

Editorial Note

This volume was produced under difficult conditions. The publication of articles was not only very slow; the number of articles was also reduced due to circumstances beyond our control - the heavy flood in Thailand during October to December 2011. So we ask the reader's indulgence for any effects this may have on the volume.

For this volume, we are pleased to present articles focused on the following languages: Jieyang-Hakka, Jowai-Pnar, Lai, Pumi, Ten-edn, Tai and Viet-Mường; these papers make contributions to language documentation, especially in phonetics and lexicography, and better understanding the historical processes of language diversification. Additionally there are typological papers on phonetics and narrative in Mon-Khmer languages which address important general issues.

Graceful acknowledgement should be made to Paul Sidwell for seeing the final volume through to press, and to Brian Migliazza for facilitating the publication of his volume.

The Mon-Khmer Studies (MKS) was first published by the Linguistic Circle of Saigon and the Summer Institute of Linguistics in 1964. After nearly 50 years, the print edition will be discontinued. From the volume 41 onward, the MKS is going completely digital and open access. The journal will move to a continuous online publication model, consistent with trends in academic publishing internationally. Also, arrangements will be made for print-on-demand delivery, although we expect electronic distribution to become normal.

We thank our readers, authors, reviewers and editors for their continuing support of the journal, now and into the future.

Naraset Pisitpanporn
for MKS Editorial Board
April 2012

Table of Contents

Editorial Note.....iv

Articles

John D. PHAN
Mường is not a subgroup: Phonological evidence for a
paraphyletic taxon in the Viet-Muong sub-family.....1

Mary M. PETERSON
Notes on Ten-edn (Tonga-Mos) and Kensiw Borrowings.....19

NITASAKORN Shiwaruangrote
A description of Jiēyáng Hakka phonology as spoken in
Bangkok, Thailand.....35

PHANINTRA Teeranon
The Interaction between Pitch and Vowel Length in
Mon-Khmer Languages.....60

SOMSONGE Burusphat
A comparison of sequential strategies in Mon-Khmer narratives.....86

Xiaohang QIN & Fanglan LI
The status quo and trend of language use by Lai people.....107

Xinyuan HE
Plant and wildlife naming system in southern Pumi.....116

Hiram RING
A phonetic description and phonemic analysis of Jowai-Pnar.....133

Notes

Xiaohang QIN
Origin of Tai numeral ha³ (five).....176

Mường is not a subgroup: Phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family*

John D. PHAN
Cornell University, USA

Abstract

The nature of the language or languages termed “Mường” has not been well understood. In the struggle to clarify the inter-branch relationships of the larger Vietic family, Mường has generally been regarded as one of two major subgroups of the Viet-Muong subfamily (opposite Vietnamese). This paper presents new data collected in north-central Vietnam on three varieties of Mường: the Trám variety of Thanh Hóa province, the Chôi variety of Phú Thọ province, and the Vang variety of Hòa Bình province. Preliminary analysis reveals salient diversity in their onset inventories, with no evidence of shared innovation. This suggests that a Proto-Muong level of diversification is not reconstructable. In other words, Mường is not a subgroup.

Keywords: Viet-Muong, Historical Phonology, subgrouping

0.0 Introduction

The status of the “Mường” language has not been well understood. The term itself—a Tai loanword—was once applied liberally to a number of relatives of modern Vietnamese. Following Michel Ferlus’ description of a number of conservative Vietic languages spoken in the central highlands, “Mường” came to refer to those related languages which—like Vietnamese—exhibit pervasive monosyllabicity, and lexical tone systems. For this reason, “Mường” and Vietnamese are now treated as forming the “Viet-Muong” subgroup of a Vietic language family.

The innovations which formed a distinct Vietnamese language from Proto-Viet-Muong (henceforth, PVM) are fairly well understood.¹ These

* Many thanks to my consultants for their critical contributions. Special thanks to Professor Tran Tri Doi of Vietnam National University, for his essential help and guidance on the project. Thanks also to Nguyen Van Duc, for his extensive assistance in many efforts.

¹ These were described by Ferlus (1986) as a kind of drag chain involving the mutation of initial consonants, plus the spirantization of medial consonants followed by the loss of minor

innovations mean that a “Proto-Vietnamese” level is reconstructable—indicating in evolutionary terms, that Vietnamese represents the evolution of a subgroup of PVM dialects. In the articulation of this Vietnamese subgroup, the assumption has crept in that “Mường” also represents a subgroup. This is not the case. Novel fieldwork conducted in 2009-2010 reveals pervasive diversity in the initial consonant inventories of three “Mường” varieties, which contradict the possibility of a “Proto-Mường” stage of linguistic development. This diversity of innovations describes, rather, the distinct evolution of an already diversified array of PVM dialects. In other words, “Mường” is not a subgroup.

And yet, the languages called “Mường” today seem to exist in mutual intelligibility with one another. Does this not constitute evidence that “Mường” is a single language, and as such, a single clade? As I will demonstrate in this study, the one does not necessarily require the other. The “Mường” languages may be mutual intelligible, but even leaving aside the deeply impressionistic nature of such a criterion, the fact remains that defining a language in terms of mutual intelligibility is not the same thing as defining it terms of genetic history. There are other paths by which a language may obtain this phenomenon of unity. As Andrew Garrett has recently proposed for Ancient Greek, convergence is one possibility (Garrett, 2006). In their study of Western Nomic, Babel, Garrett, et al. (2009) also proposed the term *apomorphic taxon* to describe those linguistic groups which demonstrate a unity of features obtained through areal diffusion, rather than shared innovation.

There remains another possibility. In their defense of phylogenetic methods, Greenhill & Gray (2009) describe the Formosan languages as bearing no cladistical unity, but rather representing several distinct lineages evolving directly out of Proto-Austronesian. These lineages were “left behind” by the mutation of other Austronesian languages after the migration out of Taiwan, and thus bear a conservative resemblance to one another despite the lack of a “proto-Formosan” stage of evolution.

The condition of the Formosan languages resembles the biological concept of *paraphyly*, which describes groups of organisms whose most recent common ancestor is also the ancestor of an innovating group no longer counted within the taxon. The easiest example of this is the class of Reptilia, which excludes the monophyletic clades of Aves (birds) and Mammalia, with whom their most recent common ancestor is shared. In the same manner, the most recent shared ancestor of the “Mường” languages is *Proto-Viet-Muong*—which is, of course, also the shared ancestor of Vietnamese.

syllables. I have modified Ferlus’ claims somewhat, and will discuss these innovations later in the study.

Through an examination of the stop onset inventories of three “Mường” varieties, this study demonstrates that “Mường” is not a subgroup, but a paraphyletic taxon. Consequent to this argument, I also present a model for Proto-Việt-Mường diversification, in which a subgroup of PVM dialects branches off to form the modern Vietnamese language.

0.1 Organization of the paper

Section 1 briefly introduces the Mường and their language, as well as reviewing some pertinent scholarship. Section 2 examines the diverse evolution of PVM initial stops among four geographically diverse Mường varieties. The lack of shared innovations and the presence of independent mutations from the PVM level are taken as evidence for the paraphyletic nature of the Mường language. I summarize these arguments in Section 3, in which I present a model for Việt-Mường speciation that describes Vietnamese—but not Mường—as a cohesive subgroup.

1.0 Introduction to the Mường and their language

The Mường are currently the third-largest ethnicity in Vietnam (after the majority Kinh, and the highland Tay), with a population of roughly 1,140,000 spread out over an area west, southwest, and south of the Red River (1999 census; Lewis, 2009). They are most heavily concentrated in the provinces of Hoa Binh, Thanh Hóa, and Phú Thọ², with communities as far west as Yên Bái and Sơn La, and as far south as Nghệ An (Lewis, 2009). Mường communities are generally situated in low mountain valleys surrounded by peaks, which places them in geographical zones contiguous with the Kinh (Vietnamese-speaking) majority, as opposed to the higher elevations inhabited by the Hmong or Dao. The Mường themselves are subsistence farmers who cultivate rice and corn as staples, alongside a number of small cash-crops including tea (Phú Thọ), sugarcane (Thanh Hóa, Phú Thọ, Hoa Binh), and recently, acacia lumber (Phú Thọ, Hoa Binh).

The genealogical status of Vietnamese (and by association, Mường) was once a subject of some controversy. Long held to be a simple offshoot of Chinese (see Taberd, 1838), Henri Maspero later recognized an affinity between Việt-Mường and other Southeast Asian language groups (like Tai and Mon-Khmer), but argued for a principal connection with Tai because of the presence of tone (Taberd, 1838; Maspero, 1912). The issue was finally laid to rest by André Haudricourt in his seminal 1954 account of tonogenesis, which demonstrated not only that Vietnamese (and Mường) tones evolved from consonants (and not a unique genetic inheritance), but that Vietnamese (and

² Vinh Phú province was split into Phú Thọ and Vinh Phúc provinces in 1996; the *ethnologue* description of Mường still lists “Vinh Phú” (rather than Phú Thọ) as a Mường area.

Mường) was a member of the Mon-Khmer family, and completely unrelated either to Tai or Chinese languages (Haudricourt, 1954).

This genetic classification was further refined in the '90s, when work by Michel Ferlus and Gerard Diffloth argued for the placement of Vietic somewhere in the Katuic branch of the Mon-Khmer family, though they differed on the details (Diffloth, 1991; Ferlus, 1992; Ferlus, 1994). This position was supported by Alves (2005), who expanded the argument with a body of lexical evidence, but was contradicted by Sidwell (2005, 2010), who argued for a more or less flat tree structure out of which radiated as many as thirteen independent branches—including distinct Bahnaric and Vietic lines (Sidwell, 2005; Sidwell, 2010). As such, there are now two opposing views on the placement of Vietnamese within Mon-Khmer—one, following Diffloth and Ferlus, that argues for an affiliation with Katuic; and the other, following Sidwell, arguing for direct diversification (alongside a number of major branches) from Proto-Mon-Khmer itself.

What is of most interest to us, however, is what is meant by the term *Mường*—in all of its connotations. It is quite plain that “Mường” as both a language and ethnicity has suffered from a long history of ambiguity. Cuisinier described the Mường as culturally and linguistically, *des Annamites attardés*—“backward Annamese”, establishing a strong bias toward understanding the Mường as primitive, fossilized cousins of the lowland Vietnamese³ (Cuisinier, 1946, pp. 562-563). This characterization was ultimately challenged by Taylor (2001), who argued that French ethnological categories forced the Mường into an artificially distinct subgroup, necessarily subordinate to the lowland Vietnamese (Taylor, 2001). Although not linguistic in character, Taylor’s criticism is largely applicable to early scholarship on Mường and other ethnic minority languages related to Vietnamese.

Vương Hoàng Tuyên (1963) counted at least three Vietic languages, identified by place-name, all as “Mường”, which he listed alongside yet another (simply) “Mường” heading. Thompson (1976) likened this to the “presentation [of] three or four Swiss German dialects, identified by village names, besides a list headed simply ‘German’” (Thompson, 1976, p. 1115). Thompson was right to point out the ambiguity in Vương’s practice, but his German analogy only illustrates a confusion of nomenclature; the real error here lies in what both Vương and Thompson appear to accept as “Mường.” As Thompson notes, the plain “Mường” listed was probably a Hòa Bình dialect of Mường (as defined today). However, the other three “forms” (May/Rục, Arem, Tay Pong) are in fact, distinct languages more closely related to each other than either to Vietnamese, or—significantly—to Mường (Ferlus,

³ For an excellent review of early French ethnographic scholarship on the Mường, see Keith Taylor’s (2001) “On Being Mườnged”. Taylor, K. W. (2001). On Being Muonged. *Asian Ethnicity, Volume 2, Issue 1*, 25-34.

Vietnamien et proto-viet-muong, 1975). Evidently, the habit of labeling any “primitive” relatives of the Vietnamese as *Mường* led to what Taylor rightly called an artificial category, which lumped together groups of very different lineages, all in opposition to the lowland, urbanized civilization of the Kinh. In fact, the “Mường” spoken in Hoa Binh (as well as Thanh Hoa and Phu Tho) is better grouped with Vietnamese than with these other languages.

This confusion was repeated in Thompson’s 1975 attempt to reconstruct PVM, in which the author appears to use the term *Mường* to refer to all immediate relatives of the Vietnamese language. Thompson thus describes a choice between relying on the “Mường forms” of Chút, Rục, and Pong, and the “Mường” which Milton & Muriel Barker had documented in their extensive work with refugee populations in the south. Because of a lack of material on the former type of “Mường”, and in favor of “an internal consistency” found in the latter, Thompson chose to rely primarily on the Barker data for his reconstructions (Thompson, 1976, pp. 1116-1117). Note, however, that Thompson did not reject the former option because he believed they belonged to a separate subgroup of Vietic.

Ferlus eventually clarified this issue by identifying a subgroup of “conservative” (*conservatrice*) Vietic languages, including Rục, Arem, Pong, Maliêng, Thavung and others, opposed to Vietnamese and Mường, both of which notably share full-blown tone systems, pervasive monosyllabicity, and an impoverished cluster inventory (Ferlus, Vietnamien et proto-viet-muong, 1975). To this we may add the presence of large bodies of Chinese loanwords (Maspero, 1912). This typological grouping implied a cladistical divide as well, with Vietnamese and Mường occupying a single “Viet-Muong” subgroup. Unfortunately, the impression that Mường itself represents a cohesive subgroup opposed to Vietnamese has somehow also passed into modern scholarship (e.g. Rischel, 1995).

Milton and Muriel Barker were the first to work on a single language defined as *Mường*, and their substantive fieldwork was, for many decades, the only Mường data available in the west (it was this data on which Thompson based his 1975 reconstruction). Unfortunately, the data collected by the Barkers was severely contaminated by language-mixing with Vietnamese, probably because of the migrant conditions under which it was collected. The languages are so very closely related that register-switching between a natural voice, and a more “Vietnamese-like” Mường is extremely common, a habit fueled by the prestige status of Vietnamese. Even today, when it is relatively fashionable to celebrate the ethnic diversity of Vietnam, I found Mường speakers extremely prone to “Vietnamicizing” their own language.

These issues aside, the most binding constraint of the Barkers, and those (like Thompson) who used their data, is that they were forced to rely on a single dialect—the “Mường Khén” of Tân Lạc District, Hoa Binh (a sub-

dialect of the Mường Bi). In a language as diverse as Mường, no single dialect can really be taken as universally representative—and indeed, that is largely the point.

Nguyễn Văn Tài did later collect data from a very broad range of dialects, but these also suffer from contaminated tonal transcriptions, similar to the Barker data mentioned above (Nguyễn, 1982). This lack of reliable, representative Mường data was the primary motivation for my fieldwork in 2009-2010.

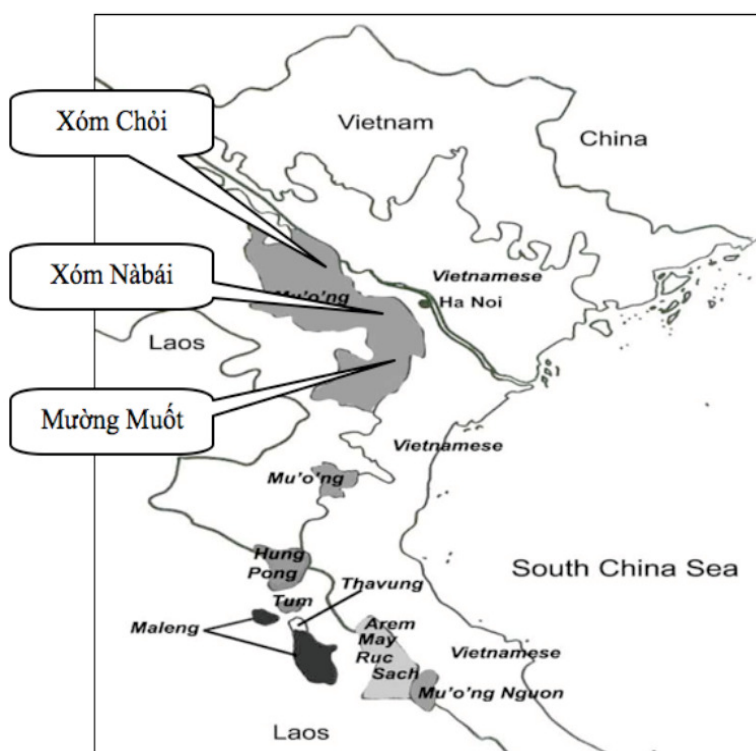


Figure 1. Relative Distribution of Field Sites⁴

1.1 Fieldwork in Thanh Hóa, Hoa Bình, and Phú Thọ

In order to redress the limitation of available data on Mường, I traveled to Vietnam from 2009-2010 to conduct descriptions of three representative Mường dialects: the Mường Châm dialect of Muốt commune in Thanh Hóa; the Mường Chòi dialect of Chòi commune in Phú Thọ, and the

⁴ Map reproduced with permission from Paul Sidwell's *mon-khmer.com*. (accessible via: <http://people.anu.edu.au/~u9907217/>)

Mường Vang dialect of Nàbái commune in Hoa Binh. For simplicity, I will hereafter refer to each dialect by commune-name.

In each of these communes, I spent one-to-two weeks acquiring a full description of the phonological inventory, after which I collected and recorded a 1,000-entry core wordlist. At all sites I relied on one male and one female consultant. The ages and gender these consultants are provided below:

Table 1. Primary Consultant Information

	Site	Sex	Age
1.	Muôt Commune	M	64
2.	Thanh Hóa	F	45
3.	Chôi Commune	M	62
4.	Phú Thọ	F	40s
5.	Nàbái Commune	M	46
6.	Hoa Binh	F	42

Preliminary analysis of these data reveals pervasive diversity in the stop onset inventories of these varieties, which contradicts the possibility of a single subgroup.

2.0 The status of Mường

In this section, I will present three salient innovations attested among the Mường dialects. None of these innovations are shared universally, and all represent mutations from the PVM level. As such, they strongly indicate that Mường represents a collection of distinct lineages, rather than a single, innovating subgroup.

The innovations under question correspond to some of the defining innovations of modern Vietnamese—which, for these very reasons, *can* be considered a subgroup. In particular, the nasalization of PVM implosives, in concert with the devoicing of regular stops, comprised an intricate drag-chain that mutated most Vietnamese onsets into forms distinct from its Mường counterparts. I have diagrammed this drag-chain below for reference:

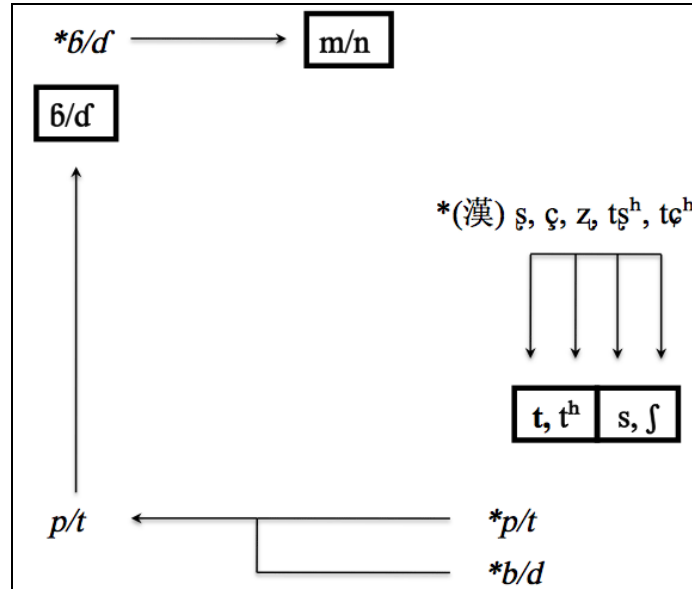


Figure 2. Viet-Muong drag chain

Ferlus (1986) argued that modern Vietnamese nasals (m-, n-) actually derived from an ancient set of implosives, while modern Vietnamese implosives evolved from ancient Proto-Việt-Mường explosives. (Ferlus, 1986) As diagrammed above, we may add a separate process through which ancient fricatives and affricates were plosivized (a process now ongoing, but not yet complete in Mường). Thus modern Vietnamese often bears nasal onsets for Mường stops.

In 2.1, I will focus on the modern correspondences for PVM implosives, as well as rhotics (not pictured above).

2.1 The problem of modern Mường diversity

This section presents an analysis of three sets of data: two sets that correspond to Proto-Việt-Mường implosives $*b-$ and $*d-$ respectively, and one set that corresponds to Proto-Việt-Mường $*r-$. As will be evident soon, each of these bears diverse reflexes in modern Mường. In the following tables, data from Muốt, Nàbái, and Chôi varieties derive from my own fieldwork introduced above, while Khên data relies on Thompson (1975), and is ultimately attributed to the Barkers (1966).

PVM $*d-$, produces either implosives, or an innovated rhotic in modern Mường:

Table 2. Mường correspondences for ancient PVM implosive *ɗ-

	Gloss	Việt orth.	Muốt	Nàbái	Chỏi	Khễn
1.	ought to, must	nên	reɲ1	ɗeɲ	rein1	ɗeɲ1
2.	deer	nai	raj1	ɗaj1	raj1	ɗaj1
3.	sunny/bright	năng	rãŋ2	ɗãŋ3	rãŋ3	ɗãŋ4
4.	water	nước	rak3	ɗak3*	rak3	ɗak3
5.	crushed	nát	raic3	ɗaic3	(pet6)	ɗac3

As Table 2 shows, the regular Vietnamese reflex for PVM *ɗ- is /n-/. This is true across all dialects of Vietnamese. However, the Mường data clearly demonstrates at least two major correspondences for Proto-Việt-Mường *ɗ-: /r-/ in Muốt and Chỏi, and either /ɗ-/ or /d-/ in Nàbái and Khễn⁵. Already we see that the innovation of *ɗ- to r- is not shared universally in Mường.

It is tempting to reconstruct *ɗ- → r- as an innovation shared between Muốt and Chỏi; however the correspondences do not always line up. For example, the verb “to cook” (Viet. <nấu>) corresponds to Chỏi /ro3/, but to Muốt /no3/; similarly, the verb “to fold, crease” (Viet. <nếp>) corresponds to Chỏi /rep3*/, but to Muốt /nep3*/. At this point the relationship between Chỏi and Muốt is unclear; however the distinctiveness of the Hoa Bình varieties seems apparent.

Reflexes for Proto-Việt-Mường labial implosive ɓ- also this division:

Table 3. Mường correspondences for ancient PVM ɓ-

	English	Việt orth.	Muốt	Nàbái	Chỏi	Khễn
1.	lose	mất	văt3	bət3*	vət3/6	bət4
2.	carry	mang1	vaŋ1	ɓaŋ1	vaŋ1	ɓaŋ1
3.	salt	muối	voj3	ɓoəj3	vwaɟ3	ɓoɟ3
4.	misshapen	méo	vew2	ɓew3	vew3	ɓew3
5.	salted fish	mắm	văm3	ɓăm3	văm3	ɓăm3

In these cases, Muốt and Chỏi demonstrate a spirantized v-, whereas Nàbái and Khễn again demonstrate a conservative ɓ/b-. Note that Vietnamese also famously demonstrates spirantization from PVM stops; however, in Vietnamese, this spirantization affected medial onsets (including *-p-), while

⁵All Khễn data in this section is taken from Thompson (1975), who in turn relied on the data of Milton & Muriel Barker. For simplicity, I have rendered the Thompson/Barker data in my own transcriptional system, numbering tones according to their correspondences. It is unclear whether or not the dentals in the Muong Khễn data are implosive or not; for this reason, I have faithfully rendered them as “d-”.

initial *p- was nasalized (Ferlus, 1975; Ferlus, 1982).⁶ Thus, the only trait grouping the four Mường dialects together so far is their lack of the nasalization seen in Vietnamese. But the lack of an innovation does not constitute a basis for subgrouping, and un-shared *d- → r and *b- → v-innovations suggest distinct, rather than shared lineages.

As noted above, it is tempting to posit a Muốt-Chỏi dialectal subgroup since these seem to share two innovations: the rhoticization of dental implosive d-, and the spirantization of labial implosive b-. As also noted above, however, there exist counter-examples to the rhoticization innovation that complicate the matter. Furthermore, each of these dialects is spoken on either side of Hoa Binh—Phú Thọ to the north, and Thanh Hóa to the south. This suggests a classic scenario of innovation expanding from a center in Hoa Binh, leaving conservative fringes around the perimeter. This scenario, however, is problematized by the fact that the perimeter appears to be innovating, while the Hoa Binh dialects manifest more conservative retentions, suggesting a migration pattern of one group out of Thanh Hóa, northward around Hoa Binh, up into modern Phú Thọ.

Let us now take a brief look at the reflexes for Proto-Việt-Mường *r-.

Table 4. Mường correspondences for ancient PVM *r-

	English	Việt orth.	Muốt	Nàbái	Chỏi	Khến
1.	fence	rào	raw2	haw2	raw2	raw2
2.	forest	rừng	ruŋ2	həŋ2	ruŋ2	rəŋ2
3.	stubble (rice)	rạ	ra4	ha4	ra4	ra-
4.	intestine	ruột	roʔc4	hwoaʔc4*	roʔc4	rɔc-
5.	spirits	rượu	raw4	haw4	raw4	raw-

We see fairly consistent r- realizations across Mường, except for the surprising h- reflexes in Nàbái. This laryngealization represents the first positive innovation we have seen in the Hoa Binh dialects. Interestingly, this innovation is not shared in the Barker data—and indeed, there is no reason to expect that it would have been, since (once again) the retentions they are shown to share in Tables 2-3 do not demonstrate a subgrouping relationship.

⁶ Note that while Ferlus' model works quite well for Vietnamese, it manifestly does not apply to any of the Mường varieties under study here. Medial *p- corresponds to non-spirantized stops in Mường, whereas the spirantized onsets in Chỏi and Muốt displayed in Table 3 result from PVM implosivized onsets. These two processes are therefore completely distinct. The fact that medial stops did not spirantize in the Mường varieties suggests either that the loss of minor syllables was part of the chain of innovations which defined a separate Vietnamese language, or that the mechanism of spirantization was not an ancient intervocalic position.

2.2 *Mường is not a subgroup*

In the case of Vietnamese, we visited a complex set of changes, a part of which transformed Proto-Việt-Mường implosives into nasals. This set of changes divided Vietnamese from Proto-Việt -Mường, and provides the basis for a Vietnamese subgroup⁷.

What we need to prove the validity of a Mường subgroup are innovations shared by all Mường dialects, but which is not shared by Vietnamese. And yet, all we seem to have that unifies Mường are retentions. To name a few, these include the preservation of medial liquid clusters tl- and kl- (cf. Muốt *tleiŋ* for Viet . “trên”, “on”; *tluj* for “trời,” “heaven”), non-spirantized plain stop p- (Muốt *pə* for Viet . “vỏ,” “to peel”), and low-front vowel –a- (Chòi *rak* for Viet . “nước”; “water”). These characteristics are shared by all the Mường dialects shown above; yet as retentions, all they really demonstrate is the validity of a Vietnamese subgroup.

In Section 2.1, we did identify three valid innovations:

Rhoticization of PVM *d-:	*d-	→	r-
Spirantization of PVM *b-:	*b-	→	v-
Laryngealization of PVM *r-:	*r-	→	h-

Yet none of these were seen to be shared universally. Rather, Mường only appears to assume unity in converse with Vietnamese, and then only by lacking the innovations that substantively define Vietnamese. In summary, we found no innovations to be shared across Mường, but we did find three innovations distributed unevenly across the dialects.

It is, maybe, impossible to prove a negative—and indeed, the possibility remains open that further research will produce innovations shared among all or even the majority of Mường dialects. However, the major innovations governing Mường onsets visited here appear to contradict that theory. Rather, Mường appears to represent the collective descendants of an already diverse array of Proto-Việt-Mường varieties.

2.3 *Clades, taxa, and the “Mường” language*

Today, the Mường dialects appear to be mutually-intelligible, whereas Vietnamese does not appear to be mutually-intelligible with any of them. Does this represent evidence that Mường is a subgroup? Until recently, the existence of a “language” defined by the mutual intelligibility of its “dialects” has

⁷ For a more thorough treatment of these shared innovations, please see my dissertation (forthcoming).

constituted evidence *prima facie* of its status as an evolutionary subgroup (or clade). However, it is becoming increasingly clear that mutual intelligibility does not require the cladistical unity of a given group of linguistic varieties.

The possibility of convergence as an alternative form of language-formation has, for example, recently been proposed as playing a larger role in the formation of languages than previously imagined. Following Thomason & Kaufman's (1988) basic observation that "any linguistic feature" can be transferred from one language system to another, Braunmueller & House argued for a larger role for diffusion and convergence in processes of language formation than prevailing views which restricted such processes to the formation of creoles (Thomason & Kaufman, 1988; Braunmuller & House, 2009). Braunmuller & House' edited volume is preceded by a number of other works on language contact and convergence, including Winford's (2001) study of contact linguistics, Heine and Kuteva's *Language Contact and Grammatical Change*, and importantly, Aikhenvald & Dixon's *Areal Diffusion and Genetic Inheritance* (2007), which established a strong foundation for our understanding of the interaction of inheritance and borrowing in the formation of languages.

Regarding Mường, it is quite plausible that convergence may be playing a role in the formation of an evolutionarily cohesive Mường language. As noted above, Garrett (2006) forwards such an argument for the formation of Ancient Greek.

However, it is crucial to recognize that even if the "Mường dialects" are not converging, they still should not be considered an evolutionary clade. Subtracting convergence as a possibility (for the moment), the evident mutual intelligibility of the Mường "dialects" only demonstrates their affiliation as varieties of Proto-Viet-Muong. The retentions that bind together the Mường varieties discussed in this article, all stem directly from PVM features, and no shared innovations were found to point toward a "Proto-Mường" stage in their evolution. Thus, innovations like the laryngealization of PVM *r- in the Nàbái variety of Mường are indeed analogous to dialectal variation—but should be understood as dialectal variation of a Proto-Viet-Muong clade, rather than of a "Mường" language that descends from it.

By contrast, the drag-chain of initial consonant mutations which occurred in Vietnamese demonstrates a "Proto-Vietnamese" evolutionary clade, which subsequently diverged into the modern dialects.

How then, do we deal with a group of varieties that have maintained mutual intelligibility, but have never apparently undergone a stage of shared evolution? In other words, what do we call these dialects of Proto-Viet-Muong that were "left behind" by the diversification of Vietnamese? We cannot call

the Mường varieties dialects of Proto-Viet-Muong, since Proto-Viet-Muong by definition, includes the ancestor(s) of modern Vietnamese.

To solve this problem, Babel, Garrett, Houser, & Toosarvandai suggested the term *taxon*, which describes groups of dialects or languages whose shared features may either be retentions or innovations.⁸ *Taxa* thus are broader than “clades,” which correspond to our understanding of evolutionary subgroups proper, and must be defined by shared innovations. As noted above, I would like to propose borrowing the term *paraphyly* to describe taxa whose most recent shared ancestor is also the ancestor of an innovating clade no longer counted within the taxon.

The issue is not merely one of nomenclature. It calls attention to a very subtle assumption embedded in the field, namely that language diversification results in clean binary branching—a fact that is manifestly untrue. The term “paraphyletic taxon” allows us to recognize groups “left behind” by the innovation of a sister group, but which do not among themselves, represent any shared innovations at the corresponding generational level. In conclusion, Mường may be a language in terms of mutual intelligibility, but it is not a subgroup. It is a paraphyletic taxon.

3.0 Việt-Mường speciation

If the Mường language does not represent a cohesive subgroup, then we must conclude that it was Vietnamese which split off from Proto-Viet-Muong. The varieties of PVM “left behind” simply continued to evolve independently of one another, and it was these varieties that were eventually named “Mường” by early French ethnographers.

⁸ From their joint article “Descent and diffusion in language diversification: A study of Western Nomic dialectology,” to appear in *International Journal of American Linguistics*.

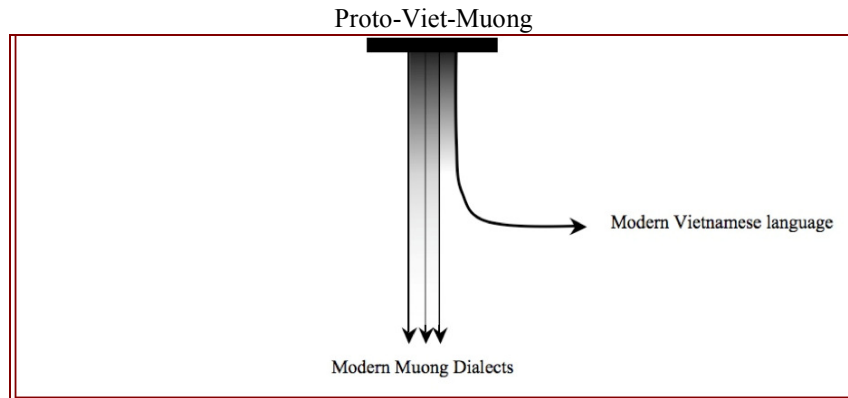


Figure 3. Model for Việt-Mường Diversification

Under this model, Vietnamese is properly understood as a subgroup, but Mường is not. Mường is represented as a paraphyletic taxon, rather than a single clade. All Mường dialects, as well as Vietnamese descend from an array of Proto-Việt-Mường dialects. However, a single lowland dialect or possibly a subset of them split off together, forming their own subgroup, and eventually evolved into modern Vietnamese. The remainder continued to evolve, but never mutually innovated into a new group (until, possibly, much, much later).

At this point, it may be important to recall Maspero's indirect claim that Mường, like Vietnamese, bears a heavily sinified lexicon. Perhaps due to the confusion discussed in 1.0, an unspoken assumption that Mường is the "un-sinicized" cousin of Vietnamese is not uncommon. It appears, however, that Maspero was correct.

Table 5. Examples of SV Grammatical Words:

#	字	Sino-Việt .	Sino-Mường (Khên) ⁹	Gloss
1.	在	tại	taj1	prep. “at”
2.	如	Như	ɲə1	“like, as”
3.	為	Vì	bi2	comp. “because”
4.	只	chỉ	ci5	“only”
5.	個	cái	kaj3	gen. classifier
6.	分	phần	fən2	classifier/“part”
7.	當	đương/đang	taŋ1	present/progressive marker
8.	來	lại	laj4	aspect marker
9.	實	thật	tʰət4	adv, “truly, really”
10.	每	mỗi	moj4	quantifier, “each”

As Table 5 shows, a large number of basic Sino-Vietnamese words also bear Sino-Mường counterparts. Note, furthermore, that the Mường tokens do not reflect the effect of the drag-chain diagrammed in Figure 4, and thus must have been borrowed directly from Chinese, rather than through an intermediate Vietnamese. This means that the speciation described in Figure 5 must have occurred after these words were borrowed, and thus, the innovations that produced a distinct Vietnamese language were not the immediate product of Chinese influence, as has sometimes been theorized. In Phan (2010), I proposed that a process of language shift might explain how such a large number of basic vocabulary found its way into PVM. If this was the case, then we may augment the speciation model in Figure 5 as follows:

⁹ All Muong Khên data is taken from L.C. Thompson’s *Proto-Viet-Muong Phonology*. Thompson, L. C. (1976). *Proto-Viet-Muong Phonology*. *Oceanic Linguistics Special Publications*, No. 13, *Austroasiatic Studies Part II*, 1113-1203.

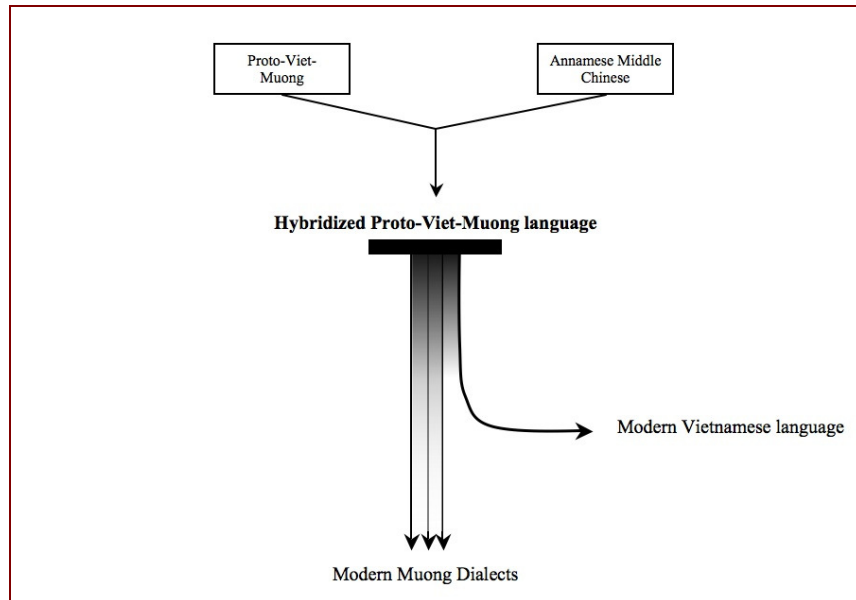


Figure 4. Complete Model for PVM Diversification

A more complete study of Mường, with particular attention to Sino-Mường, will allow us to pinpoint with more accuracy how much of the diversification process overlapped with the separate processes of borrowing and language shift that resulted in so sinified a complexion for both Mường and Vietnamese, as well as allow us to form subgroups within the Mường language, if these are to be found. In the meantime, it seems clear that Mường itself cannot be understood as a single clade, but rather the descendants of a diverse array of PVM dialects whose most recent common ancestor is shared with Vietnamese (i.e. PVM). In other words, Mường is not a subgroup, but a paraphyletic taxon.

