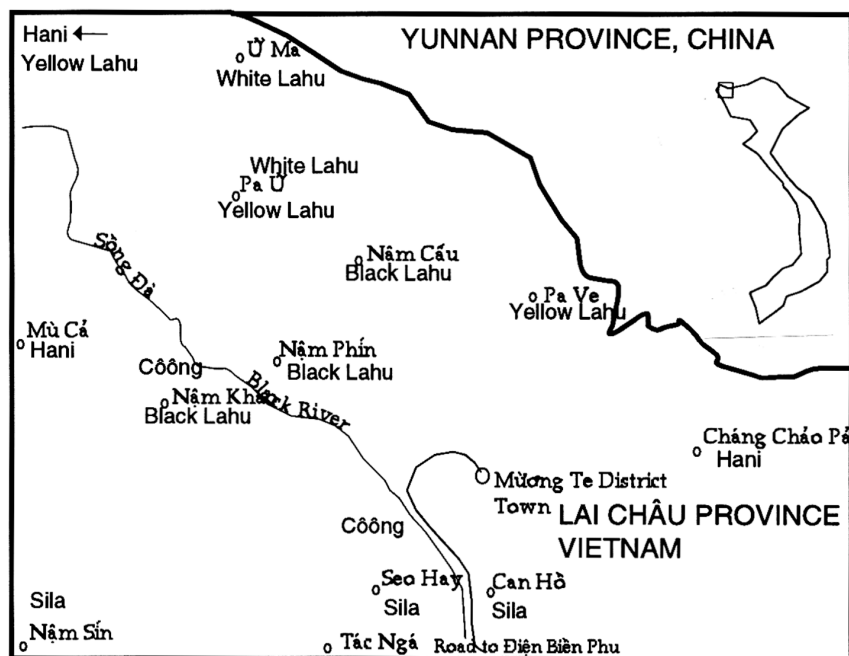
When preceded by a vowel, as in ‘wind’, this item sounds as if it were [ul³³ he⁵⁵]. This feature is found occasionally in the vocabulary of Yunnan (1959, as cited in Hansson 1989) and less in the Akha of Thailand.

Finally, it is worth noting that Southern and Central Loloish show a regular pattern of tonal development, whereby one tone value corresponds, generally, to one proto tone category. Lahu with seven tones has developed some extra splitting and thus has the most complex reflexes of the proto categories

6.0 Conclusion.

As the above sketches and comparisons have shown, the Sila, Lahu, and Hani languages of Vietnam held no real surprises. Cồông might be regarded as a bit of a surprise, as it appears to be more different from Phunoi than had been suspected. Nevertheless, it has been relatively easy to decide where to situate this group of languages within the overall Loloish pattern, unlike the languages reported on in Edmondson 2004, northern Loloish of Vietnam, which were and remain quite challenging in regard to the details of their genetic affiliation. In these southern and central Loloish languages it was important to obtain certainty that the familiar names did not conceal any unfamiliar linguistic entities. We can now confirm that Lahu and Hani are what we supposed, Sila is quite close to Hani, and Cồông is perhaps not as close to Phunoi as was thought.

⁷ I am, of course, not referring to the glottal constriction residue of original stops, which is still present in Vietnam.



Map of Mường Te District, Lai Chau Province, Vietnam

References

- Benedict, Paul K. 1972. Contributing editor James A. Matisoff. *Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.
- Bradley, David. 1978. *Lahu dialects*. Canberra.
- _____. 1979. *Proto-Loloish*. Scandinavian Institute of Asian Studies Monograph Series, No.39. London: Curzon Press Ltd.
- _____. 1996. Tibeto-Burman languages of PDR Lao. *LTBA* 19:1.19-27.
- _____. 1997. *Tibeto-Burman languages and classification in Papers in Southeast Asian linguistics, no.14: Tibeto-Burman languages of the Himalayas*, ed. by David Bradley, 1-72. Pacific Linguistics, Series A-86. Canberra: The Australian National University.
- _____. 1999. Language endangerment in China: The Yi of Kunming. *La Trobe Papers in Linguistics* 10, 1-15, La Trobe University, Department of Linguistics.
- Chazee, Laurent. 1999. *The Peoples of Laos: Rural and Ethnic Diversities*. Bangkok: White Lotus Press.
- DEMG. 1975. *Về vấn đề xác định thành phần các dân tộc thiểu số ở miền bắc Việt Nam*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội. [On determining the groupings of ethnic minorities of northern Vietnam].
- Đỗ Phương. 1996. *Việt Nam hình ảnh cộng đồng 54 dân tộc*. Hà Nội: Nhà Xuất Bản Văn Hóa Dân Tộc. [Vietnam images of the community of 54 ethnic groups. Hanoi: The Ethnic Cultures Publishing House.]
- Edmondson, Jerold A. 2002. Three Tibeto-Burman languages of Vietnam. In Bradley and Thurgood (eds.) *Festschrift for James A. Matisoff*. Pacific Linguistics Series. Canberra: Australian National University.
- EMPV. 1978. *Các dân tộc ít người ở Việt Nam (Các tỉnh phía Bắc)* [Ethnic minority peoples of Vietnam (the northern provinces)]. Abbreviated reference: EMPV.
- Hansson, Inga-Lille. 1989. A comparison of Akha, Hani, Khatu, and Pijo. *LTBA* 12:1.6-91.
- Haqniq pyulniul soqmiav niq pyu hu zzaol e nilgevnei soqhavq/Hāni-Hàn suizhào xiào cihuī*. 1959. Kūnmíng.
- Harris, Jimmy G. 1970. *Field recordings of Phunoi*. Vientiane, Laos.
- Hemmet, Christine. 1995. *Montagnards des Pays d'Indochine*. Paris: Édition Sépia.
- Lâm Tâm. 1985. Sơ lược giới thiệu người Cò Xung. *Tập chí Dân tộc học* 14:29.
- Lajonquière, Lunet de. 1906. *Ethnographie du Tonkin septentrional, redigée sur l'ordre de M. P. Beau*. Paris: E. Leroux.
- Lê Quý Đôn. 1962. *Kiến văn Tiểu Lục*. Hà Nội. [Things seen and heard, first published in the mid 1700's].
- LeFevre-Pontalis. P, 1892. Notes sur quelques populations du nord de l'Indochine. *Journal Asiatique* 8 (séries) 19:237-69
- Lǐ Yǒngsui and Wǎng Ērsōng. 1981. *Hānyǔ Jiǎnzhi*. Běijīng: Mǐnzú chūbǎnshè.
- Lục Bình Thủy and Nông Trung. 1975. Người Phù Lá ở Lào Cai. *Tập chí Dân tộc học* 1:80.
- Mai Thanh Sơn. 1995. Một vài tư liệu về lịch sử người Phù Lá (qua việc tìm hệ tộc danh). *Tập chí Dân tộc học* 3:58-64. [Some materials about the history of the

- Phù Lá people (work based on the perspective of their names.)).
- _____. 1998. Đôi nét về tập quán ăn uống của người Phù Lá. *Tạp chí Dân tộc học* 1:61-65. [Two features in the customs of eating and drinking of the Phù Lá people.]
- Matisoff, James A. 1972. *The Loloish tonal split revisited*. Research monograph no.7. Berkeley: Center for South and Southeast Asia Studies.
- _____. 1973. *The grammar of Lahu*. Berkeley: University of California Press.
- _____. 1988. *The dictionary of Lahu*. Berkeley, Los Angeles, and London: University of California Press.
- _____. 1991. Sino-Tibetan Linguistics: Present state and future prospects. *Annual Review of Anthropology*, 20:469-504. Palo Alto: Annual Reviews Inc.
- Nguyễn Ngọc Tuấn. 1974. Vài vấn đề về người La Hủ ở Việt Nam. *Tạp chí Dân tộc học* 3:71-72. [On questions regarding the Lahu nationality in Vietnam.]
- Nguyễn Văn Huy. 1975. Bước đầu tìm hiểu mối quan hệ tộc người giữa hai nhóm Phù Lá và «Xà Phó». [A first step to understanding the differences between the Xa Pho and other Pho La groups.] In *DEMG*, 415-28.
- _____. 1985. Văn hóa và nếp sống Hà Nhi, Lô Lô. Hà Nội: Nhà Xuất Bản Văn Hóa. [Culture and life style of the Hani and Lolo.]
- PV 1998. *The peoples of Vietnam*. Asian Minorities Outreach. Darawan Printing: Chiang Mai, Thailand.
- Qiu, Fuyuan. 1998. A phonetic and phonological overview of the Yi (Lolo) language. MA thesis, The University of Texas at Arlington.
- SDSRV 1991. Tổng Cục Thống Kê “General Office of Statistics”. *Số liệu thống kê Cộng hòa XHCN Việt Nam* [Statistical data of the Socialist Republic of Vietnam 1976-1990]. Nhà Xuất bản Thống kê “Statistical Publishing House”. Hà Nội. Abbreviated reference: SDSRV 1991.
- Shintani, Tadahiko, et al. 1999. *Basic vocabularies of the languages spoken in Phongsaly, Lao P.D.R.* Tokyo: Institute for the Studies of Languages and Cultures of Asia and Africa.
- Sun Jianyi. 1992. Lăhùyǔ Kūcōnghuà de ruògān tèdiǎn. *Mínzú Yǔwén* 5.68-73. [Some features of Kucong speech of the Lahu language.]
- Thurgood, Graham. 1982. Subgrouping on the basis of shared phonological innovations: a Lolo-Burmese case study. *BLS* 8:251-60.
- VNA. 1996. Thông Tấn Xã Việt Nam. [Vietnamese News Agency.] *Việt Nam: Hình ảnh cộng đồng 54 dân tộc*. [Vietnam: Image of the community of 54 ethnic groups.] Hà Nội: Nhà Xuất Bản Văn Hóa Dân Tộc. [The Ethnic Cultures Publishing House.] [Abbreviated reference: VNA.]
- Zàngmiǎnyǔ yǔyīn hé cǐhuī. 1991. Běijīng: Zhōngguó Shèhuì Kēxué Chūbǎnshè.

Jerold A. Edmondson
Program in Linguistics
University of Texas at Arlington
Arlington, TX 76019-0559

edmondson@uta.edu

Tone and Syllable Structure in Hakha-Lai

LARRY M. HYMAN & KENNETH VANBIK

University of California, Berkeley

The purpose of this paper is to present an analysis of the tone system of Hakha-Lai, a Tibeto-Burman language of the Kuki-Chin subgroup, spoken in Chin State, Burma, and parts of Mizoram State, India. After establishing the underlying tonal representations, we turn to examine the various tone sandhi which account for their realization in different contexts. In so doing, we shall be particularly interested in the relation between tone and syllable type, specifically which syllable structures allow contour tones.¹

The different syllable structures of (largely monosyllabic) Hakha-Lai words are schematized in (1).

(1) a. “Smooth” syllables

CVV V = /i, e, u, o, a/

CVD D = sonorant, i.e. /m, n, ŋ, l, r, y, w/

CVVD D = sonorant, i.e. /m, n, ŋ, l, r, y, w/

b. “Checked” syllables

CVT T = obstruent, i.e. voiceless stop /p, t, k/ or glottalized
sonorant /m', n', ŋ', l', r', y', w'/

CVVT T = voiceless stop /p, t, k/ (but not glottalized sonorants)

c. “Reduced” syllable (grammatical proclitics or derived via compounding)

CV e.g. sg. pronominal proclitics (*ka* ‘my’ in (3), N1 in (5))

¹This is a shortened version of the paper presented at BLS and in the Séminaire Tibéto-Burmane, at Université de Paris III, February 5, 2002. We are grateful for helpful comments received from interested persons at both events, especially John Ohala and David Peterson. Previous work on Hakha-Lai includes Kathol & VanBik (2001), Melnit (1997a,b), Olawsky & VanBik (2000), Patent (1997), Peterson (1998), VanBik (2001) and VanBik & Roengpitya (2001).

As seen, Hakha-Lai syllables require an onset and can be open or closed. Coda consonants can be obstruents (T), either voiceless stops or glottalized sonorants, or plain sonorants (D). Underlying length is contrastive only in syllables closed by a sonorant or a voiceless stop, and vowels are short before a glottalized sonorant coda.

As seen in (2), smooth-syllable words carry one of two tones in isolation: a falling (F) tone from a high to a low pitch [31] or a level (L) tone on a relatively low pitch [22]:

(2) Tones of smooth syllables in isolation

		CVV		CVD		CVVD	
a.	F	<i>hmaà</i> ‘wound’ <i>zuù</i> ‘beer’		<i>lùŋ</i> ‘heart’ <i>lòw</i> ‘field’		<i>tlaàŋ</i> ‘mountain’ <i>raàl</i> ‘enemy’	
b.	F	<i>oò</i> ‘voice’ <i>keè</i> ‘leg’		<i>hròm</i> ‘throat’ <i>tsàl</i> ‘forehead’		<i>koòy</i> ‘friend’ <i>tsaàn</i> ‘time’	
c.	L	<i>saa</i> ‘animal’ <i>hnii</i> ‘skirt’		<i>raŋ</i> ‘horse’ <i>kal</i> ‘kidney’		<i>koom</i> ‘corn’ <i>boor</i> ‘bunch’	

However, when preceded by a singular pronominal proclitic, e.g. *ka* ‘my’, the falling tone nouns in (2b) are instead realized with a mid-to-high [23] rising tone, as seen in (3).

(3) Tones of smooth syllables preceded by proclitic *ka*= ‘my’

		CVV		CVD		CVVD	
a.	F	<i>ka hmaà</i> ‘my wound’ <i>ka zuù</i> ‘my beer’		<i>ka lùŋ</i> ‘my heart’ <i>ka lòw</i> ‘my field’		<i>ka tlaàŋ</i> ‘my mtn.’ <i>ka raàl</i> ‘my enemy’	
b.	R	<i>ka oó</i> ‘my voice’ <i>ka keé</i> ‘my leg’		<i>ka hróm</i> ‘my throat’ <i>ka tsál</i> ‘my forehead’		<i>ka koóy</i> ‘my friend’ <i>ka tsaàn</i> ‘my time’	
c.	L	<i>ka saa</i> ‘my anim.’ <i>ka hnii</i> ‘my skirt’		<i>ka raŋ</i> ‘my horse’ <i>ka kal</i> ‘my kidney’		<i>ka koom</i> ‘my corn’ <i>ka boor</i> ‘my bunch’	

Our proposal is that there are three underlying tones in Hakha-Lai, falling (‘), rising (‘), and level low (unmarked), which we shall refer to as F, R, and L. In addition, as formalized in (4), there is a postlexical rule which changes a R tone to F in phrase-initial position:

(4) Initial Falling Rule (IFR)

$$\begin{array}{c} \phi [\sigma \\ | \\ R \rightarrow F \end{array}$$

Because of the preceding *ka*, the /R/ of nouns in (3b) does not undergo rule (4).

Now consider the N1-N2 noun compounds in (5).

(5) 3 x 3 tone patterns plotted in N1- N2 compounds (N1 = reduced)

	F			R			L
a.	F	<i>hna</i>	<i>hmaà</i>	<i>hna</i>	<i>oó</i>	<i>hna</i>	<i>hnii</i>
b.	R	<i>ke</i>	<i>hmaà</i>	<i>ke</i>	<i>oó</i>	<i>ke</i>	<i>hnii</i>
c.	L	<i>sa</i>	<i>hmaà</i>	<i>sa</i>	<i>oó</i>	<i>sa</i>	<i>hnii</i>

(*hnaà* + *hmaà* ‘ear wound’, *keé* + *hmaà* ‘leg wound’, *saa* + *hmaà* ‘animal’s wound’, etc.)

In these forms we observe that when CVV → CV as the N1 of a N1-N2 possessive/compound, its tone is deleted and therefore has no effect on N2. (Its vowel is pronounced on a mid-to-high pitch.) We interpret this as indicating that a syllable must have two moras to be a tone-bearing unit, i.e. to carry F, R or L tone.

Compounds whose N1 ends in a coda consonant do not undergo such reduction. When both N1 and N2 are full syllables, tone changes affect those nouns which are boxed in (6).

(6) 3 x 3 tone patterns plotted in N1-N2 compounds (N1 ≠ reduced)

	F			R			L
a.	F	<i>tlaàŋ</i>	zuu	<i>tlaàŋ</i>	<i>tsaán</i>	<i>tlaàŋ</i>	<i>saa</i>
b.	R	<i>thlaán</i>	<i>zuù</i>	<i>thlaán</i>	tsaàn	thlaan	<i>saa</i>
c.	Ø	<i>koom</i>	zuu	<i>koom</i>	<i>tsaán</i>	<i>koom</i>	<i>saa</i>
‘my’ +		‘mountain beer’		‘mountain time’		‘mountain animal’	
		‘grave beer’		‘grave time’		‘grave animal’	
		‘corn beer’		‘corn time’		‘corn animal’	

The above forms indicate the tones with which they are realized after a singular proclitic such as *ka* ‘my’ so that IFR will not apply to the initial R tone in (6b).

As seen, F alternates with L tone. Phrase-internally, an underlying /F/ will be realized F in the three contexts in (7).

(7) a. after a /R/ which is realized R

ka + *thlaán* + *zuù* → *ka thlaán zuù* ‘my grave beer’
ka + *koóy* + *lùŋ* → *ka koóy lùŋ* ‘my friend’s heart’

- b. after a /R/ which is realized F by IFR (4)
- | | | | | |
|----------------------------|---|--------------------------|------------------|------------------|
| <i>thlaán</i> + <i>zuù</i> | → | <i>thlaàn</i> <i>zuù</i> | ‘grave beer’ | (i.e. R-F → F-F, |
| <i>koóy</i> + <i>lùŋ</i> | → | <i>koòy</i> <i>lùŋ</i> | ‘friend’s heart’ | φ-initially) |
- c. after a reduced syllable (toneless CV)
- | | | | |
|---------------------------|---|------------------------|----------------|
| <i>ka</i> + <i>zuù</i> | → | <i>ka</i> <i>zuù</i> | ‘my beer’ |
| <i>hnaà</i> + <i>hmaà</i> | → | <i>hna</i> <i>hmaà</i> | ‘ear wound’ |
| <i>saa</i> + <i>hmaà</i> | → | <i>sa</i> <i>hmaà</i> | ‘animal wound’ |

On the other hand, a F tone is simplified to L in the two environments in (8).

- (8) a. after a full syllable with F or L tone
- | | | | |
|---------------------------|---|-------------------------|-----------------|
| <i>tlaàŋ</i> + <i>zuù</i> | → | <i>tlaàŋ</i> <i>zuu</i> | ‘mountain beer’ |
| <i>koom</i> + <i>zuù</i> | → | <i>koom</i> <i>zuu</i> | ‘corn beer’ |
- b. after two (or more) reduced CV syllables
- | | | | |
|---------------------------------------|---|----------------------------------|-------------------|
| <i>ka</i> + <i>hnaà</i> + <i>hmaà</i> | → | <i>ka</i> <i>hna</i> <i>hmaa</i> | ‘my ear wound’ |
| <i>ka</i> + <i>saa</i> + <i>hmaà</i> | → | <i>ka</i> <i>sa</i> <i>hmaa</i> | ‘my animal wound’ |

As (9a) shows, the F simplification rule (FSR) may affect more than one input F:

- (9) a. *kàn* + *tlaàŋ* + *zuù* → *kàn* *tlaaŋ* *zuu* ‘our mountain beer’
raàl + *lòw* + *hmaà* → *raàl* *low* *hmaa* ‘enemy field time’
- b. *ka* + *raŋ* + *hnaà* + *hmaà* → *ka* *raŋ* *hna* *hmaà* ‘my horse’s ear wound’

The example in (9b) shows, however, that even phrase-internally, a F will not be simplified if it is preceded by exactly one reduced CV syllable.

Our analysis is to group syllables into (largely iambic) tonal feet (f) within the phonological phrase (φ), according to the algorithm in (10).

- (10) a. each full syllable must be in a separate foot, e.g.
- | | | | |
|---------------------------|---|--|-----------------|
| <i>tlaàŋ</i> + <i>zuù</i> | → | [[<i>tlaàŋ</i>] _f [<i>zuu</i>] _f] _φ | ‘mountain beer’ |
| F F | | F L | |
| <i>koom</i> + <i>zuù</i> | → | [[<i>koom</i>] _f [<i>zuu</i>] _f] _φ | ‘corn beer’ |
| L F | | L L | |
- b. a sequence of two or more CV syllables will group together as a foot
- | | | | |
|--------------------------------------|---|--|-------------------|
| <i>ka</i> + <i>saa</i> + <i>hmaà</i> | → | [[<i>ka</i> <i>sa</i>] _f [<i>hmaa</i>] _f] _φ | ‘my animal wound’ |
| L F | | L | |

- c. otherwise, a single reduced CV syllable foots with a following full syllable, ϕ -initially...

$ka + zu\grave{u}$ → [[$ka\ zu\grave{u}$]_f]_φ 'my beer' (poss + N)
F F

saa + hmaà → [[*sa hmaà*]_F]_φ 'animal wound' (N1-N2)

- d. as well as ϕ -internally

$$\begin{array}{ccc} ka + ray + hna\grave{a} + hma\grave{a} & \rightarrow & [[ka ray]_f [\underline{hna} hma\grave{a}]_f]_\phi \\ \text{F} & & \text{F} \end{array}$$
 ‘my horse’s ear wound’

Assuming the footing structure in (10), the following generalization emerges: If not preceded by a R, a F will be deleted in a non- \emptyset -initial monosyllabic foot, as indicated in the formulation of the FSR in (11).

(11) F-Simplification Rule:
$$\frac{[\dots [\dots]_f \quad [\sigma]_f \quad \dots]_\phi}{F \rightarrow L} \text{ (FSR)}$$

The FSR also predicts that the F of the isolation form [[hmaà]_f]_φ ‘wound’ will surface because the indicated monosyllabic foot is φ-initial. Since the rule in (11) depends on footing, and since footing is based on the distinction between full vs. reduced syllables, FSR will be sensitive to how the vowel shortening rule applies. As seen in (16), strings of multiple CVV syllables show some variation:

(12) a. $\underset{\text{L}}{ka} + \underset{\text{R}}{saa} + \underset{\text{F}}{ke'è} + hma\grave{a} \rightarrow [[\underset{\text{L}}{ka saa}]_{\text{f}} [\underset{\text{F}}{ke hma\grave{a}}]_{\text{f}}]_{\phi}$

b. *ka + saa + keé + hmaà* → [[*ka sa ke*]_f [*hmaa*]_f]_ϕ
 L R F L

Both output forms mean ‘my animal’s leg wound’. In (12a), vowel-shortening applies only to *keé*. Two bisyllabic feet are thus constructed. Since the F of *hmaà* is protected by the reduced syllable *ke* in the second foot, FSR does not apply. In (12b), however, vowel-shortening applies to both *saa* and *keé*. As seen, this produces an initial foot consisting of three CV syllables. FSR therefore applies.

Whereas /f/ is affected by one rule (FSR), /r/ is affected by three rules. The first is IFR, already seen in (4): $\phi[R \rightarrow \phi F]$. A second rule is illustrated in (13).

(13) a. $ka + thlaán + tsaán \rightarrow ka\ thlaán\ tsaàn$ 'my grave time'
 R R R F

- b. $ka + koóy + hróm \rightarrow ka\ koóy\ hròm$ ‘my friend’s throat’
 R R R F

As seen, an input sequence /R-R/ is realized as R-F, a case of a contour tone appearing to obey the OCP. This dissimilatory rule is formulated in (14).

- (14) R-R Rule (RRR): $\begin{array}{c} \sigma \\ | \\ R \end{array} \begin{array}{c} \sigma \\ | \\ R \end{array} \rightarrow F$

The derivations in (15) show that, if ordered, RRR would have to precede IFR, which counterbleeds it:

- (15)

	R-R Rule	Initial RF Rule	
a. $thlaán + tsaán \rightarrow thlaán\ tsaàn \rightarrow thlaàn\ tsaàn$ ‘grave time’	R R R F F F		
b. $koóy + hróm \rightarrow koóy\ hròm \rightarrow koòy\ hròm$ ‘friend’s throat’	R R R F F F		

In addition, as seen in (16), RRR applies iteratively (from right to left), each F beginning at a lower level, hence an automatic downstepping effect:

- (16) a. $ka + tlaán + zaán + tsaán \rightarrow ka\ tlaán\ zaàn\ tsaàn$ ‘my grave night time’
 R R R R F F
- b. $tlaán + zaán + tsaán \rightarrow tlaàn\ zaàn\ tsaàn$ ‘grave night time’
 R R R F F F

(16a) shows *zaán* and *tsaán* both acquiring F tone in post-R position. The same happens in (16b), although *tlaán* then undergoes IFR to become itself a F tone.

Note in this context that IFR renders both FSR and RRR opaque. Recall that FSR changes an input F-F to F-L, as in (17a).

- (17) a. $\phi[F-F \rightarrow F-L$ e.g. $tlaàŋ + zuù \rightarrow tlaàŋ\ zuu$ ‘mountain beer’
 F F F L
- b. $\phi[R-F \rightarrow F-F$ e.g. $thlaán + zuù \rightarrow thlaàn\ zuù$ ‘grave beer’
 R F F F
- c. $\phi[R-R \rightarrow F-F$ e.g. $thlaán + tsaán \rightarrow thlaàn\ tsaà$ ‘grave time’
 R R F F

(17b) shows that IFS counterfeeds FSR, since the derived F does not condition the simplification of the following F. (17c) shows that IFS counterbleeds RRR, since the derived initial F does not prevent the following R from becoming F. There are at least two ways of capturing the non-interaction between the three rules. First, in a derivational approach, we could order the rules: FSR \supset RRR \supset IFR. On the other hand, in a two-level unificational approach, we could adopt a simultaneous input-output implementation of the three “rules”.

A fourth and last rule that affects tone in Hakha-Lai is the R-Simplification Rule (RSR), which, as seen in (18), converts input /R-L/ to L-L:

- (18) a. $\begin{array}{ccc} ka + ko\acute{o}y + koom \\ R \quad L \end{array} \rightarrow \begin{array}{ccc} ka \ kooy \ koom \\ L \quad L \end{array}$ ‘my friend’s corn’
 b. $\begin{array}{ccc} k\grave{a}n + ko\acute{o}y + koom \\ F \quad R \quad L \end{array} \rightarrow \begin{array}{ccc} k\grave{a}n \ kooy \ koom \\ F \quad L \quad L \end{array}$ ‘our friend’s corn’

This is shown after toneless *ka* ‘my’ in (18a) and F tone *kàn* ‘our’ in (18b), both of which otherwise permit a following R.

The phrases in (19) now show that when a R meets both a left condition that would convert it to F, and the right condition that would convert it to L, it is always realized as F:

- (19) a. $\begin{array}{ccc} ko\acute{o}y + thla\grave{a}n + saa \\ R \quad R \quad L \end{array} \rightarrow \begin{array}{ccc} ko\grave{o}y \ thla\grave{a}n \ saa \\ F \quad F \quad L \end{array}$ ‘friend’s grave animal’
 (not *F--L--L)
 b. $\begin{array}{ccc} ko\acute{o}y + saa \\ R \quad L \end{array} \rightarrow \begin{array}{ccc} ko\grave{o}y \ saa \\ F \quad L \end{array}$ ‘friend’s animal’
 (not *L-L)

(19a) shows that RRR takes precedence over RSR, while (19b) shows that IFR takes precedence over RSR. These can easily be incorporated into a rule ordering account by ordering RSR last: FSR \supset RRR \supset IFR \supset RSR. A non-derivational input-output account requires something further to guarantee that we do not generate **ko\grave{o}y thlaan saa* and **kooy saa*. One idea might be to scan the above forms in a left-to-right fashion. However, we saw earlier in (16) that strings must be scanned right-to-left for the purpose of RRR. A more promising approach would be to invoke constraint ranking: Given the choice of a change R \rightarrow F vs. R \rightarrow L, the former has the advantage of preserving both components of the R contour tone. That is, it is preferable to re-sequence (“metathesize”) the tonal

gestures, $\{lh\} \rightarrow \{hl\}$, rather than to lose one, $\{lh\} \rightarrow \{l\}$?² However, why doesn't FSR change F to R, rather than L?³

At this point, let us consider the following generalizations concerning tone sandhi in Hakha-Lai:

- (20) a. F can be deleted
 b. R can be changed to F or L
 c. L never changes (never becomes a contour tone)

These generalizations directly reflect what has generally been accepted in work on tone, namely that rising tones are more complex than falling tones, which are more complex than level tones (Ohala 1978:30-1). Or, in terms of constraints: $*R \gg *F \gg *L$. The modifications in (20) thus convert more complex tones into less complex tones. The reverse is not found: F does not ever become R, and L does not ever become F or R.⁴ This is a reassuring result, given the next issue.

Two other universal expectations concern the remaining “checked” or “stopped” syllables, CVT and CVVT, not yet treated. First, CVT should license fewer tonal oppositions than smooth syllables (CVV, CVD, CVVD) or long stopped syllables (CVVT). Second, CVT should disprefer (or disallow) contour tones (F, R) (see Zhang 2001 and references cited therein). As shown in (21), neither CVVT or CVT allow an underlying tonal opposition:

(21) In Hakha-Lai, neither CVVT nor CVT allows an underlying tonal opposition

- a. CVVT, where T = voiceless stop (i.e. /p, t, k/)

<i>tseep</i>	‘bug’	<i>liit</i>	‘leech’	<i>hnaak</i>	‘rib’
--------------	-------	-------------	---------	--------------	-------

- b. CVT, where T = voiceless stop, glottal stop or glottalized sonorant

<i>kep</i>	‘button’	<i>mit</i>	‘eye’	<i>vok</i>	‘pig’
<i>tsop</i>	‘chisel’	<i>kut</i>	‘hand’	<i>ru?</i>	‘bone’

Except for some derived verb forms (see Hyman & VanBik 2002), all CVVT words carry L tone, i.e. they are realized on a long, relatively low level pitch.

²In the oral presentation, we treated the level (L) tone as unmarked (\emptyset), in which case, the choice would be between $R \rightarrow F$ vs. $R \rightarrow \emptyset$. MAX(T) might then be evoked on the contour as a unit. This is still a possible analysis, which we are examining in another paper in preparation.

³Similarly, unless the change of R-L to L-L is an assimilation (rather than a contour simplification), we have no explanation as to why R-L doesn't become F-L, where F would preserve both pitch levels of the input R.

⁴This pertains to phonological rules only. Hyman & Bik (2002) show that stem2 formation frequently consists of a morphological replacement of stem1 F or L by R.

Although they do not themselves alternate (since /L/ never becomes F or R), CVVT does condition RSR, as seen in (22).

- (22) $ka + kooy + tseep \rightarrow ka\ kooy\ tseep$ 'my friend's bug'
 R L L L

Since CVVT has a long vocalic nucleus, it is surprising both that there are no underlying tonal oppositions, and that the one tone that underlying CVVT morphemes carry is L.

The situation concerning CVT is even more intriguing. In isolation, CVT words are pronounced on a very short high falling tone. Given the tonal properties we have established above, it is clear that all CVT syllables are underlyingly /R/. As /R/ fails to do in general, (23a) shows that CVT does not condition FSR on the following syllable:

- (23) a. $\begin{array}{cc} \text{mit} + \text{hma}\grave{\text{a}} & \rightarrow & \text{mit hma}\grave{\text{a}} \\ \text{R} & \text{F} & \text{F} \quad \text{F} \end{array}$ 'eye wound'
- b. $\begin{array}{cc} \text{vok} + \text{ko}\acute{\text{o}}\text{y} & \rightarrow & \text{vok ko}\acute{\text{o}}\text{y} \\ \text{R} & \text{R} & \text{F} \end{array}$ 'pig's friend'
- c. $\begin{array}{cc} \text{ka} + \text{ko}\acute{\text{o}}\text{y} + \text{mit} & \rightarrow & \text{ka ko}\acute{\text{o}}\text{y mit} \\ \text{R} & \text{R} & \text{R} \quad \text{F} \end{array}$ 'my friend's eye'

On the other hand, (23b) shows that CVT conditions RRR, as /R/ generally does. Finally, as seen in (23c), CVT undergoes RRR itself.

Although there is no underlying tonal opposition on /CVT/, (24) shows that there is a contrast on the surface:

- (24) a. $\begin{matrix} ra\grave{a}l + n\acute{i}l \\ F \quad R \end{matrix} \rightarrow \begin{matrix} ra\grave{a}l \ n\acute{i}l \\ F \quad R \end{matrix}$ 'enemy + erg.'
- b. $\begin{matrix} ko\acute{o}y + n\acute{i}l \\ R \quad R \end{matrix} \rightarrow \begin{matrix} ko\acute{o}y \ ni\grave{l} \\ F \quad F \end{matrix}$ 'friend + erg.'

In (24a), the ergative marker /niʔ/ is realized on a high (non-falling) pitch. This is as we would expect if the output tone were R, with the beginning part of the contour clipped because of the shortness of the vowel. This realization contrasts with (24b), where /niʔ/ is realized with a falling pitch—which has been downstepped from the level of the preceding F. Whereas the F tones of the first word in the two examples are identical, the two realizations of /niʔ/ are strikingly different, much higher in (24a) than in (24b). The lower pitch of what we have marked as a falling CVT syllable is even more noticeable in cases where more

than on such CVT syllable occurs in sequence, e.g. *koáy vòk nì?* ‘friend’s pig + erg.’.

As seen, CVT syllables have the same behavior as smooth syllables with /R/ tone. There is one rule, however, which applies specifically to the /R/ of CVT syllables. (25a) shows that phrase-initial /vók/ conditions RRR on /tsaán/, while (25b) shows the same conditioning when /vók/ is immediately preceded by a full syllable (foot).

- (25) a. *vók + tsaán* → *vòk tsaàn* ‘pig time’
 R R F F
- b. *koáy + vók + tsaán* → *koáy vòk tsaàn* ‘friend’s pig time’
 R R R F F F
- c. *ka + vók* → *ka vòk* ‘my pig’
 R F
- d. *ka + vók + tsaán* → *ka vok tsaán* ‘my pig time’
 R R F R

However, when preceded by a reduced CV syllable, the /R/ of a CVT syllable is changed to F. We not only hear this change in (25c), but also observed in (25d) that *vòk* now does not condition RRR on *tsaán*. Since CV syllables do not generally convert /R/ to F, e.g. *ka koáy* ‘my friend’ (not **ka koáy*), the proposed rule in (26) must make specific reference to CVT syllables:

- (26) CVT-R Rule: [CV - CVT]
 |
 R → F

Since *tsaán* is realized with R tone in (25d), (26) must be ordered before RRR.

The above completes our outline study of the Hakha-Lai tone system. Although we have illustrated the various rules citing noun compounds and possessives, the same tone rules apply more generally within the noun phrase, as in (27), and in the verb phrase, as in (28).

- (27) a. *ka + koáy + heé* → *ka koáy heè* ‘with my friend’
 R R R F
 (cf. *ka tseep heé* ‘with my bug’)
- b. *hii + raàl + hií* → *hii raal hií* ‘this enemy’
 L F R L L R
 (cf. *raàl* ‘enemy’)

- (28) a. $\begin{matrix} \text{kàn} + \text{raà} \\ \text{F} \quad \text{F} \end{matrix} \rightarrow \begin{matrix} \text{kàn} \text{ raà} \\ \text{F} \quad \text{L} \end{matrix}$ ‘we come’
(cf. *ka raà* ‘I come’)
- b. $\begin{matrix} a + \text{baá} + \text{lów} \\ \text{R} \quad \text{R} \end{matrix} \rightarrow \begin{matrix} a \text{ baá} \text{ lòw} \\ \text{R} \quad \text{F} \end{matrix}$ ‘he didn’t tire’
(cf. *a raà lów* ‘he didn’t come’)
- c. $\begin{matrix} a + \text{ka} + \text{hnám} + \text{mii} \\ \text{R} \quad \text{L} \end{matrix} \rightarrow \begin{matrix} a \text{ ka} \text{ hnam} \text{ mii} \\ \text{L} \quad \text{L} \end{matrix}$ ‘the one who kissed me’
(cf. *a ka hnám* ‘he kissed me’)

As Hyman & VanBik (2002) show, tone is also implicated in the morphological process of stem2 formation in verbs: First, all stem2 verbs carry R tone except for the L of CVVT forms whose stem1 is CVV. In addition, tonal differences of stem1 largely predict segmental differences in stem2 formation.

The last issue on which we are currently working, but should like to briefly consider here concerns the featural interpretation of what we have identified as F, R and L tones. We have considered the possibility that L is actually the unmarked (Ø) tone in the language, and will simply set aside this issue to consider the following three interpretations of F and R.

The first interpretation is that F and R are unitary contour tones. The main argument for this is that they seem to be treated as units by some of the rules, e.g. IFR and RR, both of which convert /R/ to [F]. The major problem with this analysis is that it does reveal anything about the tone rules or distributions: Why should the changes indicated in (29) take place? (Cells left blank do not undergo modification.)

(29)		F	R	L
	F	F-L		
	R		R-F	L-L
	L	L-L		

In addition, why should all CVT syllables carry underlying /R/, i.e. the most marked and unexpected tone on such a short vowel nucleus (Zhang 2001)?

To remedy these problems in part, we might consider reinterpreting F as /L/ and R as /H/, where H = high tone, and L = low tone. (Our L tone would then have to be analyzed as Ø, receiving a L pitch by default.) In this case we could say that CVT syllables carry /H/, which has been identified as intrinsically shorter than /L/ in other tone languages. Whereas it is puzzling why IFR should forbid a phrase from beginning with a R tone, it is natural to prohibit initial H tone. As far as the other tone sandhi are concerned, summarized in the table in (30), RRR

would be reinterpreted as $H-H \rightarrow H-L$, i.e. equivalent to Meussen's Rule in Bantu and more transparently related to the OCP:

(30)

	L	H	Ø
L	L-Ø		
H		H-L	Ø-Ø
Ø	Ø-Ø		

The other rules would not necessarily fare any better than in the F/R account: If $F = /L/$, why should FSR change L-L and Ø-L to L-Ø and Ø-Ø, respectively? Similarly, if $R = /H/$, why should RSR change H-Ø to Ø-Ø?

We don't have answers to all of these questions, but now turn to a third interpretation, where $F = HL$ and $R = LH$.⁵ While not explaining why LH is prohibited initially in a phrase, but it does permit a major generalization with respect to the tone sandhi.⁶ With contours represented as sequences of high and low levels, the tone changes would now be expressed as in (31).

(31)

	HL	LH	L
HL	HL-L		
LH		LH-HL	L-L
L	L-L		

Compare the two sets of input sequences in (32a,b).

(32)

a.	LH-HL	b.	HL-HL	c.	\rightarrow HL-L
	HL-LH		LH-LH		\rightarrow LH-HL
	L-LH		LH-L		\rightarrow L-L
	HL-L		L-HL		\rightarrow L-L
	L-L				

The sequences in (32a) do not change, whereas those in (32b) do. What is the difference? A close examination reveals that in (32a) the second syllable begins on the same pitch level with which the first syllable ends. In (32b), the initial pitch of the second syllable is opposite to the end pitch of the first syllable. When the sequences in (32b) are modified to those in (32c), the result is like (32a): the second syllable begins at the same pitch level as the first syllable ends. The generalization is clear: In Hakha-Lai, pitch changes may not be effected between

⁵By this we do not mean that Hakha-Lai's contours are like the tautosyllabic tone "clusters" in African languages. Rather, we follow Yip's (1989) suggestion that South and Southeast Asian contours are still units, with the sequenced tonal features dominated by a single tonal node.

⁶For the prohibition against initial LH, we have thought of a phrase-initial %H boundary tone.

syllables but only tautosyllabically. That is, the only way to get a pitch change is via a contour!⁷

There is much more to say about the interpretation and significance of the Hakha-Lai tone system. For the present purpose we restrict ourselves to the following observations concerning the phonetic grounding of tone with respect to syllable structure. As we have said, the number of tonal contrasts and tonal contours should be greater on longer than on shorter sonorous rimes (Zhang 2001). Initial evidence for this position may be derived from the fact that only “full” (bimoraic, heavy) syllables carry tone in Hakha-Lai. CV syllables are toneless. However, counterevidence is found in two cases, both involving stopped syllables. CVVT syllables have a long nucleus, but no underlying tonal contrast. In addition, they are realized with a low level tone, i.e. not a contour. In the case of CVT, the lack of an underlying contrast is as expected, given the short nucleus and non-sonorant coda. However, we have seen two complications. First, their one underlying tone is a LH rising tone—the one that in principle requires the greatest duration! Second, due to the tone sandhi rules, there actually is a surface contrast between LH and HL on CVT syllables, as seen in (24). The one tone that is not allowed on CVT syllables is the one that is most expected—level L! We suspect that the rising tone of CVT syllables may derive historically from previous final glottalization, which is attested in other languages in Southeast Asia. If correct, the present study supports the notion that history may provide a more direct contribution to the understanding of the synchronic phonological distributions and rules found in Hakha-Lai than direct reference to phonetics.

References

- Hyman, Larry M. & Kenneth VanBik. 2002. Tone and stem2-formation in Hakha (Lai Chin). *Linguistics of the Tibeto-Burman Area* 25.1.
- Kathol, Andreas & Kenneth VanBik. 2001. The syntax of verbal stem alternations in Lai. Ms. University of California, Berkeley.
- Melnit, Nurit. 1997a. The sound system of Lai. *Linguistics of the Tibeto-Burman Area* 20.2, 9-19.
- Melnit, Nurit. 1997b. Verbal alternations in Lai. *Linguistics of the Tibeto-Burman Area* 20.2, 163-172.
- Ohala, John J. 1978. Production of tone. In Victoria A. Fromkin (ed.), *Tone: A linguistic survey*, 3-39. New York: Academic Press.

⁷This generalization applies only to input forms, since, as we have seen in forms such as in (15)-(17), the rules produce output F-F sequences. Note also in this regard that RSR does not apply when followed by a mid-to-high pitched toneless CV in the next foot, which does not constitute a heterosyllabic pitch change. (It is evidence that unmarked tone is not equivalent to the L we have established for level tone full syllables, however.) We have yet to explain why HL is simplified after toneless a CV-CV foot, as in (10b).