REFERENCES

- Aikhenvald, A. Y., & Dixon, R. (2001). *Areal Diffusion and Genetic Inheritance*. Oxford: Oxford University Press.
- Babel, M. A. (forthcoming). Descent and diffusion in language diversification: A study of Western Nomic dialectology. *International Journal of American Linguistics*.
- Braunmuller, K., & House, J., editors. (2009). *Convergence and Divergence in Language Contact Situations*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Cuisinier, J. (1946). *Les Muong: geographie humaine et sociologie*. Paris: Institut d'Ethnologie.

- Diffloth, G. (1991). Vietnamese as a Mon-Khmer Language. Papers from the First Annual Meeting of the Southeast Asian Linguistics Society , 125-139.
- Ferlus, M. (1994). Contact ancien entre Viet-Muong et Austroasiatic nord. *Kristina Lindell Symposium on Southeast Asian Studies* , 1-4.
- Ferlus, M. (1986). Histoire abrégée de l'évolution des consonnes initiales du Vietnamien et du Sino-Vietnamien. *Mon-Khmer Studies*, no. 20 , 111-125.
- Ferlus, M. (1982). Spirantisation des obstruantes mediales et formation du systeme consonantique du vietnamien. *Cahier de linguistique orientale* , 11 (1), 83-106.
- Ferlus, M. (1992). Sur l'origine des langues Viet-Muong. *Mon-Khmer Studies XIVIII-XIX* , 52-59.
- Ferlus, M. (1975). "Vietnamien et proto-viet-muong." *Asie du Sud-Est et monde insulindien* 6, 4: 21-55.
- Garrett, A. (2006). Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In P. a. Forster, *Phylogenetic methods and the prehistory of languages*. Cambridge: University of Cambridge.
- Greenhill, S. J. (2009). Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In Adelaar, A., and Pawley, A. (editors) *Austronesian historical linguistics and culture history: a festschrift for Robert Blust* (pp. 375-397). Canberra: Pacific Linguistics.
- Haudricourt, A. (1954). De l'origine de tons en vietnamien. *Journal Asiatique*, v. 242 , 69-82.
- Heine, B., & Kuteva, T. (2005). *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Lewis, M. P. (2009). *Ethnologue, SIL International*. Retrieved 2010 йил 16-July from Ethnologue: Languages of the World, Sixteenth Edition: http://www.ethnologue.com/show_language.asp?code=mtq
- Maspero, H. (1912). Etudes sur la phonétique historique de la langue annamite, Les initiales. *BEFEO* 12 , 12.1-126.
- Nguyễn, V. T. (1982). Ngữ âm tiếng Mường qua các phương ngôn. Hanoi.
- Phan, J. D. (2010). Re-Imagining Annam: A New Analysis of Sino-Viet-Muong Linguistic Contact. *China Southern Diaspora Studies* , 4, 3-25.
- Rischel, J. (1995). *Minor Mlabri: A hunger-gatherer language of northern Indochina*. Copenhagen: Museum Tusulanum Press, University of Copenhagen.
- Sidwell, P. (2005). Proto-Katuic Phonology and the Subgrouping of Mon-Khmer Languages. In P. Sidwell (Ed.), *Papers from the 15th Meeting of the Southeast Asia Linguistics Society. XV. SEALS*.
- Sidwell, P. (2010). The Austroasiatic Central Riverine Hypothesis. *Journal of Language Relationship* , 4, 117-134.
- Taberd, J. (1838). *Dictionarium Anamitico-Latinum*. Serampore: Marshman.

- Taylor, K. W. (2001). On Being Muonged. *Asian Ethnicity, Volume 2, Issue 1*, 25-34.
- Thomason, S. G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley & Los Angeles: University of California Press.
- Thompson, L. C. (1976). Proto-Viet-Muong Phonology. *Oceanic Linguistics Special Publications, No. 13, Austroasiatic Studies Part II*, 1113-1203.
- Winford, D. (2003). *An Introduction to Contact Linguistics*. Malden, Massachusetts: Blackwell Publishing.

Received: 2 November 2010

Cornell University
Ithaca, USA
E-mail address: jdp49@cornell.edu

Notes on Ten-edn (Tonga-Mos) and Kensiw Borrowings

Mary M. PETERSON

SIL International

Abstract

Ten-edn (Tonga-Mos) and Kensiw are both Northern Aslian languages. Ten-edn speakers are found only in Thailand; Kensiw speakers are located in both Thailand and Malaysia. The speakers of these two languages have borrowed words from the local national speech varieties for concepts and objects outside of their traditional foraging culture. While both languages share a number of borrowings that have Malay origins, the Ten-edn in recent years have borrowed either from the local Thai language variety or have coined new words in their language. The Kensiw, however, have continued to borrow words from the local Malay varieties. This paper explores the origins of the Malay borrowings in both languages and also presents the strategies used by the Ten-edn for naming concepts or objects that are not part of their culture.

Keywords: Ten-edn, Kensiw, Borrowing

Introduction

The Ten-edn [ten'ɛdn] (Tonga-Mos¹) and Kensiw languages belong to the Northern Aslian branch of the Aslian group of the Mon-Khmer language family, a subdivision of Austroasiatic. Ten-edn speakers, numbering around 200 people, are located in the Thai provinces of Trang, Satul, and Phatthalung; Kensiw speakers are located in Yala Province² (see Figure 1).

¹ Various names have been assigned to this language. The Maniq themselves call their language *Ten-edn*. There are other variations of spelling of this language name: *Ten-en* Diffloth (2002) and *Tean-ean* Phaiboon (2006). The language name *Tonga (Mos)* is the name given by linguists and/or anthropologists to the languages spoken in Trang, Satul and Phatthalung: Bradley and Benjamin (1986), Evans (1937), Schmidt (1901), Schebesta (1928) and Skeats and Blagden (1906).

² From 1989-1996 there were 20-50 Kensiw in Yala. After 1996 most moved to Malaysia. By 2005 there were only 2 speakers left in the village.

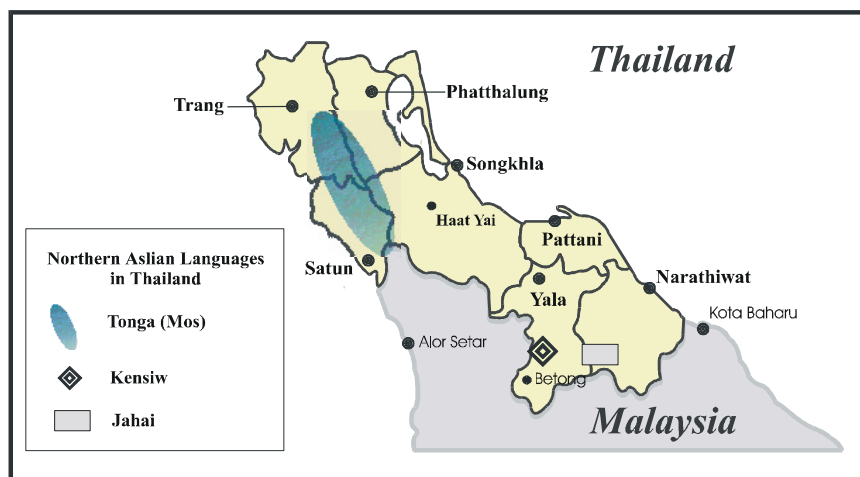


Figure 1. Location of Northern Aslian Languages in Thailand

The speakers of these two languages belong to an ethnic group known in Thailand as *Sakai*, *khon paa*, or *ngok paa*, ‘the forest people’. In Malaysia and in academic circles, the people are known as the Negritos, a subgroup of the Orang Asli (‘Original People’). The people refer to themselves as *Maniq*. They, like the Orang Asli in general, are considered the original inhabitants of Peninsular Malaysia and Thailand. In their traditional culture, the Maniq are foragers, hunting and gathering within a defined territory in the forest to meet their daily needs and living in temporary shelters. For simplicity in these notes, I will use Maniq to refer to the people group and Ten-edn and Kensiw to refer to the languages.

From the outset of my research among the Kensiw speakers in Yala Province in 1989, I have been intrigued by the loan words found in their language. They use loan words for concepts and things outside of their traditional foraging culture. As my research expanded to include the Ten-edn speakers found in Trang, Phatthalung and Satun, it became clear that while some of the same loan words were used in Kensiw, there were a number of other borrowings that were completely different in Ten-edn. In both cases, the speakers have borrowed words for things outside their traditional culture from the local speech varieties of the country where they are located or have the closest contact.

Throughout their history, the Kensiw speakers in Yala have resided as far north as Bannang Sata, Yala Province in Thailand and as far south as the states of Kedah and Perak in Malaysia. In Baansakai, Thanto District, Yala, the Kensiw speakers are surrounded by a majority of Malay speakers along with a smaller number of Southern and Central Thai speakers. The Kensiw speakers in Yala also have many relatives living in Malaysia and have regular contact with them. Historically, Malay has been the lingua franca in those areas, so it is

not surprising that they borrow readily from the local variety of Malay in either Thailand or Malaysia.

On the other hand, the Ten-edn speakers are only located in Thailand, the majority of them living in and around the Khao Bantat Mountain range in Trang, Phatthalung and Satul Provinces. They are living in small bands of 5–25 people in the forest and move frequently. All of their relatives live within the boundaries of Trang, Satul, Phatthalung and Songkhla Provinces³. They have limited contact with outsiders who speak either Central or Southern Thai. Thus, the Ten-edn speakers tend to borrow words from both Southern and Central Thai.

Both languages, Kensiw and Ten-edn, have a number of lexical items which are a part of their traditional culture but clearly have Malay origins. These lexical items do not appear to represent new or foreign concepts. It would seem that the speakers of these two languages must have had lengthy contact with Malay speakers in the past (Evans 1937:289). Historically, the area in which these two groups live was under the Kingdom of Pattani ruled by Malay sultans and the sultanate of Kedah⁴. Malay was the *lingua franca* of this area.

For the Ten-edn speakers, I would speculate that these Malay borrowings stem from the contact they had with Malay speakers during the time of the early European explorers of the Malay Peninsula in the 1700–1800s. They were also subjects of the Malays in Pattani and Songkhla Provinces in the 1800s (Brandt 1961:124). References dating back as far as 1924 show that the Maniq of Trang and Phatthalung were reported to be living totally apart from the rest of the Maniq (Negritos) (Schebesta 1928:15). Subsequent reports also documented their separation from the rest of the Maniq and from contact with Malay speakers (Brandt 1961:136). At one time there were Maniq (Negritos) living in Perlis, Malaysia, adjacent to Satul Province, Thailand; however, they are reported to have disappeared from Perlis before the 1900s (Brandt 1961:124). It is quite clear that since the early 1900s, the Ten-edn speakers have lived solely in Thailand, limiting their territory to Trang, Satul, Phatthalung and Songkhla Provinces.

³ In 2001–2003, I mapped out where relatives lived in relation to one another in order to see how much contact the various bands of Ten-edn speakers have had with each other. All the relatives of those interviewed are living in the Trang, Satul, Songkhla and Phatthalung areas. I know of only five Ten-edn speakers from Phatthalung who migrated to Malaysia in the mid-1990s, along with their relatives who used to live with the Kensiw speakers of Yala. They have not moved back to Phatthalung (Benjawan 2011).

⁴ The Kingdom of Pattani included the modern day Thai provinces of Pattani, Narathiwat, Songkhla, and Yala. Satul was originally under the Malay Sultanate of Kedah (Lazacode.Com Online Education and Creative Reference 2011)

Between 1992 and 2003, I took part in three language surveys among the Ten-edn speakers in Trang, Satul and Phatthalung. The data were gathered from 10 different band locations in these three provinces. A 400-item word list was used. The Ten-edn speakers used borrowed terms for 57 items, using various strategies for borrowing objects or concepts that are not a part of their traditional culture.

The examples given in these notes are written phonetically, using the IPA system. The phonetic inventory of sounds in Ten-edn are similar to those used in Kensiw⁵ as seen in Table 1 and 2. Primary stress and secondary stress have been marked, as in [bi'ʔeh] 'money' and [tə ,ma'gaʔ] 'silver'.

Table 1. Ten-edn and Kensiw Vowel Phones

Tongue Height	FRONT		CENTRAL		BACK	
	Oral	Nasal	Oral	Nasal	Oral	Nasal
close	i	ĩ	u	ũ	u	ũ
near-close	ɪ	ĩ				
close-mid	e̞	ẽ	ə̞		ɔ̞	õ̞
mid	e	ẽ	ə		o	õ
open-mid	ɛ	ẽ	ʌ	ã	ɔ	õ
open			a	ã		
Diphthongs	ie	ie				

Table 2. Ten-edn and Kensiw Consonant Phones⁶

		bilab.	alveo.	retro	palat.	velar	glot.
M	Stop, vl.	p	t		c	k	ʔ
	Stop, vl. asp.	(p ^h)	(t ^h)			(k ^h)	
A	Stop, vd.	b bm	d dn		ʃ ʃɲ	g gɲ	
N	Nasal	m	n		ɲ	ŋ	
N	Fricative, vl	ɸ	s				h
E	Fricative, vd.					ɣ	
R	Lat. Approximant		l		ʎ		
	Cent. Approximant	w (ʌ)		ɹ r r	j		

⁵ Phonological information is taken from Bishop (1996).

⁶ Stops. Syllable initially, [b-, d-, ʃ-, g-] are realized as voiced bilabial, alveolar, palatal and velar stops, respectively. In the syllable final position, however, these stops are frequently realized as prestopped nasals [-bm, -dn, -ɲɲ, -gɲ] in which the final nasal assimilates to the same point of articulation as the stop. Aspirated stops are found mostly in borrowed words.

Alveolar Lateral Approximate. [l] is used syllable initial and the palatal phone [ʎ] is found syllable final.

Central Approximants. The voiced retroflexed [ɹ] is found in either syllable initial or syllable final position. The flap, [r], occurs only in syllable initial clusters. The trill [r] is found only in Kensiw for the Malay word for rabbit. (See Bishop (1996) for a short discussion regarding the central approximates in borrowings from Malay.)

Borrowing Strategies of Ten-edn Speakers

It appears that the Ten-edn speakers use several strategies when they come across an object or concept that is not a part of their traditional culture. These strategies include:

1) the use of the borrowed word without any changes, e.g.

‘to buy’ sú: Thai: sú:

2) the application of Ten-edn phonology to the borrowed word⁷, e.g.

‘to buy’ suh Thai: sú:

3) the use of the borrowed word in combination with a Ten-edn word, e.g. ‘to buy’ and ‘to pay’

‘to buy’	bah su	‘go’ ‘buy’	Thai: sú:
‘to pay’	?ek bi’?eh	‘give’ ‘money’	Thai: bí:a ‘silver/money’

4) the use of a description composed of words exclusively from their own language. It can be either a phrase or a single word, e.g.

‘to buy’	bah gedn	‘go’ ‘get’
‘to pay’	?ek	‘give’

5) the use of one lexical item for 1) identifying the object and 2) identifying the substance the object is made of. For example, the word *boʌ* is used for both the object ‘hooked knife’ and ‘iron’, the substance the hooked knife is made of.

‘iron’	boʌ
‘hooked knife’	boʌ

6) the use of an attribute of the borrowed term, as in ‘ginger’ being ‘spicy hot’, e.g.

‘ginger’ pa'dih ~ pə'dij ‘spicy hot’ Malay ‘pedas’

⁷ The application of Ten-edn phonology varied among the different speakers, i.e. some closed the vowel at the end of a word with a consonant while others did not, e.g. *su* ~ *suh* ‘to buy’. Likewise, most Ten-edn speakers dropped tones from the Thai words.

The remainder of these notes are examples of the borrowings used in both Ten-edn and Kensiw. I have included the Kensiw data to show the contrast in the sources of the borrowed words.

The data are given in tables arranged by semantic domains. The Ten-edn words displayed are representative of 10 different speakers from the 10 locations surveyed⁸. They are ordered by the type of strategy used to borrow the word. (See above.) The Kensiw words are given last. (See Table 3. Organization of the Tables).

In the tables, the elicited Ten-edn and Kensiw words are given first, followed by their glosses. If the Ten-edn word was borrowed from Thai or Malay, the Thai or Malay word is given with its gloss. Phonetic forms are given in the Thai glosses. Unless otherwise noted, Central Thai has been used. Malay glosses (in Malay orthographic form) are predominately from northern Malay.

Table 3. Organization of the Tables

to buy				
	Ten-edn	Gloss	Thai	Malay
Strategy 1	sú:	'buy'	sú: 'buy'	
Strategy 2	suh	'buy'	sú:	
Strategy 3	bah su	'go' 'buy'	sú:	
	su: ?ək	'buy' 'give'	sú:	
Strategy 4	bah gedn	'go' 'get'		
Kensiw				
	bə'li?	'buy'		beli 'buy'

Exchange of Money: Buying, Selling, Earning

Traditionally the Maniq do not have a concept of money or the exchange of money in their culture. It should be noted that the Ten-edn seem to use the Thai words for 'to buy', 'to sell' interchangeably as seen in both Tables 4 and 5. Their exclusive use of Thai terms implies that their economic contact has been with the Thais and not with the Malay speaking people in Thailand or Malaysia. In contrast, the Kensiw have borrowed *beli* from Malay, implying their economic contact was with the Malays.

⁸ When all 10 speakers gave the same response, the word is listed once. If there was more than one response, each word/phrase is listed. Thus, the number of Ten-edn entries varied between items.

Table 4. Buying—Paying

to buy			
Ten-edn	Gloss	Thai	Malay
sú:	'buy'	sú: 'buy'	
suɦ		sú:	
bah su	'go' 'buy'	sú:	
su: 'ʔek	'buy' 'give'	sú:	
bah gedn	'go' 'get'		
Kensiw			
bə'liʔ	'buy'		beli 'buy'
to pay			
Ten-edn	Gloss	Thai	Malay
ka'caj	'pay'	k ^h á: te ^h áj teá:j 'expense'	
suɦ	'buy'	sú: 'buy'	
bah su	'go' 'buy'	sú:	
ʔek bi'ʔeh	'give' 'money'	bî:a 'silver/money' ⁹	
ʔek	'give'		
Kensiw			
ba'ja	'pay'		bayar 'to pay'

Table 5. Selling—Earning

to sell			
Ten-edn	Gloss	Thai	Malay
k ^h aj	'sell'	k ^h ǎ:j 'sell'	
bah k ^h aj	'go' 'sell'	k ^h ǎ:j	
bah su	'go' 'buy'	sú: 'buy'	
k ^h aj ʔek	'sell' 'give'	k ^h ǎ:j	
bəh ʔek	'go' 'give'		
ʔek	'give'		
Kensiw			
ceʌ	'sell'		jual 'sell'
to earn money			
Ten-edn	Gloss	Thai	Malay
gedn bi'ʔeh	'get' 'money'	bî:a 'silver/ money'	
tə'bas hac jik bi'ʔeh	'cut undergrowth' 'knife' 'receive' 'money'	bî:a	
tə'baʔ gedn bi'ʔeh	'cut undergrowth' 'get' 'money'	bî:a	
ʔam'jit lə'waj	'received' '(from) outside'		

⁹ Southern Thai

sə'wej bi'ʔeh	'outside' 'money'	bɪ:a
sə'wej jit	'outside' 'receive'	
Kensiw		
ko'cit	'receive'	

Clothing

The Maniq have traditionally made their clothing from jungle materials, using bark cloth and various types of plants. These natural forest materials were replaced over time with cloth (Evans 1937:71-73). With the introduction of cloth to their culture, the Ten-edn speakers use both Thai and Malay words. It is interesting to note that most of the clothing borrowings have come from Malay, presumably from the time they were subject to the Malay sultanates, as noted above. As the Ten-edn speakers have had contact with the Thais as well, they have added Thai words to their lexicon and use both the Thai and Malay words interchangeably, especially for the words for blanket and men's and women's sarongs. The word *ga'ʔet* is derived from the Malay word *kain* 'cloth' (Phillips 2010).

Table 6. Clothing

men's loincloth			
Ten-edn	Gloss	Thai	Malay
ja'wat ~ jadn'wat	'men's loincloth'		cawat loincloth'
Kensiw			
ja'wat	'men's loincloth'		cawat loincloth'
male (m.) sarong			
Ten-edn	Gloss	Thai	Malay
p ^h a'k ^h aw	'sarong m.'	p ^h â: k ^h â:w má: 'sarong m.'	
pa'hom kaw'ma	'cloth cover' 'sarong m.'	p ^h â: hòm 'cloth cover' p ^h â: k ^h â:w má:	
'p ^h a 'hom ga'ʔet	'cloth cover' 'cloth'	p ^h â: hòm	kain 'cloth'
ga'ʔet 'p ^h a:'hom	'cloth' 'cloth cover'	p ^h â: hòm	kain
Kensiw			
sap'ʔâp	'length of fabric'		
female (f.) sarong			
Ten-edn	Gloss	Thai	Malay
p ^h a't ^h uŋ	'sarong f.'	p ^h â: t ^h ũŋ 'sarong f.'	
ga'ʔet	'cloth'		kain 'cloth'
ga'ʔet 'p ^h a't ^h uŋ	'cloth' 'sarong f.'	p ^h â: t ^h ũŋ	kain
ga'ʔet 'kɔŋŋ	'cloth' 'woman'		kain
ba'juʔ	'shirt'		baju 'shirt'
Kensiw			
ko'mogŋ	'sarong f.'		

blanket

Ten-edn	Gloss	Thai	Malay
p ^h a'hom	'blanket'	p ^h â: hòm 'blanket'	
ha'mɔʔ ~ ka'mɔʔ	'cover with cloth'	hòm 'cover'	
ga'ʔet	'cloth'		kain 'cloth'
ga'ʔet ta'lej	'cloth' 'swaddle'		kain
Kensiw			
suλ	'blanket'		selimu

trousers

Ten-edn	Gloss	Thai	Malay
sə'waλ ~ sə'waʔ ¹⁰	'trousers'		seluar 'trousers'
Kensiw			
sə'weλ	'trousers'		seluar
Nature			

The Maniq have a rich lexicon to identify and describe their natural world. For the Ten-edn speaking Maniq, the names for identifying natural items that have been introduced more recently into their territory, for the most part, have been borrowed from Thai, e.g. iron, papaya, kapok, etc. The Kensiw have typically borrowed these words from Malay.

However, for both the Ten-edn and Kensiw speakers, there is a small inventory of items: 'crocodile', 'fish', 'rain', 'lightning', 'betel nut', 'banana', 'sand', 'stone' and 'cave' that are indigenous to their traditional territory. These words are clearly borrowed from Malay. Tables 7-13 give examples of borrowings by the Ten-edn for various domains of nature.

Table 7. Metals

Ten-edn	Gloss	Thai	Malay
lek	'iron'	lèk 'iron'	
boλ	'hooked machete'		
ka'tɔk	'knife' ¹¹		
Kensiw			
bə'siʔ	'iron'		besi 'iron'

¹⁰ There is a consistent $\lambda > j$ sound change between some of the groups in Trang, Satul and Phatthalung.

¹¹ knife used to harvest rice

gold			
Ten-edn	Gloss	Thai	Malay
tʰɔŋ	‘gold’	tʰɔ:ŋ ‘gold’	
saj ‘kʰɔ tʰɔŋ	‘necklace’ ‘gold’	sǎ:jkʰɔ: ‘necklace’	
saj ‘kʰɔ man ‘kɔʔ ʔa ‘wi	‘necklace’ ‘type of necklace’ ‘necklace string’	tʰɔ:ŋ sǎ:jkʰɔ: ‘necklace’	
Kensiw			
mas	‘gold’		mas ‘gold’
silver			
Ten-edn	Gloss	Thai	Malay
bi ‘jɛh	‘silver’	bî:a ‘money/silver’	
bi ‘jɛh ha ‘liʔ	‘silver’ ‘leaf’	bî:a	
dɔk biə	‘silver’ ¹²	dò:k bî:a ‘(bank) interest’	
Kensiw			
tə-ma ‘gaʔ	‘metal’		tembaga ‘metal’

Table 8. Climate

rain			
Ten-edn	Gloss	Thai	Malay
ʔu ‘jadn ~ʔu ‘jadn	‘rain’		hujan ‘rain’
Kensiw			
ʔu ‘jadn	‘rain’		hujan
lightning			
Ten-edn	Gloss	Thai	Malay
ki ‘lat	‘lightning’		kilat ‘lightning’
Kensiw			
ki ‘lat	‘lightning’		kilat

Table 9. Edible Plant—Fruits¹³

papaya			
Ten-edn	Gloss	Thai	Malay
‘luk ‘kɔ	‘fruit’ ‘papaya’	lû:k máʔláʔkɔ: ‘papaya’	
ʒɛj ha ‘tīt la ‘çɛç	‘buttock’ ‘tail’ ¹⁴ ‘buttock’		

¹² Literally *dok* ‘flower’ *biə* ‘silver’. Probably a description of a coin rather than the financial concept of ‘bank interest’ as the Thai’s use this phrase.

¹³ Red pepper and papaya are originally from the Americas.

¹⁴ The Maniq perceived the round part of the fruit as the buttock and the tail as the stem that attaches it to the tree.

Kensiw			
b ^h ʌ'tɛʔ	'papaya'		buah betik 'papaya'
ginger			
Ten-edn	Gloss	Thai	Malay
k ^h iŋ	'ginger'	k ^h iŋ 'ginger'	
mə'taŋ ka'la	'stalk' 'torch ginger'	tôn da:lă: 'torch ginger'	
hait	'ginger'		halia 'ginger'
ka't ^h u	'rhizome'		
pa'dih	'spicy hot'		pedas 'spicy hot'
Kensiw			
kə'cɔj:	'type of ginger'		
red pepper			
Ten-edn	Gloss	Thai	Malay
p ^h rik	'red pepper'	p ^h rik 'red pepper'	
la'daʔ	'red pepper'		lada 'red pepper'
pə'dij	'spicy hot'		pedas 'spicy hot'
Kensiw			
ja'bɛ	'red pepper'		cabai 'red pepper'
banana			
Ten-edn	Gloss	Thai	Malay
pi'sɛŋ	'banana'		pisang 'banana'
Kensiw			
pi'sɛŋ	'banana'		pisang
<i>Table 10. Betel Nut</i>			
betel nut			
Ten-edn	Gloss	Thai	Malay
pi'naŋ	'betel nut'		pinang 'betel nut'
Kensiw			
pi'naŋ	'betel nut'		pinang
betel leaf			
Ten-edn	Gloss	Thai	Malay
si'ɛh	'betel leaf'		sirih 'betel leaf'
p ^h lu	'betel leaf'	baj p ^h lu: 'betel leaf'	
Kensiw			
ha'liʔ si'ɛh	'leaf' 'betel'		sirih

lime used with betel nut

Ten-edn	Gloss	Thai	Malay
pun	'lime for betel nut'	pu:n kin māk 'lime for betel nut'	
ka'po	'lime for betel nut'		kapur 'lime for betel nut'
Kensiw			
ka'p ^w a	'lime for betel nut'		kapur

Table 11. Grass—Trees¹⁵**grass**

Ten-edn	Gloss	Thai	Malay
nə	'grass'	jā: 'grass'	
sebm	'dry leaves'		
ha'li?	'leaf'		
ha'du?	'jungle underbrush'		
Kensiw			
jum'pōt	'grass'		ramput 'grass'

kapok

Ten-edn	Gloss	Thai	Malay
nun	'kapok'	nūn 'kapok'	
pa'sun	'kapok fiber'		
ka'bə-? nun	'pod'	nūn	
k ^h waj nun ~ k ^h waj	'cotton' 'kapok'	fā:j 'cotton'	
bə'sə ~ bu'səh	'fiber'		
Kensiw			
ga'bu	'kapok'		kapuk 'kapok'

rubber

Ten-edn	Gloss	Thai	Malay
jaŋ	'rubber'	ja:ŋ 'rubber'	
kaŋ'jaŋ	'tree rubber'	ja:ŋ	
ka'tēk	'rubber tree'		getah 'rubber'
kaŋ'tōŋŋ	'tree rubber'		
Kensiw			
kə'tēk	'tree rubber'		getah

Table 12. Animals

cat

Ten-edn	Gloss	Thai	Malay
məw	'cat'	mə:w 'cat'	
ku'jiŋŋ			kucing 'cat'

¹⁵ Kapok and rubber are non-native trees, introduced only a few generations ago.

Kensiw			
ku'jigŋ	'cat'		kucing
fish			
Ten-edn	Gloss	Thai	Malay
ʔi'kadn	'fish'		ikan 'fish'
Kensiw			
KNS: ʔi'kadn	'fish'		ikan
crocodile			
Ten-edn	Gloss	Thai	Malay
bu'jaʔ	'crocodile'		buaya 'crocodile'
Kensiw			
bu'jaʔ	'crocodile'		buaya
rabbit			
Ten-edn	Gloss	Thai	Malay
kə'taaj	'rabbit'	kràʔ tà:j 'rabbit'	
ka'toŋŋ	'rat'		
ʔa'tinʔ la'jet	'ear' 'long'		
Kensiw			
ʔar'nap	'rabbit'		jarnab 'rabbit'

Table 13. Geography

sand			
Ten-edn	Gloss	Thai	Malay
pa'si	'sand'		pasir 'sand'
Kensiw			
pa'si	'sand'		pasir
stone			
Ten-edn	Gloss	Thai	Malay
ba'tuʔ	'stone'		batu 'stone'
Kensiw			
ba'tuʔ	'stone'		batu
cave			
Ten-edn	Gloss	Thai	Malay
gu'haʔ	'cave'		gua batu 'cave'
ha'negŋ	'cave' (not rock)		
Kensiw			
gu'wa	'cave'		gua batu

Miscellaneous Words

In Table 14 there are samples of other words from different domains that have been borrowed. Again, they represent items that are not found within

their traditional culture and show that the Ten-edn speakers use Thai for these items.

Table 14. Miscellaneous

liquor			
Ten-edn	Gloss	Thai	Malay
law	'liquor'	lâw 'liquor'	
nam 'law	'water' 'liquor'	ná:m 'water'	
		lâw 'liquor'	
kɔ̄'le	'shot glass'		
bʌ'tɛw kə'muʔ	'water' 'vomit'		
Kensiw			
ke'hɛʌ	'liquor'		
medicine			
Ten-edn	Gloss	Thai	Malay
ja	'medicine'	ja: 'medicine'	
bə'lec	'forest medicine'		
ka'muʔ	'vomit' ¹⁶		
kʌ'siŋ dʌk'jo	'tuber' 'medicinal herb'		
Kensiw			
hu'bet	'medicine'		ubat 'medicine'
needle			
Ten-edn	Gloss	Thai	Malay
k ^h em	'needle'	k ^h ěm 'needle'	
'lek	'iron'	lèk 'iron'	
Kensiw			
ja'jubm	'needle'		jarum 'needle'
comb			
Ten-edn	Gloss	Thai	Malay
ʌi ~wi	'comb' (plastic)	wǐ: 'comb'	
kʌdn'ʔac	'comb' (bamboo)		
Kensiw			
ka'nat	'comb' (bamboo)		

¹⁶ They have a distinction between manufactured medicines Thai: [ja:] and the medicines they find in the jungle. When they take manufactured medicines they become dizzy and vomit, hence the use of the word 'vomit' for manufactured medicines.

roof			
Ten-edn	Gloss	Thai	Malay
lǎŋ kʰa	'roof'	lǎŋ kʰa: 'roof'	
ka'pign ha'jǎʔ ha'poj	'top' 'lean-to', 'Yellow Earth Ginger' ¹⁷		
ka'pinj ha'pəj	'top' 'Yellow Earth Ginger		
ka'pinj sɔ'loŋ	'top' 'middle'		
Kensiw			
ja'bogŋ	'on top'		

REFERENCES

- Benjawan Srithanto. 2011. Personal communication.
- Bishop, Nancy M. and Mary M. Peterson. 1993. *Maniq Language Survey Report*. Unpublished manuscript. Bangkok: Thammasat University.
- _____. 2003. Northern Aslian Language Survey: Trang, Satul and Phatthalung Provinces, Thailand. Unpublished manuscript. Bangkok: Thammasat University.
- Bishop, Nancy M. 1996. A preliminary description of Kensiw (*Maniq*) phonology. *Mon-Khmer Studies*: 25:227-253.
- Bradley, David and Geoffrey Benjamin. 1986. Language atlas of China. *Pacific Linguistics Series C. No. 102*. New York: Longman.
- Brandt, John H. 1961. The Negrito of peninsular Thailand. *Journal of the Siam Society* XLIX:122-60.
- Diffloth, Gérard. 2002. The Northern Aslian Languages within the Mon-Khmer Family. Paper presented 14 June at "The First Inter-Dialogue Conference on Southern Thailand", Prince of Songkhla University, Pattani, Thailand (June 13-15, 2002).
- Evans, Ivor H.N. 1937. *The Negritos of Malaya*. Cambridge: Cambridge University Press.
- Gregerson, Kenneth. 2010. Personal communication.
- History of Malaysia: Bangkok Treaty of 1909. *Lazacode.Com Online Education and Creative Reference*. 13 January 2011. <http://lazacode.com/event-place/history-of-malaysia-bangkok-treaty-of-1909> 2010 lazacode.com.
- Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the World*. Sixteenth edition. Dallas, TX: SIL International Online version. <http://www.ethnologue.com/>.
- Phai boon Duangchan. 2006. "Glossary of Aslian Languages: the northern Aslian languages of southern Thailand", in *The Mon-Khmer Studies Journal*, vol. 36, pp. 207-224.

¹⁷ They use the leaf of the Yellow Earth Ginger (*Achasma macrocheilos*) on top of their lean-tos.

- Phillips, Timothy. 2010. Personal communication.
- Schebesta, P.P. 1928. *Among the Forest Dwarfs of Malaya*. Hutchinson & Co.Ltd. Re-printed 1973 Kuala Lumpur: Oxford University Press.
- Schmidt, P.W. 1901. Die Sprachen der Sakei und Semang auf Malacca und ihr Verhältnis zu den Mon-Khmér-Sprachen. *Bijdragen tot de Taal-, Land- en Volkendunde* 52:399-583.
- Skeats, Walter Wm. and Charles Otto Blagden. 1906. *Pagan Races of the Malay Peninsula, Volume II*. London: MacMillian & Co.
- Williams-Hunt, P.D.R. 1952. *An Introduction to the Malayan Aborigines*. Kuala Lumpur: The Government Press.

Received: 14 July 2011

SIL International
USA

E-mail address: marym_peterson@sil.org

A description of Jiēyáng Hakka phonology as spoken in Bangkok, Thailand

NITASAKORN Shiwaruangrote
Mahidol University, Thailand

ABSTRACT

This manuscript is the study of the phonological system of Jiēyáng 揭陽 Hakka dialect as spoken in Thailand (hereafter JHT). The main focus has been on phonemes (consonants, vowels, and lexical and grammatical tones) together with syllable structures. In addition, the study includes phonological processes of consonants and vowels, e.g., assimilation, epenthesis, and variation. Tone change which is the historical process to derive the meaning of words is also briefly described. Tone sandhi, which is complicated in the dialect, is not included in this manuscript.

Key words: Hakka Chinese, phonology

1 Introduction

1.1 *The Hakka language*

The Hakka dialect family is a variety of Chinese which belongs to the Sinitic branch of the Sino-Tibetan language family. The Hànyǔ 汉语 branch contains seven major dialect families that are mutually unintelligible. These dialect families are as follows:

1. Guān huà 官话 or Běifāng huà 北方话;
 2. Wú 吴 (of which Sūzhōu 苏州 is a member);
 3. Xiāng 湘 (of which Chángshā 长沙 is a member);
 4. Gàn 赣 (of which Nánchāng 南昌 is a member);
 5. Yuè 粤 (of which Cantonese is a member);
 6. Mǐn 闽 (of which Cháo zhōu 潮州 is a member);
 7. Kèjiā 客家 or Hakka (of which Méixiàn 梅县 is a member).
- (Ramsey 1989:87, Norman 1988:181).