Larry Hyman and Kenneth VanBik

- Olawsky, Knut J. & Kenneth VanBik. 2000. Introduction to Lai tonology. Ms. University of California, Berkeley.
- Patent, Jason. 1997. Lai verb lists. *Linguistics of the Tibeto-Burman Area* 20.2, 57-112.
- Peterson, David. 1998. The morphosyntax of transitivization in Lai. *Linguistics of the Tibeto-Burman Area* 21.1, 87-153.
- VanBik, Kenneth. 2001. Three types of causative constructions in Lai. Ms. University of California, Berkeley.
- VanBik, Kenneth & Rungpat Roengpitya. 2001. An acoustical study on Lai vowel length (abstract). Acoustical Society of America Meeting, Lauderdale, Florida, December 4, 2001.
- Yip, Moira. 1989. Contour tones. *Phonology* 6:149-174.
- Zhang, Jie. 2001. *The effects of duration and sonority on contour tone distribution—typological survey and formal analysis*. Doctoral dissertation, UCLA.

Larry Hyman / Kenneth VanBik
Dept. of Linguistics
University of California
Berkeley, California 94720-2650

hyman@socrates.berkeley.edu; vanbik@socrates.berkeley.edu

Timing Tonogenesis: Evidence from Borrowing

MARTHA RATLIFF
Wayne State University

1. Introduction

There is one part of the received wisdom about tonogenesis in Southeast Asia that has puzzled me for years. I now think I have a useful way to think about it, one that may even be right. That problematic part of the received wisdom has nothing to do with Haudricourt’s story (1954) of the segmental origin of tonal contrasts (via phonation contrasts, see Thurgood to appear) in the first instance from laryngeal contrasts in consonants at the back of the syllable (A-D), subsequently doubled by the merger of voiced and voiceless segments at the front of the syllable (1-2), as represented below:

(1)

	A (-0, -N)	B (-ʔ)	C (-s)	D (-p, -t, -k)
1 (voiceless initials)	A1	B1	C1	D1
2 (voiced initials)	A2	B2	C2	D2

I think there is ample evidence for this scenario, (1) from comparative evidence, (2) from phonation type traces of both the final laryngeal contrasts and the old initial voicing contrast, and (3) from languages of the area that have undergone the first wave of tonogenesis but not the later split. The problem has rather to do with the *propagation* of this “Sinospheric” four-by-two system of tonal contrasts. It has been claimed, explicitly by Paul Benedict, but implicitly by others, that this whole system of tonal contrasts (4 tones times 2 “registers”) was borrowed from Chinese by Hmong-Mien, Tai and Vietnamese, all of which were originally atonal. Benedict writes “...Vietnamese, under direct Chinese domination lost the ... initial syllables of MK [Mon-Khmer] *while directly borrowing the tonal system* ...” (Benedict 1997: 4, emphasis added).

For those who are unused to thinking in terms of tone categories (A1, A2, B1, B2, etc.) as opposed to phonetic tones (high level, low rising, etc.), it is useful

to think of them this way: all the words in a particular tonal category have a common historical origin in terms of final and initial consonantism (A1 = *voiceless initial, open syllable or syllable with a nasal coda). This insures that when the original consonantism is transphonologized into tone that all of the words belonging to each original category as defined by syllable type will continue to pattern together tonally. Although phonetic studies have shown that the newly emergent tones will have certain properties due directly to the type of consonant lost, once tones are created, they morph quite quickly into other things: originally high tones may lower, low tones may raise, tones may merge, contours may simplify, etc. Therefore words across languages in a family which belong to a particular tonal category may have quite different phonetic realizations. For example, within Hmong-Mien, words in the A1 category have a variety of different phonetic values: they are mid rising, high level, low rising, mid falling, and mid level. This cross-linguistic variability is true of every tone category. The categories themselves, on the other hand, are remarkably stable: in all Hmong-Mien languages, for example, the members of the group of cognates which includes “to give”, “deep”, “three”, “thatch grass” and “snake” will all have *the same tone* in each language of the family (the A1 reflex), regardless of the phonetic value of that tone in any particular language.

The perfect tone category correspondences in Chinese loanwords from the Early Middle Chinese period are also somehow attributed to the fact that Hmong-Mien borrowed its four-by-two tonal system from Chinese. For example, Ying Lin (1972: 56), in an article on Chinese loans in Hmong-Mien, writes “If we compare these loans with *Qìyùn* rhyme tables, we find that the tones have been borrowed by Miao *primarily on the basis of the píng, shǎng, qù, and rù tone categories* (emphasis added).

(2)			Early Middle Chinese	Proto- Hmong- Mien	Tone category ¹
金	<i>jīn</i>	‘gold’	kim	cəm	A1
秧	<i>yāng</i>	‘seedling’	ʔiaŋ	ʔzwaə:ŋ	A1
千	<i>qiān</i>	‘thousand’	ts ^h ɛn	tθhjin	A1
蒸	<i>zhēng</i>	‘steam’	tɕiŋ	tsuə:ŋ	A1
雞	<i>jī</i>	‘chicken’	kej	qəi	A1
銅	<i>tóng</i>	‘copper’	dəwŋ	dəəŋ	A2
羊	<i>yáng</i>	‘sheep/goat’	ɟiaŋ	zwbə:ŋ	A2

¹ Middle Chinese reconstructions are from Pulleyblank 1991 and Hmong-Mien reconstructions are from Wang and Mao 1995.

Timing Tonogenesis: Evidence from Borrowing

(2) cont.			Early Middle Chinese	Proto- Hmong- Mien	Tone category ²
牛	<i>niú</i>	‘buffalo’	ɲuw	ɲoːɲ	A2
銀	<i>yín</i>	‘silver’	ɲin	ɲwɛːn	A2
桶	<i>tǒng</i>	‘bucket’	tʰəwɲ	thɛɐɲ	B1
瓦	<i>wǎ</i>	‘tile’	ɲwaɪ	ɲwa	B2
馬	<i>mǎ</i>	‘horse’	mai	mnjuːn	B2
甑	<i>zèng</i>	‘rice steamer’	tsɿŋ ^h	tʂaəɲ	C1
炭	<i>tàn</i>	‘charcoal’	tʰan ^h	thaːn	C1
灶	<i>zào</i>	‘stove’	tsaw ^h	tsʊ	C1
箸	<i>zhù</i>	‘chopsticks’	dria ^h	dəu	C2
漆	<i>qī</i>	‘lacquer’	tsʰit	tʰhjet	D1
百	<i>bǎi</i>	‘hundred’	paijk	pɛk	D1
十	<i>shí</i>	‘ten’	dzip	jap	D2

At first, this seems reasonable for Hmong-Mien, because not only is there internal evidence of the development of the Sinospheric-type four-by-two tone system within Hmong-Mien, the languages in this family show the imprint of Chinese influence at every level of structure (Downer 1973, Ratliff 1999, 2000, 2001).

But this has always been the sticking point for me: How do speakers hear and borrow tone categories? How do speakers hear and borrow whole systems? There is nothing in the speech signal. And how does a borrowed word that has a particular niche within the donor language system, the historical antecedents of which were soon lost because the rise of tones *depends* on their being lost, embed that borrowing, in a perfectly analogous place, within a whole system that mirrors that of the donor language? Or to use another metaphor, how can the donor language regenerate an entire body around the transplanted borrowing, a body which is such a perfect clone of its mother that the borrowing comes to occupy the same relative position in the new body that it occupied in the old body?

In order to support the idea that system borrowing and borrowing on the basis of tone categories happened at a time in the distant past, one would ideally like to point to a modern-day contact situation which has recently yielded, or is in the process of yielding, a pattern similar to this one: identically structured systems, and borrowed words which occupy analogous places in the systems of both donor

² Middle Chinese reconstructions are from Pulleyblank 1991 and Hmong-Mien reconstructions are from Wang and Mao 1995.

and borrower language. I would like to review some recent situations where words have been borrowed under each logically possible combination of tonal and atonal donor and borrower languages to show that in three out of the four possible contact situations between donor and borrower — atonal donor and tonal borrower, tonal donor and tonal borrower, tonal donor and atonal borrower (the dominant theory) — all the known cases have yielded patterns quite different from this one. My conclusion is that, by elimination, and with the knowledge of how easily prosodic systems have diffused through Southeast Asia, it was the fourth contact situation — atonal donor and atonal borrower — which held at the time of these borrowings. According to Sagart, tones developed after Old Chinese but before Early Middle Chinese, so somewhere between 500 BCE and 500 CE (1999: 101). On the basis of the good segmental correspondences, the Chinese borrowings in Hmong-Mien above can be dated to only slightly before Early Middle Chinese, which we know was a tonal language, or to the first five hundred years of the Christian Era.³ My argument will be that tonogenesis was ready to happen at this point, but it hadn't happened yet. And I will suggest that rather than it being the case Hmong-Mien took tone from Chinese, the languages of the Sinosphere “all went together”.

2. Patterns of tone assignment in different contact situations

What were the logical possibilities at the time of borrowing, circa 0–500 CE? Let us take the four possible contact situations in turn, and look at the patterns they have yielded in the recent past.⁴

2.1. Donor atonal, Borrower tonal

For the case where the donor is atonal and the borrower is tonal, Jim Matisoff (2001: 321–32) reports that two variations of one basic strategy are employed in Southeast Asia: one or two common tones are selected as “loan-tones”, or a rare tone is used as the loan-tone which instantly identifies the word as a borrowing (such as English borrowings in Lahu and Cantonese). In no reported case are all native tonal contrasts realized in words borrowed from an atonal language.

The “common loan-tone” strategy often involves an attempt to reflect stress or intonation contour in the donor language through tone. Christopher Court's study of Malay borrowings in Satun Thai (1975) indicates that the pitch correlates of stress and intonation are re-interpreted within a tonal context as tones. He reports that the high rising-falling tone of Satun Thai is used to represent the falling intonation on phrase-final open syllables in Malay loanwords. The high level tone

³Although Pulleyblank (1978) suggests that the development of tones was rather late, and that certain rhymes with final /s/ persisted down to the 6th century CE.

⁴Matisoff (2001) gives a very useful catalog of contact situations involving borrowing of words between tonal and atonal languages similar to this one. He also includes consideration of the relative prestige of donor and borrower.

is assigned to other syllables which do not occur at a point which corresponds to the Malay intonation peak. Jack Gandour (1979) also found a similar stress-to-tone mapping in a majority of English polysyllabic loanwords in Thai. Although the details are considerably more complex, in general borrowed polysyllabic English words are assigned a high tone on the syllable which receives stress in English.

Although Hmong speakers in the United States do not now transliterate nativized English borrowings into Hmong orthography which makes tone assignment easy to see (they simply use the English spelling), variation in early nativized loans show the same basic strategy: an attempt to mirror either English intonation or stress through tone assignment.

(3) ‘America’

amɛlikà Mid-Mid-Mid-Low (an attempt to represent English intonation)

àmɛlikà Low-Low-Rising-Low (an attempt to represent English stress)

àmɛlikà Low-Low-Low-Low (loan-tone assignment)

So if Hmong-Mien had been tonal and Chinese had not been tonal at the time of the loans listed above, presumably all pre-Early Middle Chinese loanwords into common Hmong-Mien would have been realized with one or two tones: the then-current “loan-tones”, either a common tone or two, or a rare tone. The tonal categories for all the loanwords of this stratum would then have belonged to a restricted subset of the eight across which native words are distributed, and would have been the same for all Hmong-Mien languages — which is clearly not the case.

2.2. Donor and Borrower both tonal

The “rare loan-tone” strategy may be used in the case of borrowings between tone languages (Matisoff 2001: 321). Bunu, for example, a Hmong-Mien language of Guangxi Province in China, assigns sandhi tones to Chinese and Zhuang borrowings, tones which are restricted in use in the native portion of the vocabulary. Apparently speakers feel that these minor tones are appropriate for marking all “special purpose” words — they are also used to mark baby-names such as ‘little rat’, ‘little frog’, etc. (Mao and Chou 1972, Meng 1983).

However, the strategy most often employed when one tone language borrows tones from another tone language is a phonetic mapping to the closest phonetic match in the borrowing language. Since tone categories (as identified by sets of words which all have the same tone) are remarkably stable over time, but tone values are remarkably changeable, it is easy to see that this kind of surface mapping would not yield a neat historical pattern like the one we are attempting to explain.

For example, recent Chinese loanwords in Hmong-Mien languages are borrowed by matching the tone value of the word in the local variety of Chinese with the closest tone value in the borrowing language. “Basically, modern loans use Miao initials, finals, and tones to reflect southwestern Mandarin phonology.” (Ying 1972: 64). In the varieties of local Chinese spoken in the area of three major Hmongic languages — Yǎnghāo (Eastern Hmongic), Làyǐpíng (Northern Hmongic), and Xǔyǒng (Western Hmongic) — there is one rising tone, a reflex of category C1. The tone it will have within each borrowing Hmongic language is determined by phonetics: the rising tone of each Hmongic language is chosen, even though in Yǎnghāo the rising tone is a reflex of category B1, in Làyǐpíng the rising tone is a reflex of category A1, and in Xǔyǒng the rising tone is a reflex of category D2:

(4)

丈	zhàng	‘3 1/3 meters’	上	shàng	‘to start (class)’
Yǎnghāo	tsaŋ	B1 [35]	Yǎnghāo	saŋ	B1 [35]
Làyǐpíng	taŋ	A1 [35]	Làyǐpíng	saŋ	A1 [35]
Xǔyǒng	t̚saŋ	D2 [13]	Xǔyǒng	saŋ	D2 [13]

(Miào-Yáo Yǔ Fāngyán Cíhuì Jì 1987)

Theraphan L-Thongkum (1997) presents an interesting case study of language change in progress among younger speakers of Mien, a tone language, in a village near Chiangmai, where Standard Thai, also a tone language, is used as a medium of instruction and both Northern Thai and Standard Thai are widely spoken. In this dialect of Mien, speakers are in the process of reducing their number of tones from six to five, bringing Mien in line with Standard Thai, and are also adjusting the phonetic values of the tones themselves to be more in accord with the phonetic values of Thai tones. However, tone category correspondences do not result as a by-product of this contact-induced change.

(5)

“evolving” Mien	Standard Thai, but	Mien	Thai
33	33	A1	A1a, A2
31	51	A2	B1, C2, D2L
45	45	B1	B2
34	15	B2, C1	A1b
21	21	C2	C1, D1

(L-Thongkum 1997: 158)

Finally, James Chamberlain (1972) has also given a brief account of tone assignment for Lao borrowings in five minority Tai languages of Northeast Thailand. All of these languages, both donor and borrowers, are tonal. The purpose of his paper was to support the contention of Marvin Brown that “tone shapes may be borrowed, but patterns of coalescence and splitting may not” by showing that tone borrowing in these cases is governed by judgments of phonetic similarity. Although I am interested in tone category membership correspondences and Chamberlain was interested in correspondences in patterns of tone category coalescence and splitting, all these things are footprints of earlier historical events, and his findings are consistent with the claim of this paper — that the process of borrowing does not cause borrowers to reproduce the history of the donor language within their own language.

So if both Chinese and Hmong-Mien had been tone languages at the time of the loans, then Middle Chinese loanwords in Hmong-Mien would have the phonetic values of reflexes of various Hmong-Mien tone categories just as they do in modern Chinese loans — reflecting an effort to match the tone in the donor language to the closest tone in the borrowing language — rather than reflexes of only one tonal category across the board, moreover that one which provides a perfect match to the donor language in terms of its place in the system.

2.3. Donor tonal, Borrower atonal

This is the traditional view of the state of things at the time of early contact, especially under the Austro-Tai hypothesis which holds that Hmong-Mien was originally atonal like its Austronesian relatives (Benedict 1975). But in the modern day, we only have evidence for three basic types of outcomes from contact of this sort. First, words may be borrowed without tones because the borrowers cannot hear tone, especially if contact is minimal to non-existent and the number of borrowings is small, as is the case for the handful of Chinese borrowings in English:

(6)

麻將	<i>májiàng</i>	>	‘mahjong’
烏龍	<i>wūlóng</i>	>	‘oolong (tea)’
山東	<i>Shāndōng</i>	>	‘shantung (silk)’

Second, in a situation of greater contact, we may get the odd situation that has been reported by David Filbeck (1972) for one dialect of Mal (called T’in at a higher node in the tree), an atonal Mon-Khmer language of Southeast Asia spoken in the northeast part of Thailand. It has acquired tones in two ways. First, Northern Thai numerals have been borrowed along with their tones. Although this may have arisen as a recitation effect, the numerals are pronounced with their

Northern Thai tones even in running speech. At the same time, other Thai words have been borrowed with a rising pitch profile — presumably because of its salience, because it does not correspond to the pitch contours of the borrowings as pronounced in Thai:

(7)

Thai	Mal	
<i>yâak</i>	<i>ɲăak</i>	‘difficult’
<i>lâak</i>	<i>lăak</i>	‘to drag’
<i>khěek</i>	<i>khěek</i>	‘guest’
<i>klaaŋ</i>	<i>kăaŋ</i>	‘middle’

(Filbeck 1972: 115)

This rising contour also marks words borrowed from other atonal Mon-Khmer languages, and inexplicably, a couple of native words well.

The third situation, one of intense and prolonged contact, is probably closest to what proponents of the theory that Hmong-Mien acquired tones from Chinese must assume for early Chinese/Hmong-Mien contact, given the number of loanwords and the extent of the grammatical influence Chinese has had upon Hmong-Mien languages over the centuries.

There is one good modern day case where, in the face of widespread multilingualism, the prosodic model of a dominant tonal language (or, in this case, languages) caused an atonal language to develop a complex system of tones. This is the case of the development of tones in the originally atonal Austronesian language Tsat of Hainan Island as described by Graham Thurgood in a series of publications (see Thurgood 1999 for a full account). Tones developed in Tsat under the influence of two tone languages: the local (Mǐn) varieties of Chinese, and the Tai-Kadai language Lí. The tonal system that developed in Tsat is similar in nature to the Thai-influenced Mien described by L-Thongkum: the number of tones and the tone values are almost the same as those of these two neighboring languages which the Tsat speakers know well (Thurgood 1999: 231), a case of surface convergence.

(8)

	Tsat	Chinese (Dǎnzhōu)	Lí (Tongshi)	Lí (Yuánmén)
high level	55	55	55	55
falling	42	--	43	42
mid level	33	22	33	44
rising	24	35	13	13
low level	11	11	11	11

But the key difference between this situation and the early Sinospheric contact situation that we are trying to understand is that Tsat developed tones in its own way, in a two-by-three system, where the initials started the tonogenetic process and the later split was conditioned by the finals. Thus there can be no corresponding categories in Tsat for Chinese loanwords to map onto — a four-by-two system cannot map directly onto a two-by-three system. Chinese borrowings in Tsat are pronounced with whatever tone is phonetically most similar (Thurgood, p.c. 2002). Although given the assumptions about tone spread in Southeast Asia, this should be the model we are looking for; the kind of cross-linguistic category correspondence pattern characteristic of the Sinosphere is not found here. As Thurgood writes, “it is not being argued that the Tsat tonal *system* is borrowed from one of these languages” (1999: 231).

2.4. Donor and Borrower both atonal

Given our inability to find a present-day (or recent-day) model that could account for the correspondence of tones in the Sinosphere under these three contact situations, we must turn to the last possible scenario. If neither Hmong-Mien *nor* Chinese had tones at the time these early loans in (2) above were made, we can indeed imagine how the striking cross-family correspondence of tone categories in loanwords could have come to be. Hmong-Mien could have borrowed the Chinese words with the (perhaps already decomposing) segmental material which eventually gave rise to tones intact. Then if both developed tones in the exactly same way, out of the laryngeal features of word-final consonants, as a rash of tonogenesis swept across the area (started by who knows who, not necessarily by the Chinese), then we would get these regular correspondences. It is my belief that this is the most likely account of what happened.

The following table contains a summary of the four types of contact described above.

(9)

<p><u>1. Donor atonal, Borrower tonal</u></p> <ul style="list-style-type: none"> -assignment to a common tone (interpretation of stress/intonation as tone): English > Hmong, Malay > Thai, English > Thai -assignment to a rare tone: English > Cantonese, English > Lahu 	<p><u>2. Donor and Borrower both tonal</u></p> <ul style="list-style-type: none"> -phonetic mapping: (Modern) Chinese > HmM lgs, Lao > Tai dialects -surface convergence: Thai > Mien -assignment to a rare tone: Chinese > Bunu Zhuang > Bunu
<p><u>3. Donor tonal, Borrower atonal</u></p> <ul style="list-style-type: none"> -words borrowed without tones: Chinese > English -individual words borrowed with tones: N. Thai > T'in (numerals) -assignment to a distinctive pitch profile: N. Thai > T'in (other borrowings) -systems of independent origin, surface convergence: Chinese > Tsat Li > Tsat 	<p><u>4. Donor and Borrower both atonal</u></p> <ul style="list-style-type: none"> -the languages both stay atonal <i>-the languages both become tonal</i>

3. Sinospheric languages developed tones together

One striking piece of evidence in support of the hypothesis that Chinese itself was atonal when it lent Hmong-Mien the words in (2) above is that fact that an even older stratum of Chinese loans shows regular tonal correspondences between the two families. Most Sinologists believe that Old Chinese was toneless. How else, then, can we explain correspondences in this oldest stratum of loanwords other than to say that tones developed in the two languages in a parallel fashion after the words were borrowed? And how else can we explain the *identical pattern* in the later stratum of loanwords other than to say that tones developed in the two languages families after these words were borrowed, too?

(10)

			Middle Chinese	Old Chinese	Proto- Hmong- Mien	Tone Category ⁵
廩	lǐn	‘granary’	lim’	*(C-)rəm	*ŋgljəm	B2
力	lì	‘strength’	lik	*(C-)rek	*ŋgljo(k)	D2
鐵	tiě	‘iron’	tʰɛt	*hliit (~-k)	*ljok	D1

Although Chinese contact is probably to be credited with making Hmong-Mien and other languages of the south “tone prone”, I don’t think it is possible to know who started the tonal ball rolling. The subsequent merger of initials and compensatory tone split is understood as happening in a wave across the whole area, and no one to my knowledge is concerned about identifying which languages started it. Yet for the much earlier prosodic restructuring, many seem to be willing to identify a source. Why not treat them both the same way?

In conclusion, the value of re-thinking the relative timing of the period of contact during which the early Chinese loans were made and tonogenesis is twofold: (1) it can help us do away with the need to explain how languages might borrow things like tone categories and tone systems when we know that speakers have no access to the histories of the languages they’re listening to, and (2) insofar that we can time these early loans on the basis of the segmental correspondences, it can help us time the onset of tonogenesis both in Chinese and in the languages that have been so profoundly influenced by Chinese. My hope is that future discussions of tonogenesis in Chinese will no longer rest exclusively on internal evidence, but will also make use of the external loanword evidence in attempting to date the onset of this important typological change.

References

- Benedict, Paul K. 1997. Interphyla flow in Southeast Asia. *Mon-Khmer Studies* 27: 1-11.
- Benedict, Paul K. 1975. *Austro-Thai Language and Culture with a Glossary of Roots*. HRAF Press.
- Chamberlain, James R. 1972. Tone borrowing in five Northeastern dialects. In Jimmy G. Harris and Richard B. Noss (eds.) *Tai Phonetics and Phonology*, 43-46. Bangkok: Central Institute of English Language, Mahidol University.
- Chang, Kun. 1972. The reconstruction of Proto-Miao-Yao tones. *Bulletin of the Institute of History and Philology* 44: 541-628.
- Court, Christopher. 1975. The segmental and suprasegmental representation of Malay loanwords in Satun Thai: a description with historical remarks. In

⁵ Old Chinese reconstructions are from William Baxter (p.c. 2001).

- Jimmy G. Harris and James R. Chamberlain (eds.) *Studies in Tai Linguistics in Honor of William J. Gedney*, 67-88. Bangkok: Central Institute of English Language, Office of State Universities.
- Downer, G. B. 1973. Strata of Chinese loanwords in the Mien dialect of Yao. *Asia Major* 18.1: 1-33.
- Filbeck, David. 1972. Tone in a dialect of T'in. *Anthropological Linguistics* 14.4: 111-118.
- Gandour, Jackson T. 1979. Tonal rules for English loanwords in Thai. In Theraphan L-Thongkum, Vichin Panupong, Pranee Kullavanijaya, M.R. Kalaya Tingsabadh (eds.) *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J. A. Henderson*, 94-105. Bangkok: Chulalongkorn University Press.
- Haudricourt, André. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242: 69-82.
- L-Thongkum, Theraphan. 1997. Tone change and language contact: A case study of Mien-Yao and Thai. In Arthur S. Abramson (ed.) *Southeast Asian Linguistic Studies in Honour of Vichin Panupong*, 153-60. Bangkok: Chulalongkorn University Press.
- Mao, Tsung-Wu and Chou, Tsu-yao. 1972. A brief description of the Yao language. Translated by Chang Yu-hung and Chu Kwo-ray. In Herbert C. Purnell Jr. (ed.) *Miao and Yao Linguistic Studies: Selected Articles in Chinese Southeast Asian Program, Data Paper 88*, 239-55. Ithaca: Cornell University Department of Asian Studies.
- Matisoff, James A. 2001. Genetic versus contact relationship: prosodic diffusibility in South-East Asian languages. In Alexandra Y. Aikhenvald and R.M.W. Dixon (eds.) *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, 291-327. Oxford: Oxford University Press.
- Meng, Chaoji. 1983. Yaozu Bunuyu 1' zhi 4' diao de xingcheng he fazhan [The origin and development of tones 1'-4' in the Bunu language of the Yao people]. *Minzu Yuwen* 2: 56-59.
- Miao-Yao Yu Fangyan Cihui Ji* [A Lexicon for Miao-Yao Dialects]. 1987. Zhongyang Minzu Xueyuan Miao-Yao Yanjiu Shi [Office of Miao-Yao Research of the Central Minorities Institute]. Beijing: Zhongyang Minzu Xueyuan Chubanshe [Central Minorities Institute Press].
- Pulleyblank, Edwin G. 1991. *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press.
- Pulleyblank, Edwin G. 1978. The nature of the Middle Chinese tones and their development. *Journal of Chinese Linguistics* 6: 173-203.
- Ratliff, Martha. 1999. Sino-Hmong-Mien phonology. Paper presented at 32nd International Conference on Sino-Tibetan Languages and Linguistics, Urbana-Champaign, Illinois.

- Ratliff, Martha. 2000. Numeral classifiers and classifying nominal prefixes in Hmong-Mien: One more story of contact and its grammatical consequences. Paper presented at the Tenth Annual Meeting of the Southeast Asian Linguistics Society, Madison, WI.
- Ratliff, Martha. 2001. Voiceless sonorant initials in Hmong-Mien: Sino-Tibetan correspondences. In Graham Thurgood (ed.) *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society (1999)*, 361-375. Tempe, AZ: ASU Program for Southeast Asian Studies.
- Sagart, Laurent. 1999. The origin of Chinese tones. In Shigeki Kaji (ed.) *Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*, 91-103. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
- Thurgood, Graham. To appear. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica* 20.1.
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. Oceanic Linguistics Special Publication No. 28. Honolulu: University of Hawai'i Press.
- Wang, Fushi and Mao, Zongwu. 1995. *Miaoyaoyu guyin de gouni* [The Reconstruction of Proto-Miao-Yao Pronunciation]. Beijing: Zhongguo Shehui Kexue Chubanshe [China Social Sciences Press].
- Ying, Lin. 1972. Chinese loanwords in Miao. Translated by Chang Yu-hung and Chu Kwo-ray. In Herbert C. Purnell Jr. (ed.) *Miao and Yao Linguistic Studies: Selected Articles in Chinese* Southeast Asian Program, Data Paper 88, 55-81. Ithaca: Cornell University Department of Asian Studies.

Martha Ratliff
Linguistics Program
Wayne State University
Detroit, MI 48202

martha_ratliff@wayne.edu

Different Durations of Diphthongs in Thai: a New Finding

RUNGPAT ROENGPITYA

University of California at Berkeley, CA.

1. Introduction

This paper presents the studies of the quantity and the quality of diphthongs in Thai. In a traditional Thai grammar, there are six diphthongs in Thai: short diphthongs /iaʔ, uaʔ, uaʔ/ (only with a final glottal stop) and long diphthongs /ia, ua, ua/ (without any final or with other finals than a glottal stop), as suggested by Praya Upprakitsilpasaan (1998), and Thonglor (1966).

Linguists (Abramson 1962, Naksakul 1998, and Rudaravanija 1965), on the other hand, pointed out that there are three phonological diphthongs in Thai /ia, ua, ua/. Short and long diphthongs are not phonemically distinctive. Naksakul (1998) added that diphthongs are shorter in unstressed positions but longer in stressed positions and that diphthongs are always short when followed by a final glottal stop.

2. Aim

The aim of this paper is to see (1) whether short and long diphthongs are different in quantity, (2) whether a final glottal stop is the main cue to distinguish short and long diphthongs in Thai, and (3) whether the three diphthongs in Thai are different in quality (trivially true, but specifically how?).

3. Study A. An Acoustic Study of Diphthongs in Thai

3.1. Aim

The aim is to study acoustically the quantity and the quality of the three diphthongs in Thai.

3.2. Tokens

The tokens, used in this study, had the structure of C1V(C2)T, where C1 was a voiceless unaspirated stop /p-/ , a voiced stop /b-/ , a voiceless fricative /f-/ , a nasal

/n-/ or a lateral /l-/; V was /ia/, /ua/, or /ua/; C2 was a voiceless unreleased velar stop /-k/, a glottal stop /-ʔ/, a dental nasal /-n/, a palatal glide /-j/, or a labial-velar glide /-w/; and T was a low tone. Some tokens were meaningful but some were nonsense words. All the tokens were in a sentence “*faj kham wâ _____ sǎŋ khráj*” ‘Listen to the word _____ twice.’

3.3. Speakers

The speakers for this study were three native Standard Thai speakers: two males (Speaker 1 and Speaker 2) and one female (Speaker 3: the author). All of them were graduate students at the University of California at Berkeley. All of the speakers had normal speech and hearing.

3.4. Procedure

The three speakers read all the tokens twice. The tokens were recorded on analog tapes and were subsequently digitized at a sample rate of 16 kHz. with 16 bits per sample. All the tokens were categorized into 4 sets, as follows.

1. Diphthongs without final. There were 5 initial consonants*3 diphthongs* 2 times of each token * 3 speakers. There were 90 tokens for this set.
2. Diphthong /ia/ with finals. There were 5 initial consonants * 4 final consonants * 2 times of each token * 3 speakers. There were 120 tokens for this set.
3. Diphthong /ua/ with finals. There were 5 initial consonants* 5 final consonants *2 times of each token * 3 speakers. There were 150 tokens for this set.
4. Diphthong /ua/ with finals. There were 5 initial consonants *4 final consonants * 2 times of each token * 3 speakers. There were 120 tokens for this set.

There were a total of 480 tokens in this study.

To study diphthong quantity, all the tokens were measured¹ for the diphthong duration, the first vocalic element duration, the transition duration between the first and the second vocalic elements, and the second vocalic element duration.

To study diphthong quality, all the tokens were measured for the first and the second formants at 25%, 50%, and 75% points in diphthongs. This helps to ensure that the formant values obtained more a function of the diphthong and not the adjacent consonants.

¹ Any such demarcation of V1(the first vocalic element), transition, V2 (the second vocalic element) has some arbitrariness. One has to adopt one well-defined arbitrary measure and use it consistently.

3.5. Results

In previous literature (Naksakul 1998), there are three phonemic diphthongs in Thai /ia/, /ua/, and /ua/. That is, short diphthongs appear before a final glottal stop or in unstressed syllables. Short and long diphthongs are not phonologically distinctive in length.

The results of this study on diphthong quantity and quality are shown below.

3.5.1. Diphthong Quantity

The results of this study on diphthong quantity are presented in Figures 1-3.

Figure 1 shows the average duration of three diphthongs, which was measured for the first vocalic element duration, the transition duration, the second vocalic element duration, and the total diphthong duration, in all environments. From the figure, the average duration of the 1st vocalic element is 54% of the total diphthong duration, the transition, 17%, and the second vocalic element, 29%. This can be concluded that the 1st vocalic element of a diphthong is the most prominent part of the three phonemic diphthongs. Moreover, the data shows that the total duration of the three diphthongs are in the same range.

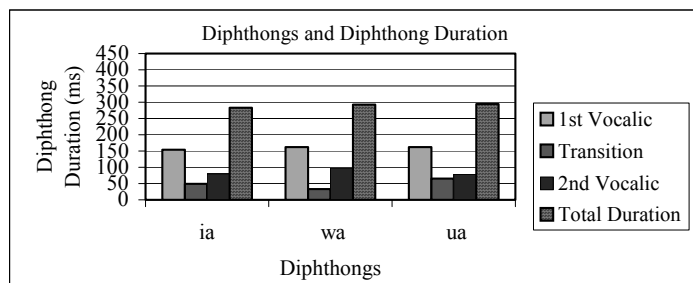


Figure 1. Duration of diphthongs in all environments (without any final, with final sonorants, with a final glottal stop, with final stops) of the three speakers.

Figure 2 presents the total durations of all three diphthongs after five initial consonants. The data shows that the durations of diphthongs after initial consonants are the same.

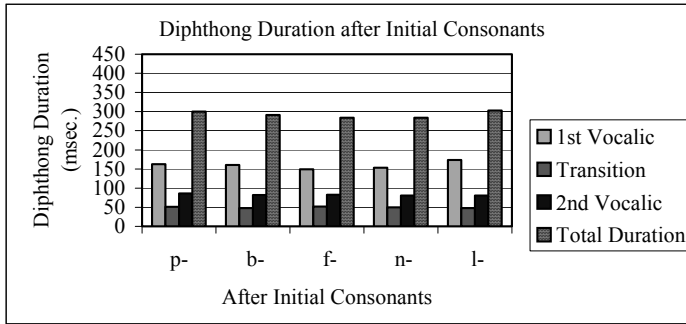


Figure 2. Duration of diphthongs and the different types of initial consonants.

The total durations of all three diphthongs before different types of final consonants are shown in Figure 3. It can be seen from the figure that diphthongs without any final (in open syllables) have the longest total diphthong duration (about 443 msec.). Diphthongs with a final (in closed syllables), regardless of whether it is a final glottal stop /-ʔ/, a final stop /-k/, a final nasal /-n/, a final glide /-w/, or a final glide /-j/, have shorter total diphthong duration (246-293 msec.).

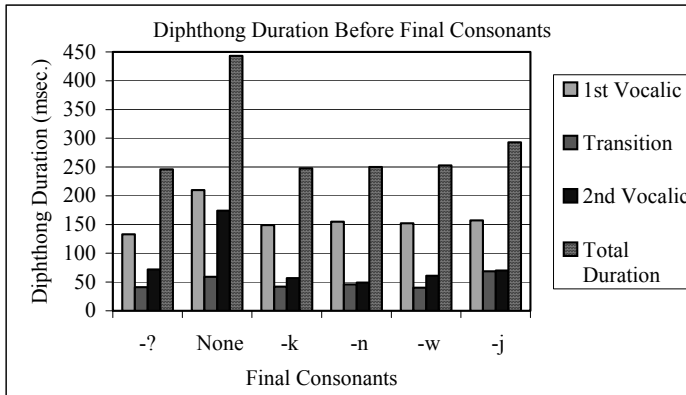


Figure 3. Duration of diphthongs and the different types of final consonants.

Moreover, diphthongs in open syllables are about 1.7 times as long as diphthongs in closed syllables. The shorter and longer duration of diphthongs is reflected by the shorter and longer duration of the first vocalic element, and, mainly, by the shorter and longer duration of the second vocalic element. The first vocalic element of diphthongs in open syllables is about 60 msec. longer than

the first vocalic element of diphthongs in closed syllables. The second vocalic element of diphthongs in open syllables is about 112 msec. longer than the second vocalic element of diphthongs in closed syllables.

3.5.2. Diphthong Quality

The results of this study on diphthong quality are shown in Figures 4-9.

Figures 4 and 5 below present the quality of three diphthongs at 25%, 50%, and 75% points in diphthongs², when diphthongs were not followed by glides /-w/ and /-j/. The data from Figures 4 and 5 show the diphthong quality after all initial consonants and before all final consonants. It was found that diphthong quality of shorter and longer diphthongs are in the same range. Moreover, there is no difference in diphthong quality after different types of initial consonants nor before different types of final consonants, except for final glides.

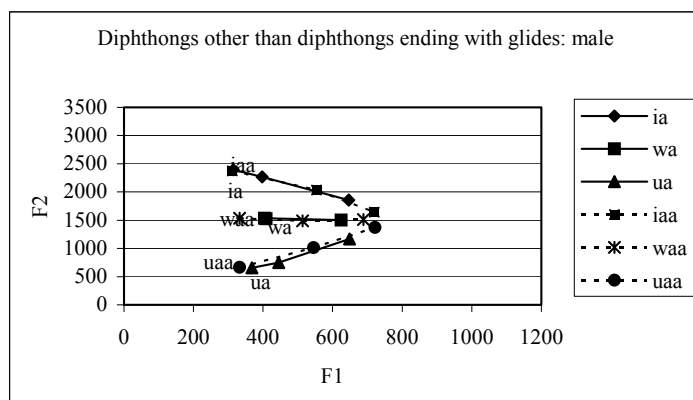


Figure 4. Diphthongs other than those ending with glides, from the tokens of male speakers.

² The diphthong quality at 25% point in diphthongs shows the value of the first vocalic element in diphthongs, the value at 50% point, the transition, and the value at 75% point, the second vocalic element.

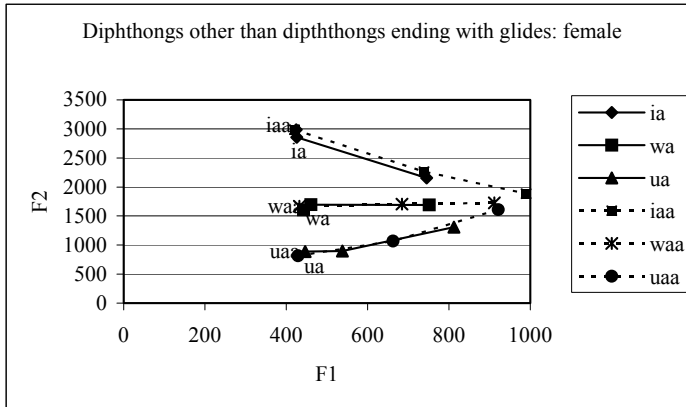


Figure 5. Diphthongs other than those ending with glides, from the tokens of female speakers.

The data show that the diphthong quality at 25% point in diphthongs, which reflects the quality of the first vocalic element in diphthongs, functions as the main cue to distinguish the three phonemic diphthongs /ia/, /ua/, and /ua/.

The diphthong quality at 50% and 75% points in diphthongs, however, is a less salient cue for differentiating the three phonemic diphthongs, as the three diphthongs share the same second vocalic element /a/ but it was found that the F2 height of the first vocalic element affects the F2 height of the second vocalic element in diphthongs. It can be seen that the F2 of the second vocalic element /a/ in diphthongs is the highest in the diphthong /ia/, higher in the diphthong /ua/, and the lowest in the diphthong /ua/.

Figures 6-7 and 8-9 show the diphthong quality before final glides /-w/ and /-j/, respectively. Figures 6-7 present the diphthong quality before a final glide /-w/ from the tokens of male and female speakers, respectively. The results show that the F2 decreases, especially at the 75% point of diphthongs, due to the coarticulation of final /-w/.

Different Durations of Diphthongs in Thai

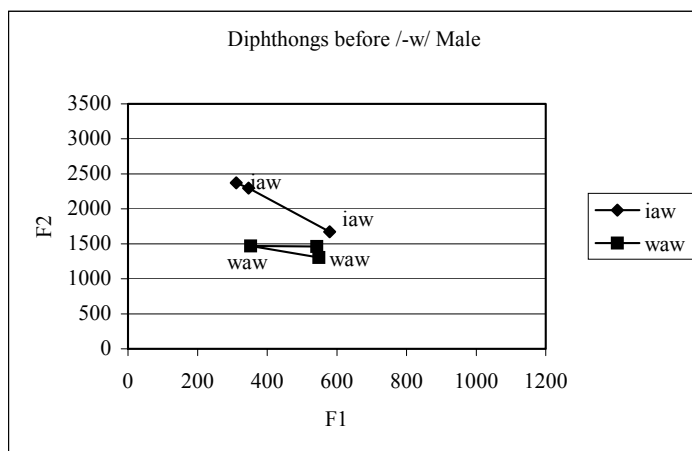


Figure 6. Diphthongs before a glide /-w/ for male at 25%, 50%, and 75% points in diphthongs.

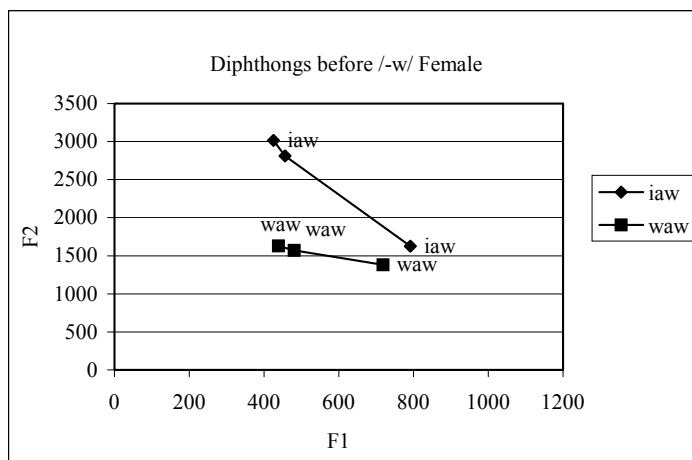


Figure 7. Diphthongs before a glide /-w/ for female at 25%, 50%, and 75% points in diphthongs.

Figures 8-9 present the diphthong quality before a final glide /-j/ from the tokens of male and female speakers, respectively. The results show that the F2 increases, especially at the 75% point of diphthongs, due to the coarticulation of final glide /-j/.

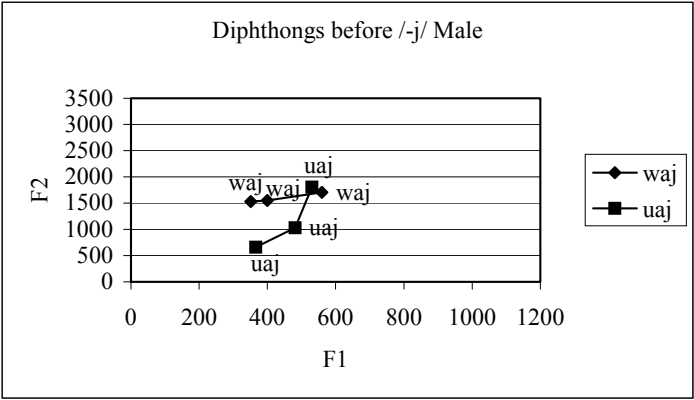


Figure 8. Diphthongs before a glide /-j/ for male at 25%, 50%, and 75% points in diphthongs.

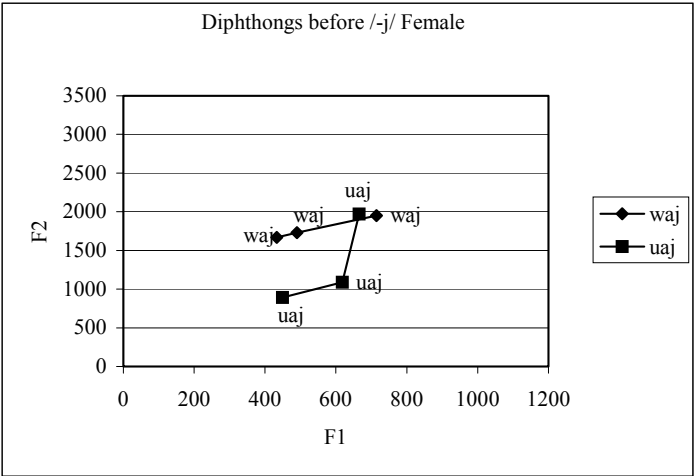


Figure 9. Diphthongs before a glide /-j/ for female at 25%, 50%, and 75% points in diphthongs.

3.6. Conclusion of Study A

From the results of study A, it can be concluded, as follows.

1. Diphthongs in Thai are not phonemically distinctive in length, as said by Abramson (1962) and Naksakul (1998).
2. In this study, it is found that diphthongs in open syllables or “longer allodiphthongs” are longer than diphthongs in closed syllables or “shorter allodiphthongs.”
3. The cue to distinguish shorter and longer allodiphthongs is whether a diphthong has a final consonant (of all types) or not.
4. The cue to distinguish three phonemic diphthongs /ia, ua, ua/ is the diphthong quality of the first vocalic element.
5. Shorter and longer allodiphthongs do not have different diphthong quality.
6. Diphthong quality is not different when preceded by different types of initial or followed by different types of final consonants, except for final glides.
7. Final glides affect the diphthong quality at 75% point in diphthongs. Final glide /-w/ decreases the F2 value at 75% point in diphthongs. Final glide /-j/ increases the F2 value at 75% point in diphthongs.

4. Study B: A Perceptual Study of Diphthongs in Thai

4.1. Aim

The aim of this study is to see how Thai listeners perceive shorter and longer allodiphthongs in Thai and to see whether diphthong duration is the main perceptual cue to differentiate shorter and longer allodiphthongs in Thai.

4.2. Stimuli

A pair of nonsense words with shorter and longer allodiphthongs [p̚iaʔ] and [p̚iaa] was chosen for this study. A native-Thai female speaker (the author) read the words in the frame sentence “*fəŋ kham wāa _____ sǎŋ khráŋ*” ‘Listen to the word _____ twice.’ Both tokens were recorded on an analog tape and were digitized at a sample rate of 16 kHz. with 16 bits per sample. Both tokens were measured for the diphthong duration.

The results from Table 1 show that the shorter allo-diphthong in [p̚iaʔ] had a shorter diphthong duration than did the longer allo-diphthong in [p̚iaa].

Table 1. Diphthong duration for a perceptual study of diphthongs.

Tokens	V1 duration /i/	transition	V2 duration /a/	Total duration
p̚iaʔ	180.94	89.88	71.75	342.57
p̚iaa	177.44	75.50	247.19	500.13

To avoid a glottal stop and creakiness after a shorter allo-diphthong [ia] being perceived as a cue for a shorter allo-diphthong in the token [piaʔ], the glottal stop and the creakiness were spliced off and the second vowel was digitally lengthened to reach the same duration as the original one. Aspiration was found at the end of the token [piaa]. To avoid any unexpected side effect from this aspiration, the aspiration was spliced off and the second vowel was digitally lengthened to reach the same duration as the original one.

The shorter allodiphthong [ia] in the token [piaʔ] was lengthened only from the transition between the vowels /i/ and /a/ to the end of the second vowel /a/ at 20 msec steps until it reached the same duration as the original long allodiphthong. The longer allodiphthong [iaa] in the token [piaa] was shortened only from the transition between the vowels /i/ and /a/ to the end of the end of the second vowel /a/ at 20msec steps until it reached the same duration as the original short allodiphthong.

All the tokens were resynthesized in the LPC analysis and resynthesis program, and were randomized. There were a total of 18 tokens for this study.

4.3. Listeners

Ten native-Thai listeners participated in this study. All the listeners were graduate students at the University of California at Berkeley. There were 8 males and 2 females with the age ranging from 22-38; the mean age was 26 years. All listeners grew up in Thailand and none had no hearing defects.

4.4. Task

The listeners were asked to listen to the processed tokens, to judge whether each token had a shorter or a longer allodiphthong, and to circle their answers on the answer sheets, provided in Thai. The listeners were told that the tokens they listened to carried no meaning.

The tokens were played by the ASPP program through a headphone for each listener, at a comfortable volume, in the Phonology Lab at the University of California at Berkeley.

4.5. Results

Figure 10 below presents the results of this study. In this figure, it can be seen that, for both tokens [piaʔ] and [piaa], the percentage of shorter-allodiphthong responses is higher when diphthongs have a shorter duration and the percentage of shorter-allodiphthong responses is lower when diphthongs have a longer duration.

Moreover, at mid range, for listeners to identify the shortened longer allodiphthongs as shorter allodiphthongs, the shortened longer allodiphthongs have to be shorter than the original shorter allodiphthongs.

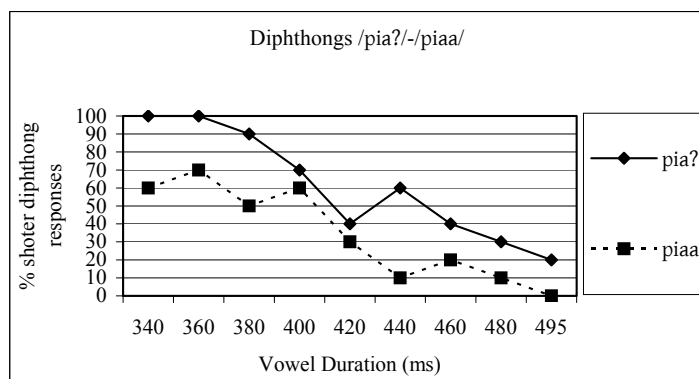


Figure 10. Diphthongs /pia?/ and /piaa/ from Study 3 Diphthongs.

4.6. Conclusion of Study B

It can be concluded that the duration of the second vocalic element functions as a main perceptual cue to differentiate between shorter and longer allodiphthongs.

5. Conclusion of Studies A and B

From previous literature (Abramson 1962, Naksakul 1998), Thai has three phonemic diphthongs /ia/, /ua/ and /ua/. From the two studies in this paper, it is found that the main cue to distinguish the three diphthongs is the quality of the first vocalic element. Moreover, the results confirmed that three diphthongs have phonetic short and long differences. It was found from the data in the two studies that shorter allodiphthongs occur in closed syllables; whereas, longer allodiphthongs occur in open syllables. Furthermore, the main acoustic and perceptual cue for shorter and longer allodiphthongs is the duration of the second vocalic element of diphthongs. Shorter and longer allodiphthongs do not have differences in diphthong quality.

ACKNOWLEDGMENTS

The author would like to thank Professor John J. Ohala for his guidance, advice, and comments. The author appreciates the help from Thai friends at Berkeley: both speakers and listeners.

References

- Abramson, Arthur S. 1962. *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington IN: Indiana University Center in Anthropology, Folklore, and Linguistics, Pub. 20.
- Naksakul, Kanchana. 1998. *Rabop siang phaasaa Thai*. (The Thai Sound System). Chulalongkorn University Press, Bangkok, Thailand. [In Thai].
- Roengpitya, Rungpat. 2001. A Study of Vowels, Diphthongs, and Tones in Thai. Ph.D.dissertation. University of California at Berkeley.
- Thonglor, Kamchai. 1977. *Lak phaasaa Thai* (Thai Grammar). Amorn Printing, Bangkok, Thailand. [In Thai].
- Upprakitsilpasaan, Praya. 1998. *Lak phaasaa Thai* (Thai Grammar). Thaiwattanaphanit Publishings, Bangkok, Thailand. [In Thai].

Dr. Rungpat Roengpitya
1229 Baromrachkumaree Building
Department of Linguistics
Faculty of Arts
Chulalongkorn University
Bangkok, 10320, Thailand

Rungpat.R@chula.ac.th

Relativization and nominalization in Bodic

SCOTT DELANCEY
University of Oregon

While the Tibeto-Burman (T-B) languages show considerable typological diversity along some dimensions—most notably in the structure and morphological complexity of the verb—in terms of most of the major typological parameters they show an impressive consistency. With the exception of the Karen branch, T-B languages are resolutely SOV, with all of the attendant stigmata (postpositions, postverbal auxiliaries, clause-final subordinators, etc.) Like many other SOV languages T-B languages are clause-chaining. And while, when considered from a T-B internal perspective, certain languages or branches do show some isolating (Loloish) or inflectional (Tibetan) behaviors, in a larger perspective the fundamental typology of the family is essentially agglutinative.

Besides this basic SOV-agglutinative structural plan, the synchronic syntactic organization and diachronic tendencies of these languages can be largely understood in terms of two syntactic tendencies: verb serialization (see Matisoff 1969, 1974, 1991, Smeall 1975, DeLancey 1991) and nominalization. The importance of both of these was first recognized by Matisoff in his analysis of Lahu,¹ and brought to the attention of the Tibeto-Burmanist world by two seminal papers (Matisoff 1969, 1972), which between them have informed the majority of the work in T-B syntax which has appeared since. The determining role of nominalization in T-B syntactic structure and diachrony is far too large a topic to do justice to here; my purpose is only to survey one manifestation, relativization, which throughout the family is based in the grammar of nominalization.

1.0 Nominalization in T-B syntax and diachrony

Early investigators of Tibeto-Burman languages seem to have had a sense of the centrality of nominalization to their syntactic organization, but were never really

1. In fact Matisoff clearly had a strong sense of the importance of verb serialization in T-B even before he encountered Lahu, as Matisoff 1974 was actually written ten years before that, prior to his first fieldwork on Lahu.

able to come to grips with its resolutely non-Indo-European manifestations. But the serious study of nominalization and relativization in Tibeto-Burman begins quite recently, with Matisoff's seminal paper "Lahu nominalization, genitivization, and relativization" (1972). In the ensuing 20 years the connection between these two phenomena has become received wisdom in the field; Bickel (1999) has recently referred to the "Standard Sino-Tibetan Nominalization" pattern (see also Herring 1991, Genetti 1992, Noonan 1997), and grammatical descriptions now routinely have a section on nominalized clauses in which relativization is discussed along with other uses of nominalization.

In Lahu, as described in detail by Matisoff (1972, 1982), a single morpheme, *ve*, functions as a nominalizer, complementizer, relativizer, and genitive marker. A similar complex of functions revolving around a single morpheme occurs in other T-B languages, e.g. Jingphaw, Southern Chin (Houghton 1892), and Chantyal (Noonan 1997). While the inclusion of genitivization on this list of functions is far from a universal T-B feature (though impressively widespread, as the list above suggests), the identity of relativization with nominalization constructions does seem to be nearly universal throughout the family, the only clear exceptions being a handful of Himalayan languages which are reported as having developed Indo-European-style relative pronoun constructions. Other apparent exceptions may have more interesting histories. One example is Burmese, which is a *prima facie* exception to the claim that relative clauses are universally nominalizations in Tibeto-Burman. The Burmese relative clause is an ordinary finite clause, ending in one of the finite declarative final particles *te* or *me*, and marked as subordinate by "induced" grammatical creaky tone (Okell 1969:18, 59-61, see also Wheatley 1982:104-9, Soe 1999:50-6, Herring 1991). But while induced creaky tone is not synchronically a form of nominalization, there is ample evidence to support Benedict's (1972) suggestion that it reflects the old PTB genitive morpheme **ki* (Thurgood 1981), and thus that the modern Burmese relative clause reflects an older construction more in line with the "Standard Sino-Tibetan Nominalization" pattern.

2.0 Variations on a theme

The fundamental relativization pattern is the same throughout the family: relativization is a subspecies of clausal nominalization. The modifying clause is nominalized, and then stands in either a genitive or an appositive relation to the head noun. Beyond this basic commonality, however, there is a substantial amount of variation in the complexity and organization of the nominalization—relativization system. In this section I will present three quite different patterns in some detail; we will see further variations on the basic theme in the following section.

2.1 A simple system: Classical Tibetan

Classical Tibetan exemplifies a very simple system, which I will argue can be ascribed to Proto-Bodic. The basic nominalizer, *pa*, is the only relative marker, and the relative construction is indifferent to distinctions of case, tense/aspect, and animacy.

2.1.1 Nominalization and relativization in Classical Tibetan

The relative construction is a nominalized clause functioning as either prenominal modifier of a head noun or a postnominal appositive (see Mazaudon 1978, Beyer 1992, DeLancey 1999). Pre-head relatives are always marked as genitive, indicating their subordinate status within the NP; post-head relatives never are. Relativization can be off of any grammatical role:

- (1) *shi-ba-'i* *ro*
die-NOM-GEN body
'a dead body'
- (2) *slob=dpon* *med-pa-'i* *brtul=zhugs-chan* *'di*
teacher not.exist-NOM-GEN ascetic DEM
'this ascetic who has no teacher'
- (3) *rgyal-po* *ngan-pa-'i* *g.yog=po* *rnams-kyis* *snying=rje*
king bad-GEN servant PLU-ERG compassion
 med-pa-s *bsad-pa='i* *bram=ze*
 not.exist-NOM-INSTR killed-NOM-GEN Brahmin
'the Brahmin whom the servants of the evil king mercilessly killed'²
- (4) *ltad=mo* *lta-ba-'i* *lam* *du*
sights see-NOM-GEN road LOC
'on the way to see the sights'

Although both the Classical Tibetan and the Lahu system involve a single nominalizer in a single construction, they differ significantly in that the Tibetan prenominal relative clause is marked as a dependent by genitive case, while in Lahu the nominalizer is itself the genitive morpheme. The motivation for the Tibetan construction is self-evident. The function of the genitive is to mark one nominal as subordinate to another, typically as a modifier. Since a nominalized clause is, by definition, nominal, it can, like any other nominal, be subordinated to a head noun using the genitive. The postnominal construction appears to be an appositive, a common construction in Tibetan and some other Bodic languages (see Watters 2002).

In older forms of Tibetan, *-pa* is a ubiquitous, general-purpose nominalizer:

2. This example is cited by Hahn (1974:74), who notes that it is typical of the extremely long and complex modifying clauses which are often found in Classical texts.

The infinitive, as well as the participles, makes use of the terminations in *pa* and *ba*, as do the substantives and adjectives. Very often it is practically impossible to decide at once, whether a certain word is meant to be an infinitive or a participle; *mthong-ba* may mean: to see, seeing, or seen; *zer-ba* may mean: to say, saying, or said. (Francke 1929:146)

By “participle” Francke, like some modern authors, is referring to a range of uses including what I have been discussing as the relative clause construction. For examples see Bacot 1948:72-4. Francke explicitly notes that *-pa*, both as a nominalizer (“infinitive”) and relativizer (“participle”) is indifferent to tense, which is to say that the aspectual restrictions on its use which we will see in modern Central Tibetan are not apparent in the Classical language.

2.1.2 *-pa* and *-ma*

Scholars generally distinguish this *-pa* from another identical morpheme which alternates with *-ma*, *-po*, and *-mo* in a rough system of gender marking in nouns and adjectives. In the modern language, many nouns referring to human beings or conspicuous animals are specified for gender by one of these suffixes: *rgyal-po* ‘king’, *rgyal-mo* ‘queen’, *btsun-pa* ‘monk’, *btsun-ma* ‘nun’, *chen-po* ‘elder brother’, *chen-mo* ‘elder sister’, etc. The suffixes appear in many other nouns, where they have no gender reference: *nyung-ma* ‘turnip’, *mjing-pa* ‘neck’, *bra-bo* ‘buckwheat’, *ri-mo* ‘picture’; the classic problem example is *bla-ma* ‘lama’, with the “feminine” *-ma* suffix but unambiguously masculine reference. Francke notes that many nouns which in the modern language have no suffix occur with one in the older texts, from which he infers an earlier system of grammatical gender. A few adjectives inflect for gender through alternation of these suffixes, e.g. *rgad-po*. *rgad-mo* ‘old’, *mdzes-po*, *mdzes-ma* ‘beautiful’, but for the most part the use of these suffixes in adjectives is lexically fixed and invariant.

While grammatical marking of gender is not widespread in Bodic, the traditional attribution of the Tibetan pattern to older levels is supported by its attestation in Limbu (sec. 2.2). The Limbu evidence supports the idea that the gender system is closely related to the *-pa* nominalization system; in Limbu they are not distinct. The forms in *o* are probably originally demonstratives, related to such forms as the Tibetan 3rd person feminine pronoun *mo*, and a definite suffix *po* ~ *bo* which still occurs in the western dialects. In Ladakhi this occurs “with singular nouns when the person or object being referred to ... has to be particularized” (Koshal 1979:108), as in *i-mi-rde=mo-bo* “this-man-good-SPEC” = ‘this good man’. (Note the lexically specified “feminine” suffix in *rde=mo* ‘good’, although the reference of this example is masculine). Francke finds this form more widely used in earlier forms of the language (cf. also Csoma de Kőrös 1834:108):

The demonstrative pronoun is often replaced or strengthened by the optional article *bo*, which may be added to any noun, furnished with an article (sexual determination) or not. It emphasizes the word to which it is added: *khyi-bo*, that dog; *myi-bo*, that man; *'adi-khang-pa-bo*, this house. (Francke 1929:112)

2.2 Case distinctions in Limbu

Limbu has a more elaborate system than Classical Tibetan, but it is built around the same *-pa* morpheme. There are three relative constructions in Limbu, distinguished by the grammatical role of the head noun in the relative clause. The simplest is almost exactly like the Tibetan—a clause nominalized with *-pa*. However, in Limbu the relative clause does not have genitive marking:

- (5) *na:m-ille* *cə-ba* *ho:rik*
 sun-ERG eat-NOM skin
 ‘skin which has been burned (‘eaten’) by the sun’ (van Driem 1987:198)

Van Driem (1987) does not describe the distribution of this construction in terms of case or role, but Genetti (1992:412) notes that in his data it occurs only as nominalization or relativization off a non-subject. Subject relativization is the function of the “active participle” construction, in which the verb is marked with a prefix *ke-* as well as the *-pa* nominalizer:

- (6) *na:-nu* *ke-bhem-be-n* *ke-ndzum-in*
 there-from AP-come-NOM-ABS 2sg-friend-ABS
 ‘your friend who is coming from over there’ (van Driem 1987:203)

This construction is found also in the closely-related Athpare, where the “agentive participle” *ka-STEM(-ba)*, as in *ka-phu-ba* ‘tailor’, *ka-khub-ba* ‘thief’, can be “used as a noun or as an attribute” (Ebert 1997:79).³ The Athpare form can be inflected for the person of the object argument by means of possessive prefixes: *a-ka-lem* ‘who beats **me**’, *ka-ka-pik* ‘who speaks to **you**’. A third construction, deriving Patient nominalizations, involves a passive participle in *-mna* plus the *-pa* nominalizer.

Limbu also shows the same adjective-forming function of **-pa* that we find in Classical Tibetan. Indeed, there can be little doubt simply on the basis of these two languages that some version of this formation existed in Proto-Bodic; cf.

3. The prefix is also strikingly similar to the Angami (Naga) general nominalizer (and relativizer) *ka-* (Herring 1991:56-9). Thus while Genetti's suggestion (1992:413) that simple nominalization with **-pa* is older in Limbu than the composite form is undoubtedly correct, the prefix might itself be quite old in some function which allowed its development into a modifying construction in more than one language.

Limbu *təŋ-pa*, Tib. *thung-ba* ‘short’, which presumably represents a direct inheritance from a Proto-Bodic form in both languages. In Limbu, as in CT, these formations are marked for gender:

- (7) *kheŋ nu-ba co:k*
 3rd good-NOM be
 ‘he/it is good’
- (8) *kheŋ nu-ma co:l*
 3rd good-NOM/FEM be
 ‘she is good.’

While van Driem does not discuss this phenomenon at length, it appears that gender marking in Limbu is much more productive and consistent than in any attested variety of Tibetan.

2.3 Animacy and number in Newari

Newari has three nominalizers which occur in relative clauses. The system distinguishes animacy and number: *gu(li)*⁴ has inanimate reference; animate reference uses *-mha* for singular and *-pi* (< *-pa-ni*) for plural.

These show the standard Sino-Tibetan range of functions (Kölver 1977). They function as a nominalizers and markers of relative clauses:

- (9) *ji-nə khun-a-gu*
 I-ERG cook-PART-NOM
 ‘what I cooked’, ‘my cooking’
- (10) *ji-nə khun-a-gu la*
 I-ERG cook-PART-NOM meat
 ‘the meat which I cooked’
- (11) *ji-nə nya:n-a:-mha*
 I-ERG buy-PART-AN.NOM
 ‘the [living] thing that I bought’
- (12) *ji-nə nya:n-a:-mha nya:*
 I-ERG buy-PART-AN.NOM fish
 ‘the [live] fish that I bought’
- (13) *ji-nə nya:n-a:-pi nya:*
 I-ERG buy-PART-PL.NOM fish
 ‘the [live] fishes that I bought’

4. Newari has undergone considerable attrition of final consonants, and in a few cases final syllables; however the eroded segments are retained in certain case forms in the nominal declension. *gu(li)* has the nominative form *gu*, but the second syllable surfaces in the irregular locative and ablative/instrumental forms *guli* and *guli*.

They also sometimes occur in genitive constructions:

- (14) *ra:m-ya:-gu* *ta:sbir*
 Ram-GEN-NOM picture
 ‘Ram’s picture’
- (15) *ra:m-ya:-mha* *khica:*
 Ram-GEN-NOM dog
 ‘Ram’s dog’

However, they are not obligatory:

- (16) *ra:m-ya:* *kala:*
 Ram-GEN woman
 ‘Ram’s wife’

It turns out, not surprisingly, to be impossible to specify in formal terms the precise conditions under which *gu(li)*-series marking varies. The alternation is discussed at length by Kölver (1977), who demonstrates a set of semantic conditioning factors which can be subsumed under a general statement that the presence of a *gu(li)* morpheme indicates a greater, and the absence a lesser, degree of conceptual independence between the dependent and the head NP—recalling a functionally similar distinction in Lahu between “*ve*-full” and “*ve*-less” genitive constructions. I cannot here enter into the interesting question of the relationship between this system and the Lahu phenomenon, in which the nominalizer has apparently simply replaced an older genitive form. But the facts of the Newari genitive construction are relevant to the problem of the origins of the innovative nominalization-relativization system, to which we will return in section 3.3.

2.4 Role and aspect in Lhasa Tibetan

In contrast to the simplicity of the relative construction in Lahu and Classical Tibetan, spoken Lhasa Tibetan has an unusually complex system of relativization (Mazaudon 1978, DeLancey 1999). There are four distinct relative markers, the choice being determined by the semantic role of the head NP in the relative clause, and to some extent also by the time reference of the relative clause. The four in Lhasa are *mkhan* for actor, *sa* locative/dative, and the default nominalizers *yag*, used for patients and instruments in non-perfective relative clauses, and *pa*, used in perfective relative clauses when the head noun is not the actor. All four are clearly nominalizers, occurring in the productive derivation of nouns like *nyo=mkhan* ‘buyer’, *nyo=yag* ‘goods, stuff to buy’. With all but *mkhan* the relative clause is or may be also marked as genitive (see DeLancey 1999).

The inherited Classical Tibetan construction with *-pa* + genitive can be used in Lhasa only in relative clauses with perfective time reference where the

head noun is coreferential with a non-actor NP in the relative clause, as in:

- (17) *kho-s bsad-pa-'i stag pha=gi red*
 he-ERG kill(PF)-NOM-GEN tiger that be
 'That is the tiger which he killed.'

Compare *kho-s bsad-pa* 'what he killed', as in:

- (18) *kho-s bsad-pa stag red*
 he-ERG kill-NOM tiger be
 'What he killed is a tiger.'

If the relative clause has other than perfective reference, a different nominalizer must be used; with patient head noun this is =yag:

- (19) *kho-s gsod=yag-gi stag*
 he-ERG kill/FT-NOM-GEN tiger
 'the tiger that he will kill.'

(This aspectual distinction also characterizes the use of *-pa* and =yag as nominalizers; see Goldstein 1973).

When the NP head is coreferential with the actor of the relative clause, the relative clause is marked with the agentive nominalizer =mkhan, without genitive marking:

- (20) *stag gsod=mkhan mi*
 tiger kill-NOM man
 'the man who killed/kills/will kill the tiger.'

(cf. *stag gsod=mkhan* 'one who killed the tiger, tiger-killer').

Locative, dative, and benefactive nominalizations and corresponding relative clauses are formed with =sa:

- (21) *kho sdod=s-'i khang=pa*
 he stay-NOM-GEN house
 'the house where he lives.'

(cf. *kho sdod=s* '(the place) where he lives').

- (22) *nga-s deb sprod=s-'i mi*
 I-ERG book give-NOM-GEN person
 'the person who I gave the book to.'
- (23) *nga-s kha=lag bzo=s-'i mi*

I-ERG food cook-NOM-GEN person
'the person who I cooked food for.'

(the nominalizer =sa requires the imperfective stem; the resulting relative clause is neutral as to time reference).

3.0 Bodic relative clauses in diachronic perspective

3.1 Tamangic

The Tamangic languages reflect minor variations on what must originally have been exactly the Classical Tibetan pattern. There is one relative construction, indifferent to case or tense/aspect, built with a reflex of the *-pa nominalizer. The one difference is that in many of these languages the nominalized clause appears prenominally as a relative clause with no trace of genitive marking. This is the case, for example, in Thakali (Georg 1996):

rástá¹ so³-pa mĩ³
nation create-NOM person
'founder of a nation'

kĩ³-se t'uy²-pa kju²
2sg-ERG drink-NOM water
'the water that you drink'

And the same situation is found in Chantyal, described in detail by Noonan (1997).

But other languages of the branch show the Classical Tibetan pattern, or at least traces of it. Noonan notes that the relative construction with genitive is preserved unchanged in Gurung (Glover 1974:99-100):

(24) *cú-r yú-ba*
this-LOC come.down-NOM
'the one who came down here'

(25) *cú-r yú-ba-e mxi*
this-LOC come.down-NOM-GEN person
'the person who came down here'

Manange (Hildebrandt to appear) seems to represent a transitional stage between the older Gurung-type pattern and the genitive-less construction of Chantyal and Thakali. Manange has a single relative clause construction, with a clause marked with *-pc* preceding the head noun:

(26) *³srə ³se-pə ²mi=ko*
goat kill-NOM person-DEF

‘the man who killed the goat’

However, Hildebrandt reports that the nominalizer/relativizer is in apparent free variation with a form *-pe*, with a raised vowel, which obviously correlates with Gurung *-ba-e*, both cognate with the Tibetan genitivized form *-pa-i*.

Thus all of the Tibetan and Tamangic languages have, or can be shown to have had, a specific relative clause structure: a clause nominalized with **-pa*, marked as genitive, and preposed to a head noun, and we can easily reconstruct this construction for the common ancestor of Tibetan and Tamangic.

3.2 Kiranti

The Kiranti languages consistently show the typical Tibeto-Burman association between nominalization and relativization. However, the actual morphemes involved, and the types of distinction found in the system, show a fair amount of variation from one language to the next. Some languages show a distinction of number. In Athpare (Ebert 1997a) relative clauses are formed from either a finite or non-finite verb form plus one of the nominalizers *-na* ‘singular’ and *-ga* ‘plural’. The latter must certainly be related to the plural agreement suffix *-ga* which occurs in the negative and progressive paradigms of Athpare (and, sometimes restricted to 1st person plural or inclusive, in a number of other Kiranti languages). Since it does not occur in a nominalizing function in closely related languages, and has a ready language-internal etymology, we can take this distinction in Athpare to be a quite recent development.

We do find clear attestation in several languages of this branch of the **-pa* construction, which is thus shown to be of at least Proto-Bodic provenience. At a glance, it seems that this is the only morpheme which can be reconstructed in a relativizing function for Proto-Kiranti, although some suggestive similarities among other forms might turn out, with more comprehensive data from this branch, to show that some innovation from the simple Proto-Bodic system had already begun around Proto-Kiranti times. We will look here at only two additional Kiranti languages, Sunwar and Thulung.

Sunwar has two relative clause markers, the familiar *-ba* and *-šo*, both basically nominalizers. (The *-ba* form is identical to the non-past 3rd person singular form (DeLancey 1992a), including the non-past stem augment). As in Central Tibetan, the Sunwar relative formation is sensitive to the aspect of the RC and to the grammatical role of the head NP. For relativization of an intransitive subject, *-ba* tends to occur in perfective, and *-šo* in imperfective, relatives:

- (27) *ŋàl-ba ?al*
 cry-NOM child
 ‘child who cried’

- (28) *ŋàʔ-šo ʔal*
 cry-NOM child
 ‘the child who is crying’

However, while the restriction of *-šo* in intransitive clauses seems to be consistent, *-ba* can occur in non-perfective as well as perfective relatives:

- (29) *ʔaî-ba mur*
 go-NOM person
 ‘the person who will go’

For transitive clauses, the basic system is that *-ba* forms subject relative clauses, and *-šo* non-subject, irrespective of tense/aspect:

- (30) *ʔal-kal khændi kye-ba mur*
 child-DAT candy give-NOM person
 ‘the person who gave the child candy’
- (31) *ʔa khændi kye-šo ʔal*
 3rdPOSS candy give-NOM child
 ‘the child who s/he gave candy to’
- (32) *ʔa ʔal-kal kye-šo khændi*
 3rdPOSS child-DAT give-NOM candy
 ‘the candy which s/he gave to the child’

In the non-subject *šo*-relative of a transitive clause, the actor is indexed by a possessive pronominal prefix. These are proclitic to the clause (not the verb): */ã-tup-šo/* ‘the one who I hit’, */i-tup-šo/* ‘the one who you (sg.) hit’. Since the possessive clitic attaches to the clause, not the verb, it can cliticize to the verb only when there is no intervening object NP (see exx. 31-32).

In Thulung (Lahaussais 2002), finite clauses can be nominalized, and used as relative clauses, by two suffixes, *-m* and *-mim*. Although the distribution of these in the modern language is partly phonologically conditioned, Lahaussais reconstructs an earlier stage in which they contrasted in aspectual reference, with *-m* forming perfective, and *-mim* imperfective, relative clauses. Both of these appear to be derived somehow from the certainly older “participial” *-ma*.

Of deeper historical interest is the “participial” construction, with a non-finite verb and one of two relativizers: *-ma* for perfective and *-pa* for imperfective relative clauses. This *-pa* is our familiar **-pa* construction:

- (33) *nem bane-pa a-luak*
 house make-NOM 1st-younger.brother
 ‘my brother who builds houses’

This two-way distinction, based on aspect, is similar to what we observed in Sunwar (ignoring the effect of subjecthood in Sunwar), with two significant differences—the aspectual orientation of the *-pa form, and the identity of the form which contrasts with it. Both differences would require somewhat imaginative explanation if we were to try to ascribe the aspectual distinction to the common ancestor of the two languages, but make perfect sense if we assume that Proto-Kiranti had only the *-pa construction, indifferent to aspect, and that the two languages later independently (or perhaps as part of an areal tendency to which both were subject) innovated an aspectual distinction. Each would then have innovated a new relativizer in one or the other aspectual category, leaving the original *-pa form in the other. The fact that, in Sunwar, the use of -šo is more clearly defined and restricted than that of -ba is perfectly consistent with this scenario.

However, while I see no evident Bodic etymology for Sunwar -šo, the alternate Thulung form may be less mysterious. An alternation between -pa and -ma in the formation of modifying constructions immediately brings to mind the Tibetan “gender” suffixes, in particular their use with adjectives. And, indeed, Thulung -ma occurs in a past participial construction which looks very much like the Tibetan adjectival construction:

- (34) *khok-ma* *jam*
 cook-NOM rice
 ‘cooked rice’

The apparent reanalysis of an opposition which originally marked gender into one reflecting aspect is a bit mysterious, but examples like (34) make the comparison of Thulung -ma with its Limbu and Tibetan resemblants unavoidable.

3.3 The development of innovative systems

We have seen, even in this superficial survey of a handful of languages, that there is substantial variation among Bodic languages in their relativization systems. One thing which does not vary is the close relationship between relativization and nominalization. While the distinctions encoded by the different nominalizers in Newari and Tibetan are quite distinct, the existence in both languages of several morphemes with exactly the same set of functions further demonstrates that the overlap of these apparently distinct functions is not accidental, but integral to the grammar of the language. On the basis of such data we can argue that the nominalization function is chronologically and systemically prior to relativization, which is merely one specialized function of nominalization.

We have seen comparative data demonstrating the relative recency of the modern Kathmandu Newari and Lhasa Tibetan nominalization-relativization

systems—a conclusion which, indeed, seems fairly clear on simple inspection. For both of these languages, documentary historical evidence further confirms this conclusion, showing the development of lexical nouns into new nominalizers, which then quickly begin to show up as relative clause constructions. The data from both languages constitute strong evidence that there is a principled basis to the nominalization-relativization syncretism, since we see new nominalizers being quickly and consistently used as relativizers.

Classical Tibetan has semantically more specific nominalizers than *-pa*, one of which is the agentive nominalizer *=mkhan* which is the Lhasa agentive nominalizer/relativizer. According to Jäschke (1881; cf. Beyer 1992), in earlier texts this has a clearly derivational function, occurring with both nouns and verbs to produce items like *shing=mkhan* ‘carpenter (*shing* ‘wood’), while in later texts it is used productively to nominalize clauses, as in Jäschke’s example:

- (35) *nga-i bu=mo ‘dod-mkhan*
 I-GEN daughter desire-NOM
 ‘such as are courting my daughter’

Jäschke gives the older lexical sense of *mkhan* as ‘one who knows a thing thoroughly, making a trade or profession of it’; the stem occurs also in the noun *mkhan=po* ‘clerical teacher, professor, doctor of divinity, abbot’, and is etymologically related to the adjective *mkhas=pa* ‘skilled, skillful’. Thus the earliest nominalizing function of the morpheme is a lexical one analogous to English *-ist* or the *-er* of *hatter*, *teacher*; the more general functions of agent (and later subject, see below) nominalization developed relatively recently.

The origin of *sa* is similar. The original function of *sa* is as a noun meaning ‘earth, ground’, with a secondary sense of ‘place’, attested in derivational nominalizations such as *yod-sa* ‘place of residence’ (cp. *sdod-sa* in (21) above). The development is thus parallel to that of *mkhan*, from a derivational function with specific semantic content to a semantically bleached general nominalizing, and hence relativizing, function.

In the case of Kathmandu Newari, Kölver (1977) infers a (presumably rather shallow) nominal origin for the nominalizers on the basis of the alternations in the genitive construction. She points out that a comparison of the genitive construction with and without a *gu(li)* morpheme suggests that these morphemes were originally, and might still be analyzed as, noun heads, thus explaining the genitive marking on the noun to which they are suffixed. Indeed, *mha* still has, besides its grammatical uses, an independent function as a lexical noun meaning ‘body’. This hypothesis is also consistent with the fact that the *gu(li)* morphemes inflect for case just like nouns (albeit with some morphological irregularities) and the participation of all three of the *gu(li)* morphemes in the nominal classifier

system. Hargreaves (1989) shows that in Late Classical Newari texts we commonly find relative clauses with only a participial form of the verb, and no nominalizer at all:

- (36) *ji-n bisya hayâ ratn*
 I-ERG give bring/PASTCONJUNCT jewel
 ‘the jewel which I brought’

In these texts the conjunct/disjunct opposition (Hale 1980, Hargreaves 1991, DeLancey 1992b) in relative clauses does not mark the person-oriented conjunct/disjunct distinction, but correlates with a distinction between subject vs. non-subject relative clauses.

Thus, while we cannot provide an etymology for *-gu(li)*, we have at least a shallow etymology for *-mha*, and structural evidence pointing to a fairly recent nominal origin for the entire category. Thus the elaboration of the nominal-relative system seems to have involved essentially the same diachronic mechanisms in Newari as in Tibetan, although the direction of elaboration is quite different. The one problematic element in this account is the plural animate form. In older texts this is consistently *-pani*, which has an obvious Bodic etymology: our familiar **-pa* nominalizer plus a pluralizing *-ni*, corresponding to the plural *ni* found in the verb agreement systems of both West Himalayish and Kiranti languages (Bauman 1975). This is primarily a 2nd person plural form in Kiranti, and this seems to be its original sense (Bauman 1975:140-1), but in Pattani and other western languages it has lost any association with person, and so it could easily have done in Newari.

3.4 The **-pa* construction

Benedict (1972:96) reconstructs a “verbal noun (infinitive) suffix” *-pa* ~ *-ba*, as well as ‘gender’ (scare quotes are Benedict’s) suffixes masculine *-pa*, feminine *-ma*, for PTB; he notes the likely connection between the nominalizer and the masculine gender suffix. There is no doubt that the Classical Tibetan nominalization/relativization system can be ascribed to Proto-Tibetan. As we have seen, the same prenominal, genitive-marked *-pa* relative construction can be reconstructed for Proto-Tamangic. Thus we can confidently reconstruct this much of the Classical Tibetan system for Proto-Bodish.

In that light, it is interesting to note the rapidity with which the system has changed, and diversified, within Tibetan. Looking only at Central Tibetan, we can see in the variety of Lhasa described here a very innovative system, still incorporating the original *-pa* construction, but with the overall system drastically expanded and elaborated by the incorporation of several new nominalizers. Francke (1929:146-7) states that the “participial” function of *mkhan* is found

already in *Mi-la-ras-pa*, which would place the origins of this construction back to at least the 14th century. However, in my admittedly rather superficial acquaintance with that text it appears that the Classical construction with *-pa'i* is still predominant. From this I would infer that the transition from the Classical to the modern Lhasa system occurred within the span of a millennium. But some Central and other dialects seem to have moved even further in this time: in Tshangla (Andvik 1999), in the western Central dialect of Kyirong (Huber 2002), and reportedly also in some varieties of Lhasa, *mkhan* is becoming the default relativizer, used without regard for role or aspect, and thus replacing *-pa* in most of its remaining functions. Extrapolating this path of development, we can easily imagine a Central dialect of the near future (or the present, for that matter), in which we have something closer to the original Classical system, but with the new form *mkhan* completely replacing the original *-pa*. Note that in Newari we can still unearth a relic of the old *-pa* construction in the modern system (although this is made much easier by the existence of textual materials documenting an earlier stage of the language), but the overall system really must be thought of as completely innovative, simply replacing rather than elaborating the original simple pattern.

And the same story is to be seen in Kiranti. Every Kiranti language for which I have seen documentation has a more elaborate system than that of Classical Tibetan. However, the elaboration has taken place in different directions, involving different etyma, in the different languages. On the basis of the comparative data so far available to me, I can see no basis for attributing any of the attested elaborations even to Proto-Kiranti, much less Proto-Bodic. (Of course, some of the elements of the modern systems, such as the *ke-* prefix in Limbu and the *-ga* plural in Athpare, clearly had some grammatical function considerably earlier on, but their incorporation into the nominalization/relativization system must postdate Proto-Kiranti).

However, we can see from the Limbu, Sunwar, and Thulung data that the *-pa* nominalizer was part of the Proto-Kiranti system, and, since it is the only etymon in this small sample which has any claim to PK status, we may infer, at least provisionally, that PK inherited a version of the Proto-Bodic system built entirely around *-pa*, and that all additional complexities of the modern systems are secondary developments. The inference that the basic PK relative construction was the *-pa* construction is buttressed by the fact that in all of these languages we find examples of *-pa* relative clauses which exactly parallel some of its distribution in Classical Tibetan. But exactly *which* uses varies from one language to another. In Limbu it is non-subject relatives, in Thulung imperfective relative clauses; in Sunwar the distribution of *-ba* is less clearly defined than that of the contrasting *-šo*, but is associated with perfective relative clauses. The obvious inference is exactly the conclusion that we have already drawn—that in the

common ancestor of these languages the *-pa* construction was used for all kinds of relativization, and that each language individually has innovated one or more newer constructions which have taken over one or another part of the original range of *-pa*.