The word Hakka is derived from the spelling and pronunciation of the Chinese characters 客家 [k^hak^{35→11}.ka⁵²] (or [hak^{35→11}.ka⁵²] according to the dialect). The word [k^hak³⁵] 客 means ‘guest, visitor’. The word [ka⁵²] 家 can be glossed by various meanings. Firstly, it means ‘house or family’. However, the common word for ‘house’ in Hakka language is [vuk³⁵] 屋, not [ka⁵²] 家. The word [ka⁵²] 家 is used mostly as a prefix or suffix in compound words, for example, [ka⁵².ŋi:n³⁵] 家人 ‘members of the house’, [vuk^{35→11}.ka⁵²]¹ 屋家 ‘house, home’. Secondly, [ka⁵²] 家 can be used as a plural marker to mark a group of people, as in the word “Hakka”. So, the meaning of Hakka is ‘guest group, guests, visitors, or newcomers’.

The Hakka people called themselves [k^hak^{35→11}.ŋi:n³⁵] 客人. The word [k^hak³⁵] 客 means ‘guest, visitor’ and the word [ŋi:n³⁵] 人 means ‘man, people’. When these two words are put together as above it means ‘Hakka people, the people who speak Hakka language’. Hakka people call their language [k^hak^{35→11}.vɔi⁵²] 客话 and call China [t^hɔ:ŋ^{35→33}.sa:n⁵²] 唐山.

There are two other terms [t^hɔ:ŋ^{35→33}.ŋi:n³⁵] 唐人 and [t^hɔ:ŋ^{35→33}.vɔi⁵²] 唐话. Both are general terms that the Hakkas use to call Chinese people of all dialects, and all Chinese languages, respectively. Non-Chinese are called [fa:n⁵²] 番 ‘barbarian’. So, [fa:n⁵².ŋi:n³⁵] 番人 means ‘non-Chinese, barbarian people’, and [fa:n^{52→33}.vɔi⁵²] 番话 means ‘non-Chinese language, language of the barbarians’. Outside of China these two words refer to native people and the native language, such as in Thailand [fa:n⁵².ŋi:n³⁵] 番人 means ‘Thai people’ and [fa:n^{52→33}.vɔi⁵²] 番话 means ‘Thai language’; while in Cambodia the words mean ‘Khmer people’ and ‘Khmer language’, respectively.

1.2 The Hakka people in Thailand

Among the Hakka people in Thailand, the people and language are commonly classified into two groups: [pa:n³¹.sa:n⁵².k^hak³⁵] 半山客 and [ts^hi:m⁵².k^hak³⁵] 深客. The differences between these two groups are as follows: in mainland China the [pa:n³¹.sa:n⁵².k^hak³⁵] 半山客 people live halfway up the mountain, while [ts^hi:m⁵².k^hak³⁵] 深客 live higher up. In addition, the word [pa:n³¹.sa:n⁵².k^hak³⁵] 半山客 is used by Hakka people to refer to their language as a mixture of Hakka and Cháo zhōu 潮州 elements, such as Jiēyáng 揭阳 dialect, while the word [ts^hi:m⁵².k^hak³⁵] 深客 is used to refer to the dialect which is considered to be more classical than the [pa:n³¹.sa:n⁵².k^hak³⁵] 半山客 group, e.g., Méixiàn 梅县.

¹ According to the analysis of Lau (2000:26) this word is derived from [vuk^{35→11}.ha⁵²] 屋下 by means of progressive assimilation where the initial [h] of the word [ha⁵²] 下 assimilates totally with the final [k] of the word [vuk³⁵] 屋. In JHT, [vuk^{35→11}.ha⁵²] 屋下 means ‘the elevated space between the ground and the first floor’ which is opposite to [vuk^{35→11}.ti:n¹⁴] 屋顶 ‘roof’.

1.3 Orthographic conventions

To avoid confusion with Hakka and other languages which are presented in IPA phonetic, Mandarin will be transcribed in standard pinyin without phonetic notification.

Some Hakka words have no corresponding Chinese characters. In this case the symbol ☐ will be used to represent them. Although the empty square symbol is traditionally used in Chinese dialectology to represent words that have no corresponding Chinese characters (Chappell and Lamarre 2005:153), it is avoided in this manuscript since it looks similar to the character [k^heu¹⁴] ☐ ‘mouth’. The same reason applies to the empty circle symbol, which can be confused with the Chinese number [k^hu:ŋ³¹] ○ ‘zero’.

2 Phonemes

In JHT there are three major classes of phonemes: consonant, vowel, and tone.

2.1 Consonant phonemes

2.1.1 Numbers and features of consonant phonemes

There are 18 consonant phonemes in JHT: /p, p^h, t, t^h, k, k^h, ʔ, ts, ts^h, f, v, s, h, m, n, ŋ, j, l/. These phonemes are grouped according to their places and manners of articulation as shown in Table 1.

Table 1 JHT consonant phoneme chart

Point of articulation / Manner of articulation	Bilabial	Labio-dental	Apico-alveolar	Fronto-palatal	Dorso-velar	Glottal
Stop: voiceless unaspirated	p		t		k	ʔ
voiceless aspirated	p ^h		t ^h		k ^h	
Affricate: voiceless unaspirated			ts			
voiceless aspirated			ts ^h			
Fricative: voiceless		f	s			h
voiced		v				
Nasal	m		n		ŋ	
Semivowel				j		
Lateral			l			

All 18 consonant phonemes can occur at the initial position of the syllable. Only six of them, /p, t, k, m, n, ŋ/, can also occur at the final position.

2.1.2 Formational statement of consonant phonemes

/p/ is realized as [p̚], an unreleased voiceless unaspirated bilabial stop when it occurs at the final position of the syllable. Elsewhere it is realized as [p], a voiceless unaspirated bilabial stop, e.g., [pui⁵²] 飞 ‘to fly’, [pa:n¹⁴] 板 ‘cake’.

/p^h/ is realized as [p^h], a voiceless aspirated bilabial stop, e.g., [p^hi:³⁵] 皮 ‘skin’, [p^hɔ:n¹¹] 饭 ‘dry cooked rice’.

/t/ is realized as [t̚], an unreleased voiceless unaspirated apico-alveolar stop when it occurs at the final position of the syllable. Elsewhere it is realized as [t], a voiceless unaspirated apico-alveolar stop, e.g., [tɛ:n¹⁴] 等 ‘to wait’, [tiɛt̚³⁵] 跌 ‘to fall’.

/t^h/ is realized as [t^h], a voiceless aspirated apico-alveolar stop, e.g., [t^hi:¹¹] 地 ‘place’, [t^huk̚⁵⁵] 毒 ‘poisonous’.

/k/ is realized as [k̚], an unreleased voiceless unaspirated dorso-velar stop when it occurs at the final position of the syllable. Elsewhere it is realized as [k], a voiceless unaspirated dorso-velar stop, e.g., [ki:m⁵²] 金 ‘gold’, [kui¹⁴] 鬼 ‘ghost’.

/k^h/ is realized as [k^h], a voiceless aspirated dorso-velar stop, e.g., [k^hɔ:ŋ³⁵] 狂 ‘mad’, [k^hiam¹⁴] 俭 ‘to economize’.

/ʔ/ is realized as [ʔ], a glottal stop, e.g., [ʔau¹⁴] 拗 ‘to break’, [ʔɔi³¹] 爱 ‘to want’.

/ts/ is realized as [ts], a homorganic voiceless unaspirated apico-alveolar affricate, e.g., [tsi:m⁵²] 针 ‘needle’, [tsɔi³¹] 嘴 ‘mouth’.

/ts^h/ is realized as [ts^h], a homorganic voiceless aspirated apico-alveolar affricate, e.g., [ts^hɔ:ŋ³⁵] 长 ‘long (in length)’, [ts^hut̚³⁵] 出 ‘to go out’.

/f/ is realized as [f], a voiceless labio-dental fricative, e.g., [fa:⁵²] 花 ‘flower’, [fɔ:¹⁴] 火 ‘fire’.

/v/ is realized as [v], a voiced labio-dental fricative, e.g., [vɔ:ŋ³⁵] 黄 ‘yellow’, [vɔi¹⁴] 会 ‘capable, to be able to’.

/s/ is realized as [s], a voiceless unaspirated apico-alveolar fricative, e.g., [su:⁵²] 书 ‘book’, [sɔ:n³¹] 算 ‘to count’.

/h/ is realized as [h], a voiceless glottal fricative, e.g., [hi:³¹] 去 ‘to go’, [hɔ:¹¹] 号 ‘number’.

/m/ is realized as [m], a voiced bilabial nasal, e.g., [mu:n⁵²] 蚊 ‘mosquito’, [mɔi³¹] 妹 ‘daughter’.

/n/ is realized as [n], a voiced apico-alveolar nasal, e.g., [nɔ:n⁵²] 暖 ‘warm (weather)’, [nu:n¹¹] 嫩 ‘soft, young’.

/ŋ/ is realized as [ŋ], a voiced dorso-velar nasal. When followed by [i] it varies freely with [ɲ], a voiced fronto-palatal nasal, e.g., [ŋian³⁵] ~ [ɲian³⁵] 迎 ‘to parade’, [ŋit³⁵] ~ [ɲit³⁵] 日 ‘day’.

/j/ is realized as [j], a voiced fronto-palatal semivowel, e.g., [jɔ:ŋ¹¹] 样 ‘type, kind’, [jɔk⁵⁵] 浴 ‘to bathe’.

/l/ is realized as [l], a voiced apico-alveolar lateral, e.g., [lɔ:n¹⁴] 卵 ‘egg’, [la:n¹¹] 烂 ‘to be torn’.

2.2 Vowel phonemes

2.2.1 Numbers and features of vowel phonemes

There are six simple vowel phonemes /i, ɛ, a, i̯, u, ɔ/. All of them are oral vowels. They are shown according to the tongue positions and lip shapes in Table 2.

Table 2 JHT vowel phoneme chart

Horizontal tongue position \ Vertical tongue position		Front	Central	Back
		Unrounded	Unrounded	Rounded
Simple vowels	High	i	i̯	u
	Mid	ɛ		ɔ
	Low		a	

These simple vowels, except /i̯/, can be combined into 14 sets of gliding vowels, of which 11 are diphthongs /iu, ui, iɔ, ɔi, ia, ai, uɛ, ɛu, uɔ, ua, au/, and three are triphthongs /iau, uai, iui/.

The vowel [i̯] occurs only in open syllables and is found in limited words with initials [t], [ts], [ts^h], and [s]. All other simple vowels can occur in both open and closed syllables.

In any position of the compound word or utterance, vowels in JHT are generally pronounced short for dead syllables and long for live syllables without a final consonant (open syllable). For live syllables with a final consonant [m, n, ŋ], vowels are also pronounced long in isolation and at the end of a compound word or utterance, but may be shorter in other positions. Length, however, is not phonemic in this language.

2.2.2 Formational statement of vowel phonemes

Simple vowels or monophthongs

- /i/ is realized as [i], a high front unrounded vowel.
- /ɛ/ is realized as [ɛ], a mid front unrounded vowel.
- /a/ is realized as [a], a low central unrounded vowel.
- /i/ is realized as [i], a high central unrounded vowel.
- /u/ is realized as [u], a high back rounded vowel.
- /ɔ/ is realized as [ɔ], a mid back rounded vowel.

Diphthongs

/ai/ is realized as [ai], a low central unrounded glided to high front unrounded vowel.

/au/ is realized as [au], a low central unrounded glided to high back rounded vowel.

/ɔi/ is realized as [ɔi], a mid back rounded glided to high front unrounded vowel.

/ɛu/ is realized as [ɛu], a mid front unrounded glided to high back rounded vowel.

/ia/ is realized as [iɛ], a high front unrounded glided to mid front unrounded vowel when followed by velars [t] and [n]. In other environments it is realized as [ia], a high front unrounded glided to low central unrounded vowel.

/iɔ/ is realized as [iɔ], a high front unrounded glided to mid back rounded vowel.

/iu/ is realized as [iu], a high front unrounded glided to high back rounded vowel.

/ua/ is realized as [ua], a high back rounded glided to low central unrounded vowel.

/uɔ/ is realized as [uɔ], a high back rounded glided to mid back rounded vowel.

/uɛ/ is realized as [uɛ], a high back rounded glided to mid front unrounded vowel.

/ui/ is realized as [ui], a high back rounded glided to high front unrounded vowel.

Triphthongs

/iau/ is realized as [iau], a high front unrounded glided to low central unrounded and then to high back rounded vowel.

/uai/ is realized as [uai], a high back rounded glided to low central unrounded and then to high front unrounded vowel.

/iui/ is realized as [iui], a high front unrounded glided to high back rounded and then back to high front unrounded vowel.

Examples of all vowel phonemes are shown in section 3.2.

2.2.3 Vowel phoneme contrasts

All the six simple and 14 gliding vowel phonemes of JHT mentioned above are contrastive. The minimal pairs that contrast in identical environments of the six simple vowels can be shown in the following set of examples.

[tshi:⁵²] 鲜 ‘fresh’
 [tshɛ:⁵²] 筛 ‘to sift, a sieve’
 [tsha:⁵²] 车 ‘car, cart’
 [tshi:⁵²] 粗 ‘rough’
 [tshu:⁵²] 苧 ‘hemp’
 [tshɔ:⁵²] 坐 ‘to sit’

2.3 Lexical tone

2.3.1 Numbers and features of lexical tone phonemes

Lexical tone is defined as significant pitch that distinguishes the meanings of words (Pike 1976:3). In JHT, there are seven citation tones of lexical tone phonemes which are arranged into two groups:

1. Five tones occur with live syllables, and
2. Two tones occur with dead syllables.

All the seven lexical tone phonemes in JHT are shown in Figure 1. The approximate level of the tone contours are represented by the sequence of numbers; that is, the tone space is divided into five levels, starting from 1 as the lowest to 5 as the highest. The first number is the tone contour's starting point and the second number is its ending point. There is no tone category number four since Shǎng 上 did not split into two tones, and it is omitted from the consecutive numbering so that the tone numbers conform to the general Chinese tone system as shown in Figure 2.

Register	Píng 平 Level	Shǎng 上 Rising	Qù 去 Departing	Rù 入 Entering
Yīn 阴 Upper	/1/ [˥˥] ㄨ [si˥˥] 西 'West'	/3/ [˨˨] ㄨ [si˥˥] 屎 'feces'	/5/ [˥˥] ㄨ [si˥˥] 四 'four'	/7/ [˥˥] ㄨ [sit˥˥] 识 'to know'
Yáng 阳 Lower	/2/ [˥˥] ㄨ [si˥˥] 时 'time'		/6/ [˥˥] ㄨ [si˥˥] 誓 'to swear'	/8/ [˥˥] ㄨ [sit˥˥] 实 'true'

Figure 1 Development of JHT tones from the historical tone categories

2.3.2 Formational statement of lexical tone phonemes

Tone /1/ is realized as [˥˥], a high falling tone. The pitch pattern starts at high pitch, falls to mid-low pitch and remains level. It occurs only on live syllables. The phonation type of this tone phoneme is normal and the length is long, e.g., [ti˥˥] 知 'to know', [kɔ˥˥] 歌 'song'.

Tone /2/ is realized as [˥˥], a mid rising tone. The pitch pattern starts at mid level pitch, rises to high pitch and remains level. It occurs only on live syllables. The phonation type of this tone phoneme is normal and the length is long, e.g., [tʂa˥˥] 茶 'tea', [ŋɔ˥˥] 鹅 'goose'.

Tone /3/ is realized as [˨˨], a low rising tone. The pitch pattern starts at low pitch, rises to mid-high pitch and remains level. It occurs only on live syllables. The phonation type of this tone phoneme is normal and the length is long, e.g., [sɛ˥˥] 洗 'to wash', [pu˥˥] 补 'to darn'.

Tone /5/ is realized as [˥˥], a mid falling tone. The pitch pattern starts at mid level pitch, falls to low pitch and remains level. It occurs only on live syllables. The phonation type of this tone phoneme is normal and the length is long, e.g., [kʰu˥˥] 裤 'trousers, pants', [ŋɛ˥˥] 蚁 'ant'.

Tone /6/ is realized as [˥˥], a low tone. The pitch pattern starts at low pitch and remains level. It occurs only on live syllables. The phonation type of this tone phoneme is normal and the length is long, e.g., [lu˥˥] 路 'road, way', [ja˥˥] 夜 'night'.

Tone /7/ is realized as [˥˥], a mid rising checked tone. The pitch pattern starts at mid level pitch, rises to high pitch and ends with stop closure [p, t, k]. The phonation type of this tone phoneme is normal and the length is short, e.g., [tap˥˥] 搭 'to go along with', [sɛt˥˥] 色 'color', [kɔk˥˥] 角 'angle, corner'.

Tone /8/ is realized as [˥˥], a high checked tone. The pitch pattern starts at high pitch and ends with stop closure [p, t, k]. The phonation type of this tone phoneme is normal and the length is short, e.g., [hap˥˥] 盒 'box', [tʂit˥˥] 直 'straight', [jɔk˥˥] 药 'medicine'.

2.4 On analyzing tone

The Ancient Chinese language of the 7th century A.D. had four tones which were called *Píng* 平 ‘Level’, *Shǎng* 上 ‘Rising’, *Qù* 去 ‘Going or Departing’, and *Rù* 入 ‘Entering’. Syllables in *Rù* 入 tone had a specific feature in that they always ended in a stop [p, t, k] (i.e., dead syllable), while *Píng* 平, *Shǎng* 上, and *Qù* 去 ended in either a vowel or a nasal (i.e., live syllable). Later, these four primary tones were affected by the devoicing of voiced initial consonants, and this caused the split into two registers which are called *Yīn* 阴 and *Yáng* 阳. The result was that the number of tones doubled to eight, as shown in Figure 2 (Norman 1988:52-54).

Register	<i>Píng</i> 平 Level	<i>Shǎng</i> 上 Rising	<i>Qù</i> 去 Departing	<i>Rù</i> 入 Entering
Yīn 阴 Upper	1	3	5	7
Yáng 阳 Lower	2	4	6	8

Figure 2 Tonal system of Chinese languages

At present, the tone system of each Chinese dialect may or may not have a full range of eight tones corresponding to Figure 2 but may vary from three to ten (Chen 2000:13), e.g., standard Mandarin has four. Some languages have more, such as Cantonese, where *Yīn Píng* 阴平 and *Yīn Rù* 阴入 further split into two categories, forming nine tones in total (Matthews and Yip 1994:20) or 10 (Bauer and Benedict 1997:121-122, Yip 2002:174) depending on the area where the language is spoken. For the Hakka dialect, the number of tones can vary from five to seven.

In the traditional Chinese analysis, including JHT, the two *Rù* 入 tones are separate categories. Since syllables with the two *Rù* 入 tones always end with a stop [p, t, k] while the others do not, these two groups cannot be minimal pairs with each other. Then, the two *Rù* 入 tones can be eliminated. If the two *Rù* 入 tones are not considered, there are only five tones in JHT. In this case, each of the two *Rù* 入 tones can be included in the tone that has the same tone value.

In fact, there may be only one *Rù* 入 tone that has the same pitch as one of the five main tones while the other one does not, or none of them conforms to any tone. The first case is found in JHT where tone /7/ has the same tone value (mid rising) as that of tone /2/, while the tone value of tone /8/ is not identical with any of the others.

However, when considering assimilative tone sandhi as shown in Table 8.2 from Nitasakorn (2008:149), tones /7/ and /8/ have the same tone values as tones /3/ and /1/, respectively. So, it is reasonable to move tone /7/ to the same category of tone /3/ and tone /8/ to tone /1/. If this is done, then some

minor changes must also be made to the formational statement of tones in section 2.3.2.

Register	Píng 平 Level	Shǎng 上 Rising	Qù 去 Departing	Rù 入 Entering
Yīn 阴 Upper	/1/, /8/	/3/, /7/	/5/	/7/
Yáng 阳 Lower	/2/		/6/	/8/

Figure 3 Categories to which Rù 入 tones of this study could be moved

Furthermore, the two Rù 入 tones (/7/ and /8/) are classified into separate categories. The reason is that, although tone /7/ pairs with /3/ and tone /8/ with /1/ in the same pattern of assimilative tone sandhi, their positional variant tone sandhi is not identical as can be seen in chapter IX from Nitasakorn (2008:179-238).

From this point of view, it can be seen that tone value itself may not be the only tool for identifying tone categories. In some cases, other phenomena such as tone sandhi can be involved, and also provide help in supporting the classification.

2.5 Tone change (or tonal change)

Tone change is one of the phonological processes to modify the meaning of words in Chinese. In early Chinese, the Qù 去 tone, unlike other tones, had a special function, to create derived words.

For JHT, tone /5/ [³¹], the mid falling tone (Yīn Qù 阴去), is used as the changed tone to derive words. However, at present, the process of tone change in JHT is no longer productive. That is, one cannot use this process to modify the meaning of words at will. It is now for historical study only. Some examples of morphological tone change of JHT are listed below. By means of tone change, the class of the derived words may change, e.g., from noun to verb, or vice versa. In addition, the vowels of the derived words may also change together with tone change.

Original words

[pien¹⁴] 变 ‘to overturn’
 [tsu:ŋ¹⁴] 种 ‘breed, heredity’
 [kau⁵²] 教 ‘to teach’
 [ho:¹⁴] 好 ‘good’
 [to:¹⁴] 到 ‘capable to reach’

Derived words by tone change

[pien³¹] 变 ‘to become’
 [tsu:ŋ³¹] 种 ‘to plant’
 [kau³¹] 教 ‘doctrine, religion’
 [hau³¹] 好 ‘like to’
 [tau³¹] 到 ‘to arrive, to reach’

2.6 Grammatical tone

Grammatical tone is defined by Schuh (1978:252) as the tone that is placed, for morphological purpose, on a particular word or morpheme as the marker of a particular grammatical meaning. For JHT, tone /6/ [11], low tone (Yáng Qù 阳去), is used as a grammatical tone. The use of this tone for grammatical purpose in JHT is limited. It is found in the following cases:

1. Changing from indicative (perfective) to subjunctive mood;
2. Changing from active to passive voice;
3. Changing from some narrative to interrogative forms.

2.6.1 Changing from indicative (perfective) to subjunctive mood

2.6.1.1 The construction of subjunctive mood in JHT

In JHT, changing from indicative to subjunctive mood can be made by adding the word [hɔ:14] 好 ‘good’ to the final part of the sentence thereby changing its tone to a low tone /6/ [11]. The following examples show the utterances in subjunctive mood compared with indicative mood.

Indicative mood (perfective)	Subjunctive mood
[sɔi11] 睡 ‘to sleep’	[sɔi11.hɔ:14→11] 睡好 ‘Sleep better!, Let’s sleep!’
[m:35.mɔi31.khɔ:n31] 唔爱看 ‘don’t look’	[m:35.mɔi31.khɔ:n31.hɔ:14→11] 唔爱看好 ‘It’s better not to look at it.’

2.6.1.2 Contrast of [hɔ:14] 好 and [hɔ:14→11] 好

The contrast between [hɔ:14] 好 using a citation tone and [hɔ:14→11] 好 using a grammatical tone can be seen from the following examples:

Indicative mood (perfective)	Subjunctive mood
[sit55.hɔ:14] 食好 ‘to eat good (food), to eat neatly (tidy, not greedy)’	[sit55.hɔ:14→11] 食好 ‘It’s better to eat now.’
[tso:31.hɔ:14] 做好 ‘to make well, to do good’	[tso:31.hɔ:14→11] 做好 ‘It’s better to do now.’

2.6.2 Changing from active to passive voice

2.6.2.1 Passive construction of JHT

The construction of passive voice in JHT can be made by using the word [pu:n52] 分 ‘to give’ as the main verb and by putting the word [tɔ:14] 到 ‘capable to reach’ at the end of the sentence with the change of tone to low tone /6/ [11].

[ki:³⁵.pu:n⁵².kia^{52→33}.tia⁵².p^ha:n⁵².tɔ:^{14→11}] 佢分𠵼爹拌到
 ‘He was beaten by his father.’

The word [tɔ:^{14→11}] at the end of some passive sentences can be omitted without changing the meaning. For example, the above sentence can also be said as follows:

[ki:³⁵.pu:n⁵².kia^{52→33}.tia⁵².p^ha:n⁵²] 佢分𠵼爹拌
 ‘He was beaten by his father.’

Normally, the meaning differs between sentences with and without the word [tɔ:^{14→11}] at the end of the sentence, e.g.,

[ki:³⁵.pu:n^{52→33}.ts^ha:⁵².ts^hɔ:^η¹¹] 佢分车撞
 ‘He let a car hit him.’ (by jumping intentionally in front of the moving car)

[ki:³⁵.pu:n^{52→33}.ts^ha:⁵².ts^hɔ:^η¹¹.tɔ:^{14→11}] 佢分车撞到
 ‘He was hit by a car.’ (by accident while walking on the street)

And also, the word [pu:n⁵²] 分 may be omitted in some sentences, e.g.,

[li:³⁵.kai³¹.sa:m⁵².pu:n^{14→11}.su:⁵².pu:n⁵².ŋai³⁵.mai⁵².tɔ:^{14→11}.liau³¹]
 这个三本书分我买到了 ‘These three books were bought by me already.’

[li:³⁵.kai³¹.sa:m⁵².pu:n^{14→11}.su:⁵².ŋai³⁵.mai⁵².tɔ:^{14→11}.liau³¹]
 这个三本书我买到了 ‘These three books were bought by me already.’

2.6.2.2 Contrast of [tɔ:¹⁴] 到 and [tɔ:^{14→11}] 到

The contrast of [tɔ:¹⁴] 到 using a citation tone and [tɔ:^{14→11}] 到 using a grammatical tone can be shown in the following pair of sentences. Both of them comprise the same words; only the tones distinguish them.

Active voice: lexical tone

[ŋai:³⁵.pu:n⁵².ki:^{35→33}.tsuk^{-35→11}.tɔ:¹⁴]
 我分佢捉到 ‘I let him catch me.’
 (such as by reaching out my hand)

Passive voice: grammatical tone

[ŋai:³⁵.pu:n⁵².ki:^{35→33}.tsuk⁻³⁵.tɔ:^{14→11}]
 我分佢捉到 ‘I was caught by him.’
 (I tried to escape but I couldn’t.)

2.6.3 Changing from narrative to interrogative forms

Some words in JHT acquire their interrogative forms by changing the tone of narrative forms to tone /6/ [11]. These words are shown in Table 3.

Table 3 Words that acquire their interrogative forms by using the grammatical tone

Narrative: lexical tone	Interrogative: grammatical tone
[li: ³⁵] ~ [lɛ: ³⁵] 这 ‘this’	[li: ^{35→11}] ~ [lɛ: ^{35→11}] 哪 ‘which’
[ʔa:n ^{35→33} .ŋiəŋ ³⁵] 咁娘 ‘like this’	[ʔa:n ³⁵ .ŋiəŋ ^{35→11}] 咁仰 ‘how’
[ʔa:n ^{35→33} .ŋiəŋ ^{35→33} .hi:n ³⁵] ~ [ʔa:n ^{35→33} .ŋiəŋ ³⁵ .hi:n ^{35→33}] ² 咁娘形 ‘like this’	[ʔa:n ³⁵ .ŋiəŋ ^{35→11} .hi:n ³⁵] ³ 咁仰形 ‘how’

2.7 Lexical tone phoneme, tone change, and grammatical tone contrast

It can be seen that the words [hɔ:¹⁴] 好 ‘good’ and [tɔ:¹⁴] 到 ‘capable to reach’ are two words in JHT that have their basic citation tone, are words derived by tone change, and their tones can also be changed by applying the grammatical tone as shown in Table 4.

Table 4 Meanings, types, and functions of the words [hɔ:¹⁴] 好 and [tɔ:¹⁴] 到 using basic citation tones, using changed tone, and overridden by grammatical tone

Basic citation tone	Changed tone /5/ [³¹]	Grammatical tone /6/ [¹¹]
Indicative mood: [hɔ: ¹⁴] 好 ‘good’	Derived word: [hau ³¹] 好 ‘like to’	To form subjunctive mood: [hɔ: ^{14→11}] 好 ‘It’s better to, Let’s’
Active voice: [tɔ: ¹⁴] 到 ‘capable to reach’	Derived word: [tau ³¹] 到 ‘to arrive, to reach’	To form passive voice: [tɔ: ^{14→11}] 到 ‘indicating passive’

3 Syllable

3.1 Syllable structure

The structure of the syllable in JHT can be represented as in Figure 4.

² In colloquial usage, the word [ʔa:n^{35→33}.ŋiəŋ^{35→33}.hi:n³⁵] ~ [ʔa:n^{35→33}.ŋiəŋ³⁵.hi:n^{35→33}] 咁娘形 is contracted to [ʔa:n³⁵.hi:n^{35→33}] 咁形 and becomes [ʔa:n³⁵.ŋi:n^{35→33}] 咁形 by means of assimilation in which the initial [h] of the word [hi:n^{35→33}] 形 assimilates totally with the final [ŋ] of the word [ʔa:n³⁵] 咁.

³ Colloquially, the word [ʔa:n³⁵.ŋiəŋ^{35→11}.hi:n³⁵] 咁仰形 is shortened to [ŋiəŋ^{35→11}.hi:n³⁵] 仰形 and becomes [ŋiəŋ^{35→11}.ŋi:n³⁵] 仰形 by means of assimilation in which the initial [h] of the word [hi:n³⁵] 形 assimilates totally with the final [ŋ] of the word [ŋiəŋ^{35→11}] 仰.

According to this formula, a minimal syllable in JHT requires the initial consonant plus vowel plus tone ($C_i V^T$) and a maximal syllable corresponds to the complete structure shown in the above formula. There are no syllables consisting only of a nuclear vowel. However, any combination of the elements is possible, except G_2 and C_f which are in complementary distribution.

Since both G_1 and G_2 are limited to /i/ and /u/, these two vowels are always either on-glide or off-glide depending on their position in the syllable. In case /i/ and /u/ occur close together in a sequence, such as /iu/ or /ui/, which one of them will be either the on-glide or off-glide depends on whether or not the syllable is open or closed. That is, if the syllable is open without a final consonant, such as [siu⁵²] 梳 ‘comb’ and [sui⁵²] 祟 ‘unlucky, bad luck’, the latter /u/ in /iu/ and /i/ in /ui/ are the off-glides. On the other hand, if the syllable is closed with a final consonant, such as [k^hiu³⁵] 裙 ‘skirt’, the preceding /i/ in /iu/ is the on-glide. The sequence of /uiC_f/ was not found to occur in this study.

The second formula of the syllable structure of JHT can also be a nasal syllabic accompanied by a tone as in the following formula. Only three words are found with this kind of syllable structure.

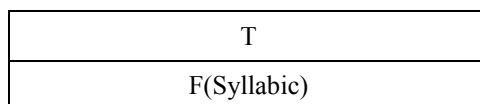


Figure 7 Structure of the simple JHT syllable

Examples of possible structure of JHT are as follows:

$C_i V^T$	[se: ³¹] 细 ‘small’, [vu: ¹¹] 芋 ‘taro’
$C_i V G_2^T$	[siu ¹⁴] 手 ‘hand’, [mai ¹¹] 卖 ‘to sell’
$C_i V C_f^T$	[si:n ⁵²] 新 ‘new’, [jap ⁵⁵] 叶 ‘tree leaf’
$C_i G_1 V^T$	[tia ⁵²] 爹 ‘father’, [ts ^h io ³⁵] 锄 ‘hoe’
$C_i G_1 V G_2^T$	[t ^h iau ¹⁴] 柱 ‘pole’, [kuai ³¹] 怪 ‘strange’
$C_i G_1 V C_f^T$	[piɔŋ ⁵²] 板 ‘wood sheet’, [ŋiun ¹¹] 朋 ‘hard, stiff’
$C_{nas.syl.}^T$	[m: ³⁵] 唔 ‘not’, [ŋ: ³⁵] 鱼 ‘fish’, [ŋ: ¹⁴] 五 ‘five’

3.2 Formation of finals through combination of vowels and final consonants

Vowels combine with final consonants in JHT to form 60 finals as follows:

- [a:], [ai], [ak^ˀ], [a:m], [a:n], [a:ŋ], [ap^ˀ], [at^ˀ], [au];
- [ɔ:], [ɔi], [ɔk^ˀ], [ɔ:n], [ɔ:ŋ], [ɔt^ˀ];
- [ɛ:], [ɛ:m], [ɛ:n], [ɛp^ˀ], [ɛt^ˀ], [ɛu];
- [i:], [ia], [iak^ˀ], [iam], [iaŋ], [iap^ˀ], [iau], [iɔ], [iɔk^ˀ], [iɔŋ], [iɔŋ], [iɛn], [iet^ˀ], [i:m], [i:n], [ip^ˀ], [it^ˀ], [iu], [iui], [iuk^ˀ], [iun], [iuŋ];
- [i:];
- [u:], [ua], [uai], [uan], [uaŋ], [uat^ˀ], [uo], [uɛ], [uet^ˀ], [ui], [uk^ˀ], [u:n], [u:ŋ], [ut^ˀ];
- [m:], [ŋ:].