References

- Andvik, Erik. 1999. *Tshangla grammar*. unpublished Ph.D. dissertation, University of Oregon.
- Bacot, Jacques. 1946. *Grammaire du tibétain littéraire*. Paris: Librairie d'Amérique et d'Orient.
- _____. 1948. *Grammaire du tibétain littéraire: Index morphologique*. Paris: Librairie d'Amérique et d'Orient.
- Bauman, James. 1975. *Pronouns and pronominal morphology in Tibeto-Burman*. unpublished Ph.D. dissertation, University of California at Berkeley.
- Benedict, Paul. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.
- Beyer, Stephan. 1992. *The Classical Tibetan language*. Albany: State University of New York Press.
- Bickel, Balthasar. 1999. Nominalization and focus in some Kiranti languages. pp. 271-96 in Yadava and Glover.
- Chhange, Lalnunthangi. 1993. *Mizo syntax*. unpublished Ph.D. dissertation, University of Oregon.
- Csoma de Körös, Alexander. 1834. *A grammar of the Tibetan language, in English*. Calcutta: Asiatic Society of Bengal.
- DeLancey, Scott. 1991. The origins of verb serialization in Modern Tibetan. *Studies in Language* 15.1:1-23.
- _____. 1992a. Sunwar copulas. *Linguistics of the Tibeto-Burman Area* 15:1.31-38.
- _____. 1992b. The historical status of the conjunct/disjunct pattern in Tibeto-Burman. *Acta Linguistica Hafniensia* 25:39-62.
- _____. 1999. Relativization in Tibetan. pp. 321-49 in Yadava and Glover.
- Ebert, Karen. 1997. *A grammar of Athpare*. Munich: LINCOM Europa.
- Francke, A.H. 1929. Addenda to Jäschke's *Tibetan Grammar*. pp. 105-60 in Jäschke 1929.
- Genetti, Carol. 1992. Semantic and grammatical categories of relative clause morphology in the languages of Nepal. *Studies in Language* 16.405-427.
- _____. 1994. *A descriptive and historical account of the Dolakha Newari dialect*. (Monumenta Serindica 24). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Georg, Stefan. 1996. *Marphatan Thakali: Untersuchungen zur Sprache des Dorfes Marpha im Oberen Kali-Gandaki-Tal/Nepal*. (LINCOM Studies in Asian Linguistics 02). Munich: LINCOM Europa.
- Goldstein, Melvyn. 1973. *Modern Literary Tibetan*. Occ. Papers of the Wolfenden

- Society on Tibeto-Burman Linguistics, 5. Urbana, IL: Center for Asian Studies, University of Illinois.
- Hahn, Michael. 1974. *Lehrbuch der klassischen tibetischen Schriftsprache*. Bonn: Michael Hahn. (repr. Swistal-Odendorf: Indica et Tibetica Verlag, 1996).
- Hale, Austin. 1980. Person markers: Finite conjunct and disjunct verb forms in Newari. pp. 95-106 in R. Trail, ed., *Papers in South-East Asian Linguistics* no. 7. (Pacific Linguistics series A, no. 53). Canberra: ANU.
- Hargreaves, David. 1989. The morphosyntax of relative clauses in Late Classical and Kathmandu Newari. Presented at the 22nd International Conference on Sino-Tibetan Languages and Linguistics.
- _____. 1991. *The concept of intentional action in the grammar of Kathmandu Newari*. Ph.D. dissertation, University of Oregon.
- Herring, Susan C. 1991. Nominalization, relativization, and attribution in Lotha, Angami, and Burmese. *LTBA* 14.1:55-72.
- Hildebrandt, Kristine. to appear. A sketch grammar of the Manange language. in Carol Genetti, ed., *Tibeto-Burman languages of Nepal: Manange and Sherpa*.
- Huber, Brigitte. 2002. *The Lende subdialect of Kyirong Tibetan: a grammatical description with historical annotations*. Ph.D. dissertation, Universität Bern.
- Kölver, Ulrike. 1977. Nominalization and lexicalization in Newari. *Arbeiten des kölnener Universalien-Projekts*, no. 30.
- Jäschke, H.A. 1881. *A Tibetan-English dictionary*. London: Routledge and Kegan Paul.
- _____. 1929. *Tibetan Grammar, with addenda by A.H. Francke*. Berlin: Walter de Gruyter.
- Koshal, Sanyukta. 1979. *Ladakhi grammar*. Delhi: Motilal Banarsidass.
- Lahaussais, Aimée. 2002. Nominalization, relativization, and genitivization in Thulung. in this volume.
- Matisoff, James. 1969. Verb concatenation in Lahu: The syntax and semantics of 'simple' juxtaposition. *Acta Linguistica Hafniensia* 12.2:169-120.
- _____. 1972. Lahu nominalization, relativization, and genitivization. pp. 237-57 in J. Kimball, ed., *Syntax and Semantics I*. New York: Seminar Press.
- _____. 1974. Verb concatenation in Kachin. *Linguistics of the Tibeto-Burman Area* 1.1:186-207.
- _____. 1985. God and the Sino-Tibetan copula, with some good news concerning selected Tibeto-Burman rhymes. *Journal of Asian and African Studies* 29,1-81.
- _____. 1991. Areal and universal dimensions of grammaticalization in Lahu. pp. 383-453 in Traugott and Heine 1991.
- Mazaudon, Martine. 1978. La formation des propositions relatives en tibétain. *Bulletin de la Société de Linguistique de Paris* 73.401-414.
- Noonan, Michael. 1997. Versatile nominalizations. *Essays on language function and language type*, ed. by Joan Bybee et. al., 373-394. Amsterdam: Benjamins.
- Okell, John. 1969. *A reference grammar of colloquial Burmese*. London: Oxford

Scott DeLancey

University Press.

Soe, Myint. 1999. *A grammar of Burmese*. Ph.D. dissertation, University of Oregon.

Thurgood, Graham. 1981. *Notes on the origins of Burmese creaky tone*. (Monumenta Serindica 9). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Traugott, Elizabeth, and Bernd Heine, eds. 1991. *Approaches to grammaticalization*, Volume II. Philadelphia: Benjamins.

Wheatley, Julian. 1982. *Burmese: A grammatical sketch*. Ph.D. dissertation, University of California at Berkeley.

Yadava, Yogendra, and Warren Glover, eds. 1999. *Topics in Nepalese linguistics*. Kathmandu: Royal Nepal Academy.

Scott DeLancey
Department of Linguistics
1290 University of Oregon
Eugene, OR 97403-1290, USA

delancey@darkwing.uoregon.edu

Kinship and Spirit Terms Renewed as Classifiers of “Animate” Nouns and Their Reduced Combining Forms in Austroasiatic

ANNE DALADIER
CNRS, France

I take up the hypothesis proposed by Schmit (1904) and Henderson (1976) and analyse the initial velar in initial clusters of consonants in Austroasiatic (AA) languages as the initial of underlying prefixes. More precisely, War-Khasic (WK) (a conservative western MK group of languages, not Khasi dialects, see Daladier (2002b)) has preserved better than other AA languages not only initial clusters of consonants but the very use of four kinship terms as affixes which appear to have played the role of some kind of “animate” noun classifiers. Many of such “animate” nouns also meet in their word formation one of two proto-AA spirit names **zjaŋ* and **ruŋ*. This double hypothesis is based on the comparison of about one hundred words in WK with MK and Munda already published data. It appears that vestiges of the use of these kinship terms as “animate” classifiers may be reconstructed in cognate lexicons of MK and Munda languages. The complete kin terms cannot always be reconstructed, but regular similarities in what is left of the disyllabic or monosyllabic words having or not kept an initial voiced or voiceless velar consonant indicate both:

- a peculiar AA “animate” notion related to a conception where animate beings are generated through a clan conception of the world. This conception involves a mother, a cultural and a biological father. Animate nouns express edible plants (fruits, seeds or rhizomes), trees, parts of the body, totemic animals, insects and plants used in medicines, natural phenomena such as stones, rocks stars, earth, rivers, mud, mountains, and vital liquids, such as sap, blood, sperm, water, tears, oil.

- different paths of morphophonemic reductions and different paths of grammaticalization (and shifts). Reduced (*k/?*)V and (*ʒ*)(*j*)V elements have been further grammaticalized as different kinds of pronouns. AA third person pronouns may be considered as gender/number kind of anaphoric ‘classifiers’. These pronouns are used again as nominal determiners, they take the forms of autonomous clitics indicating three genders and one plural for the non inflecting

WK nouns and four affixes classifying animate/non animate (usual moving being notion) dual and plural nouns in different Munda languages (these values are found in clitics of different MK languages). Associated with distal and yonder deictics, the AA third person pronouns have produced prepositions, conjunctions, relative pronouns and interrogatives pronouns, or indefinites when agglutinated with grammaticalized forms of ‘one’. Some of these elements are used again with predicates and then indicate various possible combinations of subject and also object thematization (either in affixed forms in Munda or independent clitic association in MK see Daladier 2002a and to appear).

The understanding of the AA original system of velar “animate” classifiers may also help us to retrace AA borrowings in Tibeto-Burman and Tai-Kadai cognate names.

1. Kin terms: *ka, *kur, kñi, *koŋ and spirits *ruŋ *zjaŋ as noun formatives

*ka > k(h)a/a/ə/ε. War-Khasic ka ‘give birth’, ‘be born’, ‘bear’, ‘engender’, ‘womb’, ‘relatives on the side of the father’.

Santali: *kaka khura* ‘paternal uncle’ and *kaka baba* ‘stepfather’, have *ka* still related to the relatives on the father’s side and *kuma* (all over the Kherwari group) ‘maternal uncle’ with *ku-* < *kur* relatives on the mother’s side + *ma* as in Amwi *mama* maternal uncle, distributional alternant of *kñi*.

kən < *ka* + *-n*- deictic is found in all AA languages for animate beings who bear their children and for the names of the earth, the mud, the deep jungle. ‘Woman, female’: Khasi *kenthej*, Amwi *henthe*, Old Khmer, *kəndəw* ‘wife of an official’, *kəndər* ‘wife’, Old Mon *kəntər* ‘origin, coming into existence, source’, Old Mon, *kənteh* ‘dust from which the earth developed’, Shorto. Khasi ‘earth’ *ktē?*. Katuic: Souei has *ktē?* ‘earth’, Ferlus (1974). Bahnaric: Köho *tiah*, Stieng *tēh*, Chrau *nəteh*, Sidwell (2000). Khmer *ti*. Wa languages, Diffloth(1980): Samtao *kate?*, Tailoi *kade*, Kien Ka *kati* involve *ka and *te* as something close to an AA source. Munda names may be related to the same combining elements: Korku, Mundari, Ho have *oŋe* ‘earth, soil, ground, land’ also Korku *kheŋi* ‘field’, Sora *gade*. ‘jungle, forest’: Sora *kənrəŋ* (see *ruŋ below), Santali *zongol* ‘forest, jungle, place full of undergrowth’ corresponds to the older form *gazar* according to Bodding. *gazar* might be derived from *ka+ zahe* ‘sacred grove’, see below. Wa languages often have *-Vk* rather than, or in a reduplicated addition to, *kV-* in many of their lexical elements cognate with WK, *teak* ‘deep jungle wilderness, hill’ might contain *ka and *te* as in the names of the earth related to the notion of origin and fecundation already listed.

ka is found in ethnonyms of men and in animal names, probably originally totemic like *ka* ‘fish’ or ‘crab’ (the ancestor of the river) WK **katam*, Kherwari **kaTkom*, and in many insects used in WK and Santali medicines like the scorpion,

the centipede, different kind of worms and grasshoppers, see Bodding (1925) for Santali recipes.

ka is found in plant names especially for trees in MK and in Munda, also in the name of rice seed (to plant) in many MK and Munda languages: Khasi, *khāo*, Santali *caole* ‘the husked kernel of cereals, especially of rice’, Bodding. Other WK names of the husked rice, the rice to keep for the house, and the paddy usually contains *kur* and its variants (see below).

***kur** > *k(h)u/o(l)* , *ʔu/o*, *kl-* also *hor-*, *kor-*, Munda *g(u/o)-* War-Khasic *kur* ‘clan, relatives on the side of the mother’.

Kur is found in Munda and MK, kinship terms, ethnonyms of the men, in different crucial edible plants, in several totemic animals and medicines, in important parts of the body, in the names of the mountain.

kur appears in the names of ‘husked rice, paddy’, War-Khasic *kūbaʔ* < *ku(r)* ‘clan’ + *baʔ* ‘grandfather’. For husked rice, several Aslian languages have lost the velar prefix and kept *bāʔ(?)*: Benjamin (1976) *bāʔ*, South Bahnaric languages have *bā* for paddy or husked rice. Semang has *baʔ*, Sidwell. Korku and Mundari have *bābā* and *baba*, Kharia *baʔa*, Juang *bua* for paddy and proto North Munda **baba*, A. Zide and N. Zide (1976). Mundari *huRu*, Sora *kəɾəʔ*, Korku *koɾo*, Mundari *hoɾe* ‘large millet’ are related to the name of the men with **kur* > *kor*, *hor* and a plural pronoun in *ku* or *-j*, exactly like Santali, *horo* ‘husked rice’.

Amwi ‘betel nut’ *kuwa* < *kur* + *waʔ* ‘grand father’ parallels those formations.

k(h)(u)(r/l) are found in different names of animals probably originally totemic, especially the tiger, the horse, the eagle or other prey birds, medicine insects, worms and grasshopper.

‘horse’ War-Khasi: Khasi *kulaj*, Pnar *kulē*, Dawki War *kuru*, Sohka Amwi *kulaʔ*, Thangbuli Amwi *kurwaʔ* (*kur* ‘clan’ and *waʔ* ‘grandfather’, ‘river’). Munda: some of the Santali dialects and Kharia have ‘horse’: *ghoɾa* (< *kur*), Korku has *ghurgi*, Sora *kurtā*, *kudta*, Gadaba *kirkan*, *krutā*, Zide and Zide (1976). Palaung *imjaɲ* horse. Eastern MK people have borrowed and transformed IA *ashvin* while Northern and western AA have *kur/l-* or *ki* or *-jaɲ*.

In Sedang (North Bahnaric), *khu* is used as a classifier (in the usual sense of the term) in various animate names, especially animals like: *khu pah* snake family, *khu kla* tiger family, *khu pa koɲ* family of jungle animals, Smith (1976).

kur/l- is also found in the AA names of the part of the body, especially the head and the testicles all over Munda and MK languages (including Nicobarese). *Thakur ziu*, the old name for God in Santali according to Bodding, contains *kur* ‘relatives on the mother’s side’, *ta* ‘founder’ and *ziu* ‘spirit, soul’. This *ta-* is also found in Amwi ‘first founder of a clan’ *Thāolaɲ* literally ‘the founder who gather together’ (*laɲ* may be derived from **ruɲ* see below Khasi *thaoreɲ*). *Thāo* ‘to create, to build’ or ‘cultivated place’ in Khasi, in Pnar and in Amwi has a short

combining form -t- in WK. *kur* ‘mountain’ appears in Nyah Kur (Monic), Huffman (1990).

kñī > *ki*, *ʔi*, *kin-*. War-Khasi *kñī* ‘eldest maternal uncle,’ cultural father.

kñīam sacrifice

‘rat, mouse’ Old Mon *kniʔ*, modern Mon *kniʔ*, Bahnar *kni*, Sora *kinræped*

Sora : *kinadan* ‘crab’, *kinan* ‘tiger’, *kintəŋ* ‘brain’, ‘backbone’ *kinzəŋ*

Amwi *kintəŋ* ‘guardian spirit of the sacred grove’. In Lawa, *ki* ‘body’ is used as a classifier for humans, Diffloth (1980).

AA *ʒiəŋ ‘bone, spirit’ and its short compounding forms ʒjV

ʒV as initial in Munda languages, jV in MK, is used all over AA languages for female ancestors, grandparents, spirits, soul, bones and inside words expressing the lasting spiritual force of the ancestors (megalith stones in WK, medicines, rituals). Katuic and Bahnaric groups have also kept the ʒiəŋ root for ‘spirit, soul’: E. Mnong *jaŋ*, Stieng, Chrau, Köho *jāŋ*, Katuic *əʒēŋ*.

ʒiəŋ is found all over MK and Munda, as an affix in many compounds, to name tuber, seeds, the kernel of fruits and animals like the scorpion, the eagle (or vulture) which has peculiar divination powers. In MK *ja* has often kept *k-* (or *ʔ-*) prefixes, eventually inside further agglutinations: Amwi *tʃerkjaŋ*, ‘ancestor’, *muʃʒiəŋ* ‘rock’. ‘Ginger’: Khasi *səiŋ*, Pnar *sʔiŋ*, Amwi *ʃʒiəŋ*. The analogy between the terms for bone/spirit and for ginger is transparent all over the War-Khasic languages but more generally this relationship can be reconstructed all over the AA languages. Palaung has *ʒiəŋ* for ginger, Danaw *kātsəŋ* (the word reconstructed by Sidwell for ‘bone’ in Proto-North Bahnaric), Black Riang *kəʃiəŋ*, Luce (1965). Sora *siŋēr*, Ramamurti (1933) In Standard Khasi *lənsʔiŋ* is a plant of the ginger family with white and pink flowers growing at the base, which are eaten as vegetable. The morphological relationship between the bones and the turmeric rhizome reflects the medicinal virtues of longevity of these rhizomes, which link them to the lasting properties of the bones. In the same fashion, the bones link the spirit of their departed possessors to the lasting spirit of the clan. ‘Turmeric’ in the Munda group: Sora, Gorum, Kharia *səŋsəŋ*, Remo, Mundari *sasəŋ*. The Munda names of the turmeric contain the names of the bone in Waic. Waic ‘bone’: Samtao *sʔāŋ*, Tailoi *sa-āŋ*, Kawa *saʔāŋ*. A merger of these forms might be *sak* + *ʒiəŋ*. Bahnaric has *sak ʒan* ‘human corpse’ that is: *s-* + **ka* + **ʒiəŋ*.

Santali has *ʒan-* in many words associated with witchcraft, trance medicinal plants, and intoxicating beverages like *ʒanhe* ‘a variety of millet for brewing beer’ and bony parts of the body, like *ʒaŋa* leg, foot. The same root is found in WK names of the leg, foot (the original ʒ is kept in Khasi *kdʒat*) and seed, kernel, Stampe (1985).

AA names of the eagle or kite often combine short forms of “animate” prefixes **kur*, **ka*, or **koŋ* with **zjaŋ*. Khasi *khlieŋ*, Amwi *klīaŋ*, Bahnar *klāŋ*, Khmer *khleŋ*, Chrau *khlēŋ* Sidwell (2000), Katuic: Bru, Kui, Pakoh, Katu have *klāŋ* (Peiros 1996), Pearic *klaŋ*, Headley (1977), Danaw *ləŋtaʔ*, White Riāng *laŋ ta*, Black Riāng *klaŋ ta* (I do not indicate the tones), Luce (1965): all these may be derived from **kur+*zjaŋ*. Monic: Thai Mon has *(ʔ/k)əniəŋ*, Central Nicobarese *kaləŋ* ‘white bellied sea eagle’, Man (1889), Nancowry *kaləŋ* ‘vulture’ Radhakrisnan (1970) may contain **ka* and **zjaŋ*. Nyah Kur has *ʔoŋliəŋ*, Huffman (1990). In the Munda group, Sora *əlaŋ*, A. Zide (1976), Santali, *kurit* ‘kite’. The use of *zjaŋ* ‘spirit’ in “animate” beings seems to be related to magic and religious practices like using the feathers of the eagle or an egg in divination in Amwi, Sedang, Santal, or to medicine practices of insects or plants: Amwi *kñiaŋ* (generic term for insects), Sedang *klian* ‘leech’, Burmese and Tai Mon *kajaoʔ* ‘maggot’. Mon expresses the eagle’s visions in its name *ʔmat*, Old Mon *tinmāt*, *tam māt*, Mod. Mon *tamāt* whereas Amwi has *tmat* for ‘egg’ (egg breaking is used to divine the name of a child).

zjaŋ is found under different reduced forms and gender/alternant vowels in different Munda and MK languages to name female and male ancestors, fertile earth, sacred places: Santali *zīa*, *azī* ‘grandmother’, *azā* ‘grandfather’, *zəher* ‘sacred grove’ whereas khasi has *zəhar* for a muddy place. Both can be reconstructed as *zjaŋ + ka + -r*, where *-r* has the value of ‘inhabitant, people’ in War and in Munda languages (War, Pnar, Mnar, khara, Kor(ku), Kherwar). Amwi *kintaŋ* ‘guardian spirit of the sacred grove’ < *kñi+ -t- +*zjaŋ* has a word formation which parallels that of Sora: *kinzəŋ* ‘backbone’; the *-t-* infix has a foundation value all over Munda and War languages (see *Thaolaŋ* supra).

za-/ja- are short forms of **zjaŋ*. Santali *zagao* is a kind of blessing involved to make fruitful the cattle or some medicines in the course of ceremonies. Ho *zai gunku* (*kun* child *ku* plur. *za-i* ancestors) ‘descendants’. MK, ‘grand mother *jā* in Pnar, a War-Khasi language. War *jāobej* is the great grandmother of a clan. Old Khmer has *jā* ‘grand mother’, Jenner (1980). Katuic, *jā* is found in names of the relatives on the mother’s side: Bru *kə jā*, Pacoh *ku jā*. Katu has *jəjə* ‘mother in law’. *jā / jaʔ/ jāo* related to grand mother, ancestor are also found in Kmuic and Bahnaric languages (for example *jaʔ* ‘grand mother’ in Stieng and *jāo* ‘ancestor’, ‘tiger’ in Chrau). Mod. Mon has *jā* ‘mother, parents’ from *jaj* Old Mon, Shorto (1962 and 1971), early Middle Mon has *ju* ‘great grand mother’, Bauer (1984). *jā(?)* ‘grandmother’ is found in Monic: Danaw, Black and White Riāng, Palaung and Wa, Luce (1965).

jā is also grammaticalized as a relative and interrogative pronoun in Old Mon. Bauer (1993) analyses *ja* as the third singular personal pronoun of the Old Mon pronominal system, attested only in OM frescoes and having the values of

‘he, him’. *ja* ‘he, him’ and *jā* ‘who’ are variants (Bauer p.c.). Old Mon, *jaŋ* is grammaticalized as an emphatic particle preceding focused nouns especially the subject but also an object, Shorto (1971). As a pronoun grammaticalizing the respect distance involved in the use of *jā* ‘grandmother’, *jā* is combined with gender/number personal pronouns in War languages to produce emphatic forms of personal pronouns, especially in Amwi, for example: *jākā* ‘she, her’ *jāo* ‘he, him’, *jāhem* ‘you’, *iha ben jā* ‘you and me’. *za* is grammaticalized as an indefinite pronoun in Santali ‘whatever’, however’ and it can be associated with different elements to produce all kinds of indefinite pronouns like: *zāhāte* ‘to whatever place’, *zāhā tahā* ‘here and there’ *zāhā tin* ‘at any time’. *za* is also used as an emphatic sentential element in Sora .

In War-Khasic **ka-** very often stands for an affixed **za-** in many Munda cognate words, whereas Nicobarese has kept both affixes: WK *ka/* Santali *zanam* ‘birth, give birth, origin’, where *nam* means ‘religion’ and ‘blood’ in several MK languages. Santali *Sirzan* ‘create, creation, a creation’ (*sir* ‘vein’: any kind of tree shaped artery system, like rivers, lung with its system of air circulation, blood circulation, veins in rocks, veins containing sap in threes etc. where ‘veins’ convey life either in the human body, in plants or in the cosmos. *kV* and *jV* are morphologically associated in the corresponding vocabulary of Central or Car-Nicobarese (data of Man): Car Nicobarese, *kai jok* ‘to give birth’, *kai jok njiu* ‘to issue from the womb, to be born, birth’, *ki ka nā* ‘female person’, *ki ko nja* ‘male person’, Central Nicobarese, *kaijūa* ‘birth’.

The WK names of the cultural father: ‘eldest maternal uncle’ Khasi *kñi* < *kən-* + *ji* < **ka* + *n+* **zjiaŋ* , War *nju koŋ* (*nju* < *n+* **zjiaŋ* + *ku*_{masc.}) and *kñiam* ‘sacrifice’ < *kən* + *ja* + *-m-* involve the clan spirit **zjiaŋ* rather than a wild spirit **ruŋ*. While the biological father produces the flesh of his children, the cultural father, grounds their maternal clan spirit. The sacrifice “feeds” the clan spirit like the *bonga* stones, abode of the ancestors in the Santali *zaher* are fed by the sap of the trees.

koŋ elder maternal aunts, elder brother or sister, madam in Amwi for a woman of the same age than the speaker. Found in AA kinship terms, parts of the body, animals, edible plants, wild or large things or beings. Amwi *tkoŋ* ‘tongue’, Sora *koŋsīn* ‘elbow’, Santali *goŋo* ‘elder paternal uncle’. Within the Munda group, Santali climber plants are named *koŋat* and A. Zide and N. Zide (1976) mention *gonggai* as a name for a number of plants: millet, maize, sorghum, bajra all over: Kharia, Juang, Santali, Ho, Mundari. Munda *gonggai* might be derived from: **koŋ* + **ka* + **-j* (*j* plural) which might enable us to relate a large family of edible plants, both in Munda and in MK languages. *Koŋ* is also found in animal, part of the body, field names. Rengao (Bahnaric) *khoy* ‘wild beast’; Waic: *koŋ* ‘peacock’; *kuŋ* ‘wet rice field’ in Tailoi; ‘nose’: in Amwi *merkōŋ* (nose) in Khasi Standard *kmut* < *ka* + *mu*? ‘grandmother’ + *-t-*).

***ruŋ > rXŋ** (X=e/a/u/o) is used all over AA languages as a prefix or suffix in nouns denoting: male founders (as opposed to 3V/jV for founders and especially female founders), edible plants, horn and wild animals or wild nature (jungle, mountains, rocks, torrent, river, river with deep bed or huge caves, intoxicating preparations, magic, medicines).

Khasi: *kñi raŋba?* ‘the first maternal uncle’, *raŋba?* elder man, *thāoreŋ* the first founder of the clan, *raŋba?* elder man.

**ruŋkug* rice seed in proto-Munda, A. Zide and N. Zide (1976), Old Khmer *raŋko* Luce (1965), Khmu *reŋko?*, Cong *ruko*, Lawa *rako*, Riang *ko?*, Palaung *rākau*, Proto Wa **rŋko* Difloth (1980), proto Viet-Muong **rŋko?*, proto Katuic **rŋkaw*, Old Khmer *raŋko*, Ferlus (1996), Old Mon *sjo*, modern Mon *sju*. One might reconstruct a Proto AA **ruŋkuk* (it is usually assumed that *k* < *g* but there are reasons to deny its application in AA where some MK languages violates the dissimilation rule of Greenberg and *k/g* might be a secondary feature of AA languages in contact with IA or TB languages). Interestingly enough, this **ruŋ* is also used in MK languages to name edible seeds or vegetables which have nothing in common from a botanical taxonomic point of view, like ‘sesame’ *rəŋā* in South Bahnaric languages, Sidwell (2000). Munda: Gorum, *reŋra?* ‘eggplant’ A. Zide and N. Zide (1976).

This element may be combined with a velar prefix, for example *krəaŋ* ‘horn, tusk’ in Tai Mon and Burmese Mon. It is also used in these languages in compounds for trees and for kinds of creeper. The same formation is found for the name of ‘river’, see section 4.

In Viet-Muong the common word for forest in *reŋ*, Thompson (1976). Köho has *krəŋ* ‘thick forest, jungle’. In Munda, Sora has *dereŋ* ‘horn’, *əreŋ* ‘stone, rock’, *kənreŋ* ‘forest’, *reŋge* ‘wind’, *əraŋ* ‘kite’ (to compare with *kliəŋ* which contains **3jaŋ*), *korraŋ* ‘tree bark’ (used in medicines).

2. AA Suffixes in *kV* in “animate” AA nouns

-ka

Cent. Nicobarese *ivi ka* ‘good spirit’, see Santali *zivi* ‘spirit’. Amwi *ləŋkha* ‘field, land’. Santali, Bhiror: *buka* ‘navel’

-koŋ

Amwi : *reŋkoŋ* ‘fern’, *tkoŋ* ‘tongue’, *mərkoŋ* ‘nose’, *tarakoŋ* ‘ear’

-kur

Amwi *ladkur* ‘dove’, Lawa *rako?*. Birhor < *bir* -*hor* < *bəru* + *hor* ‘hill people’ A. Zide (1976) (*hor* > *kur*). Munda: Santali *lutkur* ‘ear’, *hadkur* ‘relatives on the father’s side, *mamakur[en]*’ relatives on the mother’s side, *marwakor[en]* ‘the first parents of mankind’, *boko* ‘head’ Juang (South Munda)

-kñi

Khasi *pukñi* ‘vulture’

-*ruŋ

In Viet-Muong the common word for forest in *reŋ*, Thompson (1976). Köho has *krəŋ* ‘thick forest, jungle’. In the Munda group, *reŋ* suffixed in many relevant Sora names: *dereŋ* ‘horn’, *əreŋ* ‘stone, rock’, *kənreŋ* ‘forest’, *reŋge* ‘wind’. *raŋ* is used in *əraŋ* ‘kite’, *korraŋ* ‘tree bark’. It is also found in TK borrowings (see below).

Suffixed metathesis in *-ak -ik -(u/o)k* in AA “animate” nouns (eventually duplicating a velar prefix)

kjaik, Old Mon, ‘sacred being or thing’, ‘pagoda’ < *k-ja-ik* (see borrowings in *kja* in TB and TK)

ikʼ, Santali, excrement (animal excrements are used in medicines), in Wa languages, Samtao *ʔaʔik* ‘tuber, taro’ (also used in medicines and everyday diet) *kruik* Mod Mon, testicle

muk, Old Mon ‘face’, Khmu *muk* ‘nose’ khasi *khmut* ‘nose’, Amwi ‘maternal grandmother, ancestor’ *mūʔ* < *muk*. in Sre, *məʔ* is used (as in Amwi) for nose, for grandmother and as a respect term of address, Manley (1972). *muʔ* is also related to verbs of knowledge all over AA languages, like Amwi *kəm uʔ* ‘to remember’, to be conscious, Santali *mūʔni* < *muk* + *kñi* ‘knowledge of magic’.

Wa has *ŋok* for neck, to be compared with other Munda and MK ‘neck’ in *reŋko* which merge with the names of the rice all over MK and Munda languages.

ʔaŋruk wild dog, Samtao (Wa).

**kjauk* ‘ear’ in Wa to be compared with *ku-* and *ka-* or *-kur* and *-koŋ* in other MK ‘ear’: Nyah Kur (Monic) *katuol*, Palaung *katua*, Danaw *kātōn* viz. Bahnar: Brao *kutūr*, Munda: Sora *ktur*, Kharia *luʔtur*, Santali *lutkur*

-ak

MK and AA ‘peacock’: South Bahnaric **bərak*, Sidwell (2000), Munda : Birhor *amrak*, Bhumij *cim mərak* A. Zide and N. Zide (1976) *raŋʔak* ‘crow’ in Samtao, Wa. *mak* Mon, male animal (see *mak ruik* testicles in Nicobarese. *mak* < *m-* + *ka* ‘water, sap, juice’ in MK, in Santali sacrifice of a fowl, Bodding (1929).

teak jungle in Samtao, Wa. *vak* insect, worm in Samtao. *kjak* or *karak* buffalo in Wa (*ja* related to the sacrifice) *k-ja-k*, *ka-r-ak*

kurrag Sora, bark, husk

sak < *s* + *ka* ‘to tear’ and also ‘human corpse’ in *sak zan*, Bahnaric

‘water’ *dak* in many MK and most Munda languages. Munda languages have: Santali, *dak* ‘water’, rain’ (Bodding (1932-7), Kharia, Mundari, Ho, Korwa ‘water’ *dak*, Juang *dā* (Grierson), Korku *daʔ/g*, Gutob, Remo *daʔ*, Zide (1999), Sora, *dā* ‘water, juice, sap (of a tree)’, Ramamurti (1933). Monic languages also have *dak* for water, Huffman (1990): Thaï Mon, Burmese Mon *dac*, Nyah Kur *dak*. In Car-Nicobarese *mak*, Central Nic. *dak*, Chowra *rak*, ‘water’, Man (1889).

Khmer has *kak* ‘to water liberally, to wash and *baŋkak* ‘to purify with consecrated water’, Jenner (1980). **dāk* is found for ‘water’ in West Bahnaric, South Bahnaric, North Bahnaric Languages, Jacq and Sidwell (2000), Sidwell (2000). Then, **dak* can be reconstructed as a proto-AA word.

Amwi *ʔam* ‘water, river’ < *ʔa-* + *-m-* < *ka* + *-m-*, *-m-* related to mothers who milk their children. Likewise, seven Aslian languages have *ʔam* for breast (which contains milk), see Benjamin (1976).

3. Shifts in the use of kin prefixes and their combining forms in AA

Some of these shifts may be due to the loss of matrilinearity.

ka viz. kur

‘tiger’ *kaʔvai* in Palaung Wa: Shinman, Samtao *kawaj*, viz. Santali *kul*, Mundari, Birhor, Ho *kula* War-Khasic *khlāʔ/ khlī*, Old Khmer *khlā / klā*, Mon, Bahnar, Sedang *kla*, Prou *klo*.

ka viz. kñi viz. kur

‘crab’ Bahnar *katām*, mod. Mon *gatām*, Danaw *kātan*, Shinman *kaʔtam*, Santali *katkam*, Korku *kaṭkom*. Sora *kin-* is used in place of *ka-* or *kur-* in many War-Khasi names and vice versa: *kinadan* ‘crab’, *kinan* ‘tiger’, *kintəŋ* ‘brain’, see also the names of the rat below. ‘Sacrifice’ War-Khasic *kñia*, Sora *kub*.

‘rat, mouse’ Aslian *kaneʔ*, Benjamin (1973), Danaw *kāne*, Luce (1965), viz. Old Mon *kniʔ*, modern Mon *kniʔ*, Shorto, Bahnar *kni*, Sora *kinrəped* viz. Car-Nicobarese *kumit* viz. Waic *kjaŋ* (Lawa and Tung Wa)

‘dog’ Sora *kənsod*, Gorum *kusəd*

‘bullock’, ‘cow’, ‘buffalo’, Palaung-Wa : ‘buffalo’ *kaʔmoi* in Shinman, *aməj* in Kontoi, Santali *kaDa* viz. Munda group, *kiʔtaŋ* ‘cow’ and Gutob *gulaʔ* (<*kur* + *j_{plur}*) ‘bullock’, *kituŋ* ‘bullock’ in Sora.

koŋ viz. kur

‘heel’ Khasi *koŋdoŋdʒat*, Car-Nicobarese *kuləj trəŋ*

‘peacock’ *koŋ* in Proto-wa, Difloth viz. Khasi *klēo* (see other names in *ka* in 2)

ka viz. koŋ

‘neck’ *kaʔ* in Modern Mon, *khamej* in Palaung viz. *koŋ* in Danaw

‘jungle’ *gaʒar* < *ka* + *zəher* in Santali viz. *koŋ* in Sedang (North Bahnaric)

‘mouth, tongue’ *kəntien* in Standard Khasi viz. *tkoŋ* < *-t-* + *koŋ* in WK: Amwi and *oŋktin* < *koŋ* + *kin* + *-t-* in Mnar

****ruŋ viz. jaŋ***

wild animals, wild plants, wild natural phenomena : mountains, rocks, wind

‘stone, rock’ Viet Muong *əreŋ*, Amwi *muʃʔjaŋ* < *muʔ* + *jaŋ*

‘forest’ Viet Muong *kənreŋ*, Santali *sacred grove* *zəher* < **zjaŋ* + **ka* + *-r*

‘kite’ Viet Muong *araj* Amwi ‘eagle’ *klāj* < **kur* + **zjaŋ*

‘tree bark’ used in medicines Viet Muong *korraj* < **kur* + **ruŋ*

4. Borrowings of the AA “animate” and spirit affixes in TB and Tai-Kadai in cognate names of animals, insects, parts of the body, stone and river.

Such names have been analyzed by Matisoff (1972 and 1986), Thurgood (1988) and Smith (1975). For example, AA affixes can be found in the proto Loloish names ‘cat, tiger’: *k-* + **ruŋ* > *kroŋ* or *kur* > *kula* or ‘chicken’ in *k-* + **ruŋ* + *ka* > *krak* (see the MK and Munda names of the peacock in section 2).

Thurgood (1988) shows that an initial *k-* probably reduced from **several** prefixes, must be reconstructed in proto-Lakkia for animal names like: flea, louse, cow, pig, rat, snake, bear, porcupine, eagle and parts or produces of the body: urine, sheet, arm-pit, ear, face, neck and bone. Thurgood’s reconstructions show AA animate prefixes: *ka*, *kə* and combining forms *k+ jīu* *k+ roŋ*, *k+ ja*. It seems really interesting that combinations of *k-* and *jV/ roŋ* appear in those very names, for example ‘intestine’ is reconstructed as *kjāi3* in proto-Lakkia. This *kja* is renewed as a classifier for parts of the body and various utensils in other languages (head, eye, nose, hand, neck, testicle, bile, breast, hear, leg, handle, brush).

Haudricourt (1974) has shown that Tai *xau* ‘rice’ is derived from MK *kao* (as Santali has *caole*, *xau* may be derived from an AA *kao*) and that Miao-Yao *klay* for eagle is a borrowing from MK (here analyzed as **kur* + **zjaŋ*).

Smith (1975) gives a number of animal names in MK with velar initial which are taken up in Tai, for example in the name of the scorpion, this important medicine element where a MK *ji* element which I consider as a reduction from *jiaŋ* is found in Khasi *kñiaŋ lartham* (*lar* ‘omen in divination’ and *tham* ‘crab’), in North Bahnaric *kjip* or *kāp*, Katuic *kahip* (**ka* + **zjaŋ* + *p*), is further transformed into a velar *ki-* in Tai: *Nùŋ kim pú* as in Santali *kidin katkom*, lit. insect-crab.

AA **zjaŋ* and *ruŋ* are widely found in TB and TK river names. Pulleyblank (1991:149) gives a Chinese etymology for *jiāng* (Tone 1; Pinyin spelling) as found in the name of the “Yangtze Kyang” river with a Late Middle Chinese and an Early Middle Chinese (with reference to the Pekingese form). E. Bruce Brooks (p.c.) reconstructs further this Chinese etymon as *gaŋ* and relates it to the second syllable of Mekong, claiming that the *jiāng* (Tone 1) should be derived from *koŋ/kuŋ* as found in ‘Mekong’. Jerry Norman and Tsu-lin Mei (1976) have independently claimed that this *kjaŋ* (Tone 1) name derives from AA, but relates it to *kruŋ*. He gives MK examples of river names taken up from Shorto (1971): written Mon *kruŋ*, Brou *kruŋ* and Katu *karuŋ* and proposes a Sinitic derivation where *jiāng* is derived from Middle Sinitic *kāŋ* < Old Sinitic **kruŋ*. These two derivations of AA *kjaŋ* given in terms of the different sources: *koŋ* and *kruŋ* are not justified on any kind of internal AA phonological grounds but instead on a few MK river names. The burden of the phonological proof of a relationship

between *kjaŋ* and *kruŋ* or between *kjaŋ* and *koŋ* relies on more or less explicit hypothesis concerning the historical phonology of Chinese. Could it be that AA languages do not deserve having their own phonological laws? Whatever the meaning of ‘Chinese’ may be in terms of a group of languages, these derivations do not look very palatable in the light of an AA etymology of these different AA words indicating previous loci of AA people.

In the light of the uses of **ruŋ* as an AA word formative, *kruŋ* should be analyzed as *k-* + **ruŋ*, where *k-* is a reduced prefix from one of the elements **ka*, **kur*, **kñi* or **koŋ*. Katuic has *karuŋ* < *ka* + *ruŋ* ‘river’. Shorto (1971) gives a list of the MK words related to *kruŋ* in which various vocalic variants, loss of *k-* or weakenings of *kroŋ* into *kloŋ* or *hoŋ* river names are found, as could be expected, and no phonetic connection with *jaŋ* is proposed. The *k-* of the AA *kruŋ* is probably the *ka* which is still found in Katu *karuŋ*. There are several other AA names for ‘river, especially *dak* in many Munda and MK languages also *ʔam*, *waʔ* in War, which can all be related to *ka-* (or its *-ak* metathesis). As already shown, there are many other meanings than that of river associated with **ruŋ* and *ʒjaŋ*, including names of the earth. In Pnar, *Kloŋ* is a big tree immersed in a natural pool to insure fertility of the earth, while Sora *ʒandrum* ‘wealth’, ‘harvest’, ‘paddy’ is associated with domesticated forces. Names of the first founders may be associated with *ruŋ* but God’s names are rather associated with *ʒjaŋ*, Sora *gade ʒaŋ* (lit. the spirit of the earth, *gade* < *kate*) and in Wa and War-Khasic with predicates of knowledge. Within kinship terms, **ruŋ* remains on a masculine side whereas *(ʒ/j)aŋ* is used to denote female founders. **ruŋ* has kept something of its wild ‘horn’ meaning in the Amwi and Sre *kraŋ*, to name impressive horizontal caves under rivers and torrents. *ʒjaŋ* and *ruŋ* are both used as river names but in most cases of their other combinations with velar prefixes they appear in names having completely different meanings, as seen earlier. **Koŋ* and **ruŋ* have different values and are certainly not phonologically equivalent as they are compounded and not merely reduplicated in various “animate” words (with *rXŋ* vocalic variants) such as Amwi *reŋkoŋ* ‘fern’ and Pnar *khoeŋ reŋ* a certain tree in a ritual. This kind of compounding parallels that of *(ʒ/j)aŋ* with prefixed or suffixed **ka*, **kur*, **kñi* or **koŋ*.

From an AA point of view, **ruŋ* ‘river’, **ʒjaŋ* ‘bone, ancestor, spirit’ and **koŋ* ‘eldest maternal aunt’ occur in complementary distributions and appear to have three distinct etymologies with corresponding distinct original values which happened to have been suited for their different uses in river names among other ‘beings’. These three roots have played an important morphologic and semantic role in word formation both in Munda and MK languages to name ancestors, bone(s), spirit(s), magic, sacred groves, medicines, and different tubers like ‘ginger’ and ‘turmeric’ assumed to embed eternal life. We have seen that AA **ruŋ* is associated with male ancestors, edible wild plants which reproduce sexually,

some bones like in Khasi cheek-bone and eyebrow (as opposed to the bones of the leg/foot) and magic practices as *rəŋəj* ‘trance’. War-Khasi have complex funeral ceremony where the bones are kept until a yearly festival in spring and then are purified in a river before being placed in the maternal clan cairns. The Santal actually spread the ashes of the bones in rivers, Carrin (1986). The use of MK *jaŋ* in names of so many rivers actually located in China might be related with AA burial ceremonies of the bones in ancestors clan cairns or rivers while **rəŋ* might be associated with a former layer of culture and magic practices. Layers of cultures often cohabit, most Christian War believe in good and evil spirits and practice exogamic marriages. Khasi has two words for sacrifice: *saŋ* and *kñiam* which might also belong to different layers of cultures: *saŋ* also means taboo, incest and should be compared with Bahnaric *sak zaŋ* ‘human corpse’ -s- + *ka* + *zaŋ* while *kñiam* < *ka* + -n- + *zaŋ* + -m-. Further comparisons are needed to know whether **ruŋ* ‘horn’ belongs to a former layer or to a complementary representation of some of the bony/spirit properties of **zaŋ*. An interesting point to be added to the discussion is the fact that MK *ja* and Munda *za* as well as the velar animate “classifiers”, but not **ruŋ*, have been renewed as many different grammatical kinds of pronouns.

It would be just as meaningless to relate phonologically all the AA river names in *zaŋ*, *ruŋ*, *koŋ*, *ka*, than to relate phonologically all the AA names for ‘men, people’ as inhabitants of a given country, where these elements plus *kur*, *kñi* recur in the very same compounding way, as in: *Katu*, *Kawa*, *War*, *Khasi*, *Kherwar*, *Korku*, *Nyah Kur*, *Juang*, *Rengao* (**ruŋ* + **ka*). AA names for men and rivers appear to be related to the cultural ways people conceive their material and spiritual generation or conceive the way a wild or a socialized spirit grounds and link men, edible plants, animals, or rivers in order to provide and insure life. It seems promising to search for the history of the different AA river names in the light of the different cultures of their speakers and within the distributional properties of some regular compounding element associated with the history of their phonemic alternations.