Examples:

[a:]	[sa: ³⁵] 蛇 ‘snake’, [ta: ¹⁴] 打 ‘to beat’
[ai]	[pai: ³¹] 拜 ‘to worship’, [fai: ¹¹] 坏 ‘to be damaged’
[ak ^ˀ]	[tʰak ^{ˀ35}] 绚 ‘to tie’, [pʰak ^{ˀ55}] 白 ‘white’
[a:m]	[sa:m ⁵²] 衫 ‘shirt’, [ja:m ³⁵] 盐 ‘salt’
[a:n]	[na:n ³⁵] 难 ‘difficult’, [ja:n ¹⁴] 远 ‘far’
[a:ŋ]	[ha:ŋ ³⁵] 行 ‘to walk’, [tʰa:ŋ ³¹] 听 ‘to listen’
[ap ^ˀ]	[ʔap ^{ˀ35}] 鸭 ‘duck’, [tʰap ^{ˀ55}] 踏 ‘to trample, to thresh’
[at ^ˀ]	[mat ^{ˀ35}] 袜 ‘sock, stocking’, [vat ^{ˀ55}] 滑 ‘slippery’
[au]	[ŋau ³⁵] 熬 ‘to simmer, to stew’, [sau: ³¹] 扫 ‘to sweep’
[ɔ:]	[tʰɔ: ¹⁴] 讨 ‘to ask for, to demand’, [mɔ: ¹¹] 帽 ‘hat’
[ɔi]	[fɔi: ⁵²] 灰 ‘ash’, [lo: ³⁵] 来 ‘to come’
[ɔk ^ˀ]	[sɔk ^{ˀ35}] 索 ‘rope’, [lɔk ^{ˀ55}] 落 ‘to go down’
[ɔ:n]	[sɔ:n ³⁵] 船 ‘boat’, [kʰɔ:n ³¹] 看 ‘to look, to see’
[ɔ:ŋ]	[tʰɔ:ŋ ³⁵] 糖 ‘sugar, candy’, [tsʰɔ:ŋ ¹¹] 撞 ‘to hit’
[ɔt ^ˀ]	[tʰɔt ^{ˀ35}] 脱 ‘to take off’, [hɔt ^{ˀ55}] 渴 ‘tired, weary’
[ɛ:]	[fɛ: ⁵²] 罟 ‘net to catch rabbits’, [hɛ: ³¹] 系 ‘to be, yes’
[ɛ:m]	[nɛ:m ⁵²] 满 ‘full’, [nɛ:m ³⁵] 囫 ‘soft’
[ɛ:n]	[tʰɛ:n ³⁵] 跟 ‘to follow’, [hɛ:n ¹⁴] 肯 ‘to consent’
[ɛp ^ˀ]	[sep ^{ˀ35}] 涩 ‘acidulous, sleepy’
[ɛt ^ˀ]	[nɛt ^{ˀ35}] 笏 ‘thorn’, [hɛt ^{ˀ55}] 囫 ‘to dwell, to live’
[ɛu]	[lɛu ³⁵] 楼 ‘floor of a building’, [tʰɛu ¹¹] 豆 ‘bean’
[i:]	[mi: ¹⁴] 米 ‘hulled rice’, [ki: ³¹] 记 ‘to note’
[ia]	[sia: ¹⁴] 写 ‘to write’, [tsia: ³¹] 借 ‘to borrow’
[iak ^ˀ]	[sia: ³⁵] 惜 ‘to love’, [kʰiak ^{ˀ55}] 屐 ‘slipper’
[iam]	[tʰiam ³⁵] 甜 ‘sweet (taste)’, [kʰiam ³¹] 欠 ‘to owe’
[iaŋ]	[mian ³⁵] 名 ‘name’, [pian ¹⁴] 饼 ‘cake, biscuit’
[iap ^ˀ]	[tʰiap ^{ˀ35}] 叠 ‘to overlay’, [ŋiap ^{ˀ55}] 业 ‘activity’
[iau]	[tiau ⁵²] 鸟 ‘bird’, [kia: ³¹] 嗷 ‘to cry, to wail’
[iɔ]	[kʰiɔ ³⁵] 茄 ‘eggplant’
[iɔk ^ˀ]	[kiɔk ^{ˀ35}] 脚 ‘foot, leg’, [ŋiɔk ^{ˀ55}] 弱 ‘weak’
[iɔŋ]	[ŋiɔŋ ⁵²] 软 ‘soft’
[iɔŋ]	[hiɔŋ ⁵²] 香 ‘smell good’, [liɔŋ ³⁵] 凉 ‘cold’

[ien]	[t ^h ien ⁵²] 天 ‘sky, heaven’, [p ^h ien ³¹] 骗 ‘to lie’
[iet’]	[t ^h iet ³⁵] 铁 ‘steel’, [liet ⁵⁵] 裂 ‘to split, to crack’
[i:m]	[ts ^h i:m ⁵²] 深 ‘deep’, [ki:m ³¹] 禁 ‘to put in jail’
[i:n]	[ŋi:n ³⁵] 人 ‘man, people’, [si:n ³¹] 信 ‘letter, to believe’
[ip’]	[tsip ³⁵] 汁 ‘juice, gravy’, [ŋip ⁵⁵] 入 ‘to enter’
[it’]	[pit ³⁵] 笔 ‘writing brush’, [sit ⁵⁵] 食 ‘to eat’
[iu]	[tsiu ¹⁴] 酒 ‘liquor’, [k ^h iu ¹¹] 旧 ‘old (for inanimate objects)’
[iui]	[ŋiui ¹¹] 女 ‘female’
[iuk’]	[kiuk ³⁵] 逐 ‘to expel, to chase’, [ŋiuk ⁵⁵] 狱 ‘prison, hell’
[iun]	[k ^h iun ⁵²] 近 ‘near’, [ŋiun ¹¹] 顽 ‘obstinate, wayward’
[iuŋ]	[hiuŋ ⁵²] 兄 ‘elder brother’, [liuŋ ³⁵] 龙 ‘dragon’
[i:]	[tsi: ⁵²] 租 ‘rent’, [si: ¹¹] 事 ‘affair, matter’
[u:]	[p ^h u: ³⁵] 焯 ‘to cook, to boil’, [su: ¹¹] 树 ‘tree’
[ua]	[kua ⁵²] 瓜 ‘gourds, melons’, [k ^h ua ³¹] 挂 ‘to suspend’
[uai]	[kuai ⁵²] 乖 ‘obedient’, [ŋuai ¹¹] 外 ‘outside’
[uan]	[kuan ⁵²] 关 ‘to shut, to close’, [k ^h uan ¹⁴] 攬 ‘to carry on hand’
[uaŋ]	[k ^h uaŋ ³¹] 矿 ‘mineral; a section, an item’
[uat’]	[k ^h uat ³⁵] 阔 ‘wide’, [kuat ⁵⁵] 罇 ‘bottle’
[uɔ]	[kuɔ ¹⁴] 果 ‘fruit’
[uɛ]	[kuɛ ³⁵] 哭 ‘loud crying’
[uet’]	[kuet ³⁵] 国 ‘country’
[ui]	[tui ¹⁴] 扯 ‘to pull’, [kui ³¹] 贵 ‘expensive’
[uk’]	[kuk ³⁵] 谷 ‘unhulled rice’, [t ^h uk ⁵⁵] 读 ‘to learn, to study’
[u:n]	[ts ^h u:n ³⁵] 存 ‘to remain’, [mu:n ³¹] 问 ‘to ask (question)’
[u:ŋ]	[ts ^h u:ŋ ³⁵] 虫 ‘worm’, [ju:ŋ ¹¹] 用 ‘to use’
[ut’]	[vut ³⁵] 熨 ‘to iron’, [fut ⁵⁵] 佛 ‘Buddha image’
[m:]	[m: ³⁵] 唔 ‘not’
[ŋ:]	[ŋ: ³⁵] 鱼 ‘fish’, [ŋ: ¹⁴] 五 ‘five’

4 Phonological processes

4.1 Assimilation

Assimilation in JHT occurs with consonantal segmentation of the syllable only. It can be found in a few words, and all of them result from progressive assimilation.

In rapid speech, the initials [ʔ] and [h] of the latter words in the following terms assimilate totally with the syllabic [m̩]. That is, they change their quality to be the same as the syllabic [m̩].

[m:³⁵.ʔai:³¹] → [m:³⁵.məi:³¹] 唔爱 ‘do not want’
 [m:³⁵.ho:¹⁴] → [m:³⁵.mɔ:¹⁴] 唔好 ‘cannot; don’t’
 [m:³⁵.he:³¹] → [m:³⁵.me:³¹] 唔系 ‘It is not.’
 [he:³¹.m:³⁵.he:³¹] → [he:³¹.m:³⁵.me:³¹] 系唔系 ‘yes or no’

The word [he:³¹.m:³⁵.he:³¹] → [he:³¹.m:³⁵.me:³¹] can further be contracted to [he:³¹.me:³¹] 系㗎 with little change of the meaning ‘That it is!’ as shown in the following sentence.

[he:³¹.me:³¹.ŋai:³⁵.kɔ:¹⁴.liɑu:³¹] 系㗎! 我讲了.
 ‘That it is! I have told you before.’

The following terms show that the initial [h] of the latter syllable assimilates totally with the final consonant of the previous syllable becoming [ŋ].

[ŋiɔŋ^{35→11}.hi:n³⁵] → [ŋiɔŋ^{35→11}.ŋi:n³⁵] 仰形 ‘how’
 [ʔa:ŋ³⁵.hi:n^{35→33}] → [ʔa:ŋ³⁵.ŋi:n^{35→33}] 咁形 ‘like this, in this way’

4.2 Epenthesis

There are two question words which are open syllables; when they are followed by another syllable to form compound terms, an appropriate homorganic final determined by the following syllable initial is added together with the change of the original tone.

Table 5 Question words with additive final

Question word	Additive final	Question words with additive final
[ma: ¹⁴]	[-k ^ˀ]	[ma: ^{14→11} .kai: ³¹] → [mak ^{ˀ35→11} .kai: ³¹] 吗个 ‘what’
[ki: ¹⁴]	[-t ^ˀ]	[ki: ^{14→11} .tɔ: ⁵²] → [kit ^{ˀ35→11} .tɔ: ⁵²] ⁴ 几多 ‘how much, how many’

4.3 Variations of consonants

Some initial consonants in JHT have their variants. These variants are rather specific. Initial consonants that vary in isolation may or may not vary in compound words and vice versa. Even in compound words, they vary in one but not in another. Variations of consonants are:

⁴ In Hong Kong, Lau (2000:26) found that the term [ki:^{14→11}.tʰai:¹¹] 几大 ‘How big it is?’ was pronounced [kit^{ˀ35→11}.tʰai:¹¹] (Lau: git5tai2) in casual speech. However, it must be [kit^{ˀ35→11}.tɔ:⁵².tʰai:¹¹] 几多大 in JHT. Another word is [ki:³¹.si:³⁵] 几时 ‘when’ which was pronounced [kit^{ˀ35→11}.si:³⁵] (Lau: git5si2) in Hong Kong (Lau 2000:46).

1. [f] ~ [h];
2. [k^h] ~ [h];
3. [l] ~ [n];
4. [v] ~ [b].

4.3.1 [f] ~ [h]

The initial [f], when followed by [u] in some words, has [h] as its variant, e.g.,

[fu:³⁵] ~ [hu:³⁵] 扶 ‘to elevate with both hands’
 [fu:³¹] ~ [hu:³¹] 附 ‘on time’
 [fut⁵⁵] ~ [hut⁵⁵] 核 ‘seed’

In the following words, [f] and [h] vary only in compound words. In isolation, they do not vary.

[hiɔŋ^{14→11}.fuk^{55 (~35)}] ~ [hiɔŋ^{14→11}.huk^{55 (~35)}] 享福 ‘happy’
 [si:n³⁵.fu:^{31 (~11)}] ~ [si:n³⁵.hu:^{31 (~11)}] 神父 ‘Catholic priest’
 [t^hɛu¹¹.fu:¹¹] ~ [t^hɛu¹¹.hu:¹¹] 豆腐 ‘bean-curd’

According to Hashimoto (1973:101) /h/ and /f/ in the Meixiàn dialect are neutralized before the rounded high vowel, and they result in [f]. However, this is not always true in JHT. The following words show that the initial [f], although followed by [u], does not have [h] as its variant.

[fu:⁵²] 夫 ‘man, husband’
 [fu:n¹⁴] 粉 ‘powder, flour’
 [fu:ŋ³⁵] 红 ‘red’

4.3.2 [k^h] ~ [h]

Variations between [k^h] and [h] can be seen in a few words as follows:

[k^hak^{35→11}.ŋi:n³⁵] ~ [hak^{35→11}.ŋi:n³⁵] 客人 ‘Hakka people’
 [k^hɔ:^{14→11}.pi:¹⁴] ~ [hɔ:^{14→11}.pi:¹⁴] 可比 ‘suppose, for example’
 [k^hɔi:⁵²] ~ [hɔi:⁵²] 开 ‘to open’
 [k^hɔk³⁵] ~ [hɔk³⁵] 熨 ‘to fry’

4.3.3 [l] ~ [n]

In some JHT words, the initial [l] is pronounced as [n]. Mercer (1930:59) stated that in the dialect of Hakka spoken in the Xīn'ān 新安 district of Guǎngdōng 广东 province, from which the larger number of Hakka Chinese in North Borneo come, the initial [n] is nearly always pronounced as [l]. MacIver (1926:372) also noted that in the southwestern parts of the Hakka-speaking region, the initial [l] is sometimes interchanged with [n].

In this study, the words in which the initial [l] has [n] as its variant are as follows:

[la:¹¹.sap³⁵] ~ [na:¹¹.sap³⁵] 垃圾 ‘filthy, dirty’
 [lai:^{14→11}.kuo¹⁴] ~ [nai:^{14→11}.kuo¹⁴] 荔果 ‘lychee’
 [ʔo:ŋ^{14→11}.la:m³⁵] ~ [ʔo:ŋ^{14→11}.na:m³⁵] ㄑ籃 ‘cradle, hammock’

In isolation the word [la:m³⁵] 籃 ‘basket’ is not pronounced [na:m³⁵] as its variant. It is pronounced [na:m³⁵] only in the compound word mentioned above.

On the contrary, the words in which the initial [n] has [l] as its variant are also found as follows:

[na:m^{35→33}.tsi:^{14→11}.ŋi:n³⁵] ~ [la:m^{35→33}.tsi:^{14→11}.ŋi:n³⁵] 男子人 ‘man’
 [nap³⁵] ~ [lap³⁵] 纳 ‘to receive’
 [nu:ŋ^{35→33}.mi:n³⁵] ~ [lu:ŋ^{35→33}.mi:n³⁵] 农民 ‘farmer’

4.3.4 [v] ~ [b]

There are no voiced stops in the phonological system of JHT. The only voiced consonant is fricative /v/. For this consonant in the Méixiàn dialect, Norman (1988:224) stated that “it seems best to classify it among the liquids.” His reason is that it lacks strong friction and that “if it were considered a voiced fricative, one would have to recognize an anomalous voiced-voiceless contrast between [the] two fricatives *f* and *v*.”

In this study /v/ has the same friction as its voiceless counterpart /f/ and is classified in the fricative group. In some words, [b], a voiced unaspirated bilabial stop, may be heard as its variant, for example:

[m:³⁵.vɔi¹⁴] ~ [m:³⁵.bɔi¹⁴] 唔会 ‘incapable, cannot’
 [si:^{14→11}.vut³⁵] ~ [si:^{14→11}.but³⁵] 屎ㄑ ‘anus’
 [vu:n^{35→33}.mi:n³⁵] ~ [bu:n^{35→33}.mi:n³⁵] 文明 ‘civilized’

4.4 Variations of vowels

Vowels in JHT also have their variants. The variations of vowels found in this study result from vowel interchange and monophthongization.

4.4.1 Vowel interchange

The vowels /i/ and /ɛ/ are interchangeable in the following words:

[ma:^{14→11}.si:⁵⁵] ~ [ma:^{14→11}.se:⁵⁵] 吗时 ‘what’
 [li:³⁵] ~ [ɛ:³⁵] 这 ‘this’
 [ŋiu³⁵] ~ [ŋɛu³⁵] 牛 ‘cattle’
 [p^{hit}:⁵⁵.p^{hɔ}:³⁵] ~ [p^{het}:⁵⁵.p^{hɔ}:³⁵] ㄙㄙ ‘a bat’
 [si:m⁵²] ~ [se:m⁵²] 参 ‘ginseng’

4.4.2 Monophthongization

The diphthongs in some JHT words may monophthongize to a simple vowel. These vowels are [ai] → [ɛ:] and [au] → [ɔ:], e.g.,

[ai] ~ [ɛ:] [kai:⁵²] ~ [kɛ:⁵²] 鸡 ‘chicken’
 [lai:³⁵] ~ [ɛ:³⁵] 犁 ‘to plough’
 [nai:³⁵] ~ [nɛ:³⁵] 泥 ‘earth soil’
 [ʔai:¹⁴] ~ [ʔɛ:¹⁴] 矮 ‘low (in position), short (in height)’
[au] ~ [ɔ:] [kau:⁵²] ~ [kɔ:⁵²] 高 ‘high, tall’
 [lau:¹⁴] ~ [lɔ:¹⁴] 老 ‘old’
 [pau:¹⁴] ~ [pɔ:¹⁴] 保 ‘to protect, to guard; to nourish’
 [ts^hau:¹⁴] ~ [ts^hɔ:¹⁴] 草 ‘weed, grass’

In some JHT words, the vowel [ɔ:] which does not have [au] as its variant may be pronounced [au] in other dialects.

Table 6 Examples of JHT [ɔ:] which are pronounced [au] in other dialects

JHT	Pronunciation in other dialects and sources (Phonetic transcriptions are as quoted in the original works)
[hɔ: ¹¹] 号 ‘number’	hàu (Chiang 1941:41) hau ⁵³ (Lau 2000:146)
[kɔ: ³⁵] 膏 ‘fat, ointment’	kau ³³ (Lau 2000:146)
[mɔ: ⁵²] 毛 ‘hair’	mau (Mercer 1930:170, Chiang 1941:23) mau ³³ (Lau 2000:145)
[sɔ: ¹⁴] 嫂 ‘wife of elder brother’	sáu (Mercer 1930:151) sau ³¹ (Lau 2000:146) sau ³ (Chappell and Lamarre 2005:202)

4.5 Elision

4.5.1 Vowel [i]

Vowel [i] in the following finals, [iɛtʰ], [iɛn], [iutʰ], [iun], [iukʰ], and [iuŋ], may be elided in some words. The first word of each example is used in careful speech while the latter is its variant and normally used, e.g.,

[iɛtʰ] ~ [ɛtʰ]	[p ^h iɛt ⁵⁵] ~ [p ^h ɛt ⁵⁵] 別 ‘other’ [t ^h iɛt ³⁵] ~ [t ^h ɛt ³⁵] 忒 ‘too much, too many’
[iɛn] ~ [ɛ:n]	[kien ⁵²] ~ [kɛ:n ⁵²] 间 ‘classifier for house, room’ [sien ⁵²] ~ [sɛ:n ⁵²] 先 ‘first, foremost’
[iutʰ] ~ [utʰ]	[k ^h iut ⁵⁵] ~ [k ^h ut ⁵⁵] 窟 ‘hole, pool’
[iun] ~ [u:n]	[k ^h iun ⁵²] ~ [k ^h u:n ⁵²] 近 ‘near’
[iukʰ] ~ [ukʰ]	[liuk ³⁵] ~ [luk ³⁵] 六 ‘six’ [ŋiuk ³⁵] ~ [ŋuk ³⁵] 肉 ‘flesh’
[iuŋ] ~ [u:ŋ]	[kiuŋ ^{52→33} .tsiau ⁵²] ~ [ku:ŋ ^{52→33} .tsiau ⁵²] 芎蕉 ‘banana’ [liuŋ ³⁵] ~ [lu:ŋ ³⁵] 龙 ‘dragon’

Other than these, the vowels [iɛtʰ] and [iɛn] in some JHT words which do not have their [i] elided may be generally pronounced [ɛtʰ] and [ɛ:n], respectively, in other dialects.

Table 7 Examples of JHT [iɛtʰ] and [iɛn] which are pronounced [ɛtʰ] and [ɛ:n], respectively, in other dialects

JHT	Pronunciation in other dialects and sources (Phonetic transcriptions are as quoted in the original works)
[hiɛt ³⁵] 血 ‘blood’	het (Mercer 1930:168)
[liɛt ⁵⁵] 列 ‘to arrange in order’	lɛt ⁵⁵ (Lau 2000:158)
[t ^h iɛt ³⁵] 铁 ‘iron, steel’	thet (Mercer 1930:169) t ^h ɛt ³² (Lau 2000:159) thet ⁶ (Chappell and Lamarre 2005:218)
[lien ¹¹] 犁 ‘sleigh’	lɛn ¹¹ (Lau 2000:158)
[sien ³¹] 线 ‘wire; line’	sɛn ⁵³ (Lau 2000:158) sɛn ⁴ (Chappell and Lamarre 2005:175)
[t ^h ien ³⁵] 钱 ‘money’	tshɛn (Mercer 1930:16) ts ^h ɛn ¹¹ (Lau 2000:158)

4.5.2 Vowel [u]

The vowel [u] is elided in some words, e.g.,

[uɔ] ~ [ɔ:] [kuɔ^{52→33}.mɔ:⁵²] ~ [kɔ:^{52→33}.mɔ:⁵²] 𠵼𠵼 ‘exterminated (term of abuse)’

[uɛ] ~ [ɛ:] [kue^{35→33}.kue³⁵.kue^{35→31}.kue^{35→31}] ~ [kɛ:^{35→33}.kɛ:³⁵.kɛ:^{35→31}.kɛ:^{35→31}] 𠵼𠵼𠵼𠵼 ‘noisy’

The vowel [u] in [uɔ] and [uɔŋ] finals are generally elided becoming [ɔ:] and [ɔ:ŋ] respectively. Among all gliding vowels in JHT, only [uɔ] is a vertical diphthong, while all others are diagonal. This phenomenon is the same as the elision of [i] as mentioned in section 4.5.1.

Table 8 Examples of JHT [ɔ:] which are derived from [uɔ]

JHT [ɔ:]	Derived from	Pronunciation in other dialects and sources (Phonetic transcriptions are as quoted in the original works)
[kɔ: ³¹]	[kuɔ ³¹] 𠵼 ‘to cross, over’	Kwò (MacIver 1926:353) kwò (Chiang 1941:42) [kɔ: ³¹] (Pratoom 1984:38)
[kɔ:ŋ ⁵²]	[kuɔŋ ⁵²] 光 ‘bright’	Kwong (MacIver 1926:359) kwong (Chiang 1941:26) [kɔ: ⁵² ŋ] (Pratoom 1984:38)
[k ^h ɔ: ³⁵ ŋ]	[k ^h uɔŋ ³⁵] 狂 ‘mad’	Khwông (MacIver 1926:361)

4.6 Other variations

[uai] ~ [ɔi] is found in only one word.

[ŋuai¹¹] ~ [ŋɔi¹¹] 外 ‘outside’

The way that [ŋuai¹¹] ~ [ŋɔi¹¹] 外 ‘outside’ can be compared with [kuan⁵².ts^hɔi³⁵] ~ [kuɔn⁵².ts^hɔi³⁵] ~ [kɔ:n⁵².ts^hɔi³⁵] 棺材 ‘coffin’ is as follows:

Table 9 Pattern of JHT [ua] varies with [ɔ:]

Base word: [ua]	[ua] → [uɔ]	[uɔ] → [ɔ:]
[kuan ⁵² .ts ^h ɔi ³⁵]	[kuɔn ⁵² .ts ^h ɔi ³⁵]	[kɔ:n ⁵² .ts ^h ɔi ³⁵] 棺材 ‘coffin’
[ŋuai ¹¹]	([ŋuɔi ¹¹])	[ŋɔi ¹¹] 外 ‘outside’

5 Summary

The results of the study have revealed that the phonological system in JHT comprises 18 consonants /p, p^h, t, t^h, k, k^h, ʔ, ts, ts^h, f, v, s, h, m, n, ŋ, j, l/; six simple vowels /i, ε, a, i, u, ə/; 11 diphthongs /iu, ui, iə, əi, ia, ai, ue, eu, uə, ua, au/; three triphthongs /iau, uai, iui/; and seven tones. The structure of the syllable is C_i (G₁) V (G₂/C_f) accompanied by a tone or simply a syllabic consonant accompanied by a tone.

There are seven lexical tone phonemes in JHT. These tones conform to traditional Chinese tone categories except the Shǎng 上 tone that had not split into two categories. The Yīn Qù 阴去 tone has been used as a tone changer to modify the meanings of the words. However, this process of tone change is no longer progressive. The Yáng Qù 阳去 tone is still used as a grammatical unit to change some grammatical structures of the utterance.

Various phonological processes are found in JHT, namely, assimilation of the initial consonants, variations among initial consonants, variations among vowels (e.g., vowel interchange, monophthongization, and vowel elision), and the addition of a final consonant to the syllable. These phonological processes are not widespread and affect only some words.

REFERENCES

- Bauer, Robert S. and Paul K. Benedict. 1997. *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter.
- Chappell, Hilary and Christine Lamarre. 2005. *A Grammar and Lexicon of Hakka: Historical Materials from the Basel Mission Library*. Paris: École des Hautes Études en Sciences Sociales.
- Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.
- Chiang, Ker Chiu. 1941. *Hakka for Beginners*. Singapore: Chin Fen Book Store.
- Hashimoto, Mantaro J. 1973. *The Hakka Dialect: A Linguistic Study of Its Phonology Syntax and Lexicon*. Cambridge: Cambridge University Press.
- Lau, Chunfat. 2000. *The Decline of the General Hakka Accent in Hong Kong: A Comparison of "Old-Style" and "New-Style" as spoken by the indigenous inhabitants*. LINCOM Studies in Asian Linguistics 40. Muenchen: LINCOM EUROPA.
- MacIver, D. 1926. *A Chinese-English Dictionary: Hakka-Dialect as Spoken in Kwang-Tung Province*. 2nd edition (Revised and rearranged with many additional terms and phrases by M.C. Mackenzie). Shanghai: The Presbyterian Mission Press.
- Matthews, Stephen and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- Mercer, Bernard. 1930. *Hakka-Chinese Lessons*. London: Sheldon Press.

- Nitasakorn Shiwaruangrote. 2008. A Phonological Description of Jiēyáng Hakka Dialect as Spoken in Thailand. Ph.D. thesis. Salaya: Mahidol University.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Pike, Kenneth L. 1976. *Renewed. Tone Languages*. Ann Arbor: The University of Michigan Press.
- Pratoom Wongwantanee. 1984. A Phonology of Hakka, with Comparison with Swatow. M.A. thesis. Salaya: Mahidol University.
- Ramsey, S. Robert. 1989. *The Languages of China*. 2nd printing, with revisions, and first Princeton Paperback printing. Oxford: Princeton University Press.
- Schuh, Russell G. 1978. "Tone Rules." In Victoria A. Fromkin (ed.), *Tone: A Linguistic Survey*, pp. 221-256. New York: Academic Press.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Received: 25 August 2009

238/19 Jaransatitwong Road,
Banchanglor, Bangkok-Noi,
Bangkok 10700,
THAILAND
E-mail address: nitasakorn@yahoo.com

The Interaction between Pitch and Vowel Length in Mon-Khmer Languages

PHANINTRA Teeranon

Mae Fah Luang University, Thailand

Abstract

The objective of this research is to investigate the interaction between vowel length and pitch in four Mon-Khmer languages: Chong, Khmer, Khmu', and Nyah Kur. Nine pairs of short and long vowel counterparts were investigated; vowel duration and semitone values were measured using Praat. For each language, the informants were two males and two females aged over sixty (> 60), thirty-five to forty-five (35-45), and under twenty (< 20). This research demonstrates that interaction between pitch (as measured in semitone values) and vowel length is language-specific; it is suggested as plausible that vowel length is a factor that conditions pitch in some languages such as Chong.

Key words: phonetics, pitch, vowel length

1. Introduction

Linguists have long been interested in the interaction between pitch and vowel length, in particular the question of whether one factor conditions the other.

There have been a number of discussions regarding pitch as a factor that conditions vowel length (Abramson, 1962; Fischer-Jørgensen, 1990; Gandour, 1977; Heijmans, 2003; Kong, 1987; Lehiste, 1970; Neweklowsky, 1975; Tsukada et al., 2004; Yu, 2006). With regard to acoustic data, vowels with higher F₀, which relates to a high rate of vocal fold vibration or high pitch, tend to be of shorter duration than vowels with lower F₀ in Thai (Abramson, 1962), English (House and Fairbanks, 1953; Lehiste, 1970), Chinese, Russian (Mohr, 1971), German (Mohr, 1971; Neweklowsky, 1975), Kui (L-Thongkum, 1989), Chong (L-Thongkum, 1991), and the Malay Dialect of Pattani and Pathumthani (Lohde, 2003). These studies conclude that vowel duration inversely relates to F₀ (Gandour, 1977: 60). The analysis of acoustic data, however, is not the sole source of support for these findings.

An analysis of auditory data yields similar results; vowels with high pitch are shorter than vowels with low pitch, e.g., Chatino (Pike, 1974), and that vowels with dynamic F0 are longer than those with level F0 in English (Yu, 2006). This finding that vowels with rising tones are longer than those with falling tones can be found in the study of other languages as well as Thai (Gandour, 1977), Cantonese (Kong, 1987), and Korean (Kwon, 2003). Short vowels can bear either high or low tones in Kunama (Connell, 2002) and Kinyarwanda (Myers, 2005), and they are associated with either high or low F0 in Khmer (Wayland, 1998), Lavue', Wa, and Plang (Teeranon, 2007). Broadly, the effect of pitch on vowel length is language-specific rather than being a natural phonetic tendency.

There are numerous discussions suggesting that the opposite is true, that vowel length effects pitch (Diffloth, 1991; Lee, 1993; L-Thongkum et al., 2007; Svantesson, 1991). U and Hu are claimed to have developed high tone (which relates to high auditory pitch) from short vowels, and to have developed low tone (which relates to low auditory pitch) from long vowels (Diffloth, 1991; Svantesson, 1991). In other words, *v conditions the birth of high tone and *vv causes the birth of low tone.

For some linguists (Svantesson, 1991), vowel length distinction signals unusual or “unorthodox” tonogenesis mechanisms. Similarly, Lee (1993) found that a high upper register tone in Cantonese occurs with a short vowel, and that a mid-upper register tone occurs with a long vowel. In their consideration of the acoustic data, L-Thongkum et al. (2007) found that vowels [a] with shorter duration have higher F0 than vowels [aa] with long duration in four languages of Southeast Asia (Bangkok Thai, Kam Muang, Khmu' Rawk and Iu Mien). Accepting that the vowel is a factor that conditions high or low pitch, L-Thongkum and her colleagues claim that this phenomenon is a natural phonetic tendency; thus, it is not an ‘unorthodox’ mechanism in their view. Theoretically, both views have been seen to have had significant impact on language development as a whole.

Pitch as a factor that conditions vowel length and vice versa of interaction between F0 and vowel duration have been offered as historical explanations for the phonological system of various language. First, Gandour (1977) indicates that a Proto-Southwestern Tai monophthong vowel system was reconstructed by Sarawit (1973) as six short vowels and six long vowels; short vowels are /i/, /ī/, /u/, /e/, /a/, /o/, long vowels are /ii/, /īī/, /uu/, /εε/, /aa/, /ɔɔ/. As with the pitch condition vowel length, across Northern Thai and Southern Thai dialects, long vowels developed under rising and non-high-level tones, whereas short vowels developed under falling and high-level tones. This is the explanation for why, at present, there are nine short and long vowels counterparts (/i/-/ii/, /ī/-/īī/, /u/-/uu/, /e/-/εε/, /ə/-/əə/, /o/-/oo/, /ε/-/εε/, /a/-/aa/, /ɔ/-/ɔɔ/) in the Tai dialects.

Conversely, the effect of vowel length on pitch indicates the plausibility of tonal evolution in the language system. From previous research, it begins to appear plausible that tones are developed from the loss of vowel length. From this, two ideas have arisen.

One idea is that the loss of vowel length causing low tones in Hu is ‘unorthodox’ (Diffloth 1991, Svantesson 1991). Another suggests that this phenomenon could be phonetically universal, as it is found in Bangkok Thai, Kam Muang, Khmu’ Rawk and Lu Mien (L-Thongkum et al., 2007). It can be said that short vowels develop high tones and long vowels develop low tones, e.g., Hu, *yam > yám ‘to die’, and *yaam > yàm ‘to cry’.

A number of studies have attempted to explain these physiological mechanisms (Honda and Fujimura, 1991; Whalen et al., 1998). With short vowels, it is possible to increase the subglottal pressure in order to compensate for shortness of duration, which relates to high pitch (Fischer-Jørgensen, 1990). The interaction between pitch and vowel lengths involves the cricothyroid muscles of the larynx. Cricothyroid activity tenses the vocal folds to stiffen them for the short vowels and thus to cause higher pitch. Correspondingly, it slackens them for long vowels thus causing lower pitch (Honda and Fujimura, 1991; Hoole et al., 2004).

Between these two types of interaction between pitch and vowel length, researchers less often suggested vowel length as a factor that conditions pitch; nevertheless, this phenomenon may shed some light on phonetic tendencies having to do with tonal evolution or tonogenesis theory.

In a previous study (L-Thongkum et al., 2007), vowel duration affecting on the vowel F₀ was attested with [a]-[aa] counterparts, while other vowel qualities in the vowel system (e.g., [i]-[ii], [e]-[ee]) were left unproven and prime for further consideration. In order to investigate further the effects of vowel duration on F₀, four languages of Mon-Khmer that contain nine pairs of short and long vowels counterparts—[i]-[ii], [ī]-[īī], [u]-[uu], [e]-[ee], [ə]-[əə], [o]-[oo], [ɛ]-[ɛɛ], [a]-[aa], [ɔ]-[ɔɔ]—were selected and analyzed for this study using measurements of F₀ on isolated monosyllables spoken by representatives of the Mon-Khmer languages. The four Mon-Khmer languages chosen, Chong, Khmer, Khmu’, and Nyah Kur, were selected because they have been classified as non-tonal languages with a tendency to become tonal languages (L-Thongkum, 1988; L-Thongkum, 1991; Premrirat, 2003; Wayland and Guion, 2005). If it is demonstrated that there is an interaction between F₀ and vowel duration in these languages, it is plausible that vowel length is another factor causing tonal evolution in these four languages.

The purposes for the study are 1) to test whether the interaction between F₀ and vowel duration is applicable to all vowel qualities in Chong,

Khmer, Khmu', and Nyah Kur, and 2) to test whether the phonetic tendency of the loss of vowel length has had anything to do with the evolution of tones.

This study tests the claim that the reduction in the vowel length contrast is the cause of the emergence of the tonal contrast. If this is true, the short vowels should have been shorter and/or the long vowels longer in the four languages. In order to settle the issue a longitudinal evidence was conducted. The present study was set into a cross-sectional design comparing older, middle age, and younger speakers of the four languages. Three age groups were recorded; the over-sixty speakers (> 60 years) to represent the past, the middle age speakers (35-45 years) to represent the present time, and the under-twenty speakers (< 20 years) to represent the future. The following pattern was anticipated: (i) The oldest generation show a large difference between long and short vowel duration with a large F0 difference. (ii) The middle generation is as observed in the present study: large pitch difference (larger than in the oldest generation) but a small duration difference (indicating that the duration difference is now large enough as a sufficient cue for what used to be the length contrast), (iii) The youngest generation has no duration difference but may have increased the tonal difference further.

2. Materials and Method

2.1 Language data

The four languages used in the present analysis are Chong, Nyah Kur, Khmer, and Khmu'. They belong to the different branches of the Mon-Khmer language family.

Chong

The Khlong Phlu dialect of Chong is spoken in the Khao Khitchakut district of Chanthaburi Province. The language has 21 consonant phonemes

/	ph	th	ch	kh	
	p	t	c	k	ʔ
	m	n	ɲ	ŋ	
	w	r, l	j		
		s			h
	i, ii	i, ĩ	u, uu,		
	e, ee	ə, əə	o, oo,		
	ɛ, ɛɛ	a, aa	ɔ, ɔɔ		
			uə	/	

There are also 4 registers:

/R1/ clear voice

/R2/ clear voice followed by glottal constriction

/R3/ breathy voice

/R4/ breathy voice followed by glottal constriction

All wordlists from Chong that are used in this study are clear voice (L-Thongkum, 1991). Chong is predicted to develop four tones from phonation type of vowels (L-Thongkum, 1988, 1991). L-Thongkum (1991: 141) states that “In fact, some dialects of Chong, such as the one spoken in Chamkhlo’ Village, Takhianthong Sub-district, Makham District, have already become tonal: presyllables are dropped; phonation types are less prominent and in some cases disappear; and pitch differences can be heard clearly, especially in slow speech. Our Chong informants also describe their language as having high, higher, mid and low tones”

Khmer

Khmer is spoken in the Nok Muang subdistrict of Surin province. The language is comprised of 21 consonant phonemes, 28 monphthongs, and 5 diphthongs (Chantrupanth and Phromjakgarin, 1978).

/	ph	th	ch	kh	
	p	t	c	k	ʔ
	b	d			
	m	n	ɲ	ŋ	
	w	r, l	j		
		s			h
	i, ii	ɨ, ɨɨ	u, uu,		
	ɪ, u	ɤ, ɤɤ	ɔ, ɔɔ,		
	e, ee	ə, əə	o, oo,		
	ɛ, ɛɛ	ʌ, ʌʌ,	ɔ, ɔɔ,		
	a, aa		ɔ, ɔɔ		
			ia, ia, ia,	ua, uua	/

Instead of developing tones from registers, Khmer has chosen the pathway of developing a large vowel system. Khmer in the past is reported to have devoiced its initial consonant registers, changing initial voiced stops /b, d, j, g/ to voiceless stops /p, t, c, k/. The devoicing process causes the vowel systems to split into two series of vowels, ‘ò series vowels’, and ‘a series vowels,’ or ‘the first register’, and ‘the second register’ (Maspero, 1915; Huffman, 1967; Wayland and Jongman, 2003). That is why there are quite a number of vowels in Khmer.

Khmu’

Khmu’ is spoken in Huay Sataeng, Thung Chang district of Nan province. The language is comprised of 22 consonant phonemes, 20 monophthongs, and 3 diphthongs. There are 2 registers: clear voice or /R1/, and breathy voice or /R2/ (L-Thongkum et al., 2008). All wordlists for, Khmu’ used in this study are in clear voice.

/	ph	th	ch	kh	
	p	t	c	k	ʔ
	b	d			
	m	n	ɲ	ŋ	
	w	r, l	j		
	f	s			h
	i, ii	i, ĩ	u, uu,		
	e, ee	ə, əə	o, oo,		
		ʌ, ʌʌ			
	ɛ, ɛɛ	a, aa	ɔ, ɔɔ		
		iə, uə, iə			/

Some Khmu dialects retain their voicing distinction in terms of the initial consonants, whereas some Khmu dialects show the register distinction of the vowels. However, the loss of these two registers in some dialects is now substituted by low tones (Premsrirat, 2003).

Nyah Kur

Nyah Kur is spoken in Ban Rai, a village in the Thepsathit district of Chaiyaphoom province. The language is comprised of 21 consonant phonemes, 18 monophthongs, and 3 diphthongs (L-Thongkum, 1988). Nyah Kur comprises of two registers, clear voice and breathy voice (L-Thongkum, 1984).

/	ph	th	ch	kh	
	p	t	c	k	ʔ
	b	d			
	m	n	ɲ	ŋ	
	w	r, l	j		
		s			h
	i, ii	i, ĩ	u, uu,		
	e, ee	ə, əə	o, oo,		
	ɛ, ɛɛ	a, aa	ɔ, ɔɔ		
		iə, uə, iə			/

Similar to Chong, it is predicted that Nyah Kur will develop two tones from phonation type of vowels (L-Thongkum, 1984, 1988). High tone is associated with clear voice, while low tone is associated with breathy voice.