As could be expected, the ontology of AA languages rely on specific cultural representations of ‘generation’ and ‘spiritual rooting’ rather than on our taxonomic and genetic distinctions among plants, animals, humans and natural phenomena.

References

- Bars, E. 1973. *Khāsi-English Dictionary*. Dom Bosco, Shillong.
 Bauer, C. 1984. “Old Mon Kinship Terminology”. Canberra Symposium.
 ———. 1993. “Epithets and subordinates: a postscript to Sukothai epigraphy”.
Journal of the Royal Asiatic Society, Cambridge U. Press.

- Benjamin, G. 1976. "Austroasiatic Subgrouping in the Malay Peninsula". In *Austroasiatic Studies I*. Jener, Thomson, Starosta eds., U. of Hawaii Press.
- Bodding, P.O. 1925. *Studies in Santali medicine and connected folklore*. The Asiatic Society, Calcutta.
- _____. 1929. *Materials for a Santali Grammar*, II. Dumka.
- _____. 1932-7. *A Santal Dictionary*. (repub. 5 vol.) Gyan.
- _____. 1925-29. *Santali Folktales*. repub. Gyan.
- Carrin, M. 1986. *La fleur et l'os*. Cahiers de l'homme, Editions de l'EHESS, Paris.
- Daladier, A. 2002 a. "Definiteness in Amwi: Grammaticalization and Syntax". Hertz and Daladier ed., *Recherches Linguistiques de Vincennes* 31, PUV.
- _____. 2002b. "Khasi" in : *Encyclopédie des sciences du langage, Dictionnaire des langues*. PUF.
- Deeney, J. 1974. *Ho-English Dictionary*. Chaibasa : Xavier Ho Pub.
- Diffloth, G. 1974. "Les langues Mon-Khmères de Malaisie". *ASEMI* VI, 4.
- _____. 1980. "The Wa languages". *LTBA* 5.2.
- Ferlus, M. 1974. "Les langues du groupe Austroasiatique-Nord", "Lexique Souei-Français". *ASEMI* V,1.
- _____. 1998. *Nhaheun French English Lexicon*. Lincom Europa.
- Grierson, G.A. 1927. *Linguistic Survey of India*, Vol. IV. Motilal, and vol.1, reprinted by Gyan.
- Haudricourt. 1965/1972. "Les mutations consonantiques des occlusives initiales" in *Problèmes de phonologie diachroniques*. SELAF.
- _____. 1974. "Limites et connections de L'Austroasiatique au NE". *ASEMI*
- Headley, R. 1977. "A Pearic vocabulary". *Mon-Khmer Studies*, VI.
- Henderson, E. 1976a. "Vestiges of Morphology in Modern Standard Khasi".
- _____. 1976b. "Khasi Initial Clusters", in *Austroasiatic Studies II*. P. Jenner, L. Thomson, S. Starosta, the University Press of Hawaii.
- Huffman, F. 1990. "Burmese Mon, Thai Mon and Nyah Kur: A Synchronic Comparison". *Mon-Khmer Studies* XVI-XVII
- Jacob, J. 1991. "A Diachronic Survey of some Khmer particles (7th to 17th centuries)". in *Cambodian Linguistics, Literature and History*, SOAS.
- _____. 1974. *Concise Dictionary Cambodian-English*. Oxford U. Press.
- Jenner, P.N. 1980-1. *A Lexicon of Khmer Morphology*. *Mon Khmer Studies* IX-X, Honolulu.
- Luce, G.H. 1965. "Danaw, a dying Austroasiatic language". *Lingua* 14.
- Man, E. H. 1889. *A Dictionary of the Car-Nicobarese Language*. repub. 1975.
- Manley, T. 1972. *Outline of Sre Structure*, Oceanic Linguistics special pub. 12.
- Matisoff, J. 1972. "The Loloish Tonal Split Revisited", *LTBA*.
- _____. 1986. "Proto-Hlai initials and tones: a first approximation", *ICSTLL*.
- Pinnow, H. J. 1965. "Personal Pronouns in Austroasiatic", *Lingua* 14.
- Rabel, L. 1961. *Khasi, a Language of Assam*. Baton Rouge: Louisiana University

- Sidwell, P. 2000. *Proto South Bahnaric*. Pacific Linguistics.
- Shorto, H. 1962. *A Dictionary of Modern Spoken Mon*. Oxford.
- _____. 1971. *A Dictionary of the Mon Inscriptions*. Oxford.
- Smith, K. 1975. "The velar animal prefix relic in Vietnam languages". *LTBA*, vol. 2, 1.
- Smith, K. 1976. "Sedang Animal Folk Taxonomy". *Mon-Khmer Studies V*
- Stampe, D. 1985. *Comparative Munda Vocabulary*. mimeo, Hawaii University.
- Thompson, L. 1976. "Proto-Viet-muong Phonology". *Austroasiatic Studies II*, ed. by Jenner, Thomson, Starosta, The University Press of Hawaii.
- Thongkum, T. 1984. *Nyah Kur -Thai- English Dictionnary*. Chulalongkorn University, Thailand.
- Thurgood, G. 1988. "K- Prefixes in Kam-Sui and Kadai: Some Notes". Festschrift for Tatsuo Nishida, Kyoto/Nishido.
- Whitehead, G. 1927. *Dictionary of the Car-Nicobarese Language*. Asian Ed. Services
- Zide N., 1965. "Gutob-Remo vocalism and glottalized vowels in Proto-Munda". *Lingua* 14.
- Zide A. and Zide N. 1976. "Proto-Munda Cultural Vocabulary: Evidence for Early Agriculture". in: *Austroasiatic Studies II*, ed. by Jenner, Thomson, Starosta. The University Press of Hawaii.

Anne Daladier

a.daladier@wanadoo.fr

Nominalization, relativization, genitivization in Thulung Rai

AIMÉE LAHAUSOIS
LACITO/CNRS, Paris, France

Matisoff showed in his seminal 1972 article that the functions of nominalization, relativization and genitivization are intimately related in Lahu: they are marked by the same particle *ve*. This is shown with a few illustrative examples.

Genitive

- (1) *ŋà* **ve** *mí-cho*
 I shoulder-bag
 ‘my shoulder-bag’

Relativization

- (2) *và?* *qhe* *chu* **ve** *Pícho-pā* *ô* *te* *γâ*
 pig as fat Shan that one person
 ‘That Shan over there who’s fat as a pig’

Nominalization

- (3) *ə-ši* *tə?* *la* **ve** *thà?* *nə* *mâ* *γa* *mə* *lâ*
 blood emerge come ACC you NEG get see Q
 ‘Didn’t you see that blood was coming out?’

In Thulung, there is a set of markers which cover these same functions and which also look etymologically related, although they are not phonologically identical: the class of markers is *-m*, *-mu*, *-mim*, *-ma*.

It is my goal to describe the use of these markers, as they relate to the three functions of nominalization, relativization, and genitivization in Thulung. While the pattern is not as neat as in languages like Lahu, the Thulung data shows participation in what has been called the Standard Sino-Tibetan Nominalization pattern (“SSTN”, Bickel 1999).

Glossing these markers is problematic, because they are a set, rather than a single marker, and also because they often participate in more than one function. I do not have evidence of grammaticalization in any direction¹, nor can I tell at this point whether this pattern is converging towards or diverging from the SSTN. In light of this, I have decided to label the markers according to what appears to be their dominant usage. The marker *-mu* is only used to nominalize, and as such I label it NOM. *-m* appears most commonly as a relativizer, thus it merits the label REL. *-mim* is an alternative to *-m*, with a more restricted distribution, and REL2 is an appropriate label (for simplicity, this label is kept even in cases where it is used for nominalization); *-ma* is discussed, under genitivization, as part of this set of markers, because of its presence on an alternative set of possessive pronouns, but it is not synchronically an independent morpheme, so it receives no label.²

Relativization

Thulung has externally-headed relative clauses, most often preposed to head. There are two relativizers, *-m* and *-mim*, with *-m* having a more general distribution (used for both past and non-past), while *-mim* can only be used for non-past clauses.³

- (4) [go khok-to-*m*/*-*mim*] dzam brəpa bai-ra
 1s cook-1s/3s.PST-REL/REL2 rice good be-3s.PST
 ‘The food I cooked was good.’
- (5) wa-lwak-mim-ka [makai py-ry-*m*/*-*mim*]
 o.brother-y.brother-PLU-ERG corn eat-3s/3s.PST-REL/REL2
 bwa bre-mri.
 pig buy-3p/3s.PST
 ‘My brothers bought a pig that ate corn.’

¹ This is unlike the situation in Chantyal, where Noonan is able to convincingly claim grammaticalization of the nominalizer into the marker for the other functions.

² Abbreviations used in the examples are as follows: verb agreement encodes both agent and patient, and I represent the combinations separated by a /, so that verbal agreement showing 1s acting on 3s is abbreviated as 1s/3s; POL preceding a pronoun indicates that it is the polite form; Other abbreviations are GEN=genitive, INSTR=instrumental, ERG=ergative, DAT=dative, ABL=ablative, LOC=locative, TOP=topic, CONTR=contrastive, OBL=obligation, NEG=negation, IRR=irrealis, SEQ=sequencer, CONV=converb, PROG=progressive, CAU=cause, PURP=purposive, HS=hearsay, DU=dual, PLU=plural.

³ There is an alternative relativization strategy, with participles *-pa* and *-ma* (for non-past and past clauses respectively). One difference from finite-relativized clauses is that the participles are non-finite and therefore do not encode participant information, unless separately specified by an independent pronoun

- (6) [go dwak-pu-**m/-mim**] kitap gani-lai gwa-gwa dwa.
 1s like-1s/3s-REL/REL2 book POL.2p-DAT give-give like.3s/3s
 ‘I want to give you a book I like.’⁴
- (7) oram je [go phəŋtasi-ŋu-**m/-mim**] hapa dzupa bu.
 this clothes 1s wear-1s-REL/REL2 very nice be.3s
 ‘These clothes I am wearing are very nice.’

Additionally, *-m* is blocked in certain instances, and what is relevant seems to be the syllable length of the finite verb. Mono-syllabic verb forms block the use of relativizer *-m*.

- (8) gui si-**mim**/**-m* din
 1pi die-1pi-REL2/REL day
 ‘the day we die’
- (9) gu-ka ra-**mim**/**-m* bela
 3s-ERG say-3s/3s-REL2/REL time
 ‘At the time when she says...’

The synchronic distribution of the relativizers thus appears to be as follows:
-m is the general relativizer, available for both past and non-past clauses (except for mono-syllabic verb forms.)
-mim must be used with mono-syllabic verb forms and can be used anytime a non-past form is being relativized.

This distribution may appear to be fairly unusual, and earlier data on Thulung shows that this was not always the case. Allen, who worked on Thulung in the 1970’s, noted that tense was relevant for the choice of relativizer: “It would seem that *mim* is to present tense forms what *-m* is to past tense ones”. (1975: 88). So the earlier distribution of relativizers seems to have been exclusively based on tense. Allen does however give one example of a non-past clause marked with *-m*⁵, suggesting that the seeds of change were already present in the 1970’s.

The current distribution whereby mono-syllabic verb forms must take the relativizer *-mim* seems to be a shift in the interpretation of the relevant factor for relativizer choice. The only mono-syllabic verb forms in the language are non-past⁶, and it is possible that speakers reinterpreted the relevant factor as being not tense but syllable-length.

⁵ mi theTpum loa koNNa reakpu
 not I-understand+m words only I-write

⁶ This is the case for verbs whose agent is a first plural exclusive or a third person singular.

Nominalization

There are a number of constructions which are counted as examples of nominalization, and they are discussed below. Interestingly it is not the same marker which appears on all of them, which may suggest different paths of development (in other words some of these nominalized forms may have developed from relativized constructions while others may have been nominalized from the start).

The first two types of nominalization are typical nominalized constructions throughout Tibeto-Burman, and they both take the same nominalizer, *-mu*.

Citation form of verbs

Matisoff states that “as a general rule of thumb applicable throughout the Tibeto-Burman family, whenever one discovers the particle used in verb citation, one can be sure of having discovered the most important nominalizer of the language.” (1972: 248). Thulung uses the marker *-mu* for this function.

- (10) *on-mu-lai tsapa bəne-mu basi*
 run-NOM-DAT strong make-NOM OBL
 ‘To run, one must make oneself strong.’
- (11) *kho-mu-kam lagi...*
 cook-NOM-GEN sake
 ‘In order to cook...’
- (12) *lamdi-mu bhandane plen-ra lo-mu thik ra-ŋro*
 walk-NOM than-TOP plane-LOC go-NOM fine say-1s/3s.PST
 ‘I said that going by plane is better than walking.’

It is the nominal status of these roots which allows the verb to take on case marking.

Verb complementation

Similarly, verb complementation, where the nominalized clause is the complement of the verb, is marked with *-mu*.

- (13) *go dika [mukli lo-mu] tsahebe-u.*
 1s tomorrow Mukli go-NOM need-1s/3s
 ‘I must go to Mukli tomorrow.’
- (14) *go [dzudzuulŋ ho-mu] dwak-pu.*
 1s mountain climb-NOM like-1s/3s
 ‘I like to climb mountains.’

There are also some instances of less prototypically “nominal” constructions which are nevertheless examples of nominalization. All of these use *-m* as a nominalizer (sometimes in complementary distribution with *-mim*, when monosyllabic forms are possible.) In other words, the nominalizer in these cases is what we have seen to be the relativizer in Thulung, indicating that perhaps these constructions have different origins from standard nominalization.

Clause nominalization

Two types of clauses are nominalized, and these are causal and temporal clauses. The causal construction consists of a nominalized clause (which expresses the cause), followed by a grammaticalized case marker (the instrumental), and finally the main clause. As for the temporal construction, it is a nominalized clause followed by a temporal expression (borrowed from Nepali), which can be either *patshi* ‘after’ or *səmma* ‘until’.

Some examples of both types are shown below.

- (15) *go pomuθhok mi-peu-wa-m-ka krym si-ŋro*
 1s food NEG-eat.1s/3s-IRR-REL⁷-CAU hunger feel-1s.PST
 ‘Because I didn’t eat, I felt hungry.’
- (16) *meram tsuutsu krym si-ra-m-ka khrap-saŋa bu*
 that child hunger feel-3s.PST-REL-CAU cry-PROG be.3s
 ‘Because that child is hungry, he is crying.’
- (17) *bju-ka lə-ry-m patshi muu-tsu-tsip-ka*
 eagle-ERG carry-3s/3sPST-REL after thather-child-DU-ERG
mal-to mal-to lək-tsi-ŋe
 search-CONV search-CONV go-3d.PST-HS
 ‘After the eagle carried her off, her two children went searching.’
- (18) *meno roypha-ŋu ma dhol kuk-pu-m səmma...*
 there arrive-1s SEQ drum beat-1s/3s-REL until...
 ‘Until I arrive there and beat the drum...’

The same distribution of markers is seen as for relativization, whereby a monosyllabic verb form must be nominalized with *-mim* (and *-m* is blocked).

⁷ As indicated in the introduction, I have assigned labels to these markers based on the function they mark most frequently, and these labels appear in the glosses even when I am giving evidence of participation in a different construction. Because of the complexity of the phenomenon, and the overlap, the labels therefore become slightly misleading in certain cases.

- (19) *meram si-mim patshi*
 that die.3S-REL2 after
 ‘After she dies...’

Perhaps the presence of what we saw to be a relativizer in these constructions which are based on nominalized clauses involves a stage where they were indeed relativized clauses, with some relevant but perhaps semantically weak head noun, which was followed by either the instrumental case marker (in the expression of cause) or the temporal expression (for temporal clauses)⁸. These head nouns could then have been dropped, with the result that the construction now looks like a nominalized one.

Verbal periphrasis

This is the combination of the nominalized past-tense finite verb and of the copula. The resulting construction conveys perfect aspect. This construction also appears in a number of other languages of Nepal, such as Hayu, Chantyal, Limbu, Yamphu, among others.

- (20) *dzetha-mim tsahi waythui babante lb-mri-m bu.*
 o.brother-PLU CONTR other where go-3p.PST-REL be.3s
 ‘The older ones went somewhere else.’
- (21) *hu grenem-ra lbs-ta-m bu*
 there nettle-LOC go-3s.PST-REL be.3s
 ‘She went out to the nettles.’

In this construction, nominalization is accomplished by means of *-m* (and because the construction calls for a past form of the verb as the input, the verb is automatically disyllabic and *-mim* does not occur), which we saw was the relativizer elsewhere.

Sentence nominalization

Matisoff refers to the marking of entire sentences as nominalized as their “reification”, and says that they can often be translated as beginning with “It’s a fact that ...”. While it is reported for a number of languages of the area (Belhare and Chantyal among others) that the result is pragmatically marked, I believe this construction to be pragmatically unmarked in Thulung. The pattern seen in

⁸ In fact there is another temporal construction in Thulung with the loan word *bela* ‘time’ in place of *patshi* or *somma*. Because *bela* is a noun, this construction looks more like relativization, with the temporal clause being relativized to the head noun *bela*. Perhaps what has really happened is that the *patshi* and *somma* clauses are built on the same model as the *bela* type, and the same markers, in other words those which are really relativizers in form, were applied to the clause, even though there was a functional difference due to the lack of head noun.

conversations is that the answers take the same form as the questions (that is, the same presence or absence of a nominalizer.)

- (22) A: *bante lamdi-nni-m?*
 where walk-POL.2p.PST-REL
 B: *go ηopsuu gruum-ḍa lamdi-ηro-m.*
 1s friend visit-PURP walk-1s.PST-REL

A: 'Where did you go?'

B: 'I went to see a friend.'

- (23) A: *gumi bante bu-mi?*
 POL.3p where live-POL.3p
 B: *gumi basbari-ra bu-mi.*
 POL.3p Basbari-LOC live-POL.3p
 A: *gumi ba-lanjka ro-mri-m?*
 POL.3p where-ABL come-POL.3p.PST-REL
 B: *gumi mukli-lanjka ro-mri-m.*
 POL.3p mukli-ABL come-POL.3p.PST-REL

A: 'Where does she live?'

B: 'She lives in Basbari.'

A: 'Where is she from?'

B: 'She is from Mukli.'

Nominalized sentences also occur in narrative, but with less frequency than in conversation, and even then, the frequency depends on the speaker and the context. Even the item marked seems to vary somewhat: while the nominalizer is generally marked on the final verb, I have also seen the marker follow the final hearsay marker (as seen in (25) below). From looking at these sentences within narratives, I do not think that pragmatic marking is the trigger. There does not seem to be anything that distinguishes the sentences which are marked, compared to those which are not.

- (24) *mu-ku u-tshoktso-ka mem bet-tsi-m-ʔe*
 that-GEN his-anger-INSTR like.that do-3d/3s.PST-REL-HS
 'They acted that way out of anger with him.'
- (25) *memma meram badzi-lanjka iki-beppap-mim glwa-mri*
 after.that that bet-ABL our-ancestor-PLU win-3p.PST
 ma tsahi gui thuluj dys-ti-ʔe-m.
 SEQ CONTR 1pi Thulung become-1pi.PST-HS-REL
 'After our ancestors won that bet we became Thulung.'

One interesting fact is that the nominalized sentences are very often past in form. This leads me to believe that nominalized sentences are sentences having perfect aspect (expressed by verbal periphrasis, as seen above) from which the final copula has been omitted. This is an idea suggested by Noonan as the origin of nominalized sentences in Chantyal (although he makes it clear that nominalized sentences in Chantyal have evolved separately into pragmatically marked situations). This would explain the past tense constraint on nominalized sentences: Thulung has not yet grammaticalized the difference between verbal periphrasis with the copula omitted and nominalized sentences.

One sentence I elicited reinforces this, as the consultant told me that the final copula could perfectly well be left out with no change in meaning, implying that speakers themselves consider nominalized sentences to be ‘the same’ as sentences with verbal periphrasis missing the copula.

- (26) *ama-neb-ra* *gani* *rok-ni-m* (*bu-ni*)
 my-house-LOC POL.2s come-POL.2s.PST-REL (come-POL.2s)
 ‘You came to my house.’

In sum, sentence nominalization occurs, as it does a great many Tibeto-Burman languages. However, the role of this construction in Thulung does not seem to correspond to the parallel in other languages, as it is pragmatically unmarked. There is some evidence that it is a result of truncation of the copula for a perfect aspect-marking periphrastic construction, so perhaps what we see is an intermediate stage, and that grammaticalization will result in pragmatic differences in nominalized sentences versus non-nominalized.

Genitivization

A possessive relationship between two nouns is usually expressed with genitive markers *-ku* or *-kam*, thus the pattern of identically (or similarly for Thulung) marked relativization, nominalization and genitivization would seem to break down here. But there are two situations in which the genitive is in fact expressed with a marker from the group we saw above: one is when the possessed noun is a time word, in which case it is marked with *-m* or *-mim*. The other case is with possessive pronouns, for which there are several variant forms, some of which also show the presence of one of these markers. These two slightly marginal cases of genitivization marked with the same class of markers seen for nominalization and relativization are perhaps remnants of an earlier stage of the language, before the genitive markers *-ku* and *-kam* came to be used⁹.

⁹ In addition to which, the final *-m* on the genitive case marker *-kam* is a subject for further study.

Possession of time words

For both native and borrowed time words, the genitive is expressed with *-m/-mim* rather than the standard genitive markers. The distribution of these two markers is as follows: *-m* appears post-vocally, while *-mim* appears post-consonantly. While this pairing of *-m* and *-mim* is familiar from relativization, the distribution is different, and for this I have no explanation for the time being.

- (27) *nemtha-m/*mim* *dzam*
 evening-REL/REL2 rice
 ‘the evening meal’
- (28) *dika-m/*-mim* *lagi*
 tomorrow-REL/REL2 sake
 ‘tomorrow’s sake, ie. for tomorrow’
- (29) *aneb-mim¹⁰/*-m* *din*
 today-REL2/REL day
 ‘today’s day, ie these days’

When the construction has no overtly marked head, the result is an NP referring to an individual by his day of birth (it is surprisingly common for people to talk about themselves or their children this way.)

- (30) *buddhabar-mim/*-m*
 Wednesday-REL2/REL
 ‘[The child] born on Wednesday’
- (31) *basta-m/*-mim*
 yesterday-REL/REL2
 ‘the one born yesterday’
- (32) *bam-din-mim*
 which-day-REL2
 ‘the one from which day (N. *kun din-ko*)’

It seems significant that the markers used for this sub-set of genitivization are the same ones used for relativization (with a different distribution though: that for relativization is based on syllable-length, whereas that for genitivization is dependent on the final phoneme of the marked word). Noonan suggests that the genitive function derives from the relative: “Once the attributive function becomes established in relative clauses, it may be extended to other sorts of attributives.”(1995)

¹⁰ Perhaps there is a possibility that the reading ‘these days’ is a result of *-mim* as pluralizer, rather than genitive marker, but that still wouldn’t explain how it can be followed by *din*.

Possessive pronouns

Thulung has four sets of possessive pronouns, clearly related, and interchangeable when used prenominally. These show a borderline case between nominalization and genitivation.

Possessive pronouns

1st person		2nd person		3rd person	
<i>a</i>	<i>aki</i>	<i>i</i>	<i>ini</i>	<i>u</i>	<i>uni</i>
<i>ama</i>	<i>akima</i>	<i>ima</i>	<i>inima</i>	<i>uma</i>	<i>unima</i>

What the table above shows is the possibilities for each of the three persons: the first row of possessive pronouns are just that, and can only be used prenominally. The second row can be used prenominally, but can also be used independently, standing in for the omitted noun.

Thus we have:

- (33) *a-khlea*
aki-khlea
ama-khlea
akima-khlea
my-dog
‘My dog’

However, only those forms ending in *-ma* can stand in for the noun.

- (34) *akima/aki-khlea* *ɖokpu* *bu*
my-dog big be.3s
‘My dog is big.’
(35) *akima/*aki* *ɖokpu* *bu*.
Mine is big.

This *-ma* is not synchronically a morpheme, yet it certainly seems to represent some earlier nominalizing suffix which turned possessive pronouns into substantives. I chose to discuss this non-productive nominalizer under genitivation because synchronically the pronouns where it appears are both genitive and nominative forms, judging from their ability to perform both functions.

DeLancey (1989) gives an interpretation of the origin of a Newari genitivizer, which proves useful for the case of Thulung.

‘Ram’s thing, a dog’

-where an empty noun is apposed to another noun, becomes reinterpreted as

‘Ram’s dog’

-and the empty noun becomes a genitive marker instead.

Perhaps this is also the case here:

akima khlea

‘mine, a dog’

gets reinterpreted as a genitive, ‘my dog.’

So if the original scenario is *akima* being only nominal (we have no such data, but it seems fairly likely that such was the case at some point), then *-ma* is reinterpreted as being a genitivizer.

I have described various constructions covering the three functions of relativization, nominalization and genitivization, all of which are expressed using markers from the same set: *-m*, *-mim*, *-ma* and *-mu*. Noonan suggests for Chantyal the following line of development: nominalization (through eventual erosion of genitive marker which is initially used to link the attributive to the head noun) turns into relativization which turns into genitive (by analogy to other attributives of use of nominalizer in relativization). The fact that we have four different markers in Thulung complicates the picture, making it difficult for the time being to trace a direction of development. Nevertheless I believe these markers to be related etymologically, and that it is not a coincidence that these three classically related functions in Tibeto-Burman languages also align in Thulung.

References

- Allen, N.A. 1975. *A sketch of Thulung grammar*. Ithaca: Cornell University.
- Bickel, Balthasar. 1999. Nominalization and focus constructions in some Kiranti languages. In Y. Yadava & W. Glover (eds.) *Topics in Nepalese Linguistics*. Kathmandu: Royal Nepal Academy
- DeLancey, Scott. 1989. Relativization and nominalization in Tibeto-Burman. Ms., University of Oregon.
- Genetti, Carol. 1992. Semantic and grammatical categories of relative clause morphology in the languages of Nepal. *Studies in Language* 16: 405-427.
- Matisoff, J.A. 1972. Lahu nominalization, relativization, and genitivization. In J. Kimball, ed. *Syntax and Semantics*, Volume I. 237-258
- Noonan, Michael. 1997. Versatile Nominalizations. In J. Bybee, J. Haiman & S.A. Thompson (eds). *Essays on language function and language type*. Amsterdam/Philadelphia: John Benjamins.

Aimée Lahaussais

Aimée Lahaussais
6 square Grangé Paris 75013 France

aimee@vjf.cnrs.fr

On Khumi Verbal Pronominal Morphology

DAVID A. PETERSON

Max Planck Institute for Evolutionary Anthropology

1. Introduction

A Tibeto-Burman language of the Kuki-Chin branch, the variety of Khumi described here is spoken by about two thousand people in two distinct but mutually intelligible dialects in the Chittagong Hill Tracts of southeastern Bangladesh. The basis for the paper is material gathered during a total of about ten months of work with the language from 1999-2001.¹

The Linguistic Survey of India (1904) judged Khumi to lack verbal pronominal morphology. Later studies (Shafer 1944, Löffler 1960) of various Khumi dialects also have not detected verbal pronominal marking.

The primary goal of this paper is to describe the distribution of verbal pronominal morphology for the variety of Khumi spoken in Bangladesh. A second goal will be to discuss the possible diachronic relevance of the Khumi phenomenon vis-à-vis verbal pronominal systems in other Kuki-Chin languages and in Tibeto-Burman languages generally.

The main claims of the paper are that verbal pronominal morphology in Bangladesh Khumi is an optional, loosely grammaticalized, speech-act participant coding device. While a full pronominal paradigm exists, and may be elicited directly, parts of it which do not refer to speech-act participants do not occur frequently in texts, and those which *do*, have a highly specialized function.

In addition, I will suggest that as a more loosely grammaticalized system than the systems of pronominal morphology usually attested in Kuki-Chin languages, the Khumi system is likely to be historically primary or an altogether independent development; other systems which resemble it are probably later grammaticalizations of idiosyncratic pronominal systems which developed after Kuki-Chin languages diverged from each other. However, at least parts of the *suffixal* system

¹ This research was funded by a Fulbright fellowship and by the Max Planck Institute for Evolutionary Anthropology. I thank Lelung and Prie'ang Khumi, and Kewsamong Khyang for data and insight. But, my greatest debt is to Jim Matisoff, for inspiring my work in countless ways.

of agreement attested throughout the subgroup are likely to be shared retentions from the Proto-Kuki-Chin stage.

2. The verbal pronominal morphology in direct elicitation

In directly elicited material verbal pronominal morphology does not occur spontaneously, although it is possible to elicit full paradigms if it is made clear to consultants that it is pronominal prefixes which are of interest, or if elicitation is based on constructed examples judged for acceptability and interpretation. This section will not give extended illustrations of the use of the morphology in elicited data, but instead will simply summarize the generalizations concerning its distribution.

(1) gives the Khumi independent pronouns. Note that these exhibit a dual/plural distinction and an inclusive/exclusive distinction.

(1) Independent pronouns:

	1st person incl/exclusive	2nd person	3rd person
singular	<i>kaay</i>	<i>naang</i>	<i>ní</i>
dual	<i>ay-ni/kaay-ni</i>	<i>naang-ni</i>	<i>ní-ni</i>
plural	<i>a-cie/kaay-cie</i>	<i>naang-cie</i>	<i>ní-cie</i>

(2) contains a tabular representation of verbal pronominal markers for intransitive and transitive roots obtained through direct elicitation; the basic generalizations concerning verbal pronominal marking are given in (3). In the tables, contexts which do not involve verbal coding, including reflexives, are indicated by -. Parenthesized items in the tables only sometimes occur under direct elicitation, although we will see in the next section that there is evidence for their use in connected discourse.

(2) Tabular representation of verbal marking:

Intransitive:

S	1 exclusive	1 inclusive	2	3
	<i>ka(ng)-</i>	-	<i>ang-</i>	<i>ang-</i>

Transitive:

O/A	1 exclusive	1 inclusive	2	3
1 exclusive	-		<i>ang-</i>	<i>ang-</i>
1 inclusive		-		<i>ang-</i>
2	<i>ka(ng)-</i>		-	<i>(ang-)</i>
3	<i>ka(ng)-</i>	-	-	<i>(ang-)</i>

(2) Basic generalizations for verbal pronominal marking:

- Number (of S, A, or O) is irrelevant
- *ka(ng)-* ‘first person exclusive S/A’
- *ang-* ‘first person/second person O’ or ‘non-first person S’
- First person A marking takes precedence over second person O marking

Further generalizing over these observations, for transitives, verbal pronominal marking always involves speech-act participants. For intransitives, verbal pronominal marking may also involve exclusively non-speech-act participants.

3. The text-distribution of pronominal marking

In texts, verbal pronominal morphology occurs in three contexts. This finding is based on consideration of the distribution of verbal marking in fifteen texts (primarily narrative and conversation) of varying lengths, totaling approximately 4,500 clauses. In this corpus, there are about 100 examples of verbal pronominal marking, which should give some impression of the text-rarity of the phenomenon.

The first and most frequent context that verbal pronominal morphology occurs in is reported speech. In Khumi, as in other Kuki-Chin languages I am familiar with, reported speech is always directly quoted conversation. Some text examples of the use of verbal pronominal morphology in reported speech are given in (4).

(4) Verbal pronominal morphology in reported speech:

a. First person A, second person O²

nayboeloe *vaay=loe* *aámcløeeyng* *kang-plaw-piee-bo* *noe=piee-te*
then now=TOP eagle 1S/A-call-BEN-PERF QUOT=say-EVID

‘Then she said to them, “Now I’ll call the eagles for you,”...’ (8.39)

² Note that the benefactive applicative marker in the verbal complex of the first verb is what makes this a case of first person A acting on second person O. For details on the morphosyntax of this construction, see Peterson 2001b.

- b. First person A, third person O

amnáay=oo vayníí tbeewng lee-noe-ra=loe khúmii-coo
 friend=VOC today banana peel-NZR-NZR=TOP person-DIM

kang-caa-noe-bo noe=thúy-te
1S/A-eat-NZR-PERF QUOT=tell-EVID

“‘Friend, today, where we were peeling banana trees, I ate (=caught) a human child,’” he told (him).’ (3:50)

- c. Second person A, first person O

naang o’á matimata kacáawy khaa ang-thúy-noe maá
 2S crow what.kind.of lie EMPH 1/2O-tell-NZR where

khaa kaáy cniwcnaáw awng-noe kaáy cniwcnaáw=loe
 EMPH 1S.GEN daughter exist-NZR 1S.GEN daughter=TOP

kewsii máng-noe doey-pawpang-bo-noe noe=te
 leprosy suffer=NZR die-MIMETIC-PERF-NZR QUOT=EVID

“‘You, Crow, what lie are you telling me? Where is my daughter living? My daughter was suffering from leprosy and has died,’” she said.’ (1:73)

- d. Third person A, first person O

kaaymóey=loe phayloeyng=moe ang-ke-tlaw-noe-te-ba
 1s eye=TOP ant=DEF 1/2O-bite-LARGEO-NZR-EVID-EVAL

noe=piee-te
 QUOT=say-EVID

“‘The ant bit me in the eye!’” he said to her.’ (1.104)

- e. Third person A, second person O

...naang=poe toeéng-boeloe ang-caa-noe-bo...
 2S=ALSO arrive-CONN 1/2O-eat-NZR-PERF

...naáng jvoó=loe uymíw kung-noe-bo noe=piee-te
 2S.GEN husband=TOP cannibal turn.into-NZR-PERF QUOT=say-EVID

“...You also have come and he’s going to eat you... ...your husband’s turned into a cannibal!” it said to her.’ (3:28)

Face-to-face conversation is a second context in which verbal pronominal morphology occurs, though here it is also relatively infrequent. Some examples of this use of the morphology from a conversational text are given in (5).

(5) Verbal pronominal morphology in face-to-face conversation:

- a. First person A, second person O

hini amoe-taeng=poe aka-a nay m hini kang-thúypiee-noe-to
 DEM REFL-OBL=also blame-FUT thus PART DEM 1S/A-say-NZR-EVAL

‘In this affair (they) may also blame you. Thus, uh, this is what I say to you.’ (9.84)

- b. Third person S

aplaa-tlaa-boeloe aplaa-a cnaáw=poe
 revoke-OBLIG-CONN revoke-FUT child-ALSO

ang-vaáwy-taeng-vuy-noe-tew-bo=ie...
3S/A-return-AGAIN-PAST-NZR-CONCESS-PERF=AND

‘If we have to revoke (our oath), we’ll revoke it, but the child has returned again, and....’ (9:85)

Finally, verbal pronominal morphology has what apparently would best be described as a perspective-shifting device. This use could be likened to the use of proximate-shifting for a similar effect described for Fox by Goddard 1990. Some instances of this use of the morphology are seen in the examples given in (6); in fact, (5b) also is probably an instance of the same use.

(6) Perspective-shifting use of verbal pronominal morphology:

a. Third person S

<i>ewkuu</i>	<i>tkhoeéyng-noe</i>	<i>khang</i>	<i>khang</i>	<i>khang</i>	<i>noe =boeloe</i>
trough	tap-NZR	tap	tap	tap	QUOT=CONN

huni *ang-jeew-noe-tlaa*
 DEM 3SA-come-NZR-OBLIG

‘She beat on the trough, tap, tap, tap, and they (the eagles) had to come.’ (8.41)

b. Third person S

<i>nay’ie</i>	<i>rekheeng</i>	<i>khúmii</i>	<i>awng-ra</i>	<i>amceng;</i>
so	Arakan	Khumi	exist-NZR	small.place

acié *pree=ma* *ang-jeew*
 1PINCL.GEN country=LOC 3S/A-come

‘So, where the Arakan Khumi live is a small place; they’re coming over to our country.’ (9:88)

c. Third person S

<i>doey</i>	<i>akhrang-cie=loe</i>	<i>nay’ie</i>	<i>amoe</i>	<i>naybo</i>	<i>mayyuung</i>
die	custom-PL=TOP	so	someone	if	ash

thiw-khoekhoe-boeloe *ang-thew* *khaá=poe* *tlaáng=a*
 mark-TRULY-CONN 3S/A-come.out time=ALSO body=GOAL

<i>kamnuung</i>	<i>thiw-doe-noe</i>	<i>anáy=hawy</i>	<i>ang-thew-boeloe</i>
black.thing	mark-some-NZR	like=COM	<u>3S/A</u> -come.out-CONN

<i>neé-khoekhoe</i>	<i>nay</i>	<i>noe =piee</i>	<i>oem-nay-tlaa</i>	<i>hini</i>
true-TRULY	thus	QUOT=say	believe-thus-OBLIG	this

‘Regarding death customs, so, if someone is marked with ashes and when they come out, when they come out with a little mark like that on the body, we have to believe that it’s really true (that they are truly a reincarnation of the dead person who was marked with ashes).’ (12:44)

- d. Third person A, third person O

vaydúeeng=ma kaay=loe húu ii-bie-noe-te noe=piee-te
tonight=LOC 1s=TOP there sleep-AGAIN-NZR-EVID QUOT=say-EVID

“‘Tonight I’ll sleep there again,’ she told him.’ (1:37)

nayboeloe ang-khieeng-bie-bo dúeeng=ma bo-noe-boeloe
then 3S/A-look.for-AGAIN-PERF night=LOC PERF-NZR-CONN
‘Then he came to look for (her) again, when it became night.’ (1:38)

- e. Third person A, second person O

...pree m’roe=a kola liee ang-ee-yo-noe
country city=GOAL Bangali paddy 3S/A-shit-AWAY-NZR

pnóe-a vaay nga’ay=loe ang-ee-yo-bo-noe
know.NEG-INTERROG now father=TOP 3S/A-shit-AWAY-PERF-NZR

pnóe-a noe=piee-te-ho
know.NEG-INTERROG QUOT=say-EVID-EVAL

“‘...they were going to shit you out over the land, over the city, and in the Bangali rice fields, don’t you know? Now they’ve already shit out father (=grandfather), don’t you know?’” he said to them.’ (9:182)

In (6a), for instance, the understood agent of the first clause is a snail, who is summoning two eagles to fight with some humans who have come to avenge the killing of their father. In the second clause, the use of the person marker apparently adds to the vividness of the scene, and it is as if the listener personally witnesses the sudden arrival of the eagles.

In summary, when it occurs at all in texts, verbal pronominal marking primarily codes speech act participants. Only in some of the cases in (6) does pronominal morphology clearly refer *only* to a non-speech-act participant. My suggestion for such cases is that here it is really marking the entrance into the discourse of a (relatively) salient third person participant, or a shift from the perspective of that of an objective narrator to that of the speaker or hearer.

4. Comparative aspects of Khumi verbal pronominal morphology

While it bears formal resemblance to systems found elsewhere in Kuki-Chin, Khumi’s system of pronominal morphology is distinct from the ones found in other Kuki-Chin languages in a number of respects. Consider, for instance, the system of verbal and independent pronominal morphology attested in Hakha Lai

(spoken in Chin State, Burma; Σ indicates the position of the verb stem with respect to pronominal markers):

(7) Lai pronominal morphology:

	Verbal A/S	Verbal O	Independent pronoun root
1s	<i>ka</i> - Σ	- <i>ka</i> - Σ	<i>kay</i>
2s	<i>na</i> - Σ	- <i>n</i> - Σ ~-? <i>in</i> - Σ	<i>naŋ</i> -
3s	<i>a</i> - Σ	- \varnothing - Σ	? <i>a</i> -
1p	<i>ka-n</i> - Σ	- <i>ka-n</i> - Σ	<i>kan</i> -
2p	<i>na-n</i> - Σ	- <i>n</i> - Σ - <i>hnaa</i> ~-? <i>in</i> - Σ - <i>hnaa</i>	<i>nan</i> -
3p	? <i>a-n</i> - Σ	- \varnothing - Σ - <i>hnaa</i>	? <i>an</i> -

First, the Khumi system differs from other attested systems in terms of its formal simplicity: two markers vs. several (Lai) or many (K'cho, discussed by Bedell 2000, Daai, discussed by Hartmann 2000, and Hyow, discussed by Peterson 2001). Second, there is a lower degree of resemblance between the verbal pronominal elements and independent pronouns in Khumi than there is in other languages (cf. the highly transparent relationship between the verbal and independent pronominal morphology in Lai). Finally, unlike the highly grammaticalized, obligatory *agreement* systems found in languages like Lai, Mizo, Hyow, K'cho, Daai, and Tedim, the Khumi verbal pronominal morphology is an essentially optional speech-participant coding device.

4.1. Kuki-Chin prefixal pronominal morphology as innovative

In this last respect (looseness of grammaticalization), the Khumi system resembles what LaPolla 1992 has argued is usual for Tibeto-Burman pronominal morphology systems; LaPolla further suggests (contra those who wish to reconstruct such a system to Proto-Tibeto-Burman, like DeLancey 1989) that such systems form a functionally transitional stage between Chinese/Lolo-Burmese-type languages without agreement and the classic “pronominalized” languages like those of the Himalayish and Kuki-Chin subgroups.

We might hypothesize, then, as LaPolla does for Tangut, that the relatively loosely grammaticalized system of verbal pronominal morphology in Khumi represents a stage more closely approximating Proto-Kuki-Chin. On such an account, other, invariably more complex systems found elsewhere in this branch of Tibeto-Burman would involve (often independent) grammaticalizations and paradigmatic reshuffling of more recent pronominal systems.

Finally, I should note that there is another, functionally distinct affix in Khumi which is formally identical to the first person S/A marker, a nominalizer *ka(ng)*:

(8) *ka(ng)*- nominalizer in Khumi:

<i>laáwng</i>	‘enough, sufficient’	<i>ka(ng)láawng</i>	‘sufficient amount’
<i>vang</i>	‘brighten, become light’	<i>ka(ng)vang</i>	‘light, brightness’
<i>láng</i>	‘dance’	<i>ka(ng)láang</i>	‘dance’
<i>hóeyng</i>	‘alive’	<i>ka(ng)hóeyng</i>	‘living thing’
<i>oem</i>	‘believe’	<i>ka(ng)oem</i>	‘belief, trust’
<i>asaáng</i>	‘high’	<i>ka(ng)sang</i>	‘summit, high point’
<i>pvuy</i>	‘drunk’	<i>ka(ng)vuy</i>	‘intoxication’
<i>ptha</i>	‘itchy’	<i>kamtha</i>	‘itchiness’
<i>pnuung</i>	‘blacken’	<i>kamnuung</i>	‘black thing’
<i>pthuú</i>	‘crazy’	<i>kamthú</i>	‘crazy person’
<i>psieng</i>	‘reddden’	<i>kamsieng</i>	‘red thing’

In light of this highly similar morphology, I would speculate that the Khumi first person S/A marker perhaps actually comes from this nominalizer, and not, as might otherwise be suspected, from the *kaay* first person pronominal element.