2.2 Wordlists

For each language, nine pairs of short and long monophthong distinctions, were selected. To avoid pitch perturbation from voicing consonants, five minimal pairs of short and long vowels surrounded by voiceless plosive consonants were prepared, e.g., [pvt]-[pvvt], [tvp]-[tvvp]. Each pair of short and long vowel distinction, therefore, contains five minimal

pairs. In the case of lacking minimal pairs, non-sense words were set to complete the wordlists (See Appendix A.).

2.3 Informants

For each language, four native informants of both sexes who were <60 years of age, 35-45 years of age, and >20 years of age were selected. The four informants included two males and the two females. The informants were asked to pronounce each word 5 times. For each language, the number of tokens for each vowel quality in each age group was 100 (4 informants x 5 words x 5 times). The overall number of test tokens was 53,200 (4 informants x 3 age groups x 18 vowels x 10 words x 5 times x 4 languages). The recording was done with a SONY IC Recorder ICD-MS515.

2.4 Acoustic parameters and the method of measurement

The whole vowel was measured with respect of the first and second formant frequencies (F1, F2)¹ and fundamental frequency. In each vowel, the time was normalized at the following points: 0%, 25%, 50%, 75% and 100%. Praat program version 4.2.09 was used for the analysis. Microsoft Excel was used to analyze mean first formant frequencies, mean second formant frequencies, mean duration, and mean fundamental frequency values. The mean fundamental frequency (F0) values were later converted to mean semitone values to eliminate the influence of male and female discrepancies on the values. For the amplitude values, all test tokens were calculated to be 75-87 db. The mean values of the duration and semitone values were also analyzed statistically using t-test, repeated measures ANOVA (RM-ANOVA), and correlation (r) with a 95% level of confidence ($p < .05$). Line graphs were also drawn.

3. Results

3.1 Duration

Table 1 shows the mean duration of nine short and long vowel counterparts in each language.

In Chong, the ratio of short to long vowel counterparts is 1.61-2.19 for the over-sixty group (> 60), 1.78-2.25 for the middle group (35-45), and 1.65-2.08 for the under-twenty group (< 20).

¹ In each language, F1 and F2 values are found to be the same for short and long vowel counterparts.

As a result, it was confirmed that in every language under investigation, the mean durations of the nine short vowels are significantly shorter than those of their long counterparts. The acoustic analysis of the vowel duration in Table 1 indicates that there is a length distinction between the nine short and long vowel counterparts in the three age groups of the four languages.

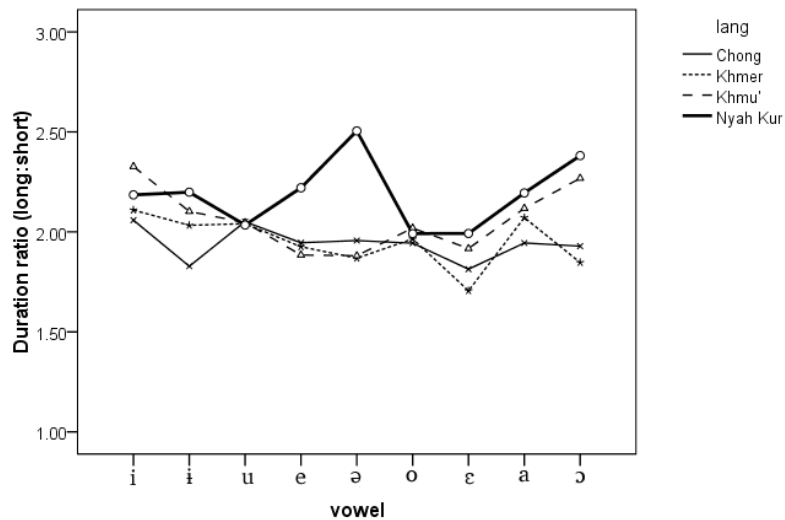


Figure 1. Long:short vowel duration ratio in Chong, Khmer, Khmu', and Nyah Kur

The long:short vowel duration ratio, in Figure 1, is generally larger in Nyah Kur than in the three languages. The difference is significant by RM-ANOVA with language as a factor, $F(3,24) = 2.37$ ($p = .000$). Simple contrasts show that Nyah Kur differs from Khmu', Khmer and Chong, $F(1,8) = 2.52$ ($p = .013$) and $F(1,8) = 2.75$ ($p = .007$), $F(1,8) = 2.60$ ($p = .011$) respectively. No other pairwise contrasts reach significance.

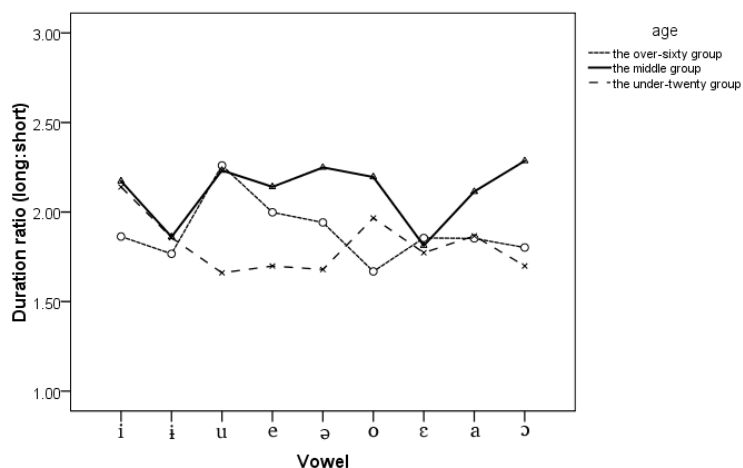


Figure 2. Long:short vowel duration ratio of the three different age groups in Chong

The graph, in Figure 2, plots the duration ratio between long:short vowel counterparts for Chong. Clearly, the RM-ANOVA shows a significant effect of age, $F(2,16)=13.09$ ($p=.000$). It is found that the duration ratio in the middle group are significantly higher than the over-sixty group and the under-twenty group, $F(1,8)=10.53$ ($p=.000$) and $F(1,8)=12.62$ ($p=.000$). The duration ratio in the over-sixty group is also significantly higher than the under-twenty group, $F(1,8)=15.40$ ($p=.000$).

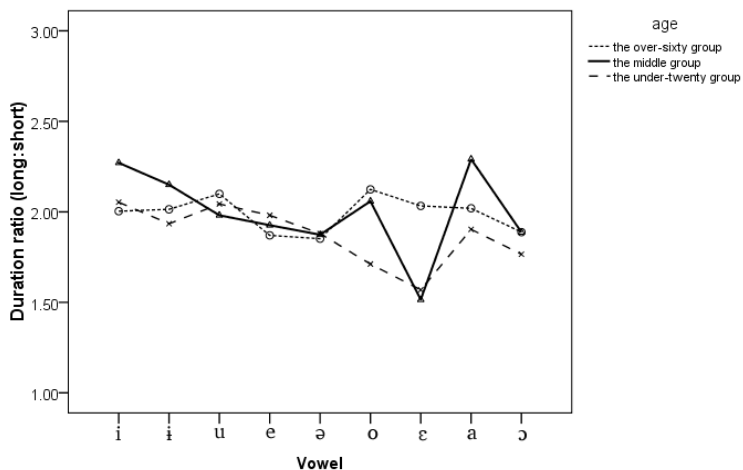


Figure 3. Long:short vowel duration ratio of the three different age groups in Khmer

In Figure 3, the graph plots the duration ratio between long:short vowel counterparts for Khmer. The RM-ANOVA shows an insignificant effect

of age, $F(2,16)=1.24$ ($p=.256$). Vowel type shows significant effect on the duration ratio $F(8,16)=2.92$ ($p=.007$).

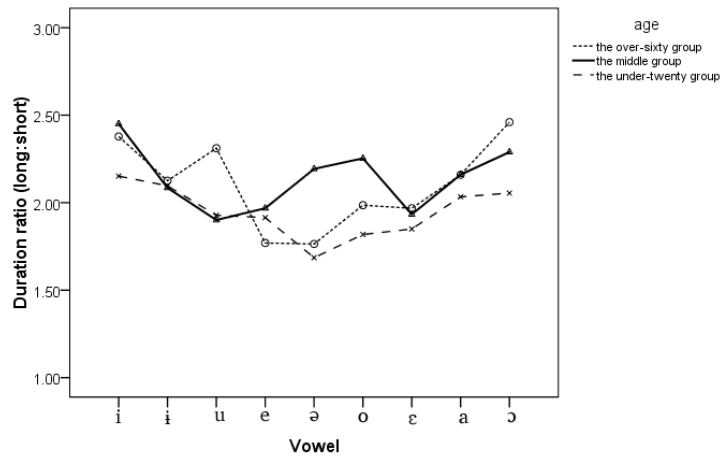


Figure 4. Long:short vowel duration ratio of the three different age groups in Khmu'

In Figure 4, the graph plots the duration ratio between long:short vowel counterparts for Khmu'. The RM-ANOVA shows an insignificant effect of age, $F(2,16)=0.89$ ($p=.583$). Vowel type shows significant effect on the duration ratio $F(8,16)=3.22$ ($p=.003$).

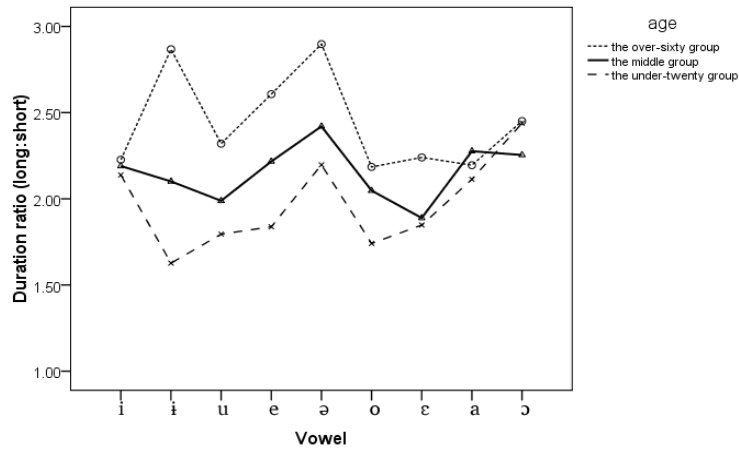


Figure 5. Long:short vowel duration ratio of the three different age groups in Nyah Kur

In Figure 5, the graph plots the duration ratio between long:short vowel counterparts for Nyah Kur. The RM-ANOVA shows an insignificant

effect of age, $F(2,16)=1.15$ ($p=.325$). Vowel type shows significant effect on the duration ratio $F(8,16)=2.39$ ($p=.023$).

3.3 Semitone values

Table 2 shows the semitone values of the nine short and long vowel counterparts in each language.

Table 2. Mean semitone values results of nine short and long vowel counterparts in Chong, Khmer, and Khmu', Nyah Kur

		i	ɨ	u	e	ə	o	ɛ	a	ɔ	
Chong	> 60	s	2.94	3.20	3.36	3.18	3.95	3.71	3.47	3.53	3.34
		l	2.84	2.96	3.06	2.87	3.37	3.17	3.03	3.00	3.00
		s-l ²	0.10	0.24	0.30	0.31	0.58	0.54	0.44	0.53	0.34
		t-test	.195	.040	.193	.014	.000	.000	.000	.000	.000
	35-45	s	3.35	3.69	4.04	3.56	3.85	4.01	3.82	3.61	3.69
		l	3.29	3.87	3.50	3.44	3.61	3.29	3.30	3.05	3.19
		s-l	0.06	-0.18	0.54	0.12	0.24	0.72	0.52	0.56	0.50
		t-test	.490	.038	.000	.019	.002	.040	.000	.000	.000
	< 20	s	4.23	4.11	4.14	3.82	4.07	4.05	4.00	3.86	3.84
		l	3.83	3.89	3.66	3.67	3.63	3.52	3.77	3.48	3.47
		s-l	0.40	0.22	0.48	0.15	0.44	0.53	0.23	0.38	0.37
		t-test	.000	.000	.000	.000	.000	.000	.000	.000	.000
Khmer	> 60	s	3.37	3.50	3.53	3.39	3.07	3.29	3.21	2.98	2.98
		l	3.27	3.59	3.64	3.04	2.90	3.32	3.16	2.72	2.93
		s-l	0.10	-0.09	-0.11	0.35	0.17	-0.03	0.05	0.26	0.05
		t-test	.287	.489	.367	.001	.185	.767	.672	.022	.618
	35-45	s	3.70	3.73	3.79	3.63	3.54	3.29	3.46	2.89	2.89
		l	3.62	3.88	3.86	3.67	3.14	3.46	3.19	3.13	3.16
		s-l	0.08	-0.15	-0.07	-0.04	0.40	-0.17	0.27	-0.24	-0.27
		t-test	.447	.188	.615	.738	.003	.173	.010	.033	.051
	< 20	s	3.09	3.53	3.41	3.16	3.27	3.19	3.05	3.04	2.95
		l	3.21	3.45	3.44	3.11	3.08	3.26	3.20	3.00	3.05
		s-l	-0.12	0.08	-0.03	0.05	0.19	-0.07	-0.15	0.04	-0.10
		t-test	.446	.633	.838	.732	.135	.594	.242	.720	.413
Khmu'	> 60	s	4.00	4.20	4.05	3.43	3.58	3.37	3.28	3.27	3.28
		l	4.07	3.69	3.72	3.30	3.18	3.27	2.99	2.92	3.09
		s-l	-0.07	0.51	0.33	0.13	0.40	0.1	0.29	0.35	0.19
		t-test	.283	.000	.000	.120	.000	.255	.001	.000	.033
	35-45	s	3.75	3.86	3.77	3.20	3.15	3.23	3.33	3.10	3.12
		l	3.52	3.69	3.55	3.19	3.27	3.09	3.04	2.97	3.03
		s-l	0.23	0.17	0.22	0.01	-0.12	0.14	0.29	0.13	0.09
		t-test	.048	.131	.056	.870	.186	.143	.001	.184	.308
	< 20	s	4.08	4.30	4.32	3.60	3.63	3.66	3.42	3.52	3.49
		l	4.92	4.32	4.33	3.61	3.48	3.45	3.45	3.31	3.39
		s-l	-0.84	-0.02	-0.01	-0.01	0.15	0.21	-0.03	0.21	0.10
		t-test	.000	.871	.964	.902	.082	.010	.724	.037	.305
Nyah Kur	> 60	s	3.75	4.65	4.90	4.21	3.78	3.76	4.04	3.60	3.86
		l	3.98	4.12	4.03	3.84	3.82	3.74	3.85	3.35	3.53
		s-l	-0.23	0.53	0.87	0.37	-0.04	0.02	0.19	0.25	0.33
		t-test	.031	.000	.000	.001	.666	.823	.036	.003	.001
	35-45	s	4.15	4.54	4.59	4.31	3.51	3.82	3.82	4.33	4.31
		l	4.54	4.17	4.22	3.77	4.05	3.85	3.47	4.09	3.85
		s-l	-0.39	0.37	0.37	0.54	-0.54	-0.03	0.35	0.24	0.46
		t-test	.052	.027	.049	.002	.001	.877	.036	.072	.008
	< 20	s	5.14	5.34	5.25	4.97	4.90	4.99	4.93	4.83	4.83
		l	5.04	5.49	5.14	4.95	4.87	4.78	4.83	4.64	4.71
		s-l	0.10	-0.15	0.11	0.02	0.03	0.21	0.10	0.19	0.12
		t-test	.179	.047	.060	.232	.230	.000	.000	.000	.000

² s represents a short vowel, l represents a long vowel, and s-l represents the size of semitone difference between short vowels and long vowels.

It is shown from Table 2 that some semitone values of the nine short vowels are significantly higher than those of their long counterparts. However, some semitone values of the nine short vowels are significantly lower than those of their long counterparts.

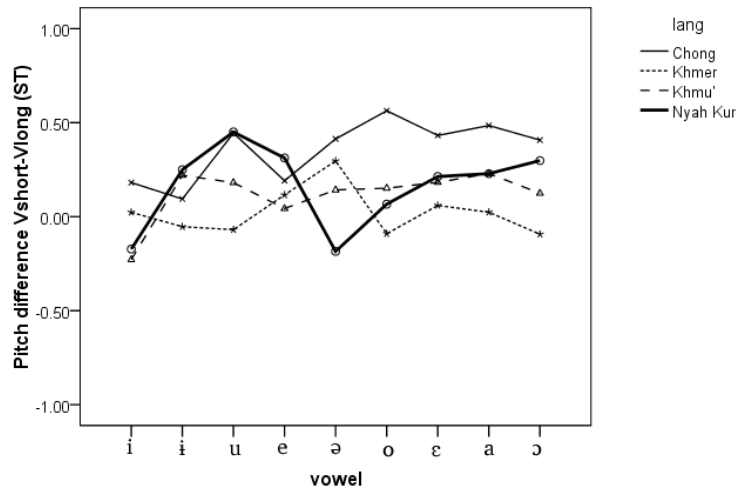


Figure 6. The size of semitone difference in Chong, Khmer, Khmu', and Nyah Kur

From Figure 6, the pitch differences are generally larger in Chong than in the three other languages. The difference is significant by RM-ANOVA with language as a factor, $F(3,24) = 1.78$ ($p = .015$). Simple contrasts show that Chong differs from Nyah Kur and Khmer, $F(1,8) = 2.25$ ($p = .027$) and $F(1,8) = 3.30$ ($p = .002$) respectively. No other pairwise contrasts reach significance.

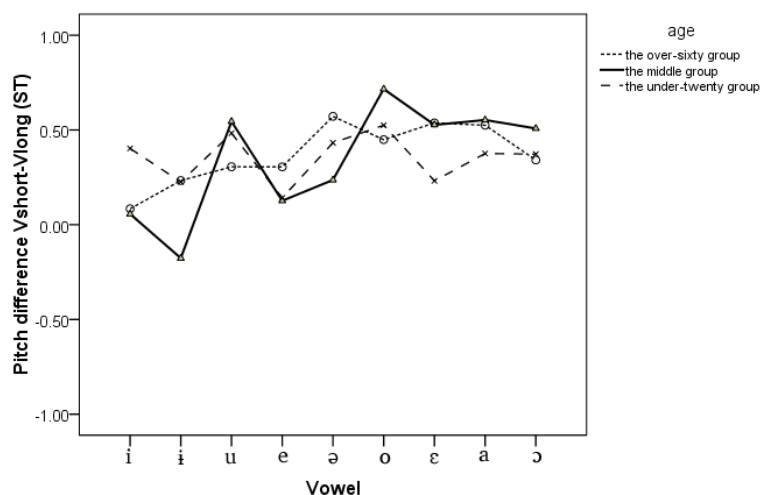


Figure 7. The size of semitone difference of the three different age groups in Chong

The graph, in Figure 7, plots the pitch differences (short vowel minus long vowel) between short and long vowel counterparts for Chong. The short vowels have higher pitch than long vowels in the three age group of Chong, except for the vowel [i] in the middle group. The RM-ANOVA shows a significant effect of age, $F(2,16)=2.39$ ($p=.006$). There is no contrast in pitch differences among the three age groups, except for the pitch differences in the middle group which are significantly higher than the under-twenty group $F(1,8)=9.89$ ($p=.000$). There is significant effect of vowel type $F(8,16)=7.02$ ($p=.000$) on pitch differences.

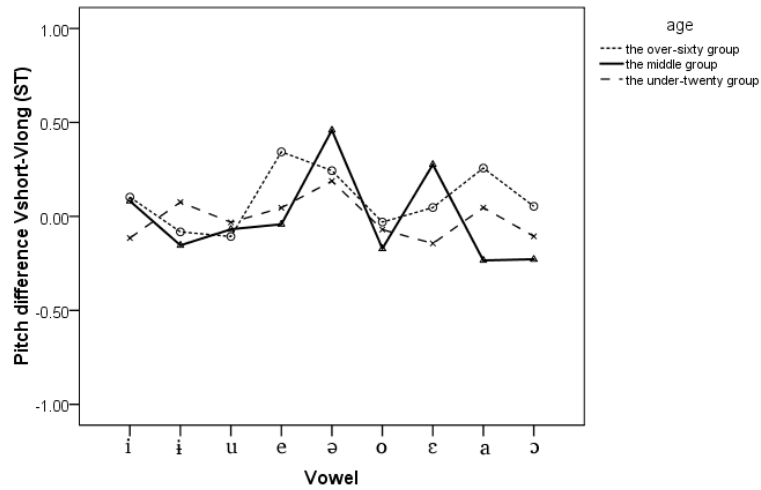


Figure 8. The size of semitone difference of the three different age groups in Khmer

In Figure 8, the graph plots the pitch differences (short vowel minus long vowel) between short and long vowel counterparts of Khmer. In the three age groups, some short vowels have higher pitch than long vowels, whilst some short vowels do not have higher pitch than long vowels. The RM-ANOVA shows an insignificant effect of age, $F(2,16)=0.79$ ($p=.689$), so there is no contrast in pitch differences among the three age groups.

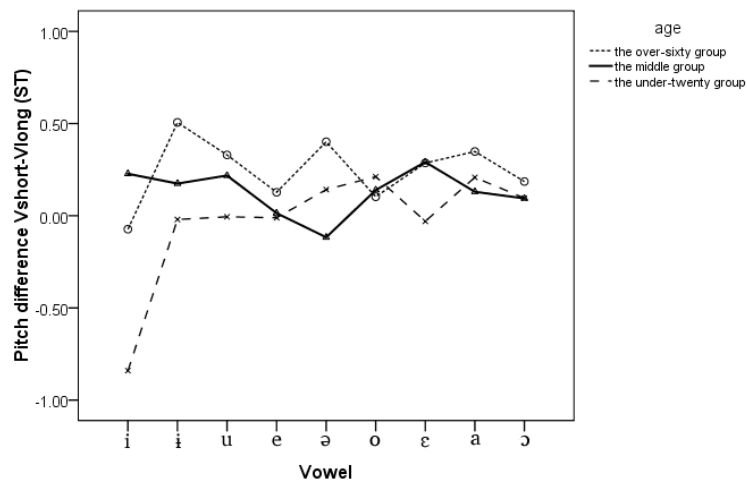


Figure 9. The size of semitone difference of the three different age groups in Khmu'

In Figure 9, the graph plots the pitch differences between short and long vowel counterparts of Khmu'. The vowel [i] in the over-sixty group and vowel [ə] in the middle group show that short vowels have lower pitch than its counterpart. Most of the short vowels in the under-twenty group have lower pitch than their long vowel counterparts. The RM-ANOVA shows an insignificant effect of age, $F(2,16) = 0.90$ ($p = .571$).

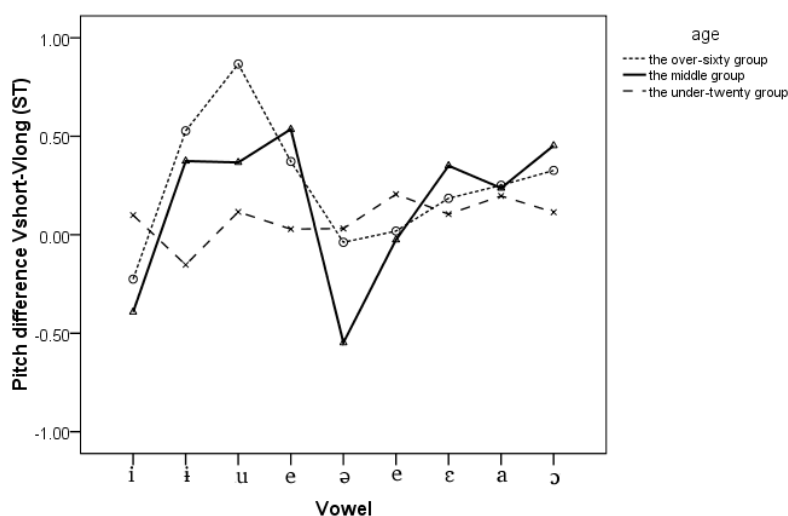


Figure 10. The size of semitone difference of the three different age groups in Nyah Kur

In Figure 10, the graph plots the pitch differences between short and long vowel counterparts of Nyah Kur. In the three age groups, some short vowels have higher pitch than long vowels, whilst some short vowels do not have higher pitch than long vowels. The RM-ANOVA shows an insignificant effect of age, $F(2,16) = 0.76$ ($p = .131$).

3.4 The interaction between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels

The relationship between pitch and vowel length of each language was studied in order to confirm the interaction between long:short vowel duration ratio and the size of semitone difference between long and short vowel. In each language, the correlation coefficients for each age group were computed.

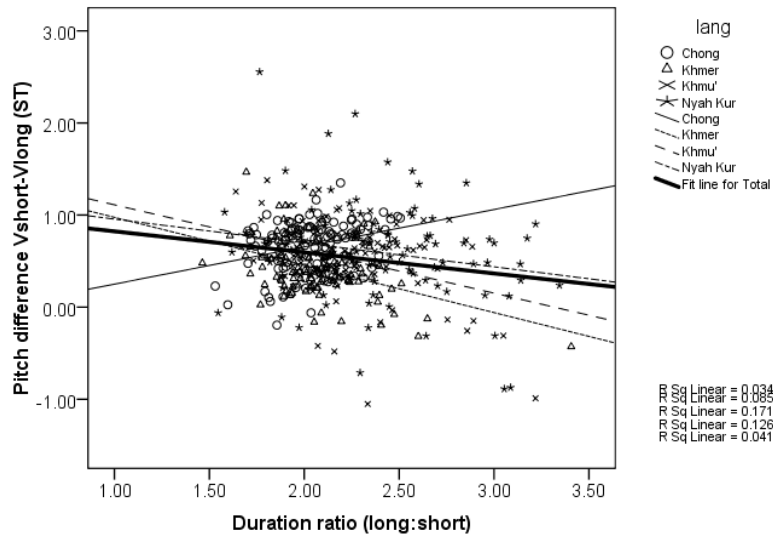


Figure 11. The relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Chong, Khmer, Khmu', and Nyah Kur

As can be seen from Figure 11, among the four languages under investigation, Chong is the only language which shows a positive correlation between the long:short vowel duration and the size of short and long vowel semitone difference.

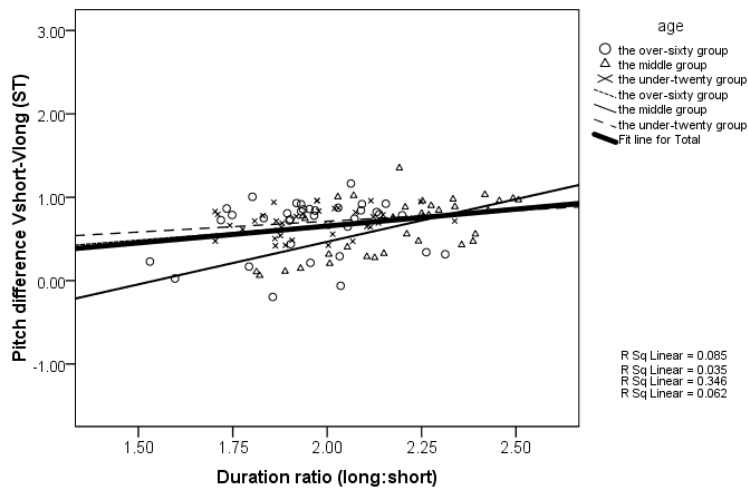


Figure 12. The relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Chong

Figure 12 shows the relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of Chong. The overall correlation is small, $r = 0.291$. When computing the correlation coefficients for each age group separately, they are practically small with the exception of the middle age group ($r=0.588$). So statistically, there is small correlation between the size of the duration contrast and the size of the semitone difference between long and short vowels ($r=0.187$ for the over-sixty group, $r=0.249$ for the under-twenty group).

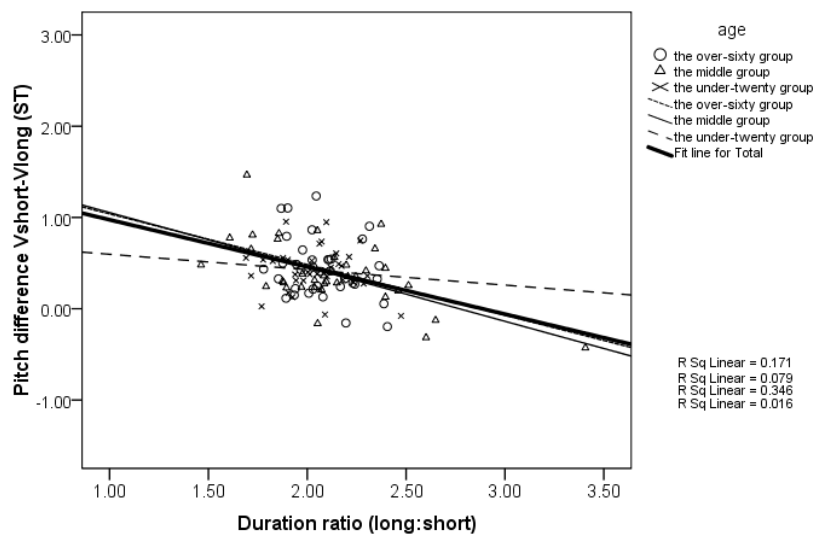


Figure 13. The relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Khmer

The graph in Figure 13 shows the relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Khmer. The overall correlation is medium and negative, $r=-0.414$. When computing the correlation coefficients for each age group, they are small negative correlation between the size of the duration contrast and the size of the pitch difference between long and short vowels in Khmer ($r=-0.281$ for the over-sixty group, $r=-0.127$ for the under-twenty group), with an exception of the middle group which shows large negative correlation ($r=-0.546$).

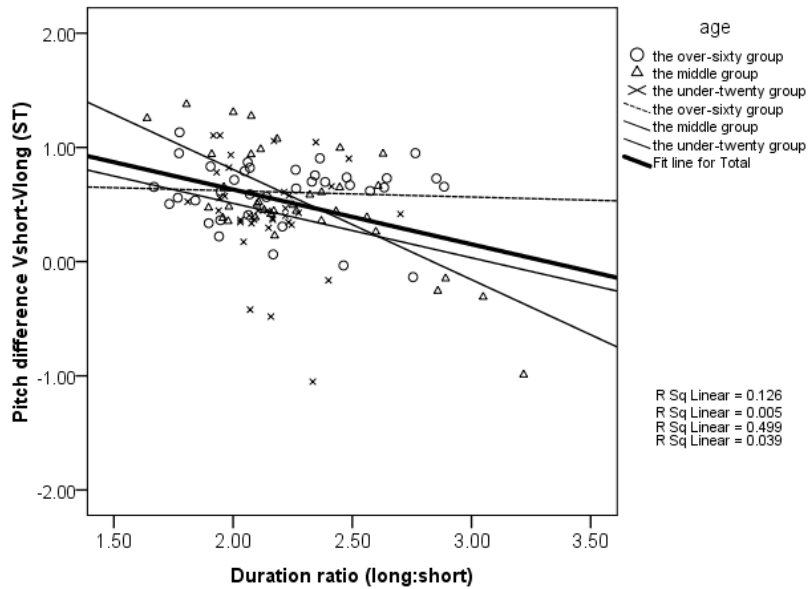


Figure 14. The relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Khmu'

Figure 14 shows the relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Khmu'. The overall correlation is negatively small, $r = -0.355$. When computing the correlation coefficients for each age group, there is no correlation between the size of the duration contrast and the size of the pitch difference between long and short vowels for the over-sixty group ($r = -0.068$) and small negative correlation for the under-twenty group ($r = -0.198$), with an exception of the middle group which shows large negative correlation ($r = -0.706$).

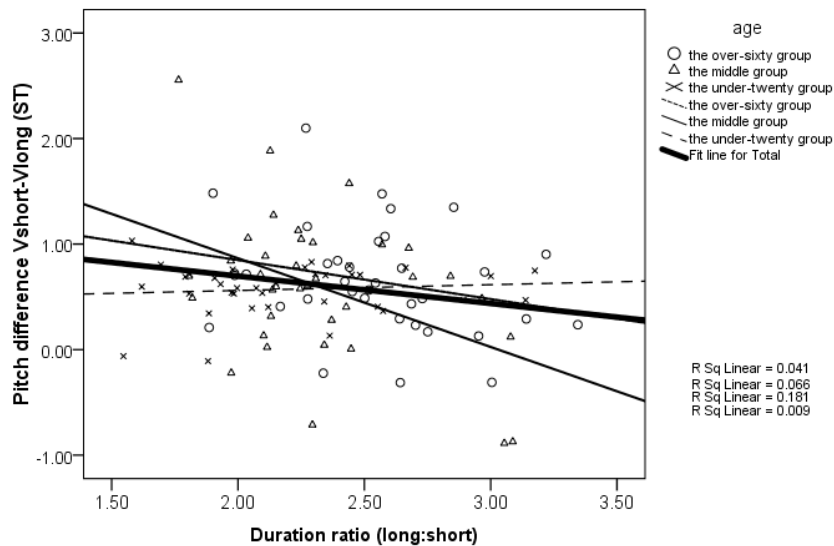


Figure 15. The relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Nyah Kur

Figure 15 shows the relationship between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels of the three age groups in Nyah Kur. The overall correlation is negative and small, $r = -0.202$. When computing the correlation coefficients for each age group, there is no correlation for the under-twenty group ($r = 0.097$) and a small negative correlation for the over-sixty group ($r = -0.257$), with an exception of the middle group which shows medium negative correlation ($r = -0.426$) between the size of the duration contrast and the size of the pitch difference between long and short vowels.

4. Discussion

In order to confirm a relationship between vowel duration on the semitone values, four languages of Mon-Khmer that contain nine short and long vowels counterparts—[i]-[ii], [ɨ]-[iɨ], [u]-[uu], [e]-[ee], [ə]-[əə], [o]-[oo], [ɛ]-[ɛɛ], [a]-[aa], [ɔ]-[ɔɔ]—were analyzed. As the results show significant differences in duration for every pair of short and long counterparts, it is argued that vowel length contrast is determined by vowel duration. This corresponds with the results found in Kunnama, which its where vowel length contrast is based only on duration (Connell, 2002).

Across the four languages, long:short vowel duration ratio is generally larger in Nyah Kur than the other three languages. However, when considering each language, only Chong shows a significant effect of age on

duration ratio. The long:short vowel duration in the middle age group is significantly higher than in the over-sixty group and in the under-twenty group.

The pitch differences are larger in Chong than in the three other languages. In each language only Chong shows a significant effect of age on pitch differences. It is shown the middle age group to have significantly larger pitch differences than the under-twenty group. Therefore, Chong seems to be the language which shows a cross-sectional design comparing the over-sixty speakers (> 60 years), the middle age speakers (35-45 years), and the under-twenty speakers (< 20 years). The youngest generation of Chong shows smaller duration differences than that of the over-sixty group and of the middle age group. However, the middle age group shows the largest duration ratio.

Table 2 shows that not all short vowels have higher semitone values than the long vowels. Across the four languages, Chong shows the most consistency of the short vowel to have higher semitone values than the long vowel. Moreover, the middle age group of Chong shows significantly larger pitch difference than the under-twenty group. Also, Chong is the only language that shows positive correlations between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels (See Figure 10). The results indicate that the effect of vowel duration on semitone values is applicable to most of the vowel qualities in Chong; however, the effect is not applicable to Khmer, Khmu' and Nyah Kur. This indicates that there might be other factor, besides vowel length, causing tonogenesis in Khmer, Khmu', and Nyah Kur. The results in Chong are congruent with the results reported in Hu (Svantesson, 1989), Kui (Suai) and Chong (L-Thongkum, 1989, 1991), where the F0s of all short vowel qualities are found to be higher than those of their long vowel counterparts. In contrast, in Khmer, Khmu' and Nyah Kur, it is found that not all short vowels have higher mean semitone values than do long vowels with small negative correlations between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels. Similar results are found in Battambang Khmer vowels (Wayland, 1998).

The three languages, Khmer, Khmu', and Nyah Kur, haven't shown any longitudinal indications as expected because age groups hasn't shown a significant effect. Whilst Chong shows longitudinal evidence; but the middle age group and the over-sixty group are found to have a larger duration and semitone differences than that of the under-twenty group. The youngest group of Chong shows smaller duration ratio and smaller semitone differences. One possible explanation of this crucial point is that Chong is classified as an endangered language; most of the under twenty years of age do not speak the Chong (Premsrirat, 2002). Those youngest speakers study Chong under a revitalization project. They are not truly native speakers of Chong anymore. Therefore, the result of longitudinal test of Chong is reasonably expected to be different from the three languages.

Each language shows that the correlations between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels are large for the middle age group. Negative correlations found across the three age groups in the three languages, Khmer, Khmu', and Nyah Kur show that, there are other factors conditioning tonal evolution in these three languages—e.g., initial consonant voicing, phonation types—it is suggested as plausible that vowel length is not a factor that conditions pitch in the Khmer, Khmu', and Nyah Kur.

As a result, the interaction between semitone values and vowel duration found in this study does not point to a natural phonetic tendency. The inconsistent results regarding vowel length as a condition of pitch, as found in this study, implicitly support the findings that correlations between vowel length and pitch are widespread cross-linguistically but are not universal. The results show a phonetic tendency—that the loss of vowel length has to do with the evolution of tone in Chong but not in Khmer, Khmu' and Nyah Kur. In other words, short vowels tend to originate high pitch and long vowels tend to condition low pitch in Chong. It is also clear that Chong is the only language under investigation which shows a significant effect of age on vowel duration and the size of semitone differences.

Regarding factors that cause tonal evolution (e.g., the loss of initial consonant voicing, the loss of phonation types, the loss of vowel length etc.), Chong shows a tendency to change to tonal language through, besides the loss of phonation types (L-Thongkum, 1988, 1991; Premsrirat, 2003), the loss of vowel length. In contrast, Khmer, Khmu' and Nyah Kur, show a tendency to change to tonal language not through the loss of vowel length but, rather, through the loss of phonation types (L-Thongkum, 1988) or the loss of consonant clusters (Henderson, 1982; Teeranon, 2008; Wayland and Guion, 2005).

Of the four languages, Chong, Nyah Kur, and Khmu' are classified as register languages (L-Thongkum, 1984, 1988, 1991; Premsrirat, 2003), while Khmer is classified as a restructured language (Huffman, 1967). A dialect of Khmer called Thung Kabin Khmer is found to preserve clear voice and breathy voice in vowels (L-Thongkum, 1991), consistent with the idea that Khmer's having had phonation types in the past. In other words, the four languages are register languages that have a tendency to develop phonation types into tones (becoming tonal languages) or to develop a large number of vowels (becoming restructured languages). In terms of register language, the term register is defined as a lexically contrastive register complex (a combination of phonation type, pitch, vowel length, vowel quality, etc.) (L-Thongkum, 1988: 319). Based on the present acoustic results, the four languages seem to have different prominent phonetic correlates. In Chong, the interaction between semitone values and vowel duration is a prominent phonetic correlate. This supports the findings of L-Thongkum (1988, 1991): that one of a prominent phonetic

correlate in Chong and Nyah Kur is the F0 (or semitones or pitch). However, the interaction between semitone values and vowel duration in Khmer is not a prominent phonetic correlate in their register complex. It is found, in Khmer, to have the largest negative correlations between the long:short vowel duration ratio and the size of the semitone difference between long and short vowels. The different results found in the four languages seem to divide the four languages into two groups; the first group is Chong, Nyah Kur, and Khmu', whereas the second group is Khmer. These divisions are congruent with the different pathways of originating tones reported in previous research; Chong, Nyah Kur, and Khmu' develop tones from the phonation types of vowels (phonation types > tones), while Khmer has chosen to develop its phonation types into a system with a large number of vowels (phonation types > a large number of vowels), the so-called restructured language.

5. Conclusions

The vowel durations condition semitone values are found to be applicable to all vowel qualities in Chong; however, the effect is applicable to only some vowel qualities in Khmer, Khmu' and Nyah Kur. The inconsistent results regarding vowel length as a condition of pitch found in this study implicitly support the findings that correlations between vowel length and pitch are widespread cross-linguistically but are not universal. The effect of pitch as conditioning vowel length tends to be a natural phonetic tendency, but not vice versa. The results show a phonetic tendency regarding the connection between the loss of vowel length and the evolution of tones in Chong, but in the same is not true for Nyah Kur and Khmer.

Acknowledgements

I would like to express my gratitude to the Thailand Research Fund (TRF) for the funding of this research project during a two-year period. I am also grateful to my mentor, Prof. Dr. Theraphan Luangthongkum, for her insightful and valuable comments.

REFERENCES

- Abramson, A. S., 1962. The vowels and tones of standard Thai: Acoustical measurements and experiments. Indiana University Research Center in Anthropology, Folklore, and Linguistics, Publication 20, Bloomington.
- Atkinson, J. E., 1973. Aspects on Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. Unpublished Ph.D. dissertation, University of Connecticut.
- Chantrupanth, D., Phromjakgarin, C., 1978. Khmer (Surin)-Thai-English Dictionary. Indigenous Languages of Thailand Research Project, Chulalongkorn University Language Institute, Bangkok.

- Connell, B., 2002. Tone languages and the universality of intrinsic F_0 : evidence from Africa. *Journal of Phonetics* 30, 101-129.
- Diehl, R. L., 1991. The role of phonetics within the study of language. *Phonetica* 48, 120-134.
- Diehl, R. L., Kluender, K. R., 1989. Reply to commentators. *Ecological Psychology* 1, 195-225.
- Diffloth, G., 1991. Paluangaic vowels in Mon-Khmer perspective. In: Davidson, J. H. C. S. (Ed.), *Austroasiatic Languages, Essays in Honour of H. L. Shorto*, School of Oriental and African Studies, London, pp. 13-27.
- Fischer-Jørgensen, E., 1990. Intrinsic F_0 in tense and lax vowels with special reference to German. *Phonetica* 47, 99-140.
- Gandour, J., 1977. On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica* 34, 54-65.
- Heijmans, L., 2003. The relationship between tone and vowel length in two neighboring Dutch Limburgian dialects. In: Fikkert, P., Haike, J. (Eds.), *Development in Prosodic Systems*. Mouton de Gruyter, Berlin, 7-45.
- Henderson, E. J. A., 1982. Tonogenesis: some recent speculations on the development of tone. *Philological Society*, 1-24.
- Honda, K., Fujimura, O., 1991. Intrinsic vowel F_0 and phrase-final F_0 lowering: Phonological vs. biological explanations. In: Gauffin, J., Hammarberg, B. (Eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*. Singular Publishing Group, San Diego, 149-157.
- Hoole, P., Honda, K., Murano, E., Fuchs, S., Pape, D., 2004. Cricothyroid activity in consonant voicing and vowel intrinsic pitch. *Proceedings of the Conference on Vocal Fold Physiology and Biomechanics*, Marseille, France, pp .
- House, A. S., Fairbanks, G., 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 105-113.
- Huffman, F. E., 1967. *An Outline of Cambodian Grammar*. Unpublished Ph.D. Dissertation, Cornell University.
- Kong, Q., 1987. Influence of tones upon vowel duration in Cantonese. *Language and Speech* 30.4, 387-399.
- Kwon, K-K., 2003. Prosodic change from tone to vowel length in Korean. In: Fikkert, P., Haike, J. (Eds.), *Development in Prosodic Systems*. Mouton de Gruyter, Berlin, pp. 67-89.
- Lehiste, I., 1970. *Suprasegmentals*. MIT Press, Cambridge.
- Lee, G. M., 1993. *Comparative, Diachronic, and Experimental Perspectives on the Interaction between Tone and Vowel in Standard Chinese*. Department of Linguistics, Columbus, The Ohio State University.
- Lohde, K., 2003. *A Comparison of the Interaction between the Fundamental Frequency and Duration of Vowels and Final Consonants in Pattani Malay Spoken in the Provinces of Pathumthani and Pattani: An Acoustic Study*. Unpublished M.A Thesis, Department of Linguistics, Chulalongkorn University.

- L-Thongkum, T., 1984. Nyah-Kur (Chaobon)-Thai-English Dictionary. Chulalongkorn University Printing House, Bangkok.
- L-Thongkum, T. 1988. Phonation types in Mon-Khmer languages. In: Fujimura, O. (Ed.), *Vocal Fold Physiology 2: Voice Production, Mechanisms and Functions*. Raven Press, New York, pp. 319-333.
- L-Thongkum, T., 1989. An acoustic study of the register complex in Kui (Suai). *Mon-Khmer Studies* 15, 1-19.
- L-Thongkum, T. 1991. An instrumental study of Chong registers. In: Davidson, J. H. C. S. (Ed.), *Austroasiatic Languages, Essays in Honour of H. L. Shorto*. School of Oriental and African Studies, London, pp. 141-160.
- L-Thongkum, T., Akharawatthanakun, P., Intajamornrak, C., Putthasatien, K., Phalipat, S., Huadsiri, Y., 2008. *Dictionary of the 15 Languages Spoken in Nan Province*. Thailand Research Fund, Bangkok.
- L-Thongkum, T., Teeranon, P., Intajamornrak, C., 2007. The interaction between vowel length and pitch in SEA languages: An implication for tonal evolution. In: Harris, J. G., Burusphat, S., Harris, J. E. (Eds.), *Studies in Tai and Southeast Asian Linguistics*. Ekphimthai, Bangkok, pp. 225-240.
- Maspero, G. 1915. *Grammaire Khmère*. Paris: Librairie nationale.
- Mohr, B., 1971. Intrinsic variations in the speech signal. *Phonetica* 23, 65-93.
- Myers, S., 2005. Vowel duration and neutralization of vowel length contrasts in Kinyarwanda. *Journal of Phonetics* 33, 427-446.
- Newklowksy, G., 1975. Spezifische Dauer und Spezifische Tonhöhe der Vokale. *Phonetica* 32, 38-60.
- Ohala, J. J., 1972. How is pitch lowered? (Abstract). Paper presented at the 83rd Meeting of Acoustical Society of America, Buffalo, New York, April 18-21, 1972.
- Pike, E. V., 1974. A multiple stress system versus a tone system. *International Journal of American Linguistics*. 40.3, 169-175.
- Premrirat, Suwilai. 2002. Report on the Chong Revitalization Workshops. *Mon-Khmer Studies* 32, 175-179.
- Premrirat, S., 2003. Khmu dialects: A case of register complex and tonogenesis. In: Shigeki, K. (Ed.), *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena: Historical Development, Phonetics of Tone and Descriptive Studies*, ILCAA, University of Foreign Studies, Tokyo, pp. 13-28.
- Sarawit, M., 1973. *The Proto-Thai Vowel System*. Unpublished Ph.D. dissertation, University of Michigan, Ann Arbor.
- Svantesson J-O., 1991. Hu-A language with unorthodox. In: Davidson, J. H. C. S. (Ed.), *Austroasiatic Languages, Essays in Honour of H. L. Shorto*. School of Oriental and African Studies, London, pp. 72-79.
- Teeranon, P., 2007., Intrinsic pitch and intrinsic length in Waic languages. *Mon-Khmer Studies* 37, 41-52.
- Teeranon, P., 2008. Tonal evolution from the influence of initial consonant clusters: Another look from Lavua', Wa, and Plang. In Ronnakiat, N., Saengmanee, A. (Eds.), *A Collection of Essays on Southeast Asian*

- Linguistics: Festschrift for Professor Theraphan Luangthongkum's 60 Birthday. Phikkhanet Printing Center, Bangkok, pp. 30-42. (in Thai)
- Tsukada, K., Burnham, D., Luksaneeyanawin, S., Krachaikiat, N., Rungrojsuwan, S., 2004. The effect of tone on vowel duration in Thai: A development study. Paper presented at the 9th Conference on Laboratory Phonology, June 24-26, 2004.
- Wayland, R., 1998., An acoustic study of Battambang Khmer vowels. *Mon-Khmer Studies* 28, 43-62.
- Wayland, R., Jongman, A., 2003. Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics* 31, 181-201.
- Wayland, R. P., Guion, S. G., 2005. Sound changes following the loss of /r/ in Khmer: A new tonogenetic Mechanism? *Mon-Khmer Studies* 35, 55-82.
- Whalen, D. H. Levitt, A. G., 1995. The universality of intrinsic F₀ of vowels. *Journal of Phonetics* 23, 349-366.
- Whalen, D. H., Gick, B., Kumada, M., Honda, K., 1998. Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F₀. *Journal of Phonetics* 27.2, 125-142.
- Yu, A. C. L., 2006. Total effects on perceived vowel duration. Paper presented at the LabPhon 10, Paris, France, June 29-July 1, 2006.

Received: 9 August 2010

Mae Fah Luang University
Chiang Rai, Thailand
E-mail address: tphanintra@hotmail.com

A comparison of sequential strategies in Mon-Khmer narratives

SOMSONGE Burusphat
Mahidol University, Thailand

Abstract

Contingent temporal succession is a significant feature of narrative. It is marked by various devices in different languages. This paper presents the sequential devices which mark the storyline of narrative texts. The data were drawn from written texts. They consist of twenty-six narrative texts of five Mon-Khmer languages, namely, Samre, Kasong, Pray, Kmhmu, and Vietnamese. These texts are analyzed using the framework of discourse analysis as expounded by Longacre (1996). The analysis is focused on the etic narrative salience scheme.

Keywords: Mon-Khmer, sequential strategies, narrative

1. Introduction

Narrative discourse is characterized by two basic etic parameters, contingent temporal succession and agent orientation defined as follows:

To begin with, we can classify all possible discourses in all languages according to two basic etic parameters: contingent temporal succession and agent orientation. Contingent temporal succession refers to a framework of temporal succession in which some of the events or doings are contingent on previous events or doings. Agent orientation refers to orientation towards agents with at least a partial identity of agent reference running through the discourse. These two parameters intersect so as to give us a four-way classification of discourse types: Narrative discourse is plus in respect to both parameters. Procedural discourse is plus in respect to contingent succession but minus in respect to the agent orientation. Behavioral discourse is minus in regard to contingent succession but plus in regard to agent orientation. Expository discourse is minus in respect to both parameters (Longacre 1996:8-9).

This paper¹ focuses on the contingent temporal succession of Mon-Khmer narratives. It aims to find sequential strategies used to identify this parameter of Mon-Khmer narratives.

2. Methodology

The Mon-Khmer narratives are all folktales. They were collected from various sources as listed in table 1.

Table 1. Sources of Mon-Khmer narratives

Languages	Location	Number of texts	Sources
Samre	Bo-rai District, Trat Province, Thailand	3	Pornsawan (2001)
Kasong	Bo-rai District, Trat Province, Thailand	5	Sunee (2002)
Pray	Thungchang District, Nan Province, Thailand	5	Mingkwan (1989)
Kmhmu' Am/Kmhmu' Ou/ Khum Ceuang	Vientiane, Luang Prabang, Chiang Khwang Lao PDR	7	Steeb (2008) Osborne (2009)
Vietnamese	Vietnam ²²	6	Suksiri (2000)

The Mon-Khmer languages under this study are classified within the Austroasiatic family as in figure 1.

¹ I would like to thank Richard Hiam for helping with the first draft of this paper. My thanks are also extended to Suksiri Danthanavanich for transcribing the folktales written in Vietnamese into phonetic symbols.

² The Vietnamese narratives were drawn from *Kho tàng Truyền Cổ Tích Việt Nam* (A Collection of Vietnamese folktales) by Nguyễn Văn Ngọc, Vũ Ngọc Phan, and Trần Thanh Mại (1998)

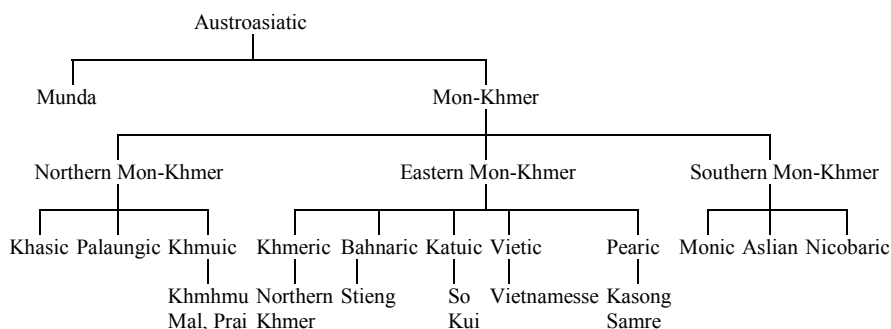


Figure 1. Classification of Mon-Khmer languages
(Adapted from Diffloth 1974 and Sidwell 2005)

The data were interpreted following a functional-typological approach, which includes Tagmemics. The discourse analysis is based on the Longacre's system.

The following steps were taken to analyze the data:

1. Find the sequential strategies focusing on the temporal auxiliaries and aspectual auxiliaries of Mon-Khmer narrative texts.
2. Compare the forms of temporal auxiliaries and aspectual auxiliaries of Mon-Khmer languages.
3. Contextualize these auxiliaries in discourse environments to find their discourse functions.
4. Compare the discourse functions of the temporal auxiliaries and aspectual auxiliaries based on the framework of salience scheme as discussed below.

3. Theoretical framework

Longacre (1986:2) proposed that for any language every type of text has a mainline of development and subsidiary material which can be conceived of as progressive degrees of departure from the mainline. Since the mainline of development outranks the subsidiary material in the structure of a discourse, both of them could be ranked internally as well. The term 'spectrum' is proposed as a metaphor for the internal ranking of verbs and clauses:

Just as a spectrographic analysis of white light separates out:

Various hues (our perception of differing wave lengths) ranging from red to violet, so the analysis of a narrative text reveals a cline of information which ranges from the most dynamic elements of the story to the most static

(depictive) elements; successive positions along the cline correlate well (as a whole) with distinctions among the verb forms of a language (i.e., with the tense/aspect/mode/voice system), but other features (word order, use of affixes, particles, or adverbs) must sometimes be invoked to round out the picture (Longacre 1981:340)

Based on this concept, Longacre (1996:28) proposed nine etic bands for narrative texts as shown in table 2.

Table 2. An etic narrative salience scheme (simplified from Longacre 1996: 28)

Band 1 Storyline	Past (S/Agent) Action, (S/Agent/Patient) Motion Past (S/Experiencer) Cognitive events (punctiliar adverbs) Past (S/Patient) Contingencies
Band 2 Background	Past Progressive (S/Agent) background activities Past (S/Experiencer) Cognitive states (durative adverbs)
Band 3 Flashback	Pluperfects (events, activities, which are out of sequence) Pluperfects (Cognitive events/states that are out of sequence)
Band 4 Setting	Stative verbs/adjectival predicates/verbs with inanimate subjects (descriptive) “Be” verbs/verbless clauses (equative) “Be”/“Have” (existential, relational)
Band 5 Irrealis	Negatives Modal/futures
Band 6 Evaluation (author intrusion)	Past tense Gnomic present
Band 7 Cohesive band (verbs in preposed/adverbial clauses)	Script determined Repetitive Back Reference

Band 1 features storyline which is a main line of development. It includes any happenings that push the story forward. The important characteristics of these happenings are punctuality, sequentiality, and (most often) volition. Punctiliar happenings include actions and events that are well articulated as to inception, terminus, or both. The sequential happenings involve the chronological succession of actions and events. Volitional happenings are actions or events that are conscious or planned (Longacre 1990).

Bands 2-7 are nonstoryline elements or non events which are supportive material. Band 2 consists of background activities and cognitive states which may temporally overlap with the storyline happenings or add more detail to the storyline. Band 3 is flashback which is similar to storyline band except for the lack of sequentiality, that is it is out of the narrative sequence. Band 4 is setting which involves expository and descriptive material. It tells when and where an event takes place. The irrealis Band consists of elements which is not part of textworld but suggests all possible events and alternatives (Longacre 1989). The next band is evaluation band which is more optional and even unnecessary to a narrative than either setting or irrealis (where alternative worlds are at least tangent to the text world)" (Longacre 1989). The final band has cohesion of various sorts. It includes such elements as adverbial clauses (script determined/repetitive back reference) and relative clauses (repetitive back reference).

Table 3. An analysis of an illustrative English narrative text using the etic narrative salience scheme

Text	The etic narrative salience scheme
Harry <u>couldn't</u> feel too excited about this.	Band 5: Irrealis (Negative)
He <u>didn't</u> think the Dursleys would like him any better in Majorca than they did in Privet Drive.	Band 5: Irrealis (Negative)
'Right – I'm off into town to pick up the dinner jackets for Dudley and me. And you,' he <u>snarled</u> at Harry, 'you stay out of your aunt's way while she's cleaning.'	Band 1: Past (S/Agent) Action (speech act)
Harry <u>left</u> through the back door.	Band 1: Past (S/Agent/Patient) Motion
It <u>was</u> a brilliant, sunny day.	Band 4 Setting (Stative verbs)
He <u>crossed</u> the lawn, <u>slumped</u> down on the garden bench	Band 1: Past (S/Agent/Patient) Motion
and <u>sang</u> under his breath, 'Happy birthday to me...happy birthday to me...'	Band 1: Past (S/Agent) Action (speech act)
<u>No</u> cards, <u>no</u> presents, and he <u>would</u> be spending the evening pretending not to exist.	Band 5: Irrealis (Negative-Modal)
He <u>gazed</u> miserably into the hedge.	Band 1: Past (S/Agent) Action
He <u>had</u> never <u>felt</u> so lonely.	Band 3: Pluperfects (Cognitive states that are out of sequence)

More than anything else at Hogwarts, more even than playing Quidditch, Harry <u>missed</u> his best friends, Ron Weasley and Hermione Granger.	Band 1: Past (S/Experiencer) Cognitive events
They, however, <u>didn't</u> seem to be missing him at all.	Band 5: Irrealis (Negative)
Neither of them <u>had written</u> to him all summer, even though Ron <u>had said</u> he was going to ask Harry to come and stay	Band 3: Pluperfects (Cognitive states that are out of sequence)

To illustrate an analysis of English narrative texts using this etic narrative salience scheme, an excerpt of *Harry Potter and the Chamber of Secrets* (Rowling 1998: 11) is analyzed as seen in table 3. The underlined words identify each band of the etic narrative salience scheme.

Longacre (1996) mentioned that storyline is the main line of development which is foregrounded in varying ways in various languages. In languages with tense-aspect systems, such as English, the verb systems facilitate discourse. That is, differing tense, aspect, mood, and voice forms have different functions in discourse. In the illustrated narrative text, for example, simple past-tense forms report successive actions and events which advance the story, whereas past-progressive forms report concomitant activities or non-storyline. On the other hand, languages which do not have much richness of structure in regard to tense-aspect distinction may distinguish storyline and non-storyline by a conspiracy of non-systemic elements.

Like other Southeast Asian languages, Mon-Khmer languages are isolating languages with no inflectional morphology and have no specific markers of past versus present tense. Clauses are ambiguous as to time reference. Therefore, storyline is distinguished from non-storyline by a conspiracy of types of verbs, perfective auxiliaries, temporal auxiliaries, temporal expressions, and context.

As mentioned earlier, sequentiality is a typical feature of foreground clauses which form storyline; it is the purpose of this paper to investigate how languages without a tense system, such as the languages of Mon-Khmer family use sequential strategies to mark storyline.

4. Findings

The sequentiality of foreground events requires that an event occur after the completion of a prior event. Mon-Khmer languages use a number of sequential strategies to mark sequential foreground events such as the use of temporal auxiliaries, expressions, or an adverbial clauses to mark the temporal succession of happenings; the use of aspectual auxiliaries to mark the

completeness of actions or event; and the juxtaposition of chronological successive clauses in iconic order, i.e., prior clauses referring to prior actions or event in a reported event sequence. The focus of this paper is the use of temporal auxiliaries and aspectual auxiliaries which is tied to the status of discourse information, i.e., storyline vs. non-storyline.

4.1 Temporal auxiliaries

The most salient happenings in the etic narrative salience scheme are reported in foreground clauses--clauses in which the action of the narrative takes up and time begins to move forward. The foreground clauses make reference to sequenced temporal points, rather than spans of time, and these points are usually the endpoints of situations (Dry 1983). The sequentially ordered clauses or temporally sequenced clauses advance the time reference of a narrative (Myhill 1992).

In Mon-Khmer narrative texts, the foreground clauses are not identified by a tense system as in English but by auxiliary words. The most dominant feature which accompanies the foreground clauses to move the story forward is use of temporal auxiliaries which are defined as preverbal elements which indicate a sequential order of happenings both in accomplished and projected time.

4.1.1. Functions of the temporal auxiliaries within the storyline

When the temporal auxiliaries mark the sequential order of happenings which take place within the narrative accomplished timeframe, they function to give prominence to happenings which advance the storyline. That is, they signal the chronological succession of happenings and the thematic or foreground events or actions. All five languages of the Mon-Khmer family use this sequential device realized in various forms as listed in table 4. These forms are glossed in various ways such as ‘then, particle, topic marker’.

Table 4. Forms of temporal auxiliaries in Mon-Khmer narratives

Languages	Forms of temporal auxiliaries
Samre	<i>kɔʔ²²</i>
Kasong	<i>kɔ</i>
Pray	<i>kɔ</i>
Kmhmu'	<i>gɔ</i>
Vietnamese	<i>thì, bèn</i>

These temporal auxiliaries are pre-verbal auxiliaries except *thì* in Vietnamese which is the sentence conjunction and occurs between two clauses. If the second clause has the same subject as the first one, it is usually left out and *thì* has a preverbal structure in the same manner as other temporal

auxiliaries. These temporal auxiliaries have the same function. They are used as sequential devices which highlight succeeding happenings which form the foreground or storyline of the narrative. They appear in the storyline clauses as illustrated in table 5.

Table 5. Comparison of temporal auxiliaries in foreground clauses

Lang.	Foreground clauses						
Samre							
	<i>chanun</i> ³³²	<i>kɔʔ</i> ²²	<i>jip</i> ³³⁴	<i>kleɛ</i> ³³²	<i>paak</i> ²²	<i>tɔŋ</i> ³³²	<i>liŋ</i> ³³²
	wife	then	come	reach	up	house	on
	<i>Ō</i>	<i>kɔʔ</i> ²²	<i>cuu.r</i> ³³²	<i>ciiw</i> ³³²	<i>neh</i> ³³²	<i>.æe</i> ⁴⁵¹	<i>njuu.r</i> ²¹
	(wife)	then	enter	go	see	in	cradle
	<i>Ō</i>	<i>cuu.r</i> ³³²	<i>ciiw</i> ³³²	<i>neh</i> ³³²	<i>kɔʔ</i> ²²	<i>ni</i> ⁴⁵¹	<i>khiin</i> ³³² <i>huuc</i> ³³⁴ <i>həj</i> ⁴⁵¹
	(wife)	enter	go	see	then	where	child die FP
	‘The wife came into the house and went to look at the child who still slept in the cradle. Then she shouted, “Oh our child is dead!”’						
Kasong							
	<i>mó:j</i>	<i>phák</i>	<i>mé:l</i>	<i>ki:</i>	<i>kɔ</i>	<i>jip</i>	<i>hó:j</i>
	one	time	fish	DEM	then	come	FP
	<i>chan</i>	<i>pé:</i>		<i>nak</i>		<i>hó:j</i>	
	I	defeated		s/he (Ingpaw)		FP	
	<i>Ō</i>	<i>ce:w</i>	<i>ròp</i>	<i>kɔ</i>	<i>pé:</i>		<i>nak</i>
	(fish)	go	fight	then	defeated		s/he
	‘A moment later the fish came. It lost the battle with Ingpaw.’						
Pray							
	<i>pak</i>	<i>ʔah</i>	<i>cak</i>	<i>mec</i>	<i>phlæʔkhæɛl</i>	<i>khæh</i>	
	happen	they	go	see	cucumber	ripe	
	<i>ʔah</i>	<i>kɔ</i>	<i>phoh</i>	<i>ʔah</i>			
	they	then	cut	they			
	<i>ʔah</i>	<i>nan</i>	<i>kɔ</i>	<i>pɔŋ</i>	<i>mi</i>	<i>ŋkrɔc</i>	
	they	that	then	eat	one	CLF	
	<i>po:k</i>	<i>kɔ</i>	<i>lɔ:t</i>	<i>dɔ:m</i>	<i>ʔah</i>	<i>ʔuʔ</i>	
	tiger	then	steal	look at	they	CON.AP	
	‘Some people went to see the ripe cucumber. Then they cut the ripe cucumber. Then they ate a piece of cucumber. The tiger was peeping at them.’						
Kmhmu’							
	<i>gə</i>	<i>gɔ</i>	<i>k^hway</i>	<i>k^hway</i>	<i>k^hway</i>	<i>k^hway</i>	
	he	then	dig	dig	dig	dig	
	<i>k^hway</i>	<i>ho:c</i>		<i>gə</i>	<i>gɔ</i>	<i>t^he:k</i>	<i>ja</i>
	dig	COM.AP		he	then	measure	grandmother
	<i>le</i>	<i>gɔ</i>	<i>dway</i>	<i>sɔ:ŋ</i>	<i>t^he:k</i>	<i>ə</i>	<i>t^he:k</i>
	and	then	take	wood	measure	right	measure
	<i>ja</i>		<i>de</i>	<i>ʔi:k</i>			
	grandmother		self	again			
	‘He then dug (and) dug (and) dug. Finished digging, he then measured grandmother, and then, (he) took (some) wood, measured, yes, measured his grandmother again.’						

Vietnamese									
thì									
<i>sa:n</i>	<i>bi:n</i>	<i>tiʔi:n,</i>	<i>hwá:</i>	<i>fě:p</i>	<i>dá:n</i>	<i>ɲaw</i>	<i>vá:j</i>	<i>can</i>	<i>tɪn,</i>
Sanh	stay	calm	use	magic	fight	together	with	python	
<i>hō:j</i>	<i>la:w</i>	thì	<i>ʔiəw</i>	<i>kwá:j</i>	<i>bj:</i>		<i>ziət</i>	<i>cé:t,</i>	
long	while	then	evil	spirits	PASS.MK		kill	die,	
<i>hwá:</i>		<i>za:</i>	<i>mɔ:t</i>	<i>kɔ:n</i>	<i>can</i>	<i>lâ:n</i>			
change		out	one	CLF	python	big			
‘Thach Sanh stayed calm and fought it with his magical powers, then the evil spirit was killed and changed into a big snake.’									
bê:n									
<i>Tin</i>	<i>là:</i>	<i>thət,</i>	<i>tám</i>	bê:n	<i>súəŋ</i>	<i>ʔa:w</i>	<i>lɔi</i>	<i>za:</i>	<i>coʔð</i>
believe	be	real	Tam	then	down	pond	swim	out	place
<i>sɬu</i>	<i>tám</i>	<i>zɬə</i>							
deep	wash								
‘Tam believed her sister’s words. Then she went down to the pond and washed her hair.’									

Vietnamese has two temporal auxiliaries, that is, the sentence conjunctive *thì* and the preverbal *bê:n*. Both of them introduce a succeeding event which builds upon the preceding event usually encoded by an adverbial clause. Besides the structural difference, *bê:n* has a slightly different use from *thì* in that *bê:n* implies a consequence of the previous event and it is no longer used in everyday conversation but in old texts.

Even though it is claimed that most storyline clauses are signaled by temporal auxiliaries, foreground clauses may also be signaled by other devices such as aspectual auxiliaries, temporal expressions, anaphoric adverbial clauses and juxtaposition of chronological successive clauses in iconic order, i.e., prior clauses referring to prior actions or events in a reported event sequence.

4.1.2 Functions of the temporal auxiliaries within the non-storyline

As mentioned in 4.1.1, the temporal auxiliaries function to mark a sequential order of happenings both in accomplished and projected time. The default timeframe of the narrative is past time or accomplished time. When the temporal auxiliaries occur along with the past timeframe, they identify the storyline. But when they occur within projected time frame as in quotation, they do not function as a storyline indicator. They may also function as an adversative connector, a conditional connector and an additive connector as in examples (1), (2) and (3).

(1) Adversative connector

Pray

luh *ɛiwaʔ* *wak* *nthan* *phɔ:t*
 angry what owl frighten barking deer
wak **kɔ** *ʔaj* *mi* *ɛwal*
 owl AVC not not answer
kɔ *ʔaj* *mi* *cɔk*
 AVC not not speak

“The horned owl frightened me (the barking deer).” “What are you angry with, horned owl?” The horned owl did not answer, did not speak.

In this Pray folktale, **kɔ** functions as an adversative connector marking an expectancy reversal which results in frustration.

(2) Conditional connector

Vietnamese

heʔẽ *kɔ:* *taibien* *zì:* **thì:** *kw:* *ʔɔi* *ta:*
 If have disaster what CONJ keep on call 1stPER
sɛʔẽ *vè:* *ɲaj*

FUT.MK return immediately

‘If any problem occurs, call me. I will be back immediately.’

Example 2 has the sentence conjunctive *thì* functioning as a conditional connector in a quotation.

(3) Additive connector

Kmhmuʔ

ja *ni* **gɔ** *ɔ* *ma:k* *nim* *ma:k* *pi*
 grandmother there also oh many year many year
le **gɔ** *tʰaw* *le* *bat* *gi* *tʰaw*
 and also old then time this old

‘The grandmother there then: Oh, many years, and so, old (is she) then, at this stage, old.’

Example (3) shows descriptive information being connected to the previous information by the additive connector *gɔ*.

4.2 Aspectual auxiliaries

Foley and Valin (1984:209) define aspect as “the category expressing the temporal structure of the event itself without regard to its participants.” Comrie (1981:3) gives a general definition of aspect as “different ways of viewing the internal temporal constituency” and classifies aspect into two main categories, i.e., perfective and imperfective. The perfective denotes a complete

situation, with beginning, middle, and end. The semantic element of perfective is the termination of the situation, that is, the situation is viewed as a single complete whole. The imperfective indicates a situation in progress and pays essential attention to the internal structure of the situation (duration, phasal sequences). It is further subcategorized into habitual and continuous.

4.2.1. Functions of the perfective auxiliaries within the storyline

The aspectual auxiliaries which play an important role in the narratives at hand are perfective auxiliaries. The perfective auxiliaries listed in table 6 are verbal elements that modify the meanings of the co-occurring verbs in terms of completive aspect.

Table 6. Perfective auxiliaries of Mon-Khmer narratives

Languages	Forms of perfective auxiliaries
Samre	<i>ɲɛc²², lɛɛw⁴⁵¹</i>
Kasong	<i>hó:j/hó:j, jðh</i>
Pray	<i>læ:w, roc</i>
Khmhu'	<i>hooc, lɛɛw</i>
Vietnamese	<i>zò:j, sɔ:ɲm, dwa:n</i>

Similar to temporal auxiliaries, perfective auxiliaries are neutral as to time reference. They mark a completion of happenings in both accomplished and projected timeframes. The narrative is represented as having already taken place therefore its timeframe is in the past. The clauses having perfective auxiliaries have perfective aspect in the past timeframe. Therefore these perfective auxiliaries function to mark the storyline which is in the past timeframe.

The presence of perfective auxiliaries indicates a completion of the preceding happening and thus these perfective auxiliaries are normally glossed as 'finished, already' and identified as terminative aspect as in the Samre language (Pornsawan 2001). Since perfective auxiliaries function to terminate the previous happening, they also signal a beginning of another foreground happening. Table 7 compares the use of perfective auxiliaries in storyline context in Mon-Khmer languages.

Table 7. Comparison of perfective auxiliaries in storyline

Lang. Foreground clauses						
Samre						
ɲac²²						
0	<i>tiək²²</i>	<i>khiin³³²</i>	<i>chaa³³²</i>	<i>luk³³⁴</i>		
(husband and wife)	persuade	children	eat	rice bran		
<i>chaa³³²</i>	<i>khajaam⁴⁵¹</i>	0	<i>chaa³³²</i>	<i>luk³³⁴</i>		
eat	chaff	(they) eat	rice bran			
<i>chaa³³²</i>	<i>khajaam⁴⁵¹</i>	ɲac²²	kɔʔ	<i>ciiw³³²</i>	<i>tɔ²¹</i>	<i>kaaɽ³³²</i>
eat	chaff	finish	then	go	do	work
‘The husband and wife persuaded the children to eat rice bran and chaff. After they had eaten rice bran and chaff, they went to work.’						
Kasong						
hó:j						
0	<i>tàŋ</i>	<i>lamə:ŋ</i>	<i>ki:</i>	<i>hó:c</i>	<i>lamə:ŋ</i>	<i>kɔ</i>
(monkey)	see	grandpa	DEM	die	grandpa	then
<i>hó:c</i>	hó:j	0	<i>na:j</i>	<i>chan</i>	<i>khə:j</i>	<i>jip</i>
die	already	(monkey)	say	I	used to come	
<i>cha:</i>	<i>rɔ:n</i>	<i>kə:</i>	<i>sih</i>			
eat	melon	he	FP			
‘The monkey saw grandpa die. After grandpa had died, the monkey said “I used to come and eat his melon.”’						
jəh						
0		<i>ce:w</i>	<i>rù:c</i>	<i>ma:l</i>	<i>rù:c</i>	jəh <i>kát</i>
(husband and wife)		go	clear	field	clear	finish cut
<i>nih</i>	<i>ʔh</i>	<i>ma:l</i>	<i>jéh</i>			
wood	give	field	dry			
‘The husband and wife went to clear the field. After they had cleared the field, they cut wood to dry the field.’						
Pray						
læ:w						
<i>po:k</i>	<i>kɔ</i>	<i>rəl</i>	<i>cak</i>	<i>phək</i>	<i>lɔ:ŋ</i>	<i>cəŋ</i> <i>mi</i> <i>naŋ</i>
tiger	then	go	go	bite	wild boar	big one CLF
<i>phək</i>	<i>ʔət</i>	<i>lɔ:ŋ</i>		<i>cəŋ</i>	<i>læ:w</i>	<i>kɔ</i> <i>pəl.liʔ</i>
bite	take	wild boar		big	already then	die
læ:w	<i>kɔ</i>	<i>cak</i>	<i>laj</i>	<i>khram</i>	<i>nan</i>	<i>niʔ</i>
already then	go	tell	man	that	that	FP
‘The tiger went to bite a big wild boar. The big wild boar died. And then the tiger went to tell that man.’						
roc						
<i>wak</i>	<i>kɔ</i>	<i>ʔaj</i>	<i>mi</i>	<i>ɛwal</i>	<i>ʔah</i>	roc <i>ʔət</i> <i>kək</i> <i>ja:</i>
owl	then	not	not	answer	he	then take pipe tobacco
<i>ŋkoh</i>	<i>kuʔ</i>	<i>wak</i>	<i>pum</i>	<i>leh</i>		
knock	head	owl	swollen	come		
‘The owl did not answer. He then used a tobacco pipe and knocked on the horned owl’s head.’						