On this account, the original element in Khumi, and in other Kuki-Chin languages by extension, would have been from this nominalizer, and not from the first person independent pronoun. This initial grammaticalization was followed by reanalysis in other Kuki-Chin languages of the marker as originating in the first person independent pronoun, which was formally quite similar; thereafter, other languages added agreement markers based on other pronominal elements (a development which never occurred in Khumi).

Alternatively, as Scott DeLancey suggested at the conference, Khumi could have undergone a development of the sort described here, but in other languages the first person marker could have come from the normal grammaticalization source for first person verbal pronominal morphology, the independent first person pronoun. Thus, while the languages would end up with fairly similar looking first person verbal morphology, the grammaticalization sources for the one in Khumi and the ones found elsewhere in the family would be different.

4.2. Kuki-Chin suffixal pronominal morphology as archaic

On the other hand, there are indications that much of the suffixal agreement morphology that occurs in Kuki-Chin is archaic. It does appear possible to reconstruct this morphology to Proto-Kuki-Chin.

Consider, for instance, the suffixal agreement found in colloquial style contexts in Tedim, as seen in Henderson’s sketch of the language:

- (9) Suffixal agreement in Tedim (Henderson 1965:109-111):
 -iŋ '1st singular' -uŋ '1st pl excl' -haŋ '1st pl incl'
 -tɛʔ '2nd singular' -uɛʔtɛʔ '2nd plural'
 -(i/?)'3rd singular' -uʔ '3rd plural'

Generalizations which should be taken away from this paradigm include the following: -ŋ marks first person, -tɛʔ marks second person, and -ʔu(?) marks plural. A highly similar system is seen in (10) for closely related Sizang.

- (10) Suffixal agreement in Sizang (Stern 1963:264):
 bo-ŋ 'I not' bua-u-ŋ 'we not'
 bua-tɛ' 'you not' bua-u-tɛ' 'y'all not'
 bo/bua' 'he/they not'

In (11), notice that much of the morphology present in Tedim and Sizang is also present in the negative agreement paradigm in Hyow, like Khumi, spoken in southeastern Bangladesh, and a language which is usually thought to be a relatively remote sister to northerly Chin languages.³

- (11) Suffixal agreement in Hyow:

affirmative:		negative:	
singular			
<i>ka-kap</i>	'I cry'	<i>kap-ŋa</i>	'I don't cry'
<i>na-kap</i>	'you cry'	<i>kap-ti</i>	'you don't cry'
<i>ʔa-kap</i>	'she cries'	<i>kap-aʔ</i>	'she doesn't cry'
dual			
<i>kihni-kap</i>	'we two cry'	<i>kap-hniʔ-ŋa</i>	'we two don't cry'
<i>ni-kap</i>	'we (incl) two cry'	<i>kap-pu</i>	'we (incl) two don't cry'
<i>hni-hni-kap</i>	'you two cry'	<i>kap-hniʔ-ti</i>	'you two don't cry'
<i>hni-kap</i>	'they two cry'	<i>kap-huiʔy</i>	'they two don't cry'
plural			
<i>kini-kap</i>	'we cry'	<i>kap-ʔu-ŋa</i>	'we don't cry'
<i>nini-kap</i>	'y'all cry'	<i>kap-cu</i>	'y'all don't cry'
<i>ni-kap</i>	'they cry'	<i>kap-ʔu</i>	'they don't cry'

³ Actually, I have argued elsewhere that traditionally Northern Chin languages, like Tedim and Sizang, and at least some of the traditionally Southern Chin languages, like Hyow, should be subgrouped together based on shared phonological and morphosyntactic innovations, in contradistinction to Central Chin languages like Lai and Mizo (Peterson 2000). However, even with this scenario, the suffixal agreement paradigm is viewed as a shared retention, for reasons outlined in what follows.

There are a number of reasons to think that much of this morphology is archaic.

To begin with, the *ŋa* first person element is clearly from the reconstructed Proto-Tibeto-Burman root for first person. The antiquity of this highly grammaticalized *ŋa* element would indicate that it existed in this agreement use already at the Proto-Kuki-Chin stage: once the *kaay* first person element was innovated, which clearly had happened by the Proto-Kuki-Chin stage since the element occurs in virtually every language in this use, it would be hard to explain a subsequent grammaticalization of a *ŋa* element (though it is feasible that these could have been two concurrent first person formatives at some stage).

Second, there are (sometimes obscure) traces of many of these pronominal elements in Central Chin languages and in Khumi. There is a first person *-ŋ* formative found in Lai singular cohortatives. The second person person *-tV* element is probably reflected in the Mizo *-te?* imperative marker (Chhange 1993:105). In addition, there is a plural marker *-u* in dual and plural cohortatives and imperatives in Lai. Khumi also has this element in plural imperatives, and rarely as a plural agreement marker in certain subordinate clause types (for older speakers), as shown in (12).

- (12) *acie* *khúmii-loe* *vay-ktí=ya* *reng-u-pyaáw*
 1P.INCL Khumi=TOP now-future=GOAL hold.festival-PL-POT
khaá=poe *tmang-u-noe* *alang-cié=moe* *thúy-noe*
 time=ALSO ert-PL-NZR other-PL=DEF say-NZR
pree=ya *yaáŋg-noe*
 country=GOAL spread-NZR
 ‘We Khumi, in the future, if we’re able to hold a festival and we make a mistake, others will talk about it (negatively).’ (13:42)

5. Conclusion

In summary, while I cannot at this point extend my observations to varieties of Khumi which are spoken in adjacent areas of Burma, Khumi as spoken in Bangladesh clearly *does* have verbal pronominal marking. The system is optional and primarily codes speech act participants.

These properties suggest that in comparison to the more highly grammaticalized agreement systems found elsewhere in the family, this system is likely to be archaic or an independent development. If either assumption is accepted, the prefixal agreement systems seen in Kuki-Chin are probably not demonstrable retentions from Proto-Kuki-Chin, although portions of the suffixal systems found in Kuki-Chin almost certainly are.

Future work in this area will have to focus on further description of agreement systems in the family. In addition, the next step will have to be an attempt to do a

genuine morphological reconstruction, using established methodology, of prefixal morphology, which I expect will be possible for certain subgroups, but will prove elusive for Kuki-Chin as a whole.

References

- Bedell, George. 2000. Agreement in Cho. Paper presented at ICSTLL 33, Bangkok/Trang.
- Chhange, Lalnunthangi. 1993. Mizo syntax. PhD dissertation, University of Oregon.
- DeLancey, Scott. 1989. Verb agreement in Proto-Tibeto-Burman. *BSOAS* 52.2:315-33.
- Goddard, Ives. 1990. Aspects of the topic structure of Fox narratives: Proximate shifts and the use of overt and inflectional NPs. *IJAL* 56.3:317-340.
- Grierson, G.A., ed. 1903-1904. *Linguistic survey of India, Vol. III: Tibeto-Burman family* (3 parts). Calcutta: Office of the Superintendent, Government Printing.
- Hartmann, Helga. 2000. The Daai pronouns and pronominal agreement system: Alternate sets and focus. Paper presented at the Fifth International Symposium on Languages and Linguistics: Pan-Asiatic Linguistics, Ho Chi Minh City.
- Henderson, E.J.A. 1965. *Tiddim Chin: A descriptive analysis of two texts*. Oxford: Oxford University Press.
- LaPolla, Randy J. 1992. On the dating and nature of verb agreement in Tibeto-Burman. *BSOAS* 55.2:298-315.
- Löffler, L. 1960. Khumi/Khumi-Vokabulare: Vorstudie zu einer sprachwissenschaftlichen Untersuchung. *Anthropos* 55:505-557.
- Peterson, David A. 2000. On the status of the Southern Chin subgroup. Paper presented at ICSTLL 33, Bangkok/Trang.
- Peterson, David A. 2001a. Agreement and grammatical relations in Hyow. Paper presented at the Southeast Asian Linguistics Society, Bangkok.
- Peterson, David A. 2001b. Transitivity and affecting valence in Khumi. Paper presented at the Workshop on Tibeto-Burman languages, Santa Barbara.
- Shafer, Robert. 1944. Khimi grammar and vocabulary. *BSOAS* 11:386-434.
- Stern, Theodore. 1963. A provisional sketch of Sizang (Siyin) Chin. *Asia Major* 10.2:222-278.

Department of Linguistics
Max Planck Institute for Evolutionary Anthropology
Inselstrasse 22
D-04103 Leipzig, Germany
peterson@eva.mpg.de

South-East Asian Features in the Munda Languages:
Evidence for the Analytic-to-Synthetic Drift of Munda

PATRICIA JANE DONEGAN AND DAVID STAMPE
Linguistics, University of Hawai`i at Mānoa

This paper, written in memory of Eugénie J. A. Henderson and A. K. Ramanujan and read at the celebration of James Matisoff on his retirement, expresses our thanks for all their deep insights into the languages of South and South-East Asia.

1. Opposite Orders of Thought

The Munda (South Asian) and Mon-Khmer (South-East Asian) branches of the Austroasiatic language family are so exactly opposite at every level of structure that Sir George Grierson in his *Linguistic Survey of India* remarked that if they were descended from a common language, the language must have been adopted by peoples with opposite orders of thought (1904: v. 2, p. 2).

In (1) is a listing of typological oppositions between Munda and Mon-Khmer, adapted from Donegan & Stampe 1983. That paper showed how their opposite synthetic vs. analytic traits might be explained as due to polar drifts driven by their opposite – falling vs. rising – phrase and word rhythms.

(1)	MUNDA	MON-KHMER
Phrase Accent:	Falling (initial)	Rising (final)
Word Order:	Variable – OV, AN, Postpositional	Rigid – VO, NA, Prepositional
Syntax:	Synthetic – subj/obj agreement on verb	Analytic – no inflectional morphology
Word Canon:	Trochaic	Iambic, monosyllabic
Morphology:	Agglutinative, Suffixing, Polysynthetic	Fusional, Prefixing or Isolating
Timing:	Isosyllabic or isomoraic	Isoaccentual
Syllable Canon:	(C)V(C)	Unaccented (C)ə, accented (C)(C)V(G)(C)
Consonantism:	Stable, Geminate clusters	Shifting, Tonogenetic, Non-geminate clusters
Tone/Register:	Level tone (Korku only)	Contour tones or registers
Vocalism:	Stable, monophthongal, harmonic	Shifting, diphthongal, reductive

We will review the polarizing effects of falling vs. rising accent in section 2.

In sections 3 and following, we discuss similarities of Munda to Mon-Khmer, and argue that these must be retentions or developments from an originally rising typology, and therefore that proto-Austroasiatic was of the rising type – that it was analytic like Mon-Khmer, not synthetic like Munda.

Some linguists view a spontaneous shift of type from analytic to synthetic as impossible, and hold that a holistic drift, as from early to modern Indo-European, must arise in the phonetic decay of suffixes, with a change from synthetic to analytic grammar, and an accompanying drift from OV to VO word order, etc. On their view, the synthetic structure of Munda would have to be reconstructed for proto-Austroasiatic and then lost in Mon-Khmer. They might even argue that the loss was due to the areal influence of the analytic languages of SE Asia.

Or they might hold that a change from analytic to synthetic can occur only under the influence of synthetic languages. It has often been asserted, e.g. in the 1978 *Encyclopedia Britannica* article on Austroasiatic languages, that Munda synthetic structure must be due to the influence of the synthetic languages of South Asia. But Indo-Aryan and Dravidian are modifier-marking, in Nichols' terminology (1992), while Munda is head-marking, and even if one does not accept this dichotomy as immutable, it is hardly likely that modifier-marking languages could induce analytic languages to become head-marking languages.

2. Polar rhythms and polar drifts

The main reason the divergent structures of Munda and Mon-Khmer cannot be explained as due to convergence with other languages in their respective South or SE Asian language areas is that Munda and Mon-Khmer, and other South and SE Asian languages, do not just differ in structure: they are *opposite* at every level of structure. Such a polarization can be explained only by a linguistic principle, not a historical one, and the fact that it pervades every level of structure, from lexicon to syntax to phonetics, points to the single opposition that pervades every level of language: the opposition of falling vs. rising rhythm.

Munda and Mon-Khmer accentuation are opposed in just this way. Munda languages have falling (initial) accent in phrases and in words, while Mon-Khmer has rising (final) accent in phrases and in words. In this section we will sketch our 1983 hypotheses about how this opposition guides the syntactic, morphological, and phonological drift.

Heads of phrases, as presupposed information, tend to be accentually backgrounded relative to modifiers, and so in consistent head-last languages, phrase accent is falling (initial), as in South Asia, while in consistent head-first languages, it is rising (final), as in South-East Asia. Perhaps head-last (left-branching) order poses problems for short-term memory, because falling languages augment word order with incorporation, as in Munda and Tibeto-Burman, or case marking, as in Indo-Aryan or Dravidian.

Word accent tends to adopt the falling or rising structures of phrases, falling (accented at or near the beginning of word) as in Munda and Dravidian, vs. rising (accented at or near the end of the word) as in Mon-Khmer and Tai. Grammatical elements are backgrounded relative to lexical elements, so they are treated as extrametrical or are affixed away from the accent, so some languages with initial accent like Dravidian or Finnic have only suffixes, and some languages with final accent like Mon-Khmer have only prefixes. VC- prefixes may be infixes before C-initial roots to avoid creating heavy syllables that would invite accent.

Rising accent gives an “iambic” word, really an anacrustic syllable plus a stressed syllable, allowing word- (stress-) timing; the initial vowel is reduced or omitted, forming monosyllables with initial clusters that invite consonant shifts and registers or contour tones on the bimoraic and highly diphthongizable final vowel (Matisoff 1973), e.g. Mon-Khmer *[bə'lu:] Khmer [p̚ləu] ‘thigh’. Falling accent gives a “trochaic” word, both syllables within the bimoraic beat, inviting harmony (Munda [b̚ulu] ‘id.’) or apocope (bimoraic [b̚ul]), but as suffixes are piled on, isochrony at the word level becomes impossible, timing focuses on the syllable or mora, and vowels and consonants are far more stable than under stress-timing (Donegan 1993).

A holistic reversal of typology seems to require a reversal of accentuation. Germanic, Italic, and Celtic, for example, originally had head-last phrases, with falling accent, as is evident in the front-rhymed (alliterative) forms of their early verse, but they shifted to head-first phrase structure, with rising accent, and end-rhymed verse. Morphology lags behind: the ordering of compounds and affixes remains head-last in English long after phrases became head-first, and it might even be argued that the order of compounds like *blackbird* is what has retarded the reversal of adjective-noun phrases like *black bird*. But this lag can preserve a hint of the history, or prehistory, of a language.

The reversal of typology in Austroasiatic has been even more profound than in Indo-European: Munda languages are more synthetic than proto-Indo-European, and Mon-Khmer languages are far more analytic even than English. Perhaps this reflects a greater time-depth for Austroasiatic than for Indo-European.

Now we will proceed to the evidence that the reversal in Austroasiatic was opposite that in Indo-European – that proto-Austroasiatic had an analytic and head-first structure like that of Mon-Khmer, but that Munda drifted to a synthetic and head-last structure due to a reversal from rising to falling accent.

3. Vocabulary

3.1. Cognates. The evidence of the original linguistic unity of Munda and Mon-Khmer has rested, and still rests, mainly on lexical cognates. Though the vocabulary that we can reconstruct as Austroasiatic is far smaller than that for proto-Munda and proto-Mon-Khmer, and those are a magnitude smaller than the shared vocabularies of Indo-Aryan or of Dravidian, the Austroasiatic vocabulary is still solid enough to leave no doubt of the unity of the family. Despite losses

due to borrowing, we have Austroasiatic cognates for the basic verbs and nouns relating to body, family, home, field, and forest, and for pronouns, demonstratives, and numerals. Agricultural vocabulary points to a very early SE Asian homeland (Zide & Zide 1976), but that does not prove that proto-Austroasiatic was of the analytic type now identified as “South-East Asian”.

3.2. Word Structure

We compared Mon-Khmer “iambic” or monosyllabic words and Munda “trochaic” words in section 2. Evidence that Munda trochaic words derive from proto-Austroasiatic originals of the Mon-Khmer iambic type include (a) The tendency for $C_1V_1C_2V_2(C_3)$ cognates of the type Mon-Khmer [bəˈluː] : Munda [ˈbulu] ‘thigh’ to show a harmonic V_1 in Munda for the unaccented neutral V_1 of Mon-Khmer, and (b) the high frequency of simple CVCV(C) words in Munda as opposed to equally admissible CVCCV(C) words.

3.3. Affixation

Mon-Khmer and other Mainland SE Asian language families have only prefixes and infixes. This is peculiar to head-first languages, just as having only suffixes is peculiar to head-last languages like Dravidian and Finnic. Munda has prefixes and infixes cognate to Mon-Khmer, but it also has even more suffixes. If these suffixes had existed in proto-Austroasiatic, and had been lost in Mon-Khmer, we would not expect to find Mon-Khmer cognates for the Munda suffixes. In fact, we do find Mon-Khmer cognates, but they are independent words in Mon-Khmer. For example, Munda languages mark the plural of nouns and 3rd plural of verbs with suffixes like *-ku*, *-ki*, *-gi*, *-ji*. Mon-Khmer languages lack number suffixes, but many have free-standing 3rd plural pronouns like Khasi *ki* ‘they’.

Munda languages mark possessive and object persons with suffixes, e.g. Sora /siʔiŋ-lən/ ‘our house (lit. house-us)’, /ərgəl-da-tə-lən/ ‘we’re thirsty (lit. thirst-affect-nonpast-us)’. Mon-Khmer languages lack person suffixes, but they have free-standing personal pronouns cognate to the Munda suffixes. Here are examples from Pinnow’s extensive 1965 study:

(3.3)	Proto-Munda	Mon-Khmer
‘1 sg.’	*-iŋ	Pear, Bahnar <i>iŋ</i> ; Mon <i>ai</i> (oa); Srê <i>ʔaŋ</i> ; Khmer <i>ən</i>
‘2 sg.’	*-me	Khmer <i>me</i> , Bahnar <i>mih</i> , Srê <i>mi</i> , Khmu’ <i>mee</i>
‘1 pl.’	*-le/-ne	(See Pinnow 1965: 23ff.)
‘2 pl.’	*-pe	Palaung <i>pe</i> , Riang <i>peʔ</i> , Mon <i>beh</i> (<i>pih</i>), Wa <i>puː-i</i> , Khasi <i>phi</i>

The change of free pronouns to clitics and affixes is commonplace, but the change of affixes or clitics to free forms is not. We conclude that Munda suffixes derive from the synthesis of independent words, as in Mon-Khmer.

3.4. Compounding

In relatively recent compounds, Munda structure is head-last, e.g. Sora /bəŋsa-im/ ‘good chicken’, but in older compounds, the structure is head-first, e.g. Sora /əḍṛɛ-im/ ‘chicken egg (lit. egg-chicken)’, /gəd-im-na/ ‘sacrificing a chicken (lit. cut-chicken-nom.)’. The latter is clearly a reflex of Mon-Khmer-type head-first compounding.

The accentuation of compounds illustrates the natural principle that heads are accentually subordinate to modifiers. Thus the head-first compounds of Mon-Khmer show the characteristic rising rhythm of Mon-Khmer words and phrases, while the head-last compounds of Munda show the falling rhythm of Munda words and phrases, and indeed in some languages like Sora the second element of the compound is synchronically limited to one syllable.

(3.4)	Khmer	/,sac - 'koo/	/,sac - 'trəy/	/,trəy - 'ŋiət/
		flesh - cow	flesh - fish	fish - dried & salted
		‘beef’	‘fish (meat)’	‘dried & salted fish’
	Sora	/'jɛlu - ,taŋ/	/'esu - ,bəb/	/'əsəŋ - ,im/
		flesh - cow	pain - head	feces - chicken
		‘beef’	‘headache’	‘chicken dung’

That falling rhythmic patterns have been imposed on older Munda compounds with head-first structure, like /'əsəŋ - ,im/ (lit. *feces chicken*, ‘chicken dung’ – compare the full form of *chicken*, /'kənsim/), shows that this word order existed before Munda adopted the falling rhythm typical of the South Asian area and of head-last languages generally.

4. Phrasing

4.1. Syntax

Munda phrase structure is consistently head-last, with SOV and AN order, and postpositions. Mon-Khmer phrase structure is just as consistently head-first, with SVO (rarely VSO) and NA order, and prepositions.

(4.1a) ‘Monosi/Saran went to the market; he bought rice.’

Sora: Mənəsi bazar -ban yer -r -ɛ; ənin rəŋko-n ni -l -ɛ
 Monosi market-to go -past -3sg; he rice -art buy -past -3sg.

Khmer: saraan tiw dal psaa; kǎət tɿŋ ʔaŋkaa
 Saran go to market; he buy rice

(4.1b) ‘I don’t want to eat all the fish.’

Sora: *ɲɛn kuddib əyɔ -n -ədəŋ ə- jəm -ben idsim -t -ay ted*
 I all fish -art-obj inf-eat -inf want -pres -1sg not
 Khmer: *khnom min caŋ ɲam trəy cap*
 I not want eat fish all

4.2. Polysynthetic Morphology

Words are more resistant to internal changes of accent and ordering than phrases. We have already noted that noun compounds in Munda retain a head-first order. The Munda verb, which is polysynthetic, likewise shows internal head-first order, as if head-first phrases of the Mon-Khmer type were fused, with no order change.

(4.2a) ‘He didn’t give me rice’:

he not give rice me (3past)
 Sora: *anin əd- tiy- dar- iŋ- ten*
 Khmer: *kəət min aoy baay khnom*
 Sre: *khay ʔàʔ ʔay pyaŋ ʔaŋ*

Similarly for the sentence cited in (4.1b) in the Sora “syntactic” style, but in (4.2b) in its more idiomatic (and older) “morphological” (polysynthetic) style:

(4.2b) ‘I don’t want to eat all the fish.’

I not want eat fish all (-pres. -intr. -1p.)
 Sora: *ɲɛn əd- məl- jəm -yɔ -aj -t -en -ay*
 Khmer: *khnom min caŋ ɲəm trəy cap*

5. Phonology

5.1. Vowels

The vowel systems of Mon-Khmer and SE Asia generally are among the most complex in the world, and even at the proto-Mon-Khmer level they share the peculiarity of having three central or back unrounded vowels (Shorto 1976). Munda vowel systems mostly appear as triangular systems of five vowels, like the typical systems of Indo-Aryan and Dravidian languages, but a striking exception is the Sora system, whose three central vowels look very Southeast Asian:

(5.1)	high	i	i	u	e.g.	[id̥-] ‘scratch’, [id̥-] ‘fan’, [-lu̥d̥-] ‘ear’
	mid tense	e		o		[-əd̥-] ‘thorn’, [-lo̥d̥-] ‘cord’
	mid lax	ɛ	ə	ɔ		[ɛd̥-] ‘roll’, [əd̥-] ‘prop’, [ɔd̥-] ‘knead’
	low		ɑ			[ɑd̥-] ‘drive (cattle)’

More striking is the fact that at the lowest levels of reconstruction, it is necessary to reconstruct three central vowels for every Munda subgroup: Sora-Gorum

(A. Zide 1982), Gutob-Remo (N. Zide 1965), Kharia-Juang (Stampe 1978), Kherwarian (Munda 1969), and Korku-Kherwarian (N. Zide 1965). These point to a proto-Munda and proto-Austroasiatic system like that of proto-Mon-Khmer.

Some South Munda languages have vowels with glottals, as in Sora, where [V_i?V_i] in free forms alternates with [V_i] in combining forms, e.g. [jaʔaŋ] ‘bone’ beside [ə-pəlaŋ-jaŋ] ‘broken bone’, and in 1965 Norman Zide proposed that these and a number of vowel and consonant puzzles in Munda history might be solved by a proto-Munda series of laryngealized vowels. In 1989 Diffloth gave evidence of creaky-voiced vowels in proto-Mon-Khmer. Vowel registers are rare in South Asia but common in SE Asia; if the correspondences can be resolved, this would be another Mon-Khmer-like feature of Munda.

5.2. Consonants

Indo-Aryan languages have released final consonants, and Dravidian languages end words with an “enunciative” vowel. In contrast, Munda languages typically have unreleased final consonants. In older handbooks these were called implosive, in the sense of “not plosive” rather than “inwardly plosive”, which led some phonological surveys to count them wrongly as ingressive; they are just unreleased, glottalized, and voiceless as in English *cat* [kætʰ], Cockney [kæʔ]. This “checking” of final stops is commonplace in Mon-Khmer and other mainland SE Asian languages. Presumably it was a proto-Austroasiatic feature, because while proto-Mon-Khmer and proto-Munda had voiced as well as voiceless stops nonfinally, there is no evidence of more than one series finally. In the absence of suffixes, as in Mon-Khmer, the invariably checked final stops are lexically voiceless. But in Munda, final stops before vocalic suffixes alternate with their voiced equivalents, as in these Sora examples:

(5.2a)	[gəʔ'.lot'.len]	but	[lo.a.dən gə.də]
	/gəʔ -loʔ -l -ən/		/loaʔ -ən gəʔ -a/
	cut -rope -pa -intr		rope -art cut -imp
	‘He cut the rope’		‘Cut the rope’
	[əp'.daj'.tay]	but	[ə.bə.dʒa]
	/əb- daj -t -ay /		/əb- əj -a/
	caus- climb -pr -lsg		caus- row -imp
	‘I make s.o. climb’		‘Make s.o. row’

Nonfinal voiceless and voiced stops contrast before vowels (e.g. [əpan] ‘easily’ vs. [əba-n] ‘mohwa tree’), but as the phonemic notations above indicate, the final voiceless stops are identified not with the voiceless non-final stops but with the voiced ones. Even a voiceless final stop in a foreign word like English *pipe*, when suffixed with the article /-ən/, is revealed as voiced: Sora [pa.ɛ.bən].

In Mundari, a similar analysis of stops in Mundari causes some speakers to produce final stops as nasally released voiced stops, e.g. [dup'] ~ [dub^m] 'to sit' (Osada 1992), and English *David* as [dɛ.biɫ'] ~ [dɛ.biɫ^m] (personal observation).

In fact, one Munda language, Juang, has lost final checking, and its previously checked and voiceless final stops have emerged as voiced, not voiceless:

(5.2b)	Juang		Kharia		Sora	
'head'	/bokob/	[bo.kob]	/bokob/	[bo.kop']	/bokob/	[bɔ.ʔɔp']
'mouth'	/tomod/	[to.mod]	/tomod/	[to.mot']	/tomod/	[tɔ.ʔɔɫ']
'water'	/dag/	[dag]	/dag/	[daʔ]	/dag/	[da.ʔa]
'die'	/goj/	[godʒ]	/goj/	[goɪ']	/goj/	[goɪ']

For the lexical representation of *all* the morpheme-final stops in a language to be perceived as voiced, even though they are voiceless except before a vowel, is quite extraordinary, but it is clearly a fact of Munda. As to how the prevocalic forms of the stops became voiced, we believe that it was by exactly the same process as in Sanskrit, where word-final stops were voiced before vowels. Proto-Austroasiatic, like Mon-Khmer, had both voiceless and voiced stops initially, but only voiceless stops finally, and the final stops were invariably voiceless because there were no affixes to block devoicing. But when Munda began to use clitics and suffixes, word-final stops must still have been syllable-final, i.e. VC#V = VC.V, and in syllable-final but intervocalic position they assimilated voicing, just as Sanskrit did in word-final (presumably syllable-final) stops in external sandhi:

(5.2c) Sanskrit: *tat aśvaḥ* → *tad aśvaḥ* 'that horse'

Kharia: /mod/ -nan/ [mɔ'ɲɲɲ], cf. /mod/ -ag/ [mo.dəʔ]
 eye -my 'my eye' eye -gen. 'of the eye'

Non-word-finally, intervocalic stops have the natural syllable division V.CV and, exactly as in Sanskrit (e.g. *pi.tah* 'father'), they do not become voiced:

(5.2d) Sora: /e.ten/ 'what?', /pə.tod/ 'hole', /ə.pan/ 'easily', /dɛ.ko/ 'stay'

Mundari: /ga.pa/ 'tomorrow', /se.taʔ/ 'morning', /ti.kin/ 'noon';
 cf. /ho.ɽo.ko/ 'person-pl., they'

What is significant for our thesis is that this voicing of stops before vocalic suffixes in Munda could only have occurred when Munda joined syllable-final checked stops to vocalic suffixes. As soon as the suffixes became integral parts of words, the syllabication of stops was naturalized to V.CV in all the Munda languages (see the examples in 5.2a–d above). So the reinterpretation of final stops as voiced must be a reflex of the moment when Munda languages crossed over from a non-suffixing Austroasiatic morphology (like that of Mon-Khmer) to a suffixing morphology.

6. Summary

We have argued that proto-Austroasiatic had the same analytic, head-first, and rising structure as its daughter Mon-Khmer and other mainland South-East Asian languages, and that the Munda languages have preserved clear evidence of that structure even as they evolved toward the synthetic, head-last, falling structure typical of other South Asian languages. That does not necessarily mean that the speakers of proto-Austroasiatic were actually in South-East Asia, or that the Munda changes took place in South Asia. But it does mean that Munda is a clear example of a drift that was exactly the opposite of the drift that is familiar from Indo-European, toward analysis. Further, the drift of Munda was more complete than that of Indo-European, since it began with one of the most analytic structures among the languages of the world, and ended with one of the most synthetic.

References¹

- Diffloth, Gérard. 1989. Proto-Austroasiatic creaky voice. *Mon-Khmer Studies* 15: 139-154.
- Donegan, Patricia J. 1993. Rhythm and vocalic drift in Munda and Mon-Khmer. *Linguistics of the Tibeto-Burman Area* 16: 1-43.
- & David Stampe. 1983. Rhythm and the holistic organization of language structure. In John F. Richardson, Mitchell Marks, & Amy Chukerman (eds.) *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*: 337-353. Chicago: Chicago Linguistic Society.
- Grierson, George A. 1904. *Mon-Khmer and Siamese-Chinese families. Linguistic survey of India, Vol. 2*. Reprinted 1966, Delhi: Motilal Banarsidass.
- Henderson, Eugénie J. A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. In Milner & Henderson 2: 400-434.
- Huffman, Franklin E. 1986. *Bibliography and index of Mainland Southeast Asian languages and linguistics. (Yale Language Series.)* New Haven: Yale University Press.
- Jenner, Philip M., Laurence C. Thompson, & Stanley Starosta (eds.). 1976. *Austroasiatic studies*. (Proceedings of the First International Conference on Austroasiatic Linguistics.) 2 vols. Honolulu: University of Hawai'i Press.
- Masica, Colin. 1976. *Defining a linguistic area*. Chicago: University of Chicago Press.
- Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed.), *Consonant types and tone (Southern California Occasional Papers in Linguistics, 1)*: 71-95. Los Angeles: UCLA.

¹ Uncredited Munda citations in this paper for languages not mentioned in the bibliography are from our own field notes and recordings. For areal and typological facts we have relied heavily on the indispensable works of Ramanujan and Masica for South Asia and Henderson for SE Asia.

- Milner, George B., & Eugénie J. A. Henderson (eds.). 1965. *Indo-Pacific linguistic studies*. 2 vols. Amsterdam: North-Holland.
- Munda, Ram Dayal. 1969. *Proto-Kherwarian sound system*. M. A. thesis, University of Chicago.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Harrassowitz.
- . 1960. Über den Ursprung der voneinander abweichenden Strukturen der Munda- und Khmer-Nikobar-Sprachen. *Indo-Iranian Journal* 4: 81-103.
- . 1965. Personal pronouns in the Austroasiatic languages: a historical study. In Milner & Henderson 2: 3-42.
- Ramanujan, A. K., & Colin Masica. 1969. Toward a phonological typology of the Indian linguistic area. In T. A. Sebeok (ed.) *Current trends in linguistics* 5: 543-577. The Hague: Mouton.
- Shorto, Harry L. 1976. The vocalism of Proto-Mon-Khmer. In Jenner et al. 2: 1041-1067.
- Stampe, David. 1978. Proto-Central Munda phonology. Second International Conference on Austroasiatic Linguistics, Mysore, 1978.
- Zide, Arlene R. K. 1982. *A reconstruction of proto-Sora-Juray-Gorum phonology*. Ph.D. thesis, University of Chicago.
- & Norman H. Zide. 1976. Proto-Munda cultural vocabulary: Evidence for early agriculture. In Jenner et al. 2: 1295-1334.
- Zide, Norman H. 1965. Gutob-Remo vocalism and glottalised vowels in Proto-Munda. In Milner & Henderson 1: 43-53.
- , ed. 1966a. *Studies in comparative Austroasiatic linguistics*. The Hague: Mouton.
- . 1966b. Korku low tone and the Proto-Korku-Kherwarian vowel system. In Zide 1966a 1: 214-229.

Linguistics Department
University of Hawai'i at Mānoa
1890 East-West Road, Honolulu, HI 96822

donegan@hawaii.edu, stampe@hawaii.edu

Parallel innovation and ‘coincidence’ in linguistic areas: on a bi-clausal extent/result construction of mainland Southeast Asia¹

N. J. ENFIELD

Max Planck Institute for Psycholinguistics, Nijmegen

The notion of a ‘linguistic area’—‘a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor’ (Thomason 2000:311)—presupposes that neighbouring languages can share structural features for three reasons: (1) due to ‘genetic inheritance’ from a common ‘ancestor’ language; (2) due to ‘diffusion’ of features from one language into the other (or indeed, from a third language into both) via language contact; (3) due to accident or coincidence. There is another way, however, in which neighbouring languages may share structure, namely due to a certain kind of parallel yet independent innovation, one which is not due *directly* to language contact, but which also is not accidental. Certain grammatical structures which languages share due to previous historical contact may cause certain innovations to be very likely for language-internal reasons, and these innovations may arise spontaneously and independently in separate languages.

The issue arises when we consider the status of grammatical constructions—i.e. morphosyntactic structures with specifiable meanings—in linguistic areas. This paper discusses a bi-clausal construction with ‘extent/result’ semantics found in a number of languages of the mainland Southeast Asia area. Mainland Southeast Asia is the home of hundreds of languages from at least six distinct language families—Austronesian, Hmong-Mien, Mon-Khmer, Sinitic, Tai, and Tibeto-Burman—which show a high degree of parallelism in a number of structural domains (Enfield in press, Chapter 2; cf. Capell 1979, Clark 1989, Bisang 1991, Matisoff 1991, 2001, Enfield 2001). While the structure and meaning of a construction can appear closely analogous across languages, the lexical ‘marker’ employed for that construction is in many cases not cognate (even when the two

¹ This paper is dedicated to Jim Matisoff, in humble appreciation of a true pioneer and a very nice fellow.

languages share cognates of a given marker). When two languages share a construction marked by non-cognate lexical material, one would normally assume they did not inherit the construction from a common ancestor. In such a case, the construction may have been calqued (i.e. where a language has borrowed an *idea* of structuring meaning in a certain way, and has not borrowed any linguistic *form* with it), or it may have been separately and independently innovated in the languages. The existence of non-cognate structural parallelism raises interesting questions regarding language contact and change. The construction and its ‘marker’ (i.e. the stable lexical material which identifies it) are logically separate with respect to the possible avenues of explanation of structural parallelism—shared genetic inheritance, contact-related diffusion, or ‘coincidence’.

1. Extent/result complement constructions

Many languages of mainland Southeast Asia share a construction with the following properties:

structure:	CLAUSE-1 “MARKER” CLAUSE-2
meaning:	‘Clause-1 was the case to such an extent that Clause-2 became the case as a result’ (or: ‘It was so Clause-1 that Clause-2’).

The semantic relationship between CLAUSE-1 and CLAUSE-2 involves both cause and temporal subsequence (namely, ‘CLAUSE-1; because of this, after this, CLAUSE-2’, with the more specific idiomatic English translation *so VI that V2*).

We first consider Lao (Southwestern Tai), Kmhmu Cwang (Eastern Mon-Khmer), Hmong (Hmong-Mien) and Vietnamese (Eastern Mon-Khmer). In these languages, the markers of the extent/result construction are derived from lexical items which are non-cognate, but which each express at an abstract level an idea of ‘coming to be in the same place’, common to concepts such as ‘reach’, ‘arrive’, ‘connect’, ‘up to’, and ‘touch’.

Speakers of Lao mark the construction with *con*³ ‘up to, until’:

- (1) *khaw*³*ten*⁴ *con*³ *mɯŋ*¹*laa*³
Lao 3PL jump until tired very
‘They jumped until they were very tired; They jumped so much they were very tired.’
- (2) *khaw*³*ha*⁵ *con*³ *phaa*⁵-*set*¹-*mɯŋ*² *piak*⁵ *met*²
Lao 3PL cry until cloth-wipe-hand wet all
‘They cried until their handkerchiefs were all wet; They cried so much their handkerchiefs got all wet.’
- (3) *dang*³ *con*³ *khon*² *nɔn*² *bɔ*⁰ *lap*²
Lao loud until person lie NEG sleep
‘(They) made so much noise people couldn’t sleep.’

The following example shows the marker *con*³ with its purely temporal meaning ‘until’:

- (4) *man*² *juu*¹ *huan*² *con*³ *haa*⁵ *moong*²
 Lao 3SG be.at houseuntil five o'clock
 ‘He was at home until five o’clock.’

In Kmhmu Cwang, the construction is marked with *cam* ‘connect to, link’:

- (5) *No tềng nan cam gôn yat da' gang leh*
 Km 3PL make loud connect person LOC be.at housebeside
ni' sih am bwan.
 this sleep NEG can
 ‘They made such a noise that the people in the next house were unable to sleep.’
- (6) *Ge 'mook nangsw cam uat tnoh.*
 Km 3MSG tell writing connect tired mouth
 ‘He taught until his mouth was tired.’

Here are two examples showing *cam* with the meaning ‘up to’, ‘touching’, ‘connecting’:

- (7) *Ô' yoh cam Viangcan.*
 Km 1 go reach V.
 ‘I went (all the way) to Vientiane.’
- (8) *cam yo'*
 Km touch RCP
 ‘touching (of two things)’

In Hmong, the extent/result complement construction is marked with *txogqhov* ‘until, up to’:

- (9) *Nws ua ntshoo txogqhov kuv pw tsis tsawgzog ib hmos.*
 Hm 3 do/make noise until I lie NEG sleep one night
 ‘They made (such a) noise that I couldn’t sleep all night.’
- (10) *Nws sau ntawv txogqhov tsaug leeg tes tas.*
 Hm 3 write writing until tired sinew hand all
 ‘He wrote so much his hands were completely tired.’

The following example shows Hmong *txog*, the first element in *txogqhov*, meaning ‘reach’:

- (11) *Kuv mus tsis txog phoosxavvam.*
 Hm 1 go NEG reach P.S.
 'I didn't get to Phone Savanh.'

In Vietnamese, the extent/result complement construction is marked by *đến-nối* 'until' (where *đến* appears elsewhere as a main verb meaning 'reach'):

- (12) *Anh ấy làm tiếng động đến-nối nhà bên cạnh*
 Vn man that make sound loud until housedirection side
không ngủ được.
 NEG sleep can
 'He made loud noise such that the (people in the) house next door couldn't sleep.'
- (13) *Anh ấy dạy đến-nối mệt.*
 Vn man that teach until tired
 'He taught until he was tired.'

The structure and meaning of these constructions are essentially the same across the languages (although of course one would expect to find minor grammatical distinctions). One explanation for this would be that the constructions have been calqued across two or more of these languages (i.e. the *idea* of the construction was borrowed across languages, with each individual language providing its own constructional marker). Another explanation would be that the similarity of these constructions is merely a coincidence. But how could such a 'coincidence' come about? A semantically and structurally specific construction does not just happen to exist in a language, but develops historically. And historical development of particular grammatical constructions is contingent upon existing grammatical and semantic structure, which constitute input to well-known processes of inference and extension which, in turn, drive semantic/grammatical change (Hopper and Traugott 1993, Harris and Campbell 1995, Traugott and Dasher 2002, *inter alia*). It is often difficult if not impossible to tell whether a construction common to two neighbouring languages has been calqued or independently innovated, and perhaps this is because the distinction is in reality not especially clear (Enfield *in press*: 368). If neighbouring languages independently innovate a semantically and structurally similar construction, chances are that the grammatical and semantic structures which provided the environment for the innovation to develop were also shared, likely due to contact at an earlier stage.²

² This point may be more forcefully made with reference to constructions of greater semantic and structural specificity. See Enfield (*in press*) for detailed discussion.

We turn now to Sinitic languages. The marker of the extent/result complement construction in Modern Standard Chinese (MSC) is *de*, a morpheme historically meaning ‘acquire, obtain’. There is a whole family of constructions in which *de* links a predicate with a complement of some kind, and these have been intensively studied in MSC (cf. Enfield in press: 268). We are here interested only in the construction expressing extent/result, as in the following examples:

- (14) *tā men tiào de hěn lèi*
 MSC 3 PL jump DE very tired
 ‘They jumped till (they) got very tired.’ (Huang 1988:274)
- (15) *tā men kū de shǒupà dōu shī le*
 MSC 3 PL cry DEhandkerchief all wet PFV
 ‘They cried so much that even the handkerchief got wet.’ (Huang 1988: 274)
- (16) *chǎo de rénjiā shuì bù zháo*
 MSC noisy DEother.people sleep NEG can
 ‘(They) made so much noise that other people couldn’t sleep.’ (Chao 1968:355)

Lamarre (2001) surveys constructions of this kind across Sinitic languages, and shows that despite these languages indisputably belonging to the same language family, they can each display a single construction with very similar semantics and very similar structure, yet use a wide range of different non-cognate lexical items to mark the construction. While the constructional marker in MSC is a verb meaning ‘obtain, acquire’, many neighbouring languages (Sinitic and otherwise) mark the same construction with other verbs, including ‘go’, ‘reach’, ‘come’, ‘attach’, and ‘arise’ (Lamarre 2001). There is a certain abstract commonality in the semantics of these markers, all of them associated with ‘reaching’ or ‘coming together’ (compare the other Southeast Asian languages discussed above).