Kmhmu'									
hooɬ									
<i>taʔ</i>	<i>niʔ</i>	<i>ʔan</i>	<i>wək</i>	<i>mɔɔt</i>					
Mr.	this_here	give	knife	take_hold					
<i>wək</i>	hooɬ	<i>gəə</i>	<i>pliat</i>		<i>sa ʔɔɔŋ niʔ</i>				
knife	and_then	3sgm	chop_up	wood	this_here				
'The man gave (him/them) the knife. Taking the knife then he/they chopped the wood.'									
lɛ:w									
<i>lanɬaak</i>	<i>gi hiʔ</i>	<i>gəə</i>	<i>pʰɔ dii</i>	<i>bian</i>	<i>mooj</i>				
after	that_one	3sgm	exactly	achieve	one				
<i>pɪi</i>	<i>pʰɔ dii maʔ</i>	<i>gəə</i>	<i>gɔʔ</i>	<i>haan</i>	<i>pəh</i>				
CLF_years	exactly	mother	3sgm	so_then	die	separate_from			
<i>maʔ</i>	<i>gəə</i>	<i>haan</i>	<i>pəh</i>	lɛw	<i>joŋ</i>	<i>gəə</i>	<i>ʔɔɔr</i>	<i>gəə</i>	
mother	3sgm	die	separate_from	already	father	3sgm	lead	3sgm	
<i>jet</i>									
stay									
'Later on, (when) he was exactly one year old, then his mother died. His mother having died, his father took him to stay (with him).'									
Vietnamese									
zɔ:j									
<i>lɔ:k</i>	<i>lɔ:ŋm</i>	<i>kwan</i>	<i>dɔŋm</i>	<i>mɔt</i>	<i>ciək</i>	<i>thwiən</i>	<i>thɔt</i>	<i>cák</i>	<i>và:</i>
Lac	Long	Quan	build	one	CLF	boat	real	strong	and
<i>thɔt</i>	<i>lɔ:n</i>	<i>zɛ:n</i>	<i>mɔt</i>	<i>kói</i>	<i>sát</i>	<i>kɔ:</i>	<i>ɲiəw</i>	<i>kə:ɲ</i>	<i>sák</i>
real	big	forge	one	CLF	iron	have	many	sharp	edge
<i>nuy</i>	<i>cɔ:</i>	<i>thɔt</i>	<i>dɔ:</i>	zɔ:j	<i>de:m</i>	<i>kói</i>	<i>sát</i>	<i>súəŋ</i>	<i>thwiən</i>
burn	to	real	red	already	carry	CLF	iron	down	boat
<i>cɛ:w</i>	<i>thəŋ</i>	<i>dɛ:n</i>	<i>ŋu:</i>	<i>tɪŋ</i>					
row	straight	to	Ngu	Tinh					
'Lac Long Quan built a big and strong boat, forged a block of iron that was full of the sharp points, heated it until it was red hot and brought it along to the evil fish's cave.'									
sɔ:ŋm									
<i>lí:</i>	<i>thoŋm</i>	<i>ŋɛ:</i>	sɔ:ŋm	<i>nɔj</i>	<i>za:</i>	<i>mɔ:t</i>	<i>mu:w</i>	<i>thɔ:m</i>	<i>dəkp</i>
Ly	Thong	listen	already	bounce	out	one	trick	wicked	
'After Ly Thong had listened, he tried a trick to intimidate Thach Sanh.'									
dwɔ:n									
<i>nɔ:j</i>	dwɔ:n	<i>ha:j</i>	<i>mɛ</i>	<i>kɔ:n</i>	<i>kwàn</i>	<i>ʔá:w</i>	<i>súŋm</i>	<i>sɪŋ</i>	<i>le:n</i>
speak	already	two	mother	child	clothes		beautiful	up	way
'After that was said, Cam and her mother went to the festival in their new dresses.'									

The examples in table 6 show that the perfective auxiliaries frequently tag an adverbial clause which repeats the previous happening as a tail-head linkage or sum up the preceding happening as a summary head linkage.

The perfective auxiliaries usually co-occur with the temporal auxiliaries. The co-occurrence of these two kinds of auxiliaries will encode foreground happenings, which comprise the most salient discourse information in the whole narrative as seen in the Samre and Kmhmu narratives below.

Samre

0	ʔam ⁴⁵¹	kluəŋ ²²	nɛh ³³²	khiin ³³²					
wife	have	husband	look after	child					
kluəŋ ²²	kɔʔ	nɛh ³³²	khiin ³³²	0	nɛh ³³²	0			
husband	then	look after	child	(husband)	look after	(child)			
kɔʔ	khiin ³³²	jaam ²¹	nah ³³²	nah ³³²					
then	child	cry	a lot	a lot					
0	kɔʔ	ʔuuc ³⁴²	ŋcuu.ɿ ²¹	ciiw ³³²	tih ²¹	phuəŋ ⁴⁵¹	0	ʔeeŋ ⁴⁵¹	
(husband)	then	take	needle	go	lance	brain	(child)	FP	
0	tih ²¹	phuəŋ ⁴⁵¹	khiin ³³²	tih ²¹	ɹuəc ²²	kɔʔ	0	ʔuuc ³⁴²	
(husband)	lance	brain	child	lance	finish	then	(husband)	take	
khiin ³³²	ciiw ³³²	huum ³³²	taak ³⁴²	huum ³³²	taak ³⁴²	ɹuəc ²²	kɔʔ	0	
child	go	bath	water	bath	water	finish	then	husband	
ʔuuc ³⁴²	0	ciiw ³³²	thiək ²²	ɹee ⁴⁵¹	njuu.ɿ ²¹				
take	(child)	go	sleep	in	cradle				

‘The wife asked her husband to look after their child. The husband stayed with the child but the child cried a lot. Thus he took a needle to lance the child’s brain then took the child to take a bath and put him in a cradle.’

Kmhmu’

ra	ʔwaajta	loh	niʔ	wat	haaw	haan	niʔ	ra	ʔwaaj
tiger	male	this_here	be_ensnared	trap	die	here	tiger		
jaaj	gaaj	kat	mah	klaa		ra	ʔwaajta	loh	niʔ
female_animal	DIR	gnaw	eat	testicle		tiger	male	this_here	
hooŋ	gɔʔ	jɔh	wat	haaw					
so_then	and_then	DIR	be_ensnared	trap					

‘The male tiger was caught in the trap (and) died there, the female tiger came (and) gnawed at the male tiger’s testicles, and then went (and) got caught in the trap.’

4.2.2 Functions of the perfective auxiliaries within the non-storyline

As perfective auxiliaries are neutral as to time reference, they may also occur in projected time as in a quotation in which they indicate temporal succession of happenings which have not yet taken place as seen in the Kmhmu’ narrative below. The quotation belongs to Band 5 in the etic narrative salience scheme. It is irrealis or collateral information which might or might not take place at the time of narration.

Kmhmu'

hooŋe ʔoo ʔiʔ tɛii ʔam bian mah lɛɛw ŋɔʔ gi
 and_then oh! 1pl IRR NEG achieve eat already rice this_one
 'And then, "Oh! we won't get to eat (it) now, this rice".'

Perfective auxiliaries may also occur in the quotation which refers back to a previous event as seen in the Samre narrative below.

Samre

sajin khe:w ke: ʔə:j pa:k lij ton thəh chan thim
 wife call you FP go up on house FP I cook
kloŋ chi:n hɔ:j
 rice cooked already
0 thim se:w chi:n hɔ:j 0 jip pa:k təh
 (I) cook curry cooked already (you) come go up FP
 'His wife called, "Come into the house, please. I have finished cooking rice and curry (the meal is ready). Please come into the house."'

In addition to quotations, perfective auxiliaries may also mark a completion of flashback happenings as seen the Kmhmu narrative below.

Kmhmu'

ra'waajpok paʔ si'haa kɔn maʔ haan
 tiger bite with 3du child mother die
 'The tiger attacked both of them, child (and) mother, (and they) died.'

ŋɔŋ jaʔ jɛʔ niʔ ŋɔʔ naa dar guut sɔɔr
 remain Ms small this_here 3sgf run DIR hide
lɔh dee daʔ lij klaak ri'haan
 body co-referent at middle clump bamboo
 '(There) remained the small girl (who) was afraid, (and) she ran in (and) hid in the middle of a bamboo clump.'

briʔ leh tɛii bah joŋ jet daʔ gaan ʔɔɔr
 environment near IRR light father located at house lead
kɔn kuŋ kɔn'gaan tuuŋ drɔʔ jɔh sɔɔk
 village_people light torch DIR seek
 '(When) it was nearly light, the father who was at the house led the village people, lit a torch (and) went (and) looked for (them).'

jɔh sɔɔk mɔʔ gɔʔ ʔam bip nɔɔ saam gon
 DIR seek INDEF so_then NEG meet 3pl three CLF_people
ra'waajhii pok mah loote lɛɛw
 tiger PstCMPL bite eat totally already
 'Wherever (they) looked (they) didn't find them, (the) three people the tiger had already attacked (and) eaten completely.'

This example shows that the last clause represents a flashback in the first clause.

5. Conclusion and discussion

Contingent temporal succession is a basic etic parameter of narrative. That is, events or doings are contingent on previous events or doings. This parameter is marked by various devices in different languages. This paper presents the sequential devices which mark the storyline of narrative texts. The data were drawn from written texts. They consist of twenty-six narrative texts of five Mon-Khmer languages, namely, Samre, Kasong, Pray, Kmhmu, and Vietnamese. These texts are analyzed using the framework of discourse analysis as expounded by Longacre (1983). The analysis is focused on the etic narrative salience scheme.

It has been found that the sequential strategies used in Mon-Khmer narrative texts include temporal auxiliaries and perfective auxiliaries. Samre, Kasong, Pray, and Kmhmu have the temporal auxiliaries which are similar in forms and thus should be regarded as cognates. They are *kɔʔ²²* (Samre), *kɔ* (Kasong), *kɔ* (Pray), *gɔ* Kmhmu'. These temporal auxiliaries function to mark the storyline status of the following happenings which are sequential to the previous ones. On the other hand, Vietnamese has two different temporal auxiliaries, i.e., *thì*, *bê:n* that have a similar function. Somsonge distinguishes these two temporal auxiliaries as follows:

The sentence conjunctive *thì* and the preverbal *bê:n* introduce

a succeeding event which builds upon the preceding event usually encoded by an adverbial clause. It can be seen that these two words have the same function. The difference is their syntactic structure: *thì* occurs between clauses and *bê:n* before a verb. Besides the structural difference, *bê:n* has a slightly different use than *thì* in that *bê:n* implies a consequence of the previous event. It should be noted also that *bê:n* is no longer used in everyday conversation. It appears only in old texts. The sentence conjunctive *thì* seems to play a more important role in the discourse.

Somsonge (2002: 73)

Each Mon-Khmer language examined here has one to three perfective auxiliaries. The main function of these auxiliaries is to signal a completion of previous happenings which are sequential relative to the following ones.

Samre, Pray, and Kmhmu' share the same perfective auxiliaries, that is, *lee^w451*, *læ:w*, and *lee^w*, respectively. Though *lé:w* is not found in Kasong narrative texts due to a limitation of data, it is found in Kasong procedural texts functioning as a sequential indicator. In addition to *lee^w451* (Samre), *læ:w* (Pray), and *lee^w* (Kmhmu'), Samre, Pray and Kmhmu share similar forms of the perfective auxiliary *luəc²²*, *roc*, and *hoo^te*, respectively. Kmhmu uses *hoo^te*

in manner similar to *lɛɛw* but it is used more frequently than *lɛ:w*. In Kasong, *hó:j* and *jðh* have a similar function. The perfective auxiliary *hó:j* marks the completion of previous happenings and is usually glossed as ‘already’ whereas *jðh* functions to terminate previous happenings and is normally glossed as ‘finish, completely’.

Vietnamese has three perfective auxiliaries, *zò:j*, *sɔ:ŋm̄*, and *dwɑ:n*. The perfective auxiliary *zò:j* has a similar form and function to *hó:j* (Kasong) and thus both of them are probably cognates. Somsonge notes some minor different usages of these three perfect auxiliaries as follows:

These three words are used in a similar manner and it is difficult to figure out when to use each word. However, it is more likely that *sɔ:ŋm̄* is used with an accomplishment verb and *dwɑ:n* is used to terminate a happening whereas *zò:j* has a wider usage with various kinds of verbs indicating the completion of a happening.
Somsonge (2002:72)

The temporal auxiliaries and perfective auxiliaries may co-occur to highlight a happening. The part of narrative that has these two kinds of auxiliaries is at the highest rank of the etic narrative salience scheme.

The narrative is recounted as happening in the past or accomplished time. Therefore occurrences of the temporal auxiliaries and perfective auxiliaries within the narrative timeframe mark happenings which are on the storyline. However, these auxiliaries are neutral as to time reference so they may occur both in accomplished and projected timeframes. When they occur in projected timeframes such as quotations, they do not mark the storyline which is in Band 1 of the etic narrative salience scheme.

The study of these auxiliaries reveals that language contact plays an important role in their usages. Samre, Kasong, Pray, and Kmhmu’ have been influenced by their Tai-speaking neighbors.

Comparing the temporal auxiliaries *kɔʔ* (Samre), *kɔ* (Kasong), *kɔ* (Pray), and *gɔ* (Kmhmu’), one might speculate that Samre, Kasong and Pray borrowed these forms from Thai and Kmhmu’ from Lao. However, when more Mon-Khmer languages are considered, we find *kɔʔ* in Northern Khmer (Somkiet 1982); *kɔʔ* (Somsonge 1992); *ka:* in So (Migliazza 1998). While this temporal auxiliary is found in all Mon-Khmer languages that have been surveyed, it is found only in Southwestern Tai languages and absent in other Tai-Kadai languages. Since all languages of the Mon-Khmer language family, except Mon, possess this temporal auxiliary, it can be concluded that this temporal auxiliary is an original Khmer form as affirmed by Somsonge:

When Thai people migrated from China to the south, they adopted this form from Khmer during the time that the Khmer empire was at the Zenith of its civilization. Not only the pronunciation of this word is similar in Thai and Khmer but also its spelling. Based on a comparative study of Tai orthography, it has been found that this word was originally pronounced with a short vowel [a] so that vowel shortening symbol 𑜀 was used to indicate this short vowel. The similarity of Thai and Khmer spelling of this word is due to the influence of Thai language on Khmer in the Ratanakosin Period. The vowel shortening symbol must have been invented by Thai and transmitted to Khmer.

Somsong (2008:442)

It is affirmed above that the temporal auxiliary was pronounced with a short vowel [a]. Thus, it would be necessary to investigate whether the initial consonant is voiced or voiceless as Mon-Khmer languages have both types of consonants. As devoicing is a natural phonological process, the original consonant was most likely voiced. Consequently, this temporal auxiliary could be reconstructed as *gɔ. Further evidence of language contact between Kmhmu' and Lao supports this reconstruction. Osborne (2009:4) states that "...there are many words common to both Lao and Kmhmu'. Some are claimed by Lao speakers to be borrowings from Lao into Kmhmu' and been incorporated into Lao." The Lao words that originally had initial voice stops are devoiced in Lao but retained in Kmhmu'. Take, for example, *gon* in Kmhmu' and *khon*³⁴ in Lao. Kmhmu' still keeps a voicing distinction of initial stops. So the initial consonant of the temporal auxiliary was originally voiced but becomes devoiced in Mon-Khmer languages, as well as in Thai and Lao.

While all Mon-Khmer languages share the same temporal auxiliary, Vietnamese has the distinct forms, *thí*, and *bè:n*. Comparing these forms to the Chinese temporal auxiliary *jiu* 'then', it is unlikely that these Vietnamese forms are borrowed from Chinese. So these Vietnamese temporal auxiliaries may be typical of Vietnamese.

In comparing the perfective auxiliaries, it has been found that Samre, Pray and Kmhmu' share the words *leew*⁴³¹, *læ:w*, and *le:w*, respectively, with the Thai perfective auxiliary *léaw*. Similar forms of these perfective auxiliaries are also found in other Tai-Kadai languages, for example *le:w*⁴ (Lue), *leu*⁴ (Bouyei), *liu*⁴/*le:u*⁴ (Northern Zhuang), *la:u*² (Kam), *leu*⁴ (LinGao), *lai*³ (Mulam), *ljeu*² (Sui), *ljeu*⁴ (Maonan). As these perfective auxiliaries have the same function as *lǎ/liau*²¹⁴ in Chinese, the Tai-Kadai forms might have been borrowed from Chinese (Somsong 2001). After Tai people settled in modern Thailand and Laos, Samre, Kasong, Pray, and Kmhmu' borrowed this form from their Tai-speaking neighbors.

Comparing other perfective auxiliaries which are typical of Mon-Khmer languages, it has been found that Kasong and Vietnamese share similar forms of *hó:j/hó:j* and *zò:j*, respectively. Kasong have two variants which are glossed the same, i.e., ‘already (past or complete aspect)’ (Sunee 2002) and also have the same discourse function. Samre also has a similar form *həj*⁴⁵¹ ‘already’ (completive aspect), which is present only once in a quotation and is thus excluded from this study. In addition to *hó:j/hó:j*, Kasong has *jəh*, which is typical of this language and not found elsewhere. Other Mon-Khmer languages also possess this perfective auxiliary such as *hři* ‘already’ in Satieng (Owen II 2002). Samre, Pray and Kmhmu’ share the similar forms *luəc*²², *roc* and *hooc*, respectively. However, Kmhmu’ also has *looc* ‘completely’, which is classed as an adverb and occurs only twice in the studied texts. It has a discourse function similar to that of *hooc* ‘be finished’ in that it marks the completion of previous happenings whereas *hooc* signals that an event or happening has been finished or completed. Similar forms of these perfective auxiliaries are also found in other Mon-Khmer languages, such as *ruəc* in Khmer and *luəc* in Stieng (Owen II 2002). In addition to *zò:j*, Vietnamese has *sə:əm*, and *dwə:n*, which are typical of this language.

Finally, there is a discourse marker which has an emphatic function and sequential function in different Mon-Khmer languages. This marker is not included in this study but is worthy of mention. This discourse marker has the following forms: *ne:* ‘link’ (Pray), *něy* ‘then’ (Stieng), *nɛ* ‘then’ (Kmhmu’), *nəeh/snəeh* ‘then’ (Northern Khmer), and *snəe/nəe* ‘then’ (Kui). In Pray, Northern Khmer and Kui, these forms usually co-occur with the temporal auxiliaries *kə*, *kə?*, and *kə?*, respectively, and function as sequential indicators. In Stieng and Kmhmu’, *něy* and *nɛ* are used alone as sequential indicators and emphatic particles as seen in the following examples.

Stieng

<i>brak</i>	<i>kməŋ</i>	<i>gur</i>	<i>kəw</i>	<i>soc</i>	<i>řn</i>	<i>sljŋ</i>	<i>bək</i> ,	<i>něy</i>	<i>əj</i>
peacock	just	paint	neck	black	have	collar	white	then	hear
<i>řn</i>	<i>bək</i>	<i>səř</i>	<i>rəw</i>	<i>wək</i>	<i>wək</i>				
have	sound	pig	squeal	oink	oink				

‘The peacock just painted the neck black with a white stripe then (he) heard the sound of a pig squealing “oink oink.”’

Kmhmu’

<i>p^hədɪ</i>	<i>ga:j</i>	<i>da</i>	<i>ga:ŋ</i>	<i>law</i>	<i>ga</i>	<i>ga:ŋ</i>	<i>kir</i>	<i>lə:j</i>	<i>nɛ</i>
Just then	return	at	home	she	climb	house	close	continue	then

‘Just then, (she) returned home, she climbed up the house, continued closely.’

As *nɛ* ‘then’ has an emphatic function which is different from *ne:* ‘link’ (Pray), *něy* ‘then’ (Stieng), *nəeh/snəeh* (Northern Khmer), and *snəe/nəe*

(Kui), it is uncertain whether *na* is a different word or the same word having a different function.³

ABBREVIATIONS AND SYMBOLS

1	first person
2	second person
3	third person
f	feminine
AVC	adversative connector
CLF	classifier
COM.AP	completive aspect
CON	conj
CON.AP	continuous aspect
DEM	demonstrative
DIR	directional
du	dual
EMP	emphatic
FP	final particle
FUT.MK	future marker
IRR	irrealis
NEG	negative
PASS.MK	passive marker
pl	plural
PstCMPL	past completed
sg	singular

REFERENCES

- Comrie, Bernard. 1981. *Aspect*. Cambridge: Cambridge University Press.
- Diffloth, Gerald. 1974. Austro-Asiatic Languages. *The New Encyclopedia Britannica* 2: 480-484.
- Dry, Helen Aristar. 1983. The movement of narrative time. *Journal of Literary Semantics* XII.2: 19-53.
- Foley, William and Robert D. Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Longacre, Robert E. 1981. A spectrum and profile approach to discourse analysis. *Text* 1:337-59.
- Longacre, Robert E. 1989. Two hypotheses regarding text generation and analysis. *Discourse Processes* 12:413-60.
- Longacre, Robert E. 1990. *Storyline concerns and word-order typology in east and west Africa*. *Studies in African Linguistics Supplement 10*. Los

³ This paper was presented at the Fourth International Conference on Austroasiatic Linguistics, hosted by SIL International, Mahidol University, and Center for Research in Computational Linguistics, Bangkok, October 29 - 30, 2009.

- Angeles: The James S. Coleman African Studies Center and the Department of Linguistics, University of California at Los Angeles.
- Longacre, Robert E. 1996. *The grammar of discourse*. New York: Plenum Press.
- Migliazza, Brian. 1998. A grammar of So-a Mon-Khmer language of northeast Thailand. Ph.D. dissertation, Mahidol University.
- Mingkwam Malapol. 1989. Pray grammar at Ban Pae Klang, Thung Chang District, Nan Province. M.A. thesis, Mahidol University.
- Myhill, John. 1992. Word order and temporal sequencing. In *Pragmatics of Word Order Flexibility*, edited by Doris L. Payne, 265-278. Amsterdam/Philadelphia: John Benjamins.
- Nguyễn Văn Ngộ, Vũ Ngộ Phan, and Trần Thanh Mai. 1996. *Kho tàng Truyền Cổ Tích Việt Nam* (A Collection of Vietnamese Folktales). Hanoi: Van Hoc.
- Osborne, Rosalind Anne. 2009. Who's who in Kmhmu': Referring expressions and participant identification in selected Kmhmu' narrative texts. M.A. thesis, Payap University.
- Owen II, James Edmond. 2002. A discourse analysis of two Stieng narratives. M.A. thesis, University of Texas at Arlington.
- Pornsawan, Ploykaew. 2001. Samre grammar. Ph.D. dissertation, Mahidol University.
- Rowling, J.K. 1998. *Harry Potter and the Chamber of Secrets*. London: Bloomsbury.
- Sidwell, Paul. 2005. Proto-Katuic phonology and the sub-grouping of Mon-Khmer languages. In *SEALSXV: Paper from the 15th Annual Meeting of the Southeast Asian Linguistics Society*, edited by Paul Sidwell, 193-204. Canberra: Pacific Linguistics.
- Somkiet Poopatwiboon. 1982. Discourse level cohesion in Northern Khmer. M.A. thesis, Mahidol University.
- Somsong Burusphat. 1992. The functions of $k\lambda?$ in oral Kui narrative. *Mon-Khmer Studies* 18-19: 223-231.
- Somsong Burusphat. 2008. An etymological speculation on $k\omega^3$ in Thai narrative. In *The Tai-Kadai Language*, edited by Anthony V.N. Diller, Jerold A. Edmondson, and Yongxian Luo, p. 431-444. London: Routledge.
- Suksiri Danthanavanich. 2000. Cohesive devices in Vietnamese folktales. M.A. thesis, Mahidol University.
- Sunee Kamnuansin. 2002. Kasong Syntax. M.A. thesis, Mahidol University.

Received: 28 October 2009

Research Institute for Languages
and Cultures of Asia
Mahidol University, Salaya
THAILAND
email address: grsbr@mahidol.ac.th

The status quo and trend of language use by Lai people

Xiaohang QIN

Minzu University of China,
Beijing, China

Fanglan LI

Guangdong University of Education,
Guangzhou, China

Abstract

This paper provides some general information about Lai People, which includes the population and distribution, the origin, the family group and the language use of Lai People. The focus of this paper is the use of Lai language use and its development trend. Questionnaires and interviews are carried out to investigate the status quo of Lai language use. The analysis of those data suggests that the number of people who speak Lai language will witness a steady increase.

Keywords: Lai language, language use, development trends

1. The population and distribution of Lai people

Lai people call themselves pa33lju13or la:i33. According to our survey in 2006, the Lai population was 1030 people. The majority of Lai population (which is more than 300 people) live in Douhong Zhai, Xinhe Hamlet, Kechang Village Longlin County. More than 50 people live in Datie Zhai. Over 300 people inhabit Luowan Xinhua Village, over 160 in Kabao and Renshang and at least another 230 are in Hengsha and Naya of the Xilin County. In addition, there are some Lai people living in Xinzhou Town, Zhelang Village, De'e Village and Shechang Village in Longlin County.

2. The origin of Lai people

Scholars differ in their opinions as to where the Lai people originated. Local Chronicles of Longlin County states that Lai people migrated from Luchu Camp in Xingren County Guizhou Province and Tieban Bridge where Anshun and Guanling converges to Longlin and Xilin County in Guangxi Province. Liang Min believes that Lai people first moved from Southeast Asia to Yunnan and Guizhou Province, then moved to Guangxi Province. Mr. Pan holds that Lai people were indigenous people who lived in the area where Dian, Gui and Qian converges.

3. The attribution of Lai language

There are different opinions on the attribution of the Lai language. Huang Caizhen (1982) believes that the Lai language is a branch of the Zhuangdong languages. Gong Yonghui (1990) believes that the Lai language is a dialect in Gelao language. Liang Min (1984), Wang Jingliu and Shi Feng (1987) believe that the Lai language belongs to Mon-Khmer branch. Some scholars, represented by Liang Min, Yan Qixiang and Zhou Zhizhi, believe that the Lai language belongs to Yuemang branch. The latter two views are more popular. In his article, *On attribution of Lai Language* (1990), Liang Min compared 200 words in the Lai language with words in the Jing, Wa and Khmer languages, three typical languages from the Yuemang and Mon-Khmer branches of the Southeast language family. He found that there are 58 similar words between the Lai language and the three languages from the Southeast language family, accounting for 29% of the 200 words. The same 200 words were also compared with the words in the Zhuang, Tai, Dong and Shui languages from the Zhuangdong branches of the Sino-Tibetan language family and there were only 17 similar words, accounting for 8.5% of the total. From this study it can be seen that Lai language is closer to Mon-Khmer branch of Southeast language family. Mr. Liang ascribed Lai language to Mon-Khmer branch. Mr. Yan and Mr. Zhou agree with Mr. Liang that the Lai language belongs to the Southeast language family, but believe it does not come from the Mon-Khmer branch. Since Lai people call themselves pa33lju13, Mr. Yan and Mr. Zhou translated the Lai language as pa33lju13 language. Building upon the foundation of the research done by Mr. Liang, Mr. Yan and Mr. Zhou compared pa33lju13 language with Jing, Mang, Wa and Khmer languages again and found the similar words accounting for 60.9%, 41.7%, 46.9% and 45% respectively. The pa33lju13 and Jing languages share the largest number of similar words. In addition, the pa33lju13, Jing and Mang are all languages with tones, whereas Wa and Khmer are without tones. The pa33lju13 language should be ascribed into Yuemang branch of the Southeast language family. Later, Li Xu lian and other scholars agreed with this view.

In summary, it's more likely that the Lai language belongs to the Southeast language family.

4. The family group of Lai people

There are 5 different views on this issue. Mr. Pan (1981) holds the view that the Lai people are an independent ethnic group, called "Lai Ethnic Group". Mr. Zhu Bin (1989) believes that the Lai people belong to the Zhuang and Li ethnic minority group. Yang Tongsu believes that the Lai people belongs to the Yi, Zhuang and the Han ethnic group. GongYonghui, whose view was officially endorsed, believes that the Lai people belong to the Gelao ethnic group. In May 1990, Autonomous Regional People's Government of

Zhuang Ethnic Group in Guangxi Province issued a document, stating that the Lai people were to be recognized as part of the Gelao ethnic group. Hence the term “Lai people” became out of date.

5. The language use of Lai people

There is no vernacular in Lai language. The Lai language is more widely used than the Gelao language. Despite mixing with the Han, Miao, Zhuang, Gelao and Yi people for a long time, the Lai people have stuck to their native languages, even though they have the ability to speak other ethnic languages. As a result, the Han, Miao, Zhuang, Yi and Gelao people learned to speak the Lai language instead. The number of people who speak the Lai language has continued to grow, from 650 in 1980s to over 1000 nowadays. According to our latest survey, there are 1000 Lai people and 200 Han, Miao, Zhuang, Yi and Gelao people in the whole county, with 80% of the 200 able to speak the Lai language. This brings the number of people who speak the Lai language to nearly 1200 people.

5.1 The Use of Mother Tongue

93% of Lai people speak the Lai language. Areas such as Datie Zhai in the Xinhe Village, Douhong Hamlet and Kabao Hamlet in the Xinhua Village, which are all inhabited by Lai people, are dominated by the Lai language. In areas where Lai and Miao live together (for example, Wenya Hamlet, Puhe Village and Xilin County), Miao people also speak Lai language. This shows that Lai language has taken firm root in these areas.

5.2 The Use of Multi-language

All the Lai people can speak several different languages because they have lived together with the Miao, Han, Zhuang and Gelao people for a long time. In order to communicate with those from other ethnic groups, Lai people must learn their languages. Lai people can speak at least two languages and sometimes as many as seven. 50% of the Lai population can speak four or five languages. The languages mastered by Lai people are Cantonese(3%), the Southwest mandarin (99%), Lai language(98%), Miao language(98%), mandarin(57%), Yi language(4%), and Zhuang language(28%).

5.3 The Use of Language at Home

Lai people use Lai languages at home. Women, from other ethnic group, who are married to Lai men will learn to speak Lai language soon after their marriage. The women who are married to men from other ethnic group must teach their children to speak the Lai language, or else they can't return to their parents' home. It's obvious that Lai people love the Lai language.

5.4 The Use of Language in Society

The Southwest mandarin is the common language for communication among people from different ethnic groups in Xilin County and Longlin County. Lai people will speak the Southwest mandarin when they leave their homes or villages to do business in markets, see a doctor, try to get something at the office of the local government or communicate with people from other villages. Mandarin is also used, although to a less frequent extent. They will use the native language of the people they speak to according to this person's ethnic group, status and native language.

5.5 Language Attitude

The language attitude of Lai people is very objective. Although they cherish the Lai language and some even advocate the use of the Lai language as the tool for teaching in primary and middle school, more than 88% of Lai people hold that the Southwest mandarin and mandarin have greater social influence and are more valuable. The use of mandarin to teach in school can help the students in primary and middle schools improve their chances to work in other parts of China instead of being bound to Longlin, Xilin and Guangxi.

5.6 The Use of Characters

There are no characters in the Lai language and so for a long time, the Lai people used Chinese characters. More than 80% of Lai people reach the level of junior middle school in the use of characters. Illiteracy doesn't exist.

6. The use of Lai language and its development trend

According to a survey in 1980, there were only 650 people who spoke Lai language. This showed that the Lai language was becoming extinct. Many predicted that it would die out within 10 years. Another survey 26 years later, in 2006, found a different situation to that predicted. The Lai language witnessed steady growth, with the number of people who speak the Lai language increasing from 650 to 1200. Guangxi Daily, the most influential newspaper in Guangxi, issued a news article which aroused concern from domestic and international scholars. The scholars are particularly interested in the factors that got the Lai language back on its feet. The following are some suggestions for consideration.

Firstly, Lai people mainly live on the top of mountains where transport is inconvenient. For instance, the altitude of Douhong hillside, the highest peak in Longlin County, is 1951 meters. It is seldom visited by strangers. The constant changes in society and culture can hardly reach the village inhabited by Lai people who enjoy a reclusive life. The closed living

environment not only hinders the influence of outside culture but also protects the traditional culture and the Lai language.

Secondly, Lai people usually live within their own ethnic group or form small villages. The Lai language is commonly used in the villages, and since they predominately interact with other Lai people they do not need to use any other language to communicate. Therefore the Lai language dominates Lai villages and people from other ethnic group also join them in using it. For example:

There are 9 households in Datie Zhai, with a population of 35. Among which, there are 7 Lai households and 2 Miao households. In the 7 Lai households, there are 5 daughters-in-law who are from other ethnic groups. One is from Han nationality and the other 4 are from the Yi and Miao ethnic groups. The majority of the people there can speak the Lai language. The 5 daughters-in-law could not speak the Lai languages when they first arrived and learned Lai language after some time. The people in the 2 Miao households learned to speak the Lai language after long-term contact with local Lai people. Although they use Miao language at home or when they speak to the daughter-in-law from the Miao ethnic group, they will use the Lai language when they are with Lai people.

The total population of Wenya Hamlet, Puhe Village, Xilin County is 1098. The Miao ethnic group is the dominant group, with a population of 984. There are 109 Lai people (25 households in total) in Hengsha Village. There are 35 Miao households in Hengsha Village. In the 25 Lai households, there are 2 daughters-in-law and one brother-in-law who are of Han nationality. There is one daughter-in-law from Zhuang ethnic group and one from Yi ethnic group. Lai people and Miao people live together and many families enjoy close association because of marriage ties. Many Miao people learned to speak the Lai language. Though outnumbered by Miao people, Lai people possess a strong national awareness. The daughter-in-law from other ethnic group will adopt Lai language soon after marriage. According to what Mr. Wen Huiming said, the secretary of the Wenya villagers committee, those who can speak Lai Language in Hengsha village are not confined to Lai families. Many Miao people in Hengsha Village can speak Lai fluently. When the Lai people host a public event in Hengsha Village, they will use the Lai language.

Lai people in Hengsha Village originated from Douhong Hillside Longlin County, with family names such as Wei and Wang. They have been living in Xilin for only 6 generations. At the time of the Qingming festival, they will worship their ancestors on the Douhong Hillside. If there is a wedding or funeral, Lai people from both Xilin and Douhong will have closer contact. In the Hengsha Village the Lai language is greatly protected due to

their close relationship with the Lai people and their correlating national sentiments. The Miao people nearby are also driven to speak the Lai language.

Thirdly, although some of them do business or work elsewhere, most Lai people make a living by farming. There has not been much change in the mode of production. The farmers mainly use cows to help the farming, which is done manually. The products they produce are mostly for themselves. They enjoy the traditional self-sufficient farming life. Various kinds of farming and foresting products are not commercialized to a large scale. Their commercial exchange with outsiders isn't frequent at all as there are few visitors to that region. Most farming households lead a slow and steady life. Changes, if any, to the Lai way of life are small and subtle. In this way, Lai language has been well preserved.