The following examples show structures analogous to the MSC structures in (14-16), from other Sinitic languages Xinyi (Yuè), Beijing Mandarin (Mandarin), and Cantonese (Yuè), respectively, in which the constructional marker is not cognate with MSC *dé* ‘obtain’, but rather with *dào* ‘reach’ (examples from Lamarre 2001:99-101):

- (17) *k'œi¹³ fun⁵³hei³⁵ tou³⁵ tai¹¹tai¹¹ seŋ⁵³ kom³⁵ ham³³*
 Xinyi3SG happy ‘reach’ loud-RDP voice PCL shout
 ‘He was so happy he shouted out.’ (Tang 1986:101)
- (18) *wǒhèn tā hèn dào tīng-jiàn tā de míng-zi jiù shēng-qì*
 BjM 1 hate 3 hate ‘reach’ hear 3 POSS name so get.angry
 ‘I hate her to the point that [or: ‘so much so that’] I get angry simply by hearing her name.’ (Chao 1926:876)

(19) *kui⁵ bei² ngoh⁵ man⁶ do³ hau² nga²nga²*
 Cant 3 by 1 ask 'reach' mouth mute

'He was questioned by me so closely he could not answer anything.' (Peng 1993:91; transliteration and translation as given in Lamarre 2001)

While the three languages possess morphemes cognate with MSC *de* (**tak* in Proto-Chinese), it happens that they do not use those morphemes to mark this construction, despite its close parallelism with the same construction in their sister language MSC (with which they are each now in contact). There is, of course, no possibility that when these languages inherited their cognate of MSC *de*, they would have inherited this modern construction with it. Proto-Chinese **tak* itself had little of the present range of functions it shows in MSC today. With extensive historical written records available, historical linguists have attempted to reconstruct the development of **tak* in Sinitic, and they are unanimous that its original meaning was 'obtain, acquire' (Sun 1996:108, Lamarre 2001), and, furthermore, that its modern 'auxiliary' functions are relatively recent developments. In many Sinitic languages the modern descendent of **tak* does not perform the same range of functions as it does in MSC. The extent/result construction described here has either been more recently borrowed across Sinitic languages or it has been independently innovated in the languages, and this distinction has no correlation with whether or not the constructional marker used is cognate. The 'typological poise' of the languages is very similar, increasing the likelihood of independent innovation toward the same outcome (Enfield 2001: 284-287, in press: 358-361).

3. Discussion and conclusion

In the absence of a single genetic origin common to two neighbouring languages, contact-related historical diffusion is often presumed to be the cause of structural parallelism. A possible assumption is that if neighbouring languages *are* known to be 'genetically related', then structural parallelism will *not* be due to diffusion but to their common genetic background. This assumption is not necessarily correct, as the examples from Sinitic languages above have shown. Rightly, the definition of linguistic area given at the beginning of this paper does not require that languages in a linguistic area be genetically *unrelated*. To the contrary, shared 'genetic' background of two languages with a common ancestor may be no more likely an explanation for modern parallels in their grammatical and semantic structure than processes of contact-related diffusion. Once a proto-language has split into two or more separate languages, speakers of those related languages may remain in contact (or later come back into contact), and resultant processes of structural diffusion will be essentially the same as those which pertain between unrelated languages. Languages may have specific constructions in common, and these constructions may or may not be marked by cognate material. But if certain analogous constructional markers in two separate languages *are* shown to be cog-

nate, one cannot conclude from this that the *constructions* they mark were inherited from a common ancestor.

Thus, we may ask: When a language calques a construction from a genetically and/or areally related language, what is the principle for selection of an appropriate constructional 'marker' to mirror the marker used in the source language? The form of the marker in the source language may of course be borrowed along with the construction, but this does not necessarily happen. The borrowing language may select from its own resources a morpheme unrelated to the constructional marker used in the source language. One expects, however, that the marker selected should be semantically appropriate, given the meaning of the construction.

To conclude, we have seen that structurally and semantically parallel constructions can be shared by areally contiguous languages, where the constructional pattern and the morpholexical marker of the construction are logically independent with respect to possible explanations for their sharedness (i.e. as due to borrowing or independent innovation). Further, the use of cognate morpholexical means to mark the same construction in two (even closely) related languages is no guarantee that the construction was not calqued through contact. It is necessary to recognise a type of explanation for shared semantic/grammatical structure in neighbouring languages ('genetically related' or not) aside from the standard options 'shared due to borrowing through contact' and 'shared due to coincidence', as follows. Two languages independently innovate the same new structure, but far from being coincidental, the process is licensed in the respective languages by common semantic and grammatical features which themselves came about as a result of common inheritance or direct diffusion in earlier contact between the languages. Shared semantic and grammatical structure in separate languages provides similar input structures for the processes of inference and extension which drive structural change. The result is parallel yet independent innovation due neither to direct borrowing nor to mere 'coincidence'.

References

- Bisang, Walter. 1991. Verb serialization, grammaticalization and attractor positions in Chinese, Hmong, Vietnamese, Thai and Khmer. In *Partizipation: das Sprachliche Erfassen von Sachverhalten*, edited by Hansjakob Seiler and Waldfried Premper, 509-562. Tübingen: Gunter Narr Verlag.
- Capell, Arthur. 1979. Further typological studies in Southeast Asian languages. In *South-east Asian linguistic studies, Volume 3*, edited by Nguyen Dang Liem, 1-42. Canberra: PL.
- Chao, Yuen Ren. 1926. Research on Beijing, Suzhou, and Changzhou auxiliaries. *Qinghua Xuebao*, 3.2, 865-917.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

- Chappell, Hilary. (ed.) 2001. *Sinitic Grammar: synchronic and diachronic perspectives*. Oxford: Oxford University Press.
- Clark, Marybeth. 1989. Hmong and areal Southeast Asia. In *Southeast Asian syntax*, edited by David Bradley, 175-230. (A-77) Canberra: PL.
- Enfield, N. J. 2001. On genetic and areal linguistics in mainland Southeast Asia: parallel polyfunctionality of 'acquire'. In *Areal diffusion and genetic inheritance: problems in comparative linguistics*, edited by Alexandra Y. Aikhenvald and R. M. W. Dixon, 255-290.
- Enfield, N. J. In press. *Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge.
- Harris, Alice C. and Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Huang, C.-T. James. 1988. *Wǒ pǎo de kuài* and Chinese phrase structure. *Language*, 64.2, 274-311.
- Lamarre, Christine. 2001. Verb complement constructions in Chinese dialects: types and markers. In *Sinitic grammar: synchronic and diachronic perspectives*, edited by Hilary Chappell. Oxford: OUP.
- Matisoff, James A. 1991. Areal and universal dimensions of grammaticization in Lahu. In Traugott, Elizabeth Closs and Bernd Heine, eds. 1991. *Approaches to grammaticalization*. Amsterdam: Benjamins, 383-453.
- Matisoff, James A. 2001. Genetic versus contact relationship: prosodic diffusability in Southeast Asian languages. In *Areal diffusion and genetic inheritance: problems in comparative linguistics*, edited by Alexandra Y. Aikhenvald and R. M. W. Dixon, 291-327.
- Peng, Xiao Chuan. 1993. The structural auxiliary 'dao' in Cantonese. In *Research and pedagogy of Cantonese*, edited by Zheng Ding Ou, 89-100. Guangzhou: Zhongshan Daxue Chubanshe.
- Sun, Chao Fen. 1996. *Word-order change and grammaticalization in the history of Chinese*. Stanford: Stanford University Press.
- Tang, Zhi Dong. 1986. Demonstratives in Xinyi dialect. *Yuyan Yanjiu*, 2, 98-108.
- Thomason, Sarah G. 2000. Linguistic areas and language history. In *Languages in contact*, edited by Dicky Gilbers, John Nerbonne, and Jos Schaecken. Amsterdam: Rodopi, 311-327.
- Traugott, Elizabeth Closs and Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Language and Cognition Group
Max Planck Institute for Psycholinguistics, Nijmegen
PB 310, 6500 AH, The Netherlands

Nick.Enfield@mpi.nl

Classifier Systems and Noun Categorization Devices in Burmese

ALICE VITTRANT

Lacito-CNRS / Université Paris VIII

This article will deal with noun categorization devices in Burmese, including the well-known numeral classifier system. It will start by a brief review of typological studies on classifiers, before focusing on the studies done especially on Burmese classifiers. Then, in a second section I will present a summary of the features and functions of NUMERAL classifiers in this language, before taking up the question of another noun classification or noun categorization device.

1. Studies on noun classification systems

1.1. Typology of classifier systems

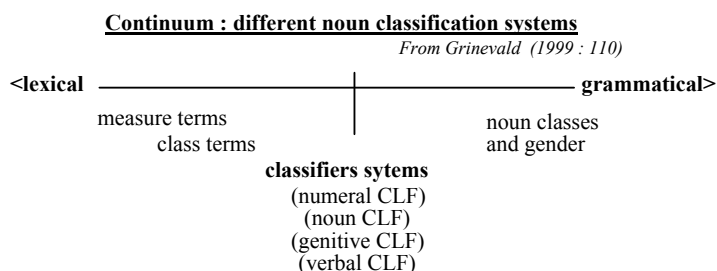
Since Greenberg's article *Nominal classifiers and substantival number* in (1972), there have been a number of proposals for a typology of noun categorization systems; Adams and Conklin (1973), Denny (1976), Allan (1977), Seiler (1980), Croft (1994) provide essentially semantic cross-linguistic criteria for classification, whereas Dixon's noun categorization analysis (1986) is based on grammatical features of classifiers, and makes a clear distinction between noun class systems and classifier systems.

Further typological studies on classifiers have been proposed by Craig (1992, 1993, 1999, 2000), Bisang (1993, 1999) and Aikhenvald (1998, 2000), extending the number, the types and the features of noun categorization systems.

Bisang (1993) focuses on the functions of classifiers — he proposes *four operations of nominal concretization* used in classifier systems — that is to say: INDIVIDUALIZATION, CLASSIFICATION, REFERENTIALIZATION, RELATIONALIZATION. While Aikhenvald's typology has seven noun categorization devices¹, Grinevald's proposal provides only four main types of classifier systems based on

¹ Aikhenvald's typology provides seven noun categorization devices: (1) NOUN CLASS, (2) NUMERAL CLASSIFIERS systems, (3) NOUN CLASSIFIERS, (4) Classifiers in POSSESSIVE construction — with three sub-types —, (5) VERBAL CLASSIFIERS, (6) LOCATIVE CLASSIFIERS, (7) DEICTIC CLASSIFIERS. For details see (1998: 430-33), or (2000: 17-18)

morpho-syntactic features: noun classifier, numeral classifier, verbal classifier, genitive classifier (1992: 281-286). Unlike Aikhenvald², she distinguishes gender or noun class systems from classifier systems, considering them as a type of noun categorization³. Regarding noun categorization as a grammatical-lexical continuum, she places noun classes (and gender) on one end and measure terms (and class terms) on the other end — the former are the most grammatical type on this continuum, whereas the latter are the most lexical type (1999: 101). Then, she considers classifiers systems to be at the mid-point of a grammatical-lexical continuum.



In this paper, we will follow Craig's typology, rather than Aikhenvald's, given that it provides a distinction between class terms (located on the lexical part of the continuum) and noun classifiers (more grammatical) that might be relevant for Burmese.

1.2. On Burmese classifier studies

NUMERAL classifiers are one of the well-known characteristics of East and Southeast Asian languages, and the Burmese classifier system (henceforth CLF system) is in fact famous in the literature thanks to Becker's often cited example of the word "RIVER"⁴ which is presented categorized by eight different classifiers that highlight different aspects of the noun meaning.

One might speak of a river in at least eight contexts. (Becker 1975:113)

myi? tə myi?	"river one river" (the unmarked case)"
myi? tə khu'	"river one conceptual unit" (rivers in general)
myi? tə ya?	"river one place" (e.g. destination for a picnic)

² Aikhenvald (2000: 3): *The term of 'classifier systems' is used to denote a continuum of methods of noun categorization.*

³ Grinevald (2000: 74): *The claim of this typology is double: on one hand that there exists a linguistic category of 'classifiers' in some languages of the world, which is distinct from other nominal classification systems of more grammatical (gender-noun classes) or more lexical nature (measure terms, class terms).*

⁴ The categorization of the word 'river' by different classifiers had been already suggested by Hla Pe (1965: 169).

myi? tə tan	“river one line” (e.g. on a map)
myi? tə mwa	“river one section” (e.g. fishing area)
myi? tə 'sin	“river one distant arc” (a path to the sea) ⁵
myi? tə θwe	“river one connection” (e.g. connecting two villages)
myi? tə 'pa	“river one sacred object (e.g. in mythology)

Often quoted, his example shows that the choice of the classifier depends upon the universe of discourse, and it does not reflect directly a classification of physical reality, but only one for linguistic purposes.

Becker (1975) puts forward a double semantic organization of the Burmese CLF system⁶, postulating that a self-other continuum underlies the whole system, which is already organized, as most of the CLF systems, according to 3 main semantic criteria: [± human], [± animacy], [shape]⁷.

Becker (1975) is not the only analysis of Burmese classifier system. As far as I know, the first attempt is due to Haas (1951), and has been followed by two major articles by Burling and Hla Pe (both published in 1965), which provide an important list of classifiers for the former, and an analysis in different morpho-syntactic types of classifiers for the latter.

To complete this brief review of previous studies of Burmese classifiers, I should mention Goral (1978) on NUMERAL CLFS of Southeast Asia, Lehman (1979) and (1990) on a formal theory of nominal classifier systems — which both devote an important part of their articles to Burmese classifiers — and a short article by Becker (1986), talking about the difficulties of translating classifier structures.

All of the surveys mentioned deal with various aspects of the Burmese NUMERAL CLF system, such as morpho-syntactic patterns, semantic features, pragmatic uses and lists of the classifiers.

The next section is a summary of what is known on Burmese NUMERAL CLF. Other noun categorization devices not discussed so far will be treated in a later section.

2. Noun Categorization Devices in Burmese

2.1. Numeral classifiers

In many languages of Southeast Asia, a number is never used without being accompanied by a special class of morphemes, known as NUMERAL CLFS. According to Hla Pe (1965: 167-68), the use of this kind of morpheme is attested in Burmese from the earliest records of the language, i.e. 12th-13th centuries. At that

⁵ Becker's translation of the classifier /'sin/ seems odd to my informant. Given that /'sin/ is also the classifier for transportation, my informant prefers to view the river as a way of transportation rather than an arc.

⁶ Becker (1975: 118): *The structure underlying classification starts with the self at the center, divides the self into head and body, and then ranges objects at four distances from the self, associating them either with the head (metaphorically top, round) or with the body (metaphorically bottom, straight).*

⁷ See Craig (1986: 5) and also Bisang (1999: 9-10).

time these morphemes are not systematically used, but they start to become more consistent later.

2.1.1. Structure

Languages of East and Southeast Asia fall into two large groups according to the structure of noun phrases involving classifiers. Jones (1970) noticed that word order within the NP follows an areal pattern. In the North, represented by Chinese, Vietnamese, as well as Hmong (Bisang, 1999: 118), the head noun follows the numeral and the classifier ([NUM-CLF]- N). Whereas in the South, represented by Thai and Khmer, the head noun precedes the numeral-classifier group (N-[NUM-CLF]). Unsurprisingly, the Burmese NUMERAL CLF construction belongs to the second group.

- (1) [N NUM - CLF]
 ခွေး နှစ် ကောင် ။
 /ˈkhwe ɲi? KɔN/
 dog two CLF: animal
 ‘two dogs’

2.1.2. Nature

From a semantic point of view, Burmese NUMERAL CLFs should be divided into two subcategories: classifiers and quantifiers — also called sortal and mensural classifiers⁸. The distinction between sortal and mensural classifiers is based on the fact that reality can be quantified by counting or measuring objects.

Therefore, SORTAL classifiers are used to specify units (and not measures of quantity) in terms of which the referent of the head noun can be counted. They categorize referents in terms of their inherent characteristics, such as animacy (example 1), humanness, shape (2), social status (3, 4) or function (5)⁹.

- (2a) ဖျာ တစ် ချပ် ။
 /phyə tə Cha?/
 mat one- CLF: flat&thin
 ‘one mat [spread out]’

⁸ On the distinction between sortal and mensural classifiers, see Craig (1992: 279) and Aikhenvald (2000: 115-18)

⁹ Denny's classification is based on three types of human interactions: physical interaction, functional interaction and social interaction (1976: 125). Allan (1977) gives a list of semantic criteria used in many classifiers systems. According to Craig (1986: 5) and Bisang (1999: 9-10), humanness, animacy and shape will be primary among the semantic features used for classification while use and consistency will be secondary criteria.

- (2b) ဖျာ တစ် လိပ် ။
 /phya tə lei?/
 mat one- CLF:cylindrical
 ‘one mat [rolled]’
- (3) ဆရာ တပည့် နှစ် ဦး ။
 /shəya təpɛ’ ɲi? `ʔu/
 teacher student two- CLF:respected.people
 ‘one teacher and one student (2 persons)’
- (4) တပည့် နှစ် ယောက် ။
 /təpɛ’ ɲi? yɑ?/
 student two - CLF:people.of.all.kind
 ‘two students’
- (5) ဂေါ်ပြဲ သုံး လက် ။
 /`go-`pya `θuN lɛ?/
 shovel three CLF: tool¹⁰
 ‘three shovels’

MENSURAL classifiers (or quantifiers), on the other hand, are used to group objects in a unit of measure that can be understood as being countable. For instance, they occur in structures of measuring mass nouns or non-discrete physical entities (6, 7), but also in arrangement of units of countable nouns (8)¹¹.

- (6) ရွှေ တစ် တုံး ။
 /ʃwe tə `ToN/
 gold one- CLF:piece
 ‘one nugget of gold’
- (7) ရိုင် နှစ် တက် ။
 /ʌChiN ɲi? tɛ?/
 ginger two- CLF:shoot/growth
 ‘two small pieces of ginger’
- (8) နွား တစ် အုပ် ။
 /nwa tə qoq/
 cow/zebu one- CLF:group
 ‘one herd of cows’

A particular structure is found for measures of time¹² as exemplified in (9).

¹⁰ The morpheme used here to classify tools means *hand* as an independent noun. It brings to the fore the functionality of the tool.

¹¹ On subtypes of mensural classifiers, see Bisang (1993: 9-11) and Hla Pe (1965: 176-80).

¹² Goral (1978: 33) noticed that time nouns had also special behavior in Thai: *Time nouns were also special cases of abstract nouns, and though they occur as CL's they do not classify themselves.*

The head noun is omitted, maybe because it is semantically redundant.

- (9) လေး ဝတ် အတွင်း ပြန် လာ မယ် ။
 / 'le paʔ ʔə-`TwɪN pyaN la Mɛ/
 Ø-four- CLF:week within be.back come MODIR¹³
 'I will come back within four weeks.'

Another particular structure, from a formal point of view, is found in both sortal and mensural CLF construction: the use of repeaters.

A repeater is *the specific object itself (or part of it) used as a numerative* (Hla Pe 1965:166)¹⁴. Repeaters are often used for otherwise 'non-classifiable' items. Notice also that if a classifier construction contains a compound noun, only the main noun is repeated as a classifier. This is the "semi-repeater" construction, e.g. (11).

- (10) အိမ် တစ် အိမ် ။
 /ʔɛiN tə ʔɛiN/
 house one- CLF:house
 'a house'
- (11) စာမေးပွဲ တစ် ပွဲ ။
 /sa-me:-pwe: tə pwe:/
 to question-party one CLF:party
 'an exam'

There does not seem to be in this language any clear grammatical (or morpho-syntactic) criterion to distinguish among classifiers or to corroborate a division into two semantic subcategories, the sortal and the mensural CLFs.

Hla Pe's attempts (1965) to find syntactic criteria to distinguish between classifiers (sortal), quantifiers and repeaters is not entirely successful¹⁵. Moreover, his three subcategories partly overlap, some morphemes being listed under both (sortal) classifiers and repeaters, or both (sortal) classifiers and quantifiers¹⁶.

¹³ Conventions used: GEN = genitive marker, MODIR = Irrealis, MV = Verbal particle, NEG = negation, NOM.Realís = nominalizer conveying realis modality, OBJ = object marker, POL= politeness, PLUR= plural, PTCL= (syntactic) particle

¹⁴ However, a distinction should be made between repeaters that are only used with one noun, called *unique CLF* by Grinevald (forthcoming) — like /ʔɛiN/ 'house' in Burmese —, and those that classify themselves but also other nouns (or compounds), like /leʔ/ or /pwe/.

¹⁵ Hla Pe (1965: 166) asserts that classifiers are not independent or cannot occur as *determinata*, i.e. as head of compound noun or main syllable of a disyllabic noun. However, we found counter-examples: the classifiers /leʔ/ for tools, /piN/ for plants and /KəN/ for animals can also occur as independent nouns, and table (12) shows classifiers occurring as main syllable of a compound.

¹⁶ The classifier for tools /leʔ/ and the quantifier for groups /su/ are also listed as repeaters. Indeed /leʔ/ may classify itself as in /leʔ tə leʔ/ *one hand*, while /su/ occurs as a semi-repeater in

Table (12): Autonomy of classifiers vs. quantifiers

(Sortal) CLF			Head noun of a compound		
clf for flat object	/chaʔ/	ချပ်	/ə-chaʔ/	အချပ်	a flat (thing)
clf for vehicles	/ˈsiN/	စင်း	/ʔə-ˈsiN/	အစင်း	a vehicle

Quantifier			Head noun of a compound		
clf for group	/su/	စု	/ʔə-su/	အစု	a group
			/mi`θa-su/	မိသားစု	a family
clf for herd/flock	/ʔoʔ/	အုပ်	/ʔə-ʔoʔ/	အအုပ်	a tightly knit group
			/sa-ʔoʔ/	စာအုပ်	a book
clf for piece of X	/ˈtoN/	တုံး	/ʔe-ˈtoN/	အတုံး	a piece

2.1.3. Function

Enumerative expressions and indefinite expressions are found among the syntactic structures in which classifiers appear in Burmese. In both types of structures, the use of classifiers involves the same functions: classification (or categorization) and individualization¹⁷.

(a) *Enumeration*

As in other classifier languages of Southeast Asia, nouns in Burmese express a mere “concept” of an object and can be viewed as denoting substance rather than body. But enumeration presupposes the isolation of natural units of the same kind. Therefore, to be used in enumeration, a noun needs to be transformed into a unit, and this individuation of units is dependent on inherent features (categorization).

Both mensural and sortal CLF in Burmese display the two functions (i.e. classification and individualization) with a slight difference: quantifiers create the unit to be counted — this is obvious with mass nouns as in (13) — whereas sortal classifiers actualize the semantic boundaries which already belong to the concept of a given noun (Bisang 1999: 3) e.g. (2).

- (13) လက်ဖက်ရည် နှစ် ခွက် ။
 /lɛʔphɛʔ-yɛ ɲiʔ Khwɛʔ/
 drinking tea two - CLF: hollow container
 ‘two cups of tea’

compound such as *family* or a *group of X* (table 12).

¹⁷ Classification (or categorization) precedes individualization according to Croft (1994: 161) and Bisang (1999: 3).

(b) Indefinite expressions

Indefinite expressions also require classifiers. The use of the numeral *one* followed by a reduplicated classifier expresses the idea of someone or any one, something or any thing as shown in examples (14). In the same way, negative indefinite expressions require the numeral *one* plus the appropriate classifier followed by the particle /mə/ as in (15). As in Lahu (Matisoff 1973: 88-93) a closely related Tibeto-Burman language, classifiers may also be used in expressing the indefinite adjective *every/each* (16) — although this is not the only possibility in Burmese¹⁸.

- (14a) တစ် ယောက်ယောက် လာ မယ် ။
/tə yəʔ yəʔ la Mɛ/
one (CLF: hum)2 come MODIR
'Someone will come.'
- (14b) (စာအုပ်) တစ် အုပ်အုပ် ယူ ပါ ။
/(sa-ʔoʔ) tə ʔoʔ-ʔoʔ yu Pa/
(letter-tied=book) one CLF- CLF (group) take POL
'Take one of the (books) [tied object].'
- (15) တောင်း တာ တစ် ခု မှ မ ပေး ဘူး ။¹⁹
/təN Ta tə Khu' mə' mə `pe `Phu/
ask NOM.Realis one (CLF: general) PTCL NEG give NEG
'He gives nothing [even not a thing] of what it is asked.'
- (16a) ကလေးတွေ ကို တစ် ယောက် ပြီး တစ် ယောက် ပေး ပါ ။
/kə'le-Twe Ko tə yoʔ `pyi tə yoʔ `pe Pa/
child-Plur OBJ one (CLF:hum) after one (CLF:hum) give POL
Give (some) to each/every child (one after the other).
- (16b) (ကလေး) တစ် ယောက် စီ စာအုပ် တစ် အုပ် ပေး ပါ ။
/(kə'le) tə yoʔ si `sa-ʔoʔ tə ʔoʔ `pe Pa/
(child) one (CLF: hum) PTCL book one (CLF: group) give POL
'Give a book to every/each (child) (one by one).'

¹⁸ Unlike other languages of Southeast Asia (Thai, Lahu), Burmese does not use reduplicated classifiers for the indefinite plural; to express the idea *roughly, approximately X items*, Burmese adds the verb /ləʔ/ *be sufficient* to the numerative construction.

¹⁹ Example from Bernot, et al. (2001: 109).

(c) *Anaphoric use or referentialization function*

In Burmese, the head noun can be omitted in CLF constructions, if it is already known from the discourse situation. In that case, which is common in numeral classifier languages²⁰, the classifier refers in anaphoric way to the deleted noun. (See examples (17) and (18)).

- (17) ဒီ ကား ရန်ကုန် ကို မ သွား ဘူး ၊
 /di `Ka yaNgon Ko mə `θwa `Phu/
 this car Rangoon OBJ NEG go NEG
 နောက် တစ်စီး စီး ပါ ။
 /nəʔ tə-`si `si Pa/
 next one-CLF:vehicle travel POL
 ‘This car does not go to Yangon. Take the following (one).’²¹
- (18a) သရက်သီး ငါး လုံး ယူ ခဲ့ ပါ ။
 /θəyəʔ`θi `ŋa `luN yu - Kheʔ Pa/
 mango five CLF: 3D take - VM (mvt) POL
 ‘Bring me back five mangos.’
- (18b) သရက်သီး ဝယ် ရင် ငါး လုံး ယူခဲ့ ပါ ။
 /θəyəʔ`θi wɛ yiN `ŋa `luN yu-Kheʔ Pa/
 mango buy if Ø - five CLF: 3D (round) take-VM (mvt) POL
 ‘If you buy mangos, bring me back five [round objects].’
- (18c) သရက်သီး ဝယ်ပေး ပါ ။ – ဘယ်နှစ် လုံး လိုချင် သ လဲ ။
 /θəyəʔ`θi wɛ-pe Pa bɛʔniʔ `luN lo-ChiN θə lɛ/
 mango take-AUX_{benef} POL how.much- CLF:3D want MODR-QU
 ‘Buy me some mangos.’ – ‘How many [round objects] do you want?’

2.2. Class terms and NOUN-CLASSIFIER system

In the previous section, I summarized the characteristics of the obvious and large NUMERAL CLF system in Burmese, which provides a conceptual and pragmatic classification. Burmese seems to have other noun categorization devices, rather based on taxonomic classification: the noun classifier (henceforth NOUN CLF) system and/or CLASS TERMS.

²⁰ See Croft (1994: 163): *Another significant function of numeral classifiers is anaphoric, and as such they also have a reference tracking function.*

²¹ Matisoff (1973) reports on the use of *one* + CLF in Lahu to express ‘the following X’. In Burmese, we consider this use of the CLF with the numeral ‘one’ simply as an anaphoric situation.

2.2.1 Definition

According to Grinevald (1999: 112-113) followed by Aikhenvald (2000: 82-84), NOUN CLFS have the following properties:

- They characterize the noun and co-occur with it in a NP (noun phrase). But every noun of the language does not necessarily take a NOUN CLF.
- NOUN CLFS correlate with inherent semantic features of noun, such as “animal”, “human”, “plant”²².
- There is often a generic-specific relationship between a NOUN CLF and a noun.
(DeLancey 1986: 438).

However, a slight difference exists between their two definitions. Unlike Aikhenvald whom uses NOUN CLFS as a uniting name, Grinevald distinguishes CLASS TERMS from NOUN CLFS. They are both related to lexicon, but the use of CLASS TERMS seems to be restricted to lexical composition, whereas NOUN CLFS may have grammatical functions such as being determinants or pronouns as in Jacaltepec (Grinevald 1999: 107 and 2000: 64-65).

Productivity, degree of lexicalization, grammatical function and semantic field are the relevant criteria in deciding whether a language has CLASS TERMS or a NOUN CLF device. However, the distinction seems hard to establish.

Therefore, I will use CLASS TERMS unless there is evidence of a clear grammatical system using the categorizing morphemes²³.

2.2.2 CLASS TERMS in Burmese

Burmese, like Garo and other Tibeto-Burman languages spoken in Northeast India²⁴, seems to have “*noun compound constructed from a categorizing initial portion [i.e. a CLASS TERM] to which is added one or more syllables that indicate the specific member of the category*” (Burling 1984: 14). The compound nouns, in which these categorizing first syllables occur, belong to particular semantic categories such as fish, birds and mushrooms.

A quick scan of Bernot’s dictionary (1979-92) revealed around 70 fish nouns starting with the CLASS TERM /ŋə/ — where we can recognize diachronically the generic name for fish /ŋə/. Concerning bird names, Bernot dictionary provides less than 30 entries having the generic noun /ŋeʔ/²⁵ as a first syllable. The generic

²² See Bisang (1993: 16-17) on the universality of splitting up the world into categories as suggested by Berlin, and the use of these particular categories.

²³ In my opinion, NOUN CLF system should be viewed as a particular kind of CLASS TERMS, a sub-category on the grammatical side rather than the lexical side of this type of classifiers, or on the path of grammaticalization.

²⁴ See Jacquesson (1998). However, unlike the Tani dialects described in this article in which almost all the nouns are disyllabic, the majority of Burmese nouns are monosyllabic.

²⁵ Notice however that there exist fish names and bird names that are not compounds with /ŋə/ and /ŋeʔ/ respectively. But, without a zoological encyclopedia, I can only give a sketch of the situation that might be confirmed (or not) by deeper studies and serious statistics.

term /mwe/ is used as the first part of the compounds in 15 names of snake, and half a dozen of turtle names start with the morpheme /lei?/, which is also the term for an unspecified turtle. (See table 20)

Table 20: Compounds with a [generic-specific] construction

ဝါး မန် /hə - `maN/	shark	ဝါး သန်း နု /hə - `θaN nu' /	catfish
ဝါး ကင် ဝါး /hə - `KiN `Pa/	mackerel	ဝါး တ ခွေ /hə - tə she/	octopus

Table 20: Compounds with a [generic-specific] construction (cont.)

ငှက် ကုလား /h̥eʔ - kə `la/	Asian stork ²⁶	ငှက် ခါး /h̥eʔ - `kha/	jay (bird)
ငှက် သူ ခိုး /h̥eʔ - θə `kho/	a kind of blackbird ²⁷	(ငှက်) စင် ရော /(h̥eʔ) - SiN ya/	seagull
ပိုး ဖြူ ဟု /m̥o - phyu hu' /	a kind of mushroom	ပိုး ဦး နက် /m̥o - `ʔu neʔ/	black-head mushroom
မြွေ ဗွား /mwe - `pwe/	viper	မြွေ လက်ပပ် /mwe - ləʔ paʔ/	whip-snake e
လိပ် ပြောက် /leiʔ - pyaʔ/	turtle (fresh water)	လိပ် ကြက်တူ ရွှေ /leiʔ - Cəʔ Tu `ywe/	sea turtle

Beside these generic-specific compounds, some compound nouns that are not synchronically analyzable can also be found in Burmese. Most of the time, their first syllable has lost tone and vowel quality [/Sə`Pwε/ ~ *table*²⁸], as well as its meaning. (In the following examples, the first syllable does not have a recognizable meaning: /Sə`Ka/ ~ *speech*, /pə`ya/ ~ *pagoda*, /Tə`she/ ~ *yeast*).

Like most Southeast Asian languages, Burmese is analytic and monosyllabic or sesquisyllabic (“syllable-and-a-half”) in structure (Matisoff 1991: 386). Therefore the occurrence of a reduced initial syllable in a compound seems to be a good clue to its antiquity.

Indeed, Thurgood (1981) mentions a pre-head classifier system regarding many of the reduced initial syllables in Burmese compounds²⁹, whereas Maspero

²⁶ *Xenorhynchus asiaticus*.

²⁷ *Monticola solitaria affinas*.

²⁸ In /Sə`Pwε/ ~ *table*, the first syllable pronounced toneless and with a schwa, can still be analyzed as /'sa/ ‘to eat.’

²⁹ Thurgood (1981: 12): *Many of the reduced initial syllables in Burmese compounds appear to*

(1947) found traces of prefixes in Burmese vocabulary³⁰.

2.2.3 Search for an old noun categorization device

Starting the search for remnants of an old noun categorization device, here is the procedure followed.

(a) *Procedure*

- The first step was to make a list of all Burmese disyllabic words having the same initial syllable, i.e. supposedly traces of prefixes. The second step was to group the words by semantic criteria, given that the purpose was to figure out if a general meaning could be found for these prefixes.
- The first syllable to be analyzed was the one written ဝ /ka'/ (with the inherent vowel /a/). All disyllabic (polysyllabic) nouns starting with this letter were listed, paying particular attention to the vowel pronunciation, to keep only vowels realized without any indication of tone or vowel quality, given that this is a good clue for old compounds words.

Identified loan words³¹ were also discarded.

- From the list obtained, nouns were then grouped by semantic domain. As suggested by Thurgood's remark (see footnote 35), I check first if the nouns refer to plants.
- The procedure was repeated with two other syllables ဝ /sa'/ and ဝ /ta'/.

(b) *Results*

- Working with the Burmese-English dictionary, for the letter ဝ /ka'/, 63 nouns with the first reduced syllable /Kə/ were found, excluding the loans from Môn, Pali, English and other languages.

What emerges from this list is that 16 of the 63 nouns are plant names, 7 animal names and 4 tribe names.

be the remnants of a pre-head classifier system, reflecting a centuries-old interest in the medicinal and culinary properties of plants

³⁰ Prefixation was a derivational process used in PTB languages (1947: 155-56), also attested in Burmese according to Maspero (1947: 155-56, 167-68). He noticed that until now, the prefix /qa/ is still used in Burmese word formation, whereas the other prefixes survive only through traces, and are not anymore productive: *D'autre part, il [le birman] forme aujourd'hui encore des noms verbaux par le préfixe a-. [...] Mais aucune autre formation par préfixe n'est restée vivante, et n'a laissé plus que des survivances dans le vocabulaire* (1947: 168).

³¹ I worked with the *English-Burmese Dictionary* (1998), which indicates the origin of loans words.

- For the syllable ၀ /saʔ/, 28 nouns with the reduced first syllable /Sə/ were found, excluding loans. Of these 28 nouns, only one was a plant name, and two were animal names.
- For the syllable ၀ /taʔ/, 33 nouns with the reduced first syllable /Tə/ were found, excluding loans. In this list, there were 3 plant names.

(b) *Conclusion of this sketch*

The number of nouns having the reduced syllable ၀ /kaʔ/ is larger than that of nouns having ၀ /saʔ/ or ၀ /taʔ/. One quarter of this list represents plant names³², whereas they are almost non-existent in the two other lists.

Nouns starting with	Total	Plant nouns	Animal nouns
prefix ၀ /kaʔ/	63	16	7
prefix ၀ /saʔ/	28	1	2
prefix ၀ /taʔ/	33	3	-

Regarding the results, ၀ /kaʔ/ seems to be a good candidate for being a trace of an old pre-head “classifying” system, as suggested by Thurgood. This result should be compared to the categorizing prefixes found in other languages of the area, even those that are not genetically related (See Bilmes 1998, Thurgood 1988). However, even with this encouraging result, we are far from giving some meaning to the prefix ၀ /kaʔ/.

2.2.4 - Conclusion about this CLASS TERM system or generic-specific compounds

This first attempt to answer the question, “*Are there CLASS TERMS in Burmese?*” leads us to the following temporary conclusion; the presence of recognizable categorizing morphemes in the Burmese lexicon had led us to postulate a second classifying process. We found good clues to the existence of an old noun categorization device in Burmese using unanalyzable prefixes. However, none of these structures — the old structure with the reduced syllable or the more recent structure with analyzable (and semantically motivated) first syllable — is productive.

2.3. **New categorization structures**

2.3.1. Categorizing the vegetal domain

According to typological studies (Grinevald 2000: 59), vegetal domain is often the most productive field for CLASS TERMS and NOUN CLF device. Indeed in Burmese, naming a plant, a fruit or a flower requires the use of a generic term. But

³² In Bernot's dictionary (1979-92), which is more complete but does not always give the origin of the words, I picked up 53 plant nouns with a first reduced syllable /Kə/.

unlike the examples given above, the categorizing morpheme in these nouns FOLLOWS the species noun and the compound noun shows the following morpho-syntactic pattern:

- (21) [Species NOUN – CATEGORIZING MORPHEME] or
[NOUN - CLASS TERM]

Examples are given in tables 22, 23 and 24.

Table 22: Class terms in the vegetal domain (fruits and plants)

Fruit		Plant	
1a	ငှက်ပျော သီး ၵဲး? pya `θi /ၵဲး? pya/-FRUIT	1b	ငှက်ပျော ပင် ၵဲး? pya PiN /ၵဲး? pya/-PLANT
2a	သရက် သီး θa ye? `θi /θa ye? `θi/-FRUIT	2b	သရက် ပင် θa ye? PiN /θa ye?/-PLANT
3a	သစ်တော သီး θi? Tə `θi /θi? Tə/-FRUIT	3b	သစ်တော ပင် θi? Tə PiN /θi? Tə/-PLANT
4a	လိမ္မော် သီး leiN mə `θi /leiN mə/-FRUIT	4b	လိမ္မော် ပင် leiN mə PiN /leiN mə/-PLANT
5a	သံလွင် သီး θaN lwiN `θi /θaN lwiN/-FRUIT	5b	သံလွင် ပင် θaN lwiN PiN /θaN lwiN/-PLANT

Moreover, what is noticeable is that it seems to be a productive process. Indeed, the Burmese nouns for daisy and dahlia (table 24) — which are imported flowers — follow the rule and occur with the CLASS TERM at the end of the compound.

Table 23: Class terms in the vegetal domain (flowers and plants)

FLOWER		PLANT	
1a	စံပယ် ပန်း zəPe `PaN /zəPe/-FLOWER	1b	စံပယ် ပင် zəPe PiN /zəPe/-PLANT
2a	စက္ကူ ပန်း se?Ku `Pan /se?Ku/-FLOWER	2b	စက္ကူ ပင် se?Ku PiN /se?Ku/-PL.
3a	(တရုတ်) စံကား ပန်း (təyo?) zə`Ka `PaN /təyo? zə`Ka/-FL.	3b	(တရုတ်) စံကား ပင် (təyo?) zə`Ka `PiN /təyo? zə`Ka/-PL.
4a	နှင်းခီ ပန်း `niN shi `PaN /`niN shi/-FLOWER	4b	နှင်းခီ ပင် `niN shi PiN /`naN shi/-PL.

³³ Also known as the Pagoda tree (apocynacée, *Plumeria acutifolia*).

Table 23: Class terms in the vegetal domain (flowers and plants) (cont.)

5a	ဒေလီယာ ပန်း deliya `PaN	dahlia /deliya/-FLOWER	5b	ဒေလီယာ ပင် deliya PiN	dahlia /deliya/-PL.
6a	ဇော်မွာ့ ပန်း zo`mwa `PaN	carnation /zo`mwa/-FLOWER	6b	ဇော်မွာ့ ပင် zo`mwa PiN	carnation /zo`mwa/-PL.
7a	ဒေစီ ပန်း desi `PaN	daisy /desi/-FLOWER	7b	ဒေစီ ပင် desi PiN	daisy /desi/-PLANT

The exceptions to this quasi-systematic categorizing process in the vegetal domain are generally phonological (or periphrastic) loans such as the one for *litchi*³⁴ (table 24). However, notice that unlike the name for the fruit, the noun for *litchi tree* follows the pattern: the CLASS TERM appears after the species noun.

Table 24: Irregularities in Class terms (Flora)

	Flower or Fruit			Plant	
1a	စတော်ဘယ်ရီ (သီး) s(ə)təbəri (ʰθi)	strawberry /s(ə)təbəri/-(FR.)	1b	စတော်ဘယ်ရီ ပင် /s(ə)təbəri PiN/	strawberry plant /s(ə)təbəri/-PL.
2a	လီချီ li `Chi-	litchi /li `Chi/-Ø	2b	လီချီ ပင် li `Chi PiN	litchi tree /li `Chi/-PLANT
3a	ပန် သီး `paN `θi	apple “flower”-FRUIT	3b	ပန် သီး ပင် `paN `θi PiN	apple tree “flower fruit”-PL.

Coming back to the animal realm, Burmese surprisingly also provides bird and fish nouns following this second morpho-syntactic pattern. For instance, as exemplified in table (25), the CLASS TERM for birds /*h̥ɛʔ*/ occurs at the end of numerous bird nouns. Notice that it is the same morpheme that occurs as a categorizing prefix in table (13). In the same way, the generic term /*ŋa*/ occurs as a CLASS TERM for fish at the end of the compound.

Table 25: Class terms in the animal realm

BIRD		FISH	
1	သစ် တောက် ငှက် θiʔ təʔ ၵɛʔ	5	ရွှေ ငါး ʃwe `ŋa
	woodpecker /θiʔ təʔ/- BIRD		gold fish /ʃwe /- FISH
2	ခင် ပုပ် ငှက် khiN poʔ ၵɛʔ	6	ယင် ပေါင် စာ ငါး yiN pəN sa `ŋa
	brown hawk-owl /khiN poʔ/- BIRD		white bellied opsarion /yiN pəN sa/- FISH
3	သိမ် ငှက် `θeiN ၵɛʔ		
	falcon /ʰθeiN /- BIRD		

³⁴ The word for *litchi* is borrowed from Cambodian, according to the *Burmese-English Dictionary* (1998). (Which is in turn a loan from Sinitic.ed.)

I also notice a few mushrooms names, insect names and one turtle name, in which the generic term appears at the end of the compound. (See table 26)

Table 26: Class terms in other compounds

Animal		Mushroom	
1	ယင် ကောင် yiN KəN /yiN /-ANIMAL	4	မြက် ကြား ဝှို myɛʔ`Ca m̥o /myɛʔ`Ca/-MUSHROOM
2	နဲ ကောင် nəN KəN /nəN/-ANIMAL	5	တောင် ဝှို təN po`m̥o /təN po`/- MUSHROOM
3	သင် လိပ် `θiN leiʔ /`θiN/-TURTLE		

But regarding some nouns ending with the generic term, such as the fish noun /ʃwe `ŋa/ (table 25-5), which is analyzable as ‘gold + fish’, we may wonder what the relationship is between this classifying process and determination, given that formally they show similar structures.

- (27) a. ရွှေ ငါး ။ b. မိုးမိုး (ရဲ) ငါး ။ c. ရွှေ ရှပ်တု ။
 /ʃwe `ŋa/ /`mo `mo (yɛ') `ŋa/ /ʃwe yoʔ-tu'/
 gold fish Name (GEN) fish gold statue
 ‘a gold fish’ ‘Mo Mo’s fish’ ‘a golden statue’

2.3.2 Relationship between CLASS TERMS and NUMERAL CLFS

We may also report on the formal relationship that exists between CLASS TERMS and NUMERAL CLFS.

CLASS TERMS have been sometimes considered as classifiers. Confusion in the terminology is partly due to the relationship that exists between CLASS TERMS and NUMERAL CLFS. Related through their semantic features and/or their forms, they may also co-occur in a language (Aikhenvald 2000: 187).

DeLancey, in his history of Tai classifier system, says that *lexically the two categories overlap to a considerable degree* (1986: 442), and suggests that CLASS TERMS [*class nouns*] provide a source for NUMERAL CLFS (1986: 445-46). Also for Bisang (1999: 41), the process of grammaticalization involved in Southeast Asian Languages (Hmong, Vietnamese, Thai) starts from a categorical system based on taxonomy, i.e. CLASS TERMS or a NOUN CLF system.

As in Thai (DeLancey 1986: 438), some CLASS TERMS in Burmese also function as numeral CLF. Table 28 shows the CLASS TERMS for plants, fruits, animals and the numeral CLF generally used for these items. Notice however that the animal CLASS TERM does not occur in many compounds. Moreover, the CLASS TERM for trees, which is used also as the numeral CLF, can be omitted in enumeration under certain circumstances, for example, when a person is in a nursery shop, and enumerates how many of each kind of plants (s)he will take.

Table 28: CLASS TERMS and NUMERAL CLFS

Item	Class term	Numeral CLF	Item	Class term	Numeral CLF
plant, tree	-ပင် /piN/	ပင် /piN/	animal(worm, fly...)	-(ကောင်) /KəN/	ကောင် /KəN/
flower	-ပန်း /ˈpaN/	ပွင့် /pwiNˈ/	fish	-ငါး /ŋa/	ကောင် /KəN/
fruit	-သီး /ˈθi/	လုံး /ˈloN/	bird	-ငှက် /ŋɛʔ/	ကောင် /KəN/

3 - Summary and Conclusion

What I intend to show in this article is that Burmese has (at least) two noun categorization devices: an already known and described NUMERAL CLF system and a CLASS TERMS system.

It was easy to show the NUMERAL CLF system used in the language. However, it was more complicated to report on CLASS TERMS (or a NOUN CLF system?).

What emerges from this study is that Burmese has superposed strata of noun classification systems. The examination of the Burmese lexicon reveals that at first this language used classifying prefixes, as did other languages of the family (and of the area?).

However nowadays, three types of classifying prefixes are found in Burmese polysyllabic nouns: reduced and non-motivated syllables phonologically reduced but diachronically analyzable syllables or plain morphemes. These different layers of noun categorization lead us to assume that this classifying structure is old (phonetic erosion of the first syllable), but that it has endured through time.

Subsequently — and we assume more recently — another classifying structure has appeared, similar to the syntactic determination construction: categorizing morphemes (superordinate or generic nouns) are placed at the end of the compound, as are head nouns at the end of the NP³⁵. This second categorizing structure conforms better to the canonic word order of the language, i.e. [MODIFIER - HEAD], and may enter into competition with the old structure, as shown by the few doublets of table (29). It may have the potential to develop into a noun CLF system, i.e. a grammatical system. However, while this process is productive, it seems restricted to the usual fields in which languages of the world categorize, i.e. the plant and animal realms.

³⁵ Determination is marked by the position of the morphemes in Burmese. The occurrence of a relator (genitive) morpheme is optional, and depends on the animacy of the possessor.

Table 29: Two names per species involving different structures

1 ဝေလ ငါး /wela'-`ɲa/	whale /wela'/-fish	~ ငါး ဝန် /ɲə-waN/	whale ³⁶ fish-/waN/
2 လိင် ကောင် ပို ငှက် /liN KəN po-ɲɛʔ/	barn howl /liN KəN po/-bird	~ ငှက် ဆိုး /ɲɛʔ-`sho/	barn howl bird-/`sho/
3 ဝန် ပို ငှက် /waN po-ɲɛʔ/	pelican /waN po/-bird	~ ငှက် ကြီးဝန်ပို /ɲɛʔ-Ci waN po/	pelican bird-/Ci waN po/

From all this, it follows that Burmese does have different layers of noun categorization using Class Terms as prefixes or suffixes but not yet a grammatical Noun clf device.

The noun categorization devices discussed here are characterized by a certain degree of semantic motivation, a clear lexical origin, and a particular morpho-syntactic behavior. However, regarding the continuum of nominal categorization proposed by Grinevald, Burmese NUMERAL CLFS can be viewed as a grammatical system, whereas the CLASS TERMS are still located on the lexical side of the continuum.

Noun categorization in Burmese warrants more work, including listing carefully all the compounds with reduced syllables, and tackling the question of the relationship between categorization and determination.

References:

- Aikhenvald, A. Y. and Diana Green. 1998. "Palikur and the typology of classifiers". *Anthropological Linguistics* 40: 429-480.
- _____. 2000. *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith. 1977. Classifiers. *Language* 53: 284-310.
- Adams, Karen L. and N. F. Conklin. 1973. "Towards a theory of natural classification". *Papers from the annual meeting of the Chicago Linguistic Society* 9: 1-10.
- Becker, A. J. 1975. "A linguistic image of nature: the Burmese numerative classifier system". *Linguistics* 165: 109-121.
- _____. 1986. "The figure of a classifier makes: describing a particular Burmese classifiers". in *Noun classes and categorization: proceedings of a symposium on categorization and noun classification*, Ed. By C. Craig, Amsterdam: J.

³⁶ Concerning the "whale" nouns, one of my informants told me that the [CLASS TERM + noun] structure "sounds" older.

- Benjamins, 327-343.
- Bernot, Denise, M.H. Cardinaud, Yin Yin Myint. 2001. *Grammaire du Birman*. vol. 2, Paris: L'asiathèque.
- _____. 1979-92. *Dictionnaire birman-français* (15 vol.). Paris, SELAF, 15 vol.
- Bilmes, Leela. 1998. "The /ka-/ and /kra-/ prefixes in Thai". *LTBA* 21.2.
- Bisang, Walter. 1993. "Classifiers, quantifiers & class noun in Hmong". *Studies in Language* 17: 1-51.
- _____. 1999. "Classifiers in East and Southeast Asian Languages: counting and beyond". *Changes in Numeral Systems*. Berlin: Mouton de Gruyter, 113-185.
- Burling, Robert. 1965. "How to choose a Burmese Numeral Classifier", *Context and Meaning in Cultural Anthropology*. ed. by Melford E. Spiro, New York: The Free Press, 243-264.
- _____. 1984. "Noun compounding in Garo". *LTBA* 8.1: 14-42.
- Craig, Colette. 1992. "Classifiers in a functional perspective". In *Layered Structure and reference in Functional perspective*, ed. by M. Fortescue et al. Amsterdam: J. Benjamins, 277-301.
- Croft, W. 1994. "Semantic universals in classifier systems". *Word* 45: 145-71.
- Denny, Peter. 1976. "What are noun classifiers good for?" *Papers from the 12th International Meeting*, Chicago linguistic Society, 122-132.
- DeLancey, Scott. 1986. "Towards a history of Thai classifiers system". in *Noun classes and categorization: proceedings of a symposium on categorization and noun classification*, Ed. By C. Craig, Amsterdam: J. Benjamins, 437-452.
- Dixon, K. M. W. 1986. "Noun classes and noun classification in typological perspective". in *Noun Classes and categorization: proceedings of a symposium on categorization and Noun classification*, Ed. By C. Craig Amsterdam: J. Benjamins, 105-12.
- Greenberg, Joseph H. 1972. "Numeral classifiers and substantival number: Problem in the genesis of a linguistic type". *Working papers In Language Universals* 9.
- Goral, Donald R. 1978. "Numerical Classifier Systems: A Southeast Asian Cross-linguistic Analysis". *LTBA* 4.1: 1-72.
- Grinevald, Colette. 1999. "Typologie des systèmes de classification nominale". In *Faits de Langue* 14, *La catégorisation dans les langues*: 101-122.
- _____. 2000. "A morphosyntactic typology of classifiers". In *Nominal classification*. ed. by G. Senft, Cambridge: Cambridge University Press, 50-92.
- _____. forthcoming. "Linguistics of classifiers". in *Encyclopedia of the Social and Behavioral Sciences*. Elsevier.
- Haas, Mary R. 1951. "The use of numeral classifiers in Burmese". in *Semitic and Oriental Studies*, edited by W. Popper, Berkeley & Los Angeles: Univ. of California Press, 191-200.
- Hla Pe. 1965. "A re-examination of Burmese Classifiers". *Lingua* 15: 163-185.
- Jacquesson, François. 1998. "L'évolution et la stratification du lexique. Contribution à une théorie de l'évolution linguistique". *BSLP* XCIII-1: 77-136.

- Jones, Robert B. 1970. "Classification in Southeast Asia". *Journal of the American Oriental Society* 90-1: 1- 13.
- Lehman, F. K. 1979. "Aspects of a formal theory of Noun classifiers". *Studies in Language* 3.2: 153-180.
- _____. 1990. "Outline of a formal syntax of numerical expressions, with especial reference to the phenomenon of numeral classifiers". *LTBA* 13.1: 89-120.
- Matisoff, James A. 1973. *The grammar of Lahu*. Linguistics Vol.75, Berkeley: University of California Press.
- _____. 1991, "Areal and universal dimensions of grammaticalization in Lahu". in *Approaches to grammaticalization VOL.II*, ed by E. C. Traugott et al, Amsterdam: J. Benjamins, 383-453.
- Maspero, H. 1947. "Notes sur la morphologie du tibéto-birman". *BSLP* 44: 155-185.
- Thurgood, Graham. 1981. "Notes on the origins of Burmese Creaky tone". Institute for the study of Languages and Cultures of Asia and Africa, *Monumenta Serindica* n°9.
- _____. 1988. "K- prefixes in Kam-Sui and Kadai: some notes". in *Languages and History in East Asia: Festschrift for Tatsuo Nishida*, ed by P.K. Eguchi, Yukio Fujimoto, Nobuyoshi Fukuhara, Masura Hashimoto,... Kyoto: Shokado, 229-235.
- Seiler, Hansjacob. 1983. *Possession and Operational Dimension of Language*. Tübingen: G. Narr.
- Myanmar-English Dictionary*. 1998. Department of the Myanmar language Commission, Ministry of Education, Yangon.

Alice Vittrant
53 rue des amandiers,
75020 Paris
France

alice.vittrant@vjf.cnrs.fr

Contact induced variation and syntactic change in the Tsat of Hainan

GRAHAM THURGOOD and FENGXIANG LI
California State University, Chico

Introduction

Tsat is an Austronesian language located on Hainan Island. The 1982 census lists 4131 Utsat people largely in the villages of Huihuī and Huixīn near Sānyá on Hainan Island (which has recently been designated as a province), 3849 of whom still speak Tsat. Virtually all the Tsat speakers also speak one or more Chinese dialects, typically Fukienese or Cantonese, the languages of business, and Mandarin, the language of school.

Genetically the closest language to Tsat is the Northern Roglai of Vietnam, a Chamic language (Austronesian) which it split off from first around 982, with a second migration probably around 1471. Despite the genetic closeness, Tsat is now radically different both phonologically and syntactically from N. Roglai. Phonologically, Northern Roglai is sesquisyllabic and atonal whereas Tsat is monosyllabic and fully tonal. Structurally, Northern Roglai is much, much more like the other Chamic languages of Vietnam which, in turn resemble the Mon-Khmer languages of the region, while Tsat, not surprisingly, is much like the Chinese dialects that surround it. Increasingly, all that remains of Tsat is the vocabulary, with the structure being Chinese, albeit with Tsat lexical items. Thus Tsat provides some exceptionally clear examples of contact-induced syntactic variation and change. Work has been done on genetic affiliations of Tsat (Benedict 1941), the history of the Chamic languages including Tsat (e.g. Thurgood 1999, 1996), and on the description of Tsat itself, Ouyang and Zheng (1983), Zheng (1986, 1997), with the later work by Zheng including numerous valuable observations on the influence of Chinese on Tsat, both identifying Chinese borrowings and commenting on Chinese structural influence.

Tsat contact

Changes in Tsat resulting from contact with neighboring languages of Hainan are quite obvious. Phonologically, it has gone from sesquisyllabic and registral to monosyllabic and tonal. Lexically, it contains four layers of borrowings reflecting contact patterns since the Tsat arrival in Hainan: a Hlai (= Li) strata, an early Chinese level reflecting early contact with speakers of Min

dialects such as Hainanese and various Cantonese dialects, a later layer of contact with the Mandarin spoken by the army and officials, and most recently the Mandarin of the schools. The intensity of the last layer of contact looks to have initiated rapid and through restructuring of the language.

Here we will restrict our examination to four constructions with extant variation, two involving word orders that do not correlate with VO order (Dryer 1992) and two involving word orders that do correlate: genitive constructions, demonstratives and head nouns, adjectives and head nouns, and comparative constructions. All show the structural influence of Chinese.

Genitive (“associative”) constructions

The genitive patterns have been divided into those with full noun phrases as the genitive and those involving pronouns, reflecting the differences in their historical paths of change.

Genitives with full noun phrases

In Northern Rglai full NPs are postposed. In Tsat, even in the most colloquial, non-Sinicized texts, full GenNPs already show Chinese influence: all full genitive NPs are preposed with the genitive construction marked by *sa*³³. In the case of locative NPs, the preposed NPs look to be developing into prepositions. Elsewhere, the genitive marker is a marker of pre-head modification.

- | | | | |
|-----|--|--|-----------------|
| | Northern Rglai: | Nh GENNP | |
| (1) | <i>ga?</i> <i>sa:k</i>
roof house
‘the roof of the house’ | | (Lee 1966:65) |
| | Tsat (colloquial): | GENNP <i>sa</i> ³³ Nh | |
| (2) | <i>ʔa¹¹ʔba¹¹</i>
father’s.older.brother
<i>bófū</i>
<i>bófū de fángzi</i>
‘father’s eldest brother’s house.’ | <i>sa</i> ³³ <i>sa:y</i> ³³
GEN house
<i>de fángzi</i> | (Zheng 1997:70) |
| (3) | <i>ʔia³³ʔbe²⁴</i> <i>sa</i> ³³ <i>ka:n</i> ³³
river GEN fish
<i>hělǐ</i> <i>de yú</i>
<i>hělǐ de yú</i>
‘the fish in the river’ | | (Zheng 1997:71) |

- (4) ...*piat*³³ *sa*³³ *za:ŋ*⁴²
 ...village GEN person
 ...*cūn* *de rén*
 ...*cūn de rén*
 ‘people of the village...’ (Zheng 1997:95)
- Tsat (Mandarinized with *ti*³³) GENNP *ti*³³ HeadNP
 -the same pattern with the genitive borrowed from Mandarin
- (5) *tan*³³ *k^hua*⁵⁵ *ti*³³ *si^lha:u*⁴², *la*³³ *piat*³³ *sa*³³ *za:ŋ*⁴²
 arrive daybreak GEN after under village GEN person
dào tiānliàng de shíhòu xià cūn de rén
dì'èrtiān tiānliàng shíhòu, cūnlǐ de nánrén hé nǚrén...
 ‘Early the next morning, the villagers...’ (Zheng 1997:4.1.4)
- (6) *sui*¹¹ *tso*⁴² *kai*⁴² *kak*²⁴ *k^ha:i*³³ *p^ha:ŋ*³³ *ti*³³ *sin*³³ *zit*²⁴
 along.with reform bloom GEN penetrate
suízhe gǎigé kāifàng de shēnrù
suízhe gǎigé kāifàng de shēnrù
 ‘As the Reform and Open-door policies continue...’ (Zheng 1997:3.3.3)
- Mandarin: Poss *de* NP
- (7) *duìzhǎng* *de* *érzi*
 captain GEN child
 ‘the captain’s son’ (Zheng 1997:71)
- (8) *lǎoshī* *de* *shū*
 teacher GEN book
 ‘the teacher’s books’ (Zheng 1997:71)

In the genitives, as with the other constructions, the more Mandarinized variants tend to co-occur with borrowed Mandarin grammatical markers (which in some cases seem to mark the construction), tend to have more Mandarin borrowings in the sentence (marked in this paper through the underlining of both the borrowed Tsat term and its corresponding Mandarin (in Pinyin)), and, if one examines the texts, occur in the more Mandarinized texts. For instance, the text on the origin of the Tsat less Mandarin influence than does the text describing the Japanese invasion of Hainan with its Mandarin influenced political content.

Table 1: Noun modification: Genitives (full NPs)

	simple modification	prehead <i>sa</i> ³³ pattern
N. Roglai	Nh GENNP	_____
Colloquial Tsat	_____	GENNP <i>sa</i> ³³ Nh
Mandarinized Tsat	_____	GENNP <i>ti</i> ³³ Nh
Mandarin	_____	GENNP <i>de</i> Nh

In the case of full noun phrases, the Northern Roglai post-head genitives have been totally replaced by Chinese-influenced prehead genitive constructions. Nonetheless, contact has resulted in variation as the colloquial pattern uses a Tsat genitive marker while the Mandarinized pattern uses a borrowed genitive marker. Both patterns show the structural influence of Chinese.

Genitives with pronouns

- Northern Roglai: Nh GENPr
- (9) *sa:k hã*
house you
'your house' (Lee 1966:65)
- Tsat (colloquial): Nh GENPr
- (10) *ɲa:n³³ kau³³ ki²⁴*
hand I painful
shõu wõ tòng
wõ de shõu tòng
'My hand hurts.' (Zheng 1997:97)
- (11) *ko²⁴bu²⁴ nau³³ sa²⁴*
head.hair she messy
tóufa tã luàn
tã de tóufa luàn
'Her hair is messy.' (Zheng 1997:92)
- Tsat (Chinese influenced, with *sa*³³)
- (12) *nau³³ sa³³ ko²⁴bu²⁴ sa²⁴*
she GEN head.hair messy
tã de tóufa luàn
'Her hair is messy.' (Zheng 1997:97)

Variation and syntactic change in Hainan Tsat

- (13) *ha*³³ *sa*³³ *ʔa*¹¹*sa:u*¹¹ *sa*³³ *ha*³³
you GEN o.bro.wife seek you
nĩ *de* *sǎo* *zhǎo* *nĩ*
nĩ de sǎo zhǎo nĩ
‘Elder brother’s wife seeks you.’ (Zheng 1997:87)
- Mandarin: GENPr Nh
(14) *wǒ* *fùqin* *shì* *tā* *bófū*
I father be he uncle
‘My father is his uncle.’ (Zheng 1997:77)
- GENPr *de* Nh
(15) *wǒ* *de* *shǒu* *tòng*
I GEN hand painful
‘My hand hurts.’ (Zheng 1997:97)

Table 2: Noun modification : Genitives (pronouns)

	simple modification	prehead <i>sa</i> ³³ pattern
N. Roglai	Nh GENPr	_____
Colloquial Tsat	Nh GENPr	_____
Mandarinized Tsat	_____	Pr <i>sa</i> ³³ Nh
Mandarin	Pr Nh	Pr <i>de</i> Nh

Demonstratives and head nouns

Like adjectives, demonstratives are postposed in Northern Roglai and the colloquial Tsat, but preposed in Chinese-influenced Tsat and Mandarin. In the Chinese-influenced Tsat, however, the demonstratives are often accompanied by a genitive marker, a pattern that matches the adjective plus genitive construction immediately above, a construction that reflects Mandarin influence.

- Northern Roglai:
(16) *sa:k* *gheŋ* *ʔunĩ*
house big this
‘this big house’ (Lee 1966:65)
- (17) *dua* *ia:k* *labu?* *ʔaŋa?* *sia:p* *ñũ* *ʔanĩ* *la* *sa:k*
two person plural child good he this in house
‘these two children of his in the new house’ (Lee 1966:66)

- (18) Tsat:
ʔai³³ ni³³ sat²⁴ ʔan³³
 water this truly cold
shǔi zhè zhēn lěng
zhè shǔi zhēn lěng
 ‘This water is very cold.’ (Zheng 1997:84)
- (19) Tsat (Chinese influenced) (this + GEN) + clf
ni³³ sa³³ ta¹¹ p^han³² pi¹¹kiau³³ lu³³
 this GEN one clf CM much
zhè de yī fēn bǐjiǎo duō
 ‘This portion is bigger.’ (Zheng 1997:75)
- (20) Mandarin this + clf
zhè lù...
 this road
 ‘This road...’ (Zheng 1997:75)
- (21) *zhè shǔi...*
 this water...
 ‘This water...’ (Zheng 1997:84)
- (22) -the classifier version
zhè ge dà fángzi
 this CLF big house
 ‘this big house’

Needless to say, the demonstrative-noun order is a result of Chinese contact. This word order change induced by extensive and prolonged contact with Chinese is quite systematic and pervasive throughout the grammatical system of Tsat. It is found in texts collected from the same speaker by Zheng Yiqing in the 80s published in Zheng (1997). It is interesting to note that the borrowed patterns are found in texts that describe more recent phenomena, whereas the native patterns are used in texts of traditional stories.

Table 3: Noun modification: Demonstratives

	simple modification	prehead <i>sa³³</i> pattern
N. Roglai	Nh Dem	_____
Colloquial Tsat	Nh Dem	_____
Mandarinized Tsat	_____	Dem <i>sa³³</i> Nh
Mandarin	Dem Nh	Dem CLF Nh

Adjectives and head nouns

N. Roglai has postposed adjectives as does the colloquial Tsat, while the Mandarinized Tsat and Mandarin itself have preposed adjectives.

- N. Roglai: postposed adjectives
- (23) *sa:k gheŋ ʔunī*
house big this
'this big house' (Lee 1966:65)
- Tsat: postposed adjectives
- (24) *na¹¹tsun³³ pioŋ^{ʔ42} poi²⁴*
bird big say
niǎo dà shuō
'The big bird said:...' (Zheng 1997:1.1.9)
- (25) *tʰun³³zau³³ pioŋ^{ʔ42} siʔ^{ʔ42}liang¹¹*
tree big relax.in.cool.place
shù dà xiēliáng
dà shù xià xiūxi de
relaxed under a big tree (Zheng 1997:1.2.21)
- Tsat (Mandarin-influenced) preposed adjectives
- (26) *hu¹¹tsa:n^{ʔ42} mi³³ san¹¹ na:i^{ʔ42} sin³³ na:i^{ʔ42} pa:u³³, ...*
Tsat we believe good heart good reward
Huízú wǒmen xìn hǎo xīn hǎo bào
wǒmen Huízú rén xiāngxìn hǎo xīn de rén yíding dédào hǎo bào...
'We Tsat people believe that people with kind hearts will be rewarded...' (Zheng 1997:4.2.1)
- (27) *...kiu³³ san³³*
...old village
...jiù cūn
...jiù cūn
'... the old village' (Zheng 1997:2.1.1)
- preposed with *sa³³*, a calque on Mandarin *de*
- (28) *na:i^{ʔ42} sa³³ san³³huat²⁴*
good GEN life
hǎo de shēnghuó
hǎo de shēnghuó
'(the) good life' (Zheng 1997:2.1.10)
- Mandarin
- (29) *dà niǎo shuō*
big bird say

‘The big bird said:’

(Zheng 1997:1.1.9)

-preposed with *de*, a ‘genitive’ marker

- (30) *hǎo de shēnghuó*
 good GEN life
 ‘(the) good life’

(Zheng 1997:2.1.10)

Note that under the influence of Chinese, the preposed adjectives of Tsat are often accompanied by a genitive marker (or, as Li and Thompson (1981:113-116) term it, an “associative” marker).

Table 4: Noun modification: Adjectives

	simple modification	prehead <i>sa</i> ³³ pattern
N. Roglai	Nh Adj	_____
Colloquial Tsat	Nh Adj	_____
Mandarinized Tsat	Adj Nh	Adj <i>sa</i> ³³ Nh
Mandarin	Adj Nh	Adj <i>de</i> Nh

The spread of the *sa*³³ construction

The *sa*³³ construction is a calque on the Mandarin *de* construction illustrated throughout this paper: X *sa*³³/*de* NP, in which the first element (X) modifies the final NP. This construction, marked by *sa*³³, is expanding in use. Its initial use appears to have been with preposed genitive NPs involving full NPs and then expanded to other parts of the grammar. The three constructions already discussed show this movement from posthead to prehead using the *sa*³³ construction (see Table 5).

Table 5: Noun modification: Spread of the *sa*³³ construction

	Genitive NP	Genitive pr	Dem	Adj
N. Roglai	Nh NP	Nh Pr	Nh Dem	Nh Adj
Colloquial Tsat	NP <i>sa</i> ³³ Nh	Nh Pr	Nh Dem	Nh Adj
Mandarinized Tsat	NP <i>sa</i> ³³ Nh	Pr <i>sa</i> ³³ Nh	Dem <i>sa</i> ³³ Nh	Adj Nh Adj <i>sa</i> ³³ Nh
Mandarin		Pr Nh	Dem Nh	Adj Nh
	NP <i>de</i> Nh	Pr <i>de</i> Nh	Dem clf Nh	GENNP <i>de</i> Nh

For full NP genitives, the older Northern Roglai postposed NPs have been completely replaced by preposed NPs using the *sa*³³ construction; for genitive

pronouns, the older postposed genitive pronouns are still used in more colloquial contexts, but in more Mandarinized speech these are now preposed. For demonstratives, the situation is more complex. All demonstratives were postposed in Northern Roglai and tend to retain this posthead position in the more colloquial Tsat. However, in the more Mandarinized speech demonstratives are variably preposed through the use of the *sa*³³ construction. The distribution between the postposed and the preposed with *sa*³³ variants appears to correlate albeit only loosely with both the register and the type of NP involved. Adjectives were postposed in Northern Roglai, are postposed in the more colloquial Tsat contexts, and even in the most Mandarinized Tsat texts still remain postposed some of the time, at other times being preposed using the *sa*³³ construction. The use of this pattern has expanded beyond the examples in this paper to include prehead relative clauses, a construction highly marked highly marked for an SVO language like Tsat.

Comparative constructions

The existence of contact-induced word order variation is obvious in the two distinct Tsat comparative patterns: the native pattern is inherited from Chamic; the other is borrowed from Chinese:

X - Adj - CM/ST	native pattern
X - CM/ST - Adj	Chinese influenced pattern

In the native pattern, Zheng (1997:75) notes that the word order is quality-marker-standard (X - Adj - CM/ST), that is, the quality being compared, followed by the preposition *la:u*⁷³² ‘CM; pass’ (which serves as the comparative marker), followed by the standard of comparison, typically a pronoun. The extent or degree of the quality may also be marked, in which case it is through modification of the quality.

When relationships are compared, the comparative marker is the preposition *la:u*⁷³² ‘CM; pass; exceed’, derived from a verb. For example,

Tsat (colloquial):

- (31) *nau*³³ *ma*⁴² *la:u*⁷³² *ha*³³
 he fat CM you
tā *pàng* *bǐ* *nǐ*
tā bǐ nǐ pàng
 ‘He is fatter than you.’ (Zheng 1997:75)
- (32) *lu*³³ *pion*⁷⁴² *la:u*⁷³² *bo*¹¹ *koi*²⁴
 coconut.palm big CM pomelo
yēzi *dà* *guò* *yòuzi*

- yēzi bǐ yòuzi dà
 ‘The coconut palm is bigger than the pomelo.’ (Zheng 1997:89)
 (33) ʔa¹¹ko³³ p^hai⁴³ k^ha:n²¹ na:i^{ʔ32} la:u^{ʔ32} ʔa¹¹t^hai¹¹
 elder.brother read book good CM younger.brother
 gēge dú shū hǎo guò dìdi
 gēge xuéxí bǐ dìdi hǎo
 ‘Elder brother studies more than younger brother.’ (Zheng 1997:75)

Tsat influenced by Chinese

However, as Zheng goes on to note, under the influence of Chinese, comparatives often follow a Chinese order, namely, comparative marker, standard, quality (X - CM/ST - Adj), using *pi*¹¹ ‘CM; compare’ borrowed from Chinese to mark the comparison. For example:

- (34) kau³³ pi¹¹ ha³³ tsat²⁴tso³³ ki³³ sun³³
 I CM you short three inch
 wǒ bǐ nǐ ǎi sān cūn
 wǒ bǐ nǐ ǎi sān cūn
 ‘I am three inches shorter than you.’ (Zheng 1997:75)
 (35) mi³³ sa³³ saŋ³³huat²⁴ ta¹¹ zai³³ pi⁴³ ta¹¹ zai³³ pu³³ na:i³²
 we GEN life, one day CM one day NEG good
 wǒmen de shēnghuó yī tiān bǐ yī tiān bù hǎo
 wǒmen de shēnghuó yītiān bǐ yītiān chà...
 ‘...our life went downhill each day,’ (Zheng 1997:2.1.4)
 (36) zin¹¹min¹¹ sa³³ saŋ³³huat²⁴ ta¹¹ zai³³ pi⁴³ ta¹¹ zai³³ na:i^{ʔ32} a^ʔ
 people GEN life one day CM one day good PART
 rénmin de shēnghuó yī tiān bǐ yī tiān hǎo a
 rénmin de shēnghuó cái yītiān bǐ yītiān hǎo yuè a
 ‘...people’s lives began to get better and better.’ (Zheng 1997:2.1.16)

Mandarin

- (37) wǒ bǐ nǐ ǎi sān cūn
 I CM you short three inch
 ‘I am three inches shorter than you.’ (Zheng 1997:75)
 (38) tā bǐ nǐ pàng
 he CM you fat
 ‘He is fatter than you.’ (Zheng 1997:75)

In these examples, both the word order and the comparative marker itself are Chinese. Instead of the native pattern of quality-standard-noun illustrated by the example in (1), we have noun-marker-standard-quality exemplified by examples in (2a) to (2c). In fact, this kind of almost wholesale borrowing from

Chinese is not confined to a limited number of grammatical structures in Tsat. In other words, it is quite pervasive throughout the grammatical system of Tsat.

Adverbs and conjunctions from Chinese

Adverbs, prepositions, and conjunctions are all borrowed from Mandarin.

Adverbs: (the examples given here are intensifiers)

- (39) *tʰa:i*³³ *ɕiɑ*²⁴. *pʰa:i*³³ *siaŋ*²¹ *na:i*³²
 very hot extremely good
tài rè fēicháng hǎo
 'very hot' 'extremely good' (Zheng 1997:76)
- (40) *na:i*³² *ket*⁴³. *sa*²⁴ *ti*⁵⁵
 good extremely really white
hǎo jí zhēn bái
hǎojí zhēn bái
 'extremely good' 'truly white'

Correlative conjunctions:

- (41) *ziu*³³ *pa*³³ *ziu*³³ *ha:i*³³
 both hungry and tired
yòu è yòu lèi
yòu è yòu lèi
 'Both hungry and tired.' (Zheng 1997:84)

Both the Mandarin and the Tsat have exactly the same structure with the key morphemes borrowed from Mandarin.

Clausal conjunctions:

- (42) *zi*¹¹ *ko*¹¹ *kʰi*⁴³ *tʰa:i*³³ *ɕai*²⁴, *kau*³³ *sau*⁴³ *pu*³³ *na:u*³² *lɑ*³³
 if tomorrow very hot, I then NEG go PERF
rúguǒ míngtiān tài rè, wǒ jiù bú qù le
rúguǒ míngtiān tài rè, wǒ jiù bú qù le
 'If tomorrow is very hot, I won't go.' (Zheng 1997:85)

What makes these examples particularly interesting is that not only are they borrowed but that for the most part their syntax in Tsat matches their syntax in Mandarin. That is, what has been borrowed is a construction still marked by its characteristic lexical item.

Other Han influenced constructions

Not all Mandarin influence has resulted in patterns of synchronic variation. However, even when this sort of syntactic variation is no longer found, it is still fairly obvious that contact has been at work. Whenever Tsat word order patterns differ from those of the Chamic languages of Vietnam, they are either identical with or close to the patterns found in Chinese. And, of course, the fact that oftentimes grammatical morphemes are borrowed together with the syntactic constructions, even serving to define the construction, is noteworthy.

Examples abound. In (43) below are three separate constructions showing Chinese syntactic influence. The first, indicated by the initial double underlining, is the extension of the prehead modification of the *sa*³³ construction to produce a prehead relative clause. This type of typologically marked prehead relative clause has developed under Chinese influence in at least three independent but parallel cases, once in Karen, one in Bai, and once in Tsat (Thurgood and Li, in preparation). This has been extended, under the influence of Mandarin, to include other constructions quite new to Tsat.

- (43) *dī*⁵⁵ *nan*³³ *sa*³³ *mo*³³ *sī*¹¹ *mai*³³ *sa*³³
 lie.down that GEN cow be female GEN
tǎng nà de huángniú shì mǔ de
tǎngzhe de nà huángniú shì mǔ de
 ‘The yellow cow lying down is female.’ (Zheng 1997:73)

The second is the use of the Mandarin borrowing *sī*¹¹ to mark the equative construction; typically the Chamic languages simply use a zero copula for such sentences. And the third is the use of a postposed *sa*³³ as a nominalization in the *mai*³³ *sa*³³ ‘female’. All three reflect Mandarin influence.

Other languages

None of this is restricted to Tsat, of course. Strikingly parallel developments are in progress throughout the Chinese dominated area of Asia. Everywhere where intense Chinese influence is manifested languages are undergoing major restructuring, resulting in word order variation and change. Contrary to the general belief that a very long period of time of persistent contact is needed for structural borrowing to occur, the wholesale restructuring is taking place quite rapidly. Intensity of contact more than duration seems to be the crucial factor. The level of bilingual proficiency and the instability of the social structure of the borrowing language seem to have a major impact on the length of time needed and the extent of the structural shift in contact induced changes.

Oroqen

For instance, in the southeastern and central dialect regions of Oroqen (a Tungusic language of the northeast; Li and Whaley 2000, Whaley, Grenoble, and 160

Li 1999; Li 2000, Sun and Li 2001; Whaley and Li 1998, 2000), Chinese contact did not occur until after the settlement in the late 1950s and early 1960s. Nonetheless, within a short span of only a few decades, we already see signs of strong contact influence on the grammatical structure of the language. A case in point is the fact that one of our informants from the central dialect region used the adverb *maʃaŋ*, which is a Chinese borrowing meaning ‘immediately’ spontaneously without noticing it. When fed the Oroqen word *diyaldʒi* meaning the same thing, he accepted it, but strongly prefers the Chinese borrowing. Interestingly enough, our informants from the western and northeastern dialect regions adamantly rejected the Chinese form insisting that it is not an Oroqen word.

Even in the northeastern dialect region where Oroqen is preserved the best, we saw strong Chinese influence, which is shown in the examples in (44).

- (44) a) *ʃi tɪmana ŋənə-ni yəʃa*
 you tomorrow go-2SG.PRES Q.2SG.PRES
 ‘Are you going tomorrow or not?’
 b) *yabuʃa haʃi yəʃa*
 walk.PAST still-be Q.PAST
 ‘Went or not?’

In (44), we have two examples of the A-not-A question formation in Oroqen. Notice that the informant produced the Chinese *háishi* ‘still be: or not’ in 44b). She did so without realizing it at all until it was pointed out to her. This informant feels at ease with both languages. In fact, she possesses native proficiency in both Chinese and Oroqen.

This kind of phenomenon suggests that when a speaker reaches a certain level of bilingual proficiency, borrowing between the languages is much more readily than is generally assumed in the literature. Thus, it does not take a very long time for a language to shift to a completely different typological pattern in its grammatical structures. Central to the rate of such structural shifts are sociolinguistic factors, particularly, the so-called intensity of contact. Our work suggests that for both Tsat and Oroqen a crucial factor has been schooling in Mandarin.

Mulam

In Mulam (Zheng 1988), a Kam-Sui language of the Guangxi area related ultimately to Thai, is undergoing many of the same changes Tsat is: borrowing of conjunctions, adverbs, and prepositions along with the introduction of constructions with new word orders under the influence of Chinese. In fact, Mulam even has its own equivalent of the *sa*³³ construction, built on a different genitive marker but, like its Tsat counterpart, resembling the *de* construction of Mandarin.

Zheng Guoqiao (1988:173), in discussing the Mulam borrowing of Chinese adverbs and conjunctions, notes that “degree and quantity adverbs are all borrowed from Han” and that “borrowed adverbs generally are subject to the same syntactic rules as Han”. That is, what is borrowed is not just a lexical item but a lexical item along with accompanying syntax – in short, a construction.

Mulam, despite being its geographical distance from Tsat, has calqued the *de* construction of Mandarin very much as Tsat has, and Mulam is borrowing many of the same constructions along the representative morpheme, leading to convergence with Mandarin. Sometimes this has produced variation with the native patterns competing with the borrowed patterns; in other cases, the native pattern has been completely replaced.

Observations

In short, under intense Chinese similar, rapid restructuring is occurring in geographically distant languages belonging to distinct language families: in Tsat, an Austronesian language of Hainan; in Oroqen, a Tungusic language much farther to the north; and in Mulam, a Tai-Kadai language found south of the Yangtze.

Although not our major focus, it is obvious that sociolinguistic factors rather than structural factors that provide the impetus for the word order changes. Although some linguists consider structural similarity and functional congruence as the most important factors in cases of grammatical borrowing (e.g. Weinreich (1953)), Tsat and Oroqen seem instead to support Thomason and Kaufman’s contention (1988:35) that “it is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact. Purely linguistic considerations are relevant but strictly secondary overall.” In the case of Tsat, a major impetus to wholesale restructuring has been the fluency brought about by schooling in Mandarin.

Most central to the data presented are the paths of diachronic change. Grammatically, one configurational grammatical structure is being replaced by another, construction-by-construction, with the older Chamic word order being replaced by its Mandarin equivalent. In some cases, the Mandarin-influenced construction, often marked by a transparently Mandarin grammatical morpheme, is simply a marked alternative, as with comparatives, but in other constructions the word order of the native Chamic construction has been completely superseded, as with the genitives. The word order changes have entered the language as borrowed constructions, marked by a characteristic often-borrowed grammatical morpheme. Still further influence is manifested in the overgeneralization of calques, cf. the spread of the *sa*³³ construction in Tsat.

The complex set of conditions responsible for the restructuring are only partly explainable by internally and externally motivated principles proposed in the literature. A profound understanding of the situation must take account of the dynamic changes that take place in not only linguistic structures but in the social

conditions as well. There is a range of social factors that contribute to borrowing and structural shifts. In the case of Tsat, encroaching bilingualism with a powerful dominant language (along with schooling and social mobility) are among the most prominent factors that lead to the massive borrowing and drastic structural shift.

The rapid changes taking place in these languages make it imperative that the nature of the speech community be specified in far more detail than it is now and as quickly as possible. We can already see that extensive language change in these particular languages whose speaker community is constituted by bi- or multi-lingual linguistic and ethnic minorities but the details need to be documented and the paths of change need to be examined. A lot more work needs to be done and done quickly before the languages in question cease to exist.

Finally, a comment on the obvious: It is the construction, rather than just its characterizing grammatical morpheme, that is the typical unit of borrowing.

gthurgood@csuchico.edu

fli@csuchico.edu

Acknowledgements

This paper is a modified version of a paper presented first at presented at the Thirty-fourth International Conference on Sino-Tibetan Languages and Linguistics, held in Kunming, Yunnan, October 2001, at the Linguistics Society of America annual meeting, held in San Francisco, January 3-6, 2002, and at the Berkeley Linguistics Society meeting, February, 2002.

References

- Benedict, Paul K. 1941. A Cham colony on the island of Hainan. *Harvard Journal of Asiatic Studies* 4:129-34.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81-138.
- Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. University of California Press.
- Li, Fengxiang. 1996. A report on the study of the consonant systems of two Manchu-Tungusic languages in China: Evenki and Oroqen, paper presented at the annual meeting of the Linguistic Society of America, San Diego.
- Li, Fengxiang and Lindsay J. Whaley. 2000. Emphatic reduplication in Oroqen and its Altaic context. *Linguistics* 38.2:355-372.
- Ouyang, Jueya and Zheng Yiqing. 1983. The Huihui language of the Hui nationality of Yaxian, Hainan. *Minzu Yuwen* 1:30-40.
- Sun, Hongkai and Fengxiang Li. 2001. Contact, attrition, and structural shift: the case of Anong and Qiang. Paper presented at the Tibeto-Burman Workshop at the University of California, Santa Barbara, July 27-29, 2001.

- Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thurgood, Graham. 1999. *From Ancient Cham to modern dialects: Two thousand years of language contact and change*. Oceanic Special Publication No. 28. Honolulu: University of Hawai'i Press.
- _____. 1996. Language contact and the directionality of internal 'drift': the development of tones and registers in Chamic. *Language* 71.1:1-31.
- Weinreich, Uriel. 1953. *Languages in contact*. The Hague: Mouton.
- Whaley, Lindsay J., Lenore A. Grenoble, and Fengxiang Li. 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75.xxx
- Whaley, Lindsay J. and Fengxiang Li. 1998. The suffix *-kan* in Oroqen. *Studies in Language* 22:181-205.
- _____. 2000. Oroqen dialects. *Central Asiatic Journal* 44.1:105-130.
- Zheng, Guoqiao. 1988. The influences of Han on the Mulam language. In Jerold A. Edmondson and David B. Solnit, eds, *Comparative Kadai: Linguistic Studies Beyond Tai*. Summer Institute of Linguistics. Pp. 167-177.
- Zheng, Yiqing. 1986. A further discussion of the position of Huihui speech and its genetic relationship. *Minzu Yuwen* 6:37-43.
- _____. 1997. *Huihui Yu Yanjiu [A Study of Tsat]*. Shanghai Yuandong Chubanshe [Shanghai Far East Publishers].

Graham Thurgood / Fengxiang Li

CSU Chico
Chico, CA 9

gthurgood@csuchico.edu,