Finally, Lai people possess strong national sentiments and self-awareness, which exert great influence on the attitude towards the Lai language. Language is not only the exterior characteristics of an ethnic group, but also the affectionate tie that holds a nation together. If an ethnic minority group can stick to its own language, it usually shows that its language has sufficient influence. Ethnic minority groups are usually surrounded by other more powerful ethnic groups. Without great sentiments and self-awareness, it is almost impossible for an ethnic minority group to keep its language.

Although its population isn't large, Lai people always keep strong national sentiments and self-awareness. It is the aim of Lai people to be recognized by mainstream society. Especially after the establishment of Guangxi Zhuang Nationality Autonomous Region in 1958, their national sentiments and self-awareness have become stronger. Since they have unique history, language, traditions and costume, Lai people believe themselves to be an independent nationality. It's generally held by Lai people that they are the offspring of King Lai and were dispersed everywhere by the invaders from stronger nations. It was not until in 1990, when Lai people were recognized as part of the Gelao nationality, that their request to become an independent nationality was granted. After this, Lai people began to accept their new identity.

It is the strong national sentiments and self-awareness that helped the Lai people continue to pass on the Lai language from generation to generation. Lai people not only cherish the name of their ethnic group and brotherhood among themselves, but also value the Lai language. The difference between Lai people and people of other ethnic groups is that Lai people take ownership of and have always used the Lai language. They never hide their deep attachment and respect for their native language. This is confirmed by the field study on language attitude. We created a survey consisting of 69 questions to

examine the attitude of participant towards the Lai language in areas inhabited by Lai people. Our findings are as follows:

	pleasant		affable		influential		useful	
	number of subjects	ratio	number of subjects	ratio	number of subjects	ratio	number of subjects	ratio
1 point	1	1.4%	0	0.0%	28	40.6%	2	2.9%
2 point	0	0.0%	0	0.0%	10	14.5%	0	0.0%
3 point	2	2.9%	0	0.0%	15	21.7%	12	17.4%
4 point	1	1.4%	2	2.9%	2	2.9%	2	2.9%
5 point	64	92.8%	66	95.7%	13	18.8%	52	75.4%

The points range from one to five, with one being the lowest score and five being the highest. Higher scores show a higher belief from the participant about whether the Lai language is pleasant, affable, influential or useful. According to the data in the table, Lai language got high scores in terms of being pleasant and affable. In terms of being useful, the score for Lai language is not bad, however participants differ in the influence of Lai language. Some believe that Lai language is influential. They think that Lai language is a precious language because very few people can speak it. Others hold the opposite view. They believe that Lai language is hardly influential if comparing Lai language with mandarin and Zhuang language or even that it is less influential than the Miao and Yi language.

No matter how the outside world views the Lai language, Lai people believe it to be the foundation of their traditional culture and the link which bonds their friends and relatives. Lai people are quite united because of their universal sense of self-identity amongst themselves. Far-sighted people very often attach great importance to the cultivation of language ability in their children. The field study materials confirm this point. Among the 69 participants whose first language is Lai language, fathers of 64 participants spoke to them in Lai language since they were very young and mothers of 62 participants spoke to them in Lai language since they were very young. It shows that under normal conditions, cultivating language skills in the home will guarantee that the native language is preserved.

Mr. Wei, owner of a grocery store in Changfa Road and an ex-serviceman, served in the army in Guilin for four years. Before joining the army, he went to school in other places. He can speak mandarin fluently and some southwest language. His wife is not from the Lai minority group, so he seldom uses Lai language at home. In order to promote his children's ability to

use Lai language, he sent his son to his old house and let him play with his cousins' sons.

Mr. Wei, a teacher in Xinhe primary school, graduated from a junior college. He received a formal education, so he can speak mandarin, the Southwest mandarin and English. When he was very young, he learned the Miao and Zhuang language while playing games with other children. Since his wife is of Yi nationality, he can speak some everyday Yi language. Mr. Wei is an experienced teacher. He thinks that the preservation of the native language needs bi-lingual teaching beginning with primary school education. He taught his students the basic knowledge and culture and trains their ability to think in the native language. He used the native Lai language to facilitate the intellectual development of his students. In this way, his students received education of modern civilization, and also of traditional fine culture as well. The students became bi-lingual or multi-lingual when they were very young so that they are able to meet the fierce economic competition in society and carry on the culture and traditions of their nationality.

The old secretary in Douhong Hillside stated clearly that to lose one's mother tongue is shameful. The offspring should never forget their origin and lose their native language. He said that Lai people in Douhong Hillside would never relinquish their native language. If married to people elsewhere, the girls should teach their children to speak Lai language, or else they would be totally turned out by everyone else should they visit their grandparents' home.

This attitude is conducive to the preservation of language. Girls who marry into other ethnic groups will teach their children to speak Lai language. Therefore it is no wonder that the number of people who use Lai language keeps increasing. Because of self-identity, affability of Lai language and the influence of its traditional culture, daughters-in-law and live-in sons-in-law from other ethnic groups will try to integrate into the community life of Lai people and become fluent in speaking Lai language eventually.

Mr. Cao, Han nationality and live-in husband in Hengsha, is another example. Since he has lived with his Lai family for a long time, he can speak and act like a native Lai person. He can use emotionally charged terms fluently and his way of conduct is exactly like that of a native. He talks with his sons at home using Lai language. Everyone both at home and in the village regard him as a native Lai person. Whenever he runs an errand somewhere, people mistake him for a native Lai person.

A more typical example is a daughter-in-law (De'e Han nationality) in Hengsha. Since there aren't many relatives left at her parents' home, she seldom visits them. She works within the village. Gradually, her ability to speak the Lai language has improved, and is even better than her mandarin.

When talking with other people, she prefers to use Lai language than mandarin.

Conclusion

It's no wonder that national sentiment and self-identification of Lai people helps to strengthen the affinity of Lai language. The rallying power of Lai language unites Lai people closer together and gets them to realize the great difference between themselves and people from neighboring ethnic groups. The rallying power of Lai language and their sense of national sentiments are major driving forces that promote the adherence to their national language. To sum up, the number of people who speak Lai language will witness a steady increase.

REFERENCES

- Compilation Committee of Autonomous County of all ethnic groups in Longlin. 2002. *Annal of Autonomous County of all ethnic groups in Longlin*, Nanning: Guangxi People's Publishing House
- Yang Guangming. 2002. Miao Language in Longlin, Printing House of Autonomous County of all ethnic groups in Longlin, Guangxi
- Yang Yijie. 2003. Yi Ethnic Group in Longlin, Printing House of Autonomous County of all ethnic groups in Longlin, Guangxi
- Chen Chaogui. 2003. Gelao Ethnic Group in Longlin, Printing House of Youjiang Medical College for Nationalities
- Gong Yonghui. 1990. Recognizing Lai through Characteristics of Lai Nationalities, Nanning: Guangxi People's Publishing House
- Yan Qixiang and Zhou Zhizhi. 1995. China's Mon-Khmer languages and the Austro-Asiatic language family, Beijing: Central Nationalities University Press
- Institute of Ethnology, Chinese Academy of Social Sciences and Cultural Division of State Ethnic Affairs Commission. 1994. The situation of language use by Ethnic Minority Groups in China, Beijing: China Tibetology Publishing House
- Ruan Qingping. 2003. Han Nationalities in Longlin, Printing House of Youjiang Times newspaper
- Cen Yingshen. 1997. Selection of Historic Interest of the Zhuang Ethnic Group in Longlin, Printing House of Youjiang Times newspaper
- Liang Min. 1983. Survey of Lai Language, *Minority Languages of China*
- Li Xulian. 1994. Study on Lai Language, Beijing: Central Nationalities University Press

Received: 4 October 2010

Minzu University of China, Beijing
PR China

E-mail address: qinxiaohang@2911.net, susanli60@163.com

Plant and wildlife naming system in southern Pumi*

Xinyuan HE

Southwest Forestry University, China

Abstract:

With a population of about 33,600 (in 2000), the Southern Pumi ethnic group, linguistically classified as Tibeto-Burman, is mainly distributed in Lanping, Lijiang, Weixi and Yongsheng counties of Yunnan Province, as well as in the Yi Autonomous County of Ninglang. This region lies on the borders of Tibetan, Yi, Naxi, Lisu, Bai and Chinese languages and coincides with a world-renowned biodiversity rich environment. In such context, the Pumi have formed an extensive wildlife naming system establishing the relationship between experience, embodied cognition and language. Based on principles of ethno-semantics (Leon, 1977), central guiding principles of cognitive linguistics (Johnson and Lakoff, 2003; Lakoff, 1987; Fillmore, 1982) and principles of categorization of plants and animals in traditional societies (Berlin, 1974, Berlin, 1992), the present research focuses on the wildlife naming system of the Pumi. Data on plant and wildlife names collected from the Pumi communities in the adjacent areas of national nature reserves are selectively analysed and processed. The study attempts to examine the taxonomy, functions and linguistic implications of the plant and wildlife terms in the Pumi language. It is found that there exists in the Pumi language a sophisticated and practical wildlife naming system. The research contributes to the reconstruction of Proto-Tibeto-Burman from the ethno-semantic and cognitive linguistic perspectives.

Keywords: ethnosemantics, cognitive semantics, ethno-botany, taxonomy

* This research is based on ecological and natural resource inventories intended for ecological conservation purposes which were conducted 2003 to 2006 in the adjacent areas of the Laojunshan Nature Reserve, part of the famous Three Parallel Rivers World Natural Heritage Site in NW Yunnan, China. The idea of using the data to be analysed from the linguistic viewpoint has always been what the author wants to do and is initiated as a result of several discussions with Prof. Randy J. LaPolla of La Trobe University and Dr. Luo Yongxian of University of Melbourne on the understanding of the target issue from an ecological linguistic perspective, based on the author's nature conservation experiences and encounters with ethnic languages in the Northwest Region of Yunnan province in China. The data is collected from two Pumi villages in Lanping county in NW Yunnan.

1. Introduction

Language, in its process of evolution, is a reflection of the worldview and perception of the environment by the people who use the language in a given context. One part of the reflection presents itself through the creation of words in different categories describing the experiences of the people concerned. In one way or another, taxonomy would be an appropriate word to summarize the process although the word is often used in biology to denote the classification of different species according to their respective properties and interrelationships. When introduced in linguistics, in this case ethno-semantic, taxonomy could mean the different word categories which incorporate the lexicon of a language. And a lexicon viewed this way can usually show some hints for people to know the different features, components and principles governing the process of concept realization into words. The exploration of such reflection is usually within the research domain of what is known as ethno-semantic and cognitive semantics, although each deals with the issue from different perspectives. From the ethno-semantic point of view, words fall into categories, and much more specifically, contexts of words and groups of words reflect the ways that different cultures categorize concepts into speech and assign meaning to their language (Ottenheimer, 2006, p.18). To put it another way, words express the ideas of people about their worlds and experiences.

From the cognitive semantics point of view, the world view and perception of a people are constructed through meaning construction and knowledge representation based on the experience of the people in the environment they live in and there is a category structure governing the meaning of a language. And this worldview and perception, in one way or another, can be traced from the analysis of the semantic components of the language. This is the case when we examine those traditional societies still practising a lifestyle that has remained much the same for thousands of years. Still, it has been widely acknowledged that every ethnic group has its own unique view of nature, its own interpretation of what the world is all about. And the language used to interpret the world and experience it is usually unique to the people themselves and thus to be understood in a specific context. Knowing a society's concept of nature is tantamount to knowing how a society thinks, acts and negotiates its relationship with the environment around it. And the environment, in this sense, not only includes the physical context, but also the cultural and societal aspects which amount to a holistic environment (language aspects included). In establishing such a relationship, the role of language is of paramount importance.

2. Pumi and Southern Pumi

Pumi is one of the officially acknowledged 56 nationalities of China. Chinese chronicles started to mention the Pumi, known in those texts as *Xifan*, as early as the 7th century A.D. Most descriptions, however, are just sketchy accounts of some ‘barbarians’ living in the frontiers of imperial China. Very few references were drawn to the culture of these people, much less on the language spoken by these people. Systematic studies on the Pumi were launched in the mid-1950s when China started a nationwide ethno-survey as a preparation to identify the different nationalities living within the boundaries of China. These surveys were carried out more or less on a sociolinguistic or cultural anthropological basis, aiming at obtaining data such as population, language, history and other socio-economic data. With these surveys, a brief history of the Pumi and their culture was roughly reconstructed. Names for the Pumi include the Han Chinese name *Xifan* and *Baju* (general pejorative name literally meaning *western barbarians*, once covering virtually every ethnic group living on the western frontiers of China), Yi name *Wotzu*, Naxi name *Ba (Bo)*, and there are also autonyms such as *Pumi*, *Pei Er Mi*, *Peimi*, *Primi*, depending on which dialect area of the spoken language, but all approximately meaning “white”, probably as comparison to the neighbouring Tibetans and Yi branch of *Nuosu* people who usually have darker complexion. But more studies need to be done to establish such a statement.

In Yunnan, where the Pumi people mostly live in Lanping, Lijiang, Weixi, Ninglang and Yongsheng counties, the Pumi are culturally influenced more by the Han, Naxi, Bai and Lisu nationalities. Although a certain form of Tibetan Lamaism is practised in some communities, the dominant population practise, however, certain forms of polytheism, shamanism, *Dongba* and Taoism. In the case of Sichuan, however, the dominant neighbouring inhabitants are Tibetan, and Lamaism is actively and widely practised across the Pumi communities. The Pumi in these communities are geographically, and possibly linguistically, closer to other Qiangic speakers in the region. The Pumi people there were thus willingly to identify themselves as Tibetan in the late 1950s and early 1960s when China identified its official nationalities.

From a typological point of view, the Pumi language is classified into Sino-Tibetan linguistic family, Tibeto-Burman subfamily, Qiangic branch. Pumi population is divided into two major groups, northern (living in SW Sichuan province) and southern (living in NW Yunnan province), with those living in Yunnan put under the name of Pumi *Zu*. This research deals with the southern group, known as Southern Pumi. The 5th national census records the population of the Southern Pumi at 33,600 with a sizable number being active users of the native tongue, especially in the villages some distance from major urban centres. The Southern Pumi (ISO 693-3; *Ethnologue* 15th Ed.), being the southern most of the Qiangic languages, maintains many ancestral traits of

Sino-Tibetan languages and is regarded as “residual type of Loloish” (Matisoff, 1997). The language still keeps a large part of its Qiangic features due to the relatively late migration of its speakers to the present region from their ancestral homes in northwest China in the 7th century AD (Yan and Wang, 1988). Unlike many of its neighbours whose languages are often influenced through language and cultural contacts by either native inhabitants or major dominant groups (i.e. Chinese), the language of the Southern Pumi, while retaining many of its original features, presents a unique semantic feature as a result of the cognitive experiences in the new environment with which the Pumi people have established a close relationship. Such experiences have left fundamental impacts on the Pumi language through modification of the modern day Southern Pumi language.

For decades, study on the language spoken by the Pumi people remained at a fledgling level. There was no in-depth study carried out on the phonology, morphology, syntax, semantics and other fields at all. Since 1980s, following the revival of academic research into ethnic groups and their languages, a series of systematic in-depth investigation into the Pumi and their language has contributed a rich literature in the Chinese language about the Pumi, its culture, society and language. One of the introductory studies of this kind is done by Lu Shaojun, who, in his *Outline of Pumi* (Lu, 1983), describes preliminarily the genetic affiliation, phonology, grammar, vocabulary and dialects of the Pumi. Later researchers on the topic include Yin Haitao, who gives a detailed study on the culture and custom aspects of the Pumi people (Yin, 1993) and Yang Zhaohui who analyses, in addition to culture and customs, some aspects of Pumi vocabulary and its features with references to other Qiangic languages (Yang, 1999). Li Hui in his attempt to design a Latinization writing for the Southern Pumi, discusses aspects of the phonology of the Pumi (Li, 2008).

International investigations on the Pumi language dates back to the 1950s when some researchers attempted to classify the different branches of the Sino-Tibetan languages. Some references were made on the Pumi people and their language when describing the Qiangic languages (Ding, 2003 in Thurgood and LaPolla, ed. 2003). Investigations specifically directed to the Pumi language include Matisoff’s study on the phonology in his *Dayang Pumi phonology and adumbrations of comparative Qiangic* (Matisoff, 1997), based on data from the Southern Pumi in Lanping county in NW Yunnan. Picus Sizhi Ding made a thorough study on the Phonology, Morphology and Syntax of the language based field work data collected in Ninglang county, also in the Southern Pumi region not far from Langping county.

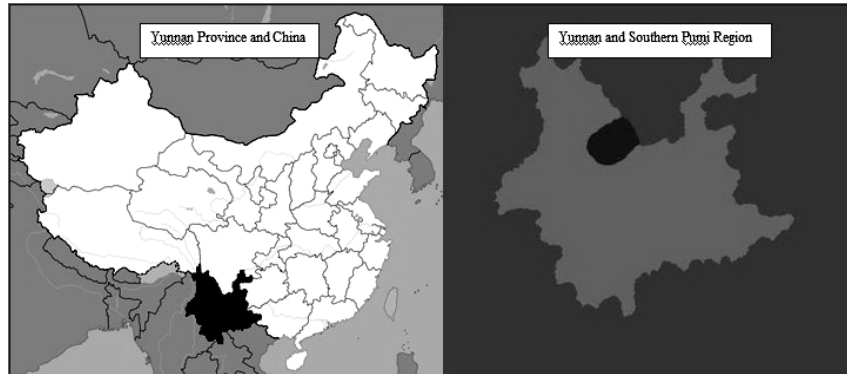


Figure 1. Yunnan province and Southern Pumi Region

3. Physical, Cultural and Linguistic Contexts of the Southern Pumi Area

Some terrain and topographical features of the Southern Pumi region are worth mentioning since they either expand or limit the Pumi view of what is to be named or classified. The Southern Pumi regions straddle the border areas of Yunnan, Sichuan and Tibet. These are mostly rugged or steep mountainous slopes at elevation between 1,300m to 3,000m above sea level. The investigated villages are situated near the Laojunshan and Liming Sections of the world renowned Three Parallel Rivers World Heritage Site and with a mix of geological, climatic, ecological and topographical conditions ranging from snow capped peaks, alpine meadows, pristine forests, serene plateau lakes and valley terraces. This physical context nourishes a rich biodiversity which has been listed as one of 25 key biodiversity hotspots in the world. The Southern Pumi region is in an area which is abundant in biodiversity, and this gives the Pumi an unlimited source from which they can differentiate the subtleties in naming things found in their natural surroundings.

From the cultural and linguistic point of view, the Southern Pumi region lies right in the heartland of a mosaic of Tibeto-Burman speakers and Sino-Tibetan speakers such as the Yi, Naxi, Lisu, Tibetan and Han (Figure 2). Other languages may have influenced the Southern Pumi language in one way or another, but the essence of the language remains original in that it still keeps many of the ancient features of Qiangic branch of Tibeto-Burman.

Language reflects, as is previously discussed, the environment inhabited by the people who use the language. Members of any ethnic group in the world, at one stage or another, respond to the diversity of plants and animals in their habitats by grouping them into labelled or named categories of greater and lesser inclusiveness. The Pumi people are no exception in classifying the things they encounter on a daily basis. Over about a millennium

since migration to the present day region, the Pumi have developed an extensive naming system of the plants and wildlife in the region.

In the last 20 years or so, biodiversity in the Pumi region has long attracted the attention of naturalists, biologists, ecologists, conservationists and policy makers in the government sector and there are a lot of inventories, documentaries and books written about the different plant and wildlife species found in the mountains slopes where the Pumi villages are located. However, instead of paying due attention to the description in the local Pumi language, nearly all the species are recorded using either Chinese names or Latin names and very few are documented using the ethno-names used by the local Pumi people, although the Pumi themselves have a rich vocabulary to describe the environment and species they encounter every day.

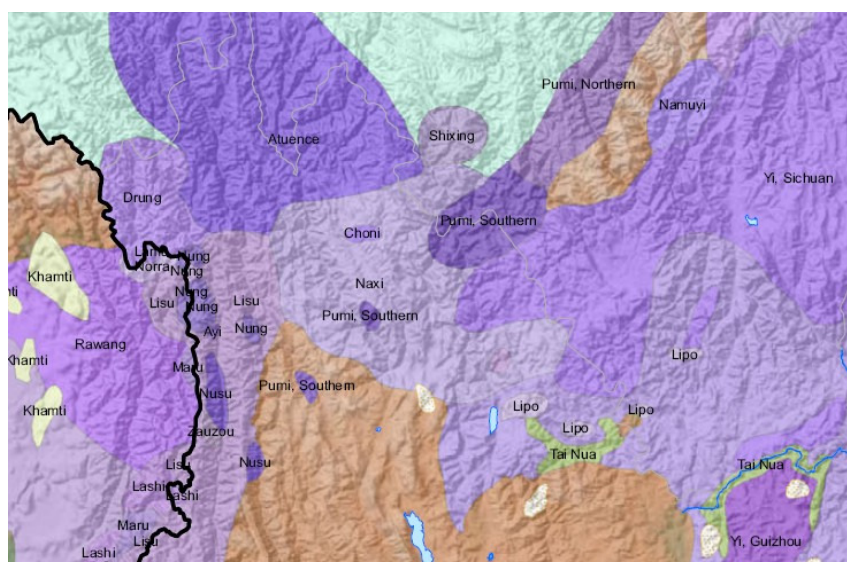


Figure 2. Cultural and linguistic location of Southern Pumi

(Source: Map by Steve Huffman, Data from World Language Mapping System v.3.2)

(Southern Pumi, as is shown on this map, occupies a cultural and linguistics region roughly following a NE to SW direction in the adjacent areas of Yunnan and Sichuan provinces. Other ethnic groups in the region include Naxi, Yi, Lisu, Nung and Tibetan)

4. Key Research Issues of this Paper

The present research focuses on plant and wildlife naming system of the Southern Pumi in their specific environment. Data of plant and wildlife names collected from the Pumi communities in the adjacent areas of national nature reserve (Laojunshan Nature Reserve) are selectively analysed from the ethno-semantic and cognitive semantic perspectives and initially intended to be used for eco-environment protection and natural resources management purposes. The author notices, from the linguistic point of view, there are interesting features in classifying the most common plants and animal life central to the livelihood of the local people, most of them being the Pumi, who inhabit an area of about 4,000 square kilometres, extending from Lanping County in a southwest to northeast direction along both banks of the Lancang (Mekong) and Jinsha (Yangtze) rivers well up into Ninglang County in Yunnan. The inhabitants of the two studied villages—Laridi (腊日底) and Zhumi (主米)—are overwhelmingly Pumi and most of them still have Pumi as their first language and their language is the least influenced by other languages in terms of lexical, syntactic and semantic patterns, although in the region as a whole, bilingualism is becoming increasingly common among the younger generation.

When analysing the existing data, the author bears in mind the following four questions: 1) Is there a wildlife naming system in the Southern Pumi with ethno-semantic and cognitive significance? 2) If there is such a system, what are the mechanisms governing the system? 3) Is this naming system accidental or systematic? 4) What are the linguistic implications of this system in relation to other relevant Tibeto-Burman studies? It is the aim of this paper to present an impressionistic insight into the Southern Pumi wildlife naming system and to contribute to, in one way or another, the study of proto-Tibeto-Burman languages from the ethno-semantic and cognitive semantic perspectives. The findings, however, may be preliminary in nature due to data availability from only two study sites.

5. Southern Pumi Plant and Wildlife Naming System

Over the last decades, quite a few researchers have approached plant and wildlife naming system of the traditional societies from biological, ecological, anthropological, linguistic, cultural and social perspectives. These investigations focus either on the structure of naming or the categories of the things they intend to describe. And a large literature has been written about the different levels of classification, semantic features and cultural implications of the naming system. The most comprehensive documentation of the plant species is *Yunnan Flora*, compiled by Wu Zhengyi, which records altogether 16,201 species found in Yunnan and the Pumi region represents a typical domain in terms of plant species in NW Yunnan (Wu, 2003). Yet, as

elaborated elsewhere in this article, most names are recorded from scientific perspective, with little reference to the commonly used local names known to the Pumi people, thus showing very little features, if any, of the Pumi language. Researchers in cognitive linguistics normally assume that language is the outcome of general properties of cognition (Lakoff and Johnson, 2003), and that conceptual representation is the outcome of the nature of the bodies humans have and how they interact with the geo-linguistic and socio-physical world (Lakoff 1987). The experience of the Pumi people demonstrates essence of these ideas.

One such researcher dealing with the structure of naming is Brent Berlin who pioneers in dealing with folk taxonomy in traditional societies. At the core of Berlin's argument is the 6 level structure of the taxa, and he proposes that there exists a specifiable and partially predictable set of plants and animal taxa that represent the smallest fundamental biological discontinuities easily recognized by any particular habitat. This may, or not, from a structural perspective, reflect the true picture of the ethno-semantic rules of governing the naming mechanisms of traditional societies in representing the world they know. But there might be cases when the rules are environment specific and thus revealing other rules in the naming mechanism (Berlin et al, 1974, Berlin, 1992). Other researchers argue, when dealing with issues of classification and categories in traditional societies, that there are contrasting ideas on whether the naming is "symbolic or arbitrary" or "scientific and technological" (Reason, 1979). Some species are called the name only because we "define" them that way, while there is no direct "scientific" reason for doing so.

In a geo-linguistic region as unique as that of the Pumi, the particular habitat hosting the abundance of biodiversity manifests an environment specific naming mechanism. The Southern Pumi plants and wildlife naming system does share some of the features proposed by these researchers, but in the author's observation there are more than that. There is a unique way in the Pumi language of naming plants and wildlife species and this will be discussed in detail in the following sections.

5.1 Ways of Naming Common Plants in the Southern Pumi Region

The bio-geographical location of the Southern Pumi region in a biodiversity rich area is of vital importance in shaping the Pumi language in terms of the expansion of the lexicon and its development in general. It offers the Pumi people invaluable possibilities and opportunities to explore the environment around them and enrich their experience in the language. Over hundreds of years' adaptation with the new environment, they have accumulated a large set of taxa specific to their living environment.

Although the local Pumi may not have scientific terms such as “ecosystem” and bio-geographical ideas in the sense as we know this type of knowledge, they do have a system of their own which is similar to what we know about, and describes similar entities. As is the situation with the two pilot villages, the landscape is divided by the local Pumi into 7 categories, namely, $\text{ɣ}^{\text{55}}\text{y}^{\text{55}}$ (settlements, villages), $\text{y}^{\text{55}}\text{j}^{\text{55}}$ (terrace farming fields), $\text{s}^{\text{v}^{\text{13}}}\text{s}^{\text{u}^{\text{13}}}\text{y}^{\text{55}}$ (orchards: cash crops such as walnut, apple and chestnuts), $\text{s}^{\text{e}^{\text{13}}}\text{p}^{\text{o}^{\text{55}}}$ (forests: timber forests, fuelwood forests, ritual and sacred forests), $\text{d}^{\text{z}^{\text{u}^{\text{13}}}}\text{s}^{\text{a}^{\text{55}}}$ (livestock grazing lands: montane meadows, river valley grasslands), $\text{f}^{\text{j}^{\text{55}}}$ (water sources: rivers, springs) and $\text{x}^{\text{i}^{\text{55}}}\text{t}^{\text{s}^{\text{u}^{\text{55}}}}\text{g}^{\text{o}^{\text{55}}}$ (sacred sites; burial grounds, religious ceremonial sites). There are subdivisions of each of these 7 major categories, known mainly among the old generations in both villages. Younger generation tends to use only general terms such as *lintzi* (which is borrowed from a local form of Chinese) to mean forests of different categories.

With the distinctions of these different landscapes made, the Pumi, however, treat the different components of the environment as whole and maintain a holistic harmony of these components, which they regard as the key to their existence and livelihood on the daily basis. Table 1 compares the scientific classification of the landscape in the Pumi region with the Pumi classification.

Table 1. Scientific Classification of landscapes vs. Pumi classification of landscapes

Scientific Classification	Pumi Classification
Coniferous forest, shrubs and alpine meadows (elevations approx. from 3,300m to 4,200m)	Forest and grazing lands, water sources (elevations approx. from 2,800m to 4,200m)
Subtropical coniferous forest and deciduous forest (elevations approx. from 2,200m to 3,300m)	Forest and seasonal farmlands (elevations approx. from 2,400m to 2,800m)
Subtropical deciduous and shrubs (elevations approx. from 1,800m to 2,200m)	Settlements, farmlands, orchards, sacred sites (elevations approx. from 1,800m to 2,400m)

(Source: unpublished biological inventory report prepared in 2006)

Speaking of plants, preliminary field work indicates that the Pumi people usually classify plants within primary consideration to their functions

and uses. They have essentially the following primary categories of plant life: trees ($s\tilde{e}^{13}b\tilde{o}^{55}$), crops ($t\acute{e}hu^{55}$), vegetables ($\acute{e}\tilde{e}^{35}r\tilde{e}^{35}$), medicinal herbs ($\acute{z}\tilde{A}s\tilde{y}\tilde{l}^{55}$) and fungus ($m\tilde{z}\tilde{l}^{55}$), and a more ambiguous term *others* (di^{13}). Of course, such classification is based on the daily encounters with the landscape and living environment. Similar to the way of subdividing the primary categories when talking about landscapes, the primary plant life categories are subdivided into numerous smaller entities which could be further broken down into even more specific categories. Take trees ($s\tilde{e}^{13}b\tilde{o}^{55}$) for example, described from the functional perspective, the Pumi subdivide them into *building material* tree ($\acute{y}\tilde{\delta}^{55}s\tilde{e}^{13}$), *fuelwood* tree ($m\tilde{v}^{13}s\tilde{e}^{13}$), *production tool* tree ($s\tilde{e}^{55}\acute{t}\tilde{s}h\tilde{a}^{55}s\tilde{e}^{13}$), *ritual or ceremonial* tree ($hi^{55}s\tilde{e}^{13}$), *fruit* tree ($s\tilde{e}^{13}s\tilde{u}^{35}s\tilde{e}^{13}$), *useless* tree ($tsu^{55}s\tilde{e}^{13}$) (see Table 2). By building tree, they mean the trees are harvested and used as timber for beams, planks or roof blocks when building houses. This category also entails trees could be used to build easy bridges over creeks and rivers and trees to be prepared for aqueducts to channel water from the rivers into the houses and farms. *Fuelwood* tree plays a key role the daily life of the Pumi which centres upon a fireplace in the centre of the living room. Fire burns all the year round to warm up the house and to cook meals. There is a huge consumption of fuelwood and collecting actually takes up a large part of the women workforce in the Pumi villages. Practicality in classifying categories is also found in the *production tool* tree category. The Pumi economy is almost entirely based on agriculture and livestock grazing to a lesser extent, and hunting which used to be widely practised. A lot of tools (i.e. pA^{35} and $t\tilde{o}^{55}$) are needed in conducting these two major activities. Certain tree species are harvested especially for making farming tools. *Ritual or ceremonial* tree actually entails two smaller categories: trees used for religious rituals and ceremonies and trees which are believed to have spirits and needs to be respected. There are subdivisions on of *ritual or ceremonial* trees, incorporating 15 different species used on different ritual and religions occasions. To increase sources of income, the Pumi people often plant some *fruit* trees which include the commonly seen walnut tree, apple tree and a special kind of spice. Interestingly, they even regard chilli pepper crop as *tree*.

Table 2. Function and Use Based Tree Categories in Southern Pumi (with examples)

Examples Categories	Examples expressed in 3 different languages		
	Southern Pumi	Chinese	English
Building material tree ($\text{ʃh}^{55}\text{s}\tilde{\text{e}}^{13}$)	$\text{ʃh}\tilde{\text{a}}^{55}\text{s}\tilde{\text{b}}\tilde{\text{o}}^{55}$	松树	Yunnan pine (<i>Pinus yunnanensis</i>)
fuelwood tree ($\text{m}\tilde{\text{e}}^{13}\text{s}\tilde{\text{e}}^{13}$)	$\text{ʃh}\tilde{\text{a}}^{55}\text{z}\tilde{\text{a}}^{13}$	麻砾树	Oak (<i>Quercus</i>)
production tool tree ($\text{s}\tilde{\text{e}}^{55}\text{t}\tilde{\text{sh}}\tilde{\text{a}}^{55}\text{s}\tilde{\text{e}}^{13}$)	$\text{ʃh}\tilde{\text{a}}^{55}\text{t}\tilde{\text{s}}\tilde{\text{o}}^{55}$	冷杉	Fir (<i>Abies fabri</i> (Mast). Craib)
Ritual and ceremonial tree ($\text{h}\tilde{\text{i}}^{55}\text{s}\tilde{\text{e}}^{13}$)	$\text{s}\tilde{\text{i}}^{13}$	柏枝树	Cypress (<i>Platycladus orientalis</i> (Linn.) Franco arborvitae)
fruit tree ($\text{s}\tilde{\text{e}}^{13}\text{s}\tilde{\text{u}}^{13}\text{s}\tilde{\text{e}}^{13}$)	$\text{k}\tilde{\text{h}}\tilde{\text{d}}\text{s}\tilde{\text{b}}\tilde{\text{o}}^{55}$	核桃树	Walnut tree (<i>uglans regia</i>)
useless tree ($\text{t}\tilde{\text{s}}\tilde{\text{u}}^{55}\text{s}\tilde{\text{e}}^{13}$)	$\text{ʃh}\tilde{\text{a}}^{55}\text{v}\tilde{\text{a}}\text{b}\tilde{\text{o}}^{55}$	倒钩树	<i>canthopanax senticosus</i> (Rupr. Et Maxim.) Harms

A major category in the plant echelon is a wide range of crops the Pumi cultivate. In fact, they subdivide this part of the plants into several groups: grains, beans and *semi-cash* crops such as potato and sunflower seeds.

On the grain division, there are: corn ($\text{q}\text{h}\text{A}^{35}\text{s}\tilde{\text{e}}^{35}$), wheat ($\text{ŋ}\tilde{\text{i}}^{53}$), highland barley ($\text{ŋ}\text{u}^{13}$). These are mainly cultivated for self-consumption, although a small amount is occasionally traded in the countryside fair which is held once in every 5 days.

Vegetable ($\text{e}\tilde{\text{e}}^{35}\text{r}\tilde{\text{e}}^{35}$) is not oftentimes treated as a real crop but an *idle crop*. It is so named because Pumi people do not actually plant vegetables in a certain sense. Most vegetable crops are found planted at the edge of farmlands or in the scattered vacant plots or small gardens around houses to enrich the Pumi kitchen, which is usually dominated by corn cakes, milk tea and barley flour mixture. Thus it is understandable that the Pumi would regard vegetable as an *idle crop*.

Medicinal herbs ($\text{z}\tilde{\text{a}}^{13}\text{s}\tilde{\text{y}}\tilde{\text{t}}^{55}$) represent a huge range of Pumi vocabulary. Traditionally, when someone is sick, Pumi people usually turn to two solutions, seeking treatment from a traditional village healer and consult a religious leader to pacify the spirit of the sick which is believed to be possessed by evil spirits. Sometimes, both solutions are done simultaneously.

Though there is never any formal documentation in the Pumi language of the different medicinal herbs used by the healer, the healing practice is conducted in a systematic way, incorporating a spectrum of herb species catering for different kind of diseases and sickness. The healer in the village is usually a hereditary title and at times, the healer and religious leader is the same person, which is rather common in a lot of traditional societies. The healer masters an incredibly large vocabulary of the medicinal herbs mostly found in the areas near the villages. And there is definitely a gap between the traditional ways of naming these herbs and the scientific way of naming them as is conducted in most biological inventories done in the area up to now. A detailed study needs to be conducted to document the vocabulary of these herbs.

Another major subcategory within plant life is fungus (*mzi*⁵⁵), which is found in abundance in the Pumi region. Fungus is important to the Pumi people in two ways—they provide a rich food and income source. Within this category, there are about 30 words to describe the most commonly seen fungus species found in the forests and shrubs near villages. The Pumi people even have different words to differentiate fungus with different shapes and colours within the same species in the scientific sense. Preliminary interviews and investigations show, however, the naming of the fungus is mostly based on flavour, taste, shape, colour or habitats in which they grow. Of course, this is still within the overall criterion of function and practicality.

Beyond these major categories of plant life the Pumi people normally classify and name, there is a wide range of plant life which is oftentimes lumped together for convenience sake under the general name *others*. Although most plants within this scope are referred to by the Pumi people using demonstrative pronouns which are equivalent to *that* and *this* in English, some other plants in this group do have Pumi names. Interviews with the local villagers show that there is a tendency that the Pumi are using more and more words from Naxi, Yi, Lisu, and especially Chinese, to name these plants they previously referred to as *this* (*tə*¹³) and *that* (*thA*¹³).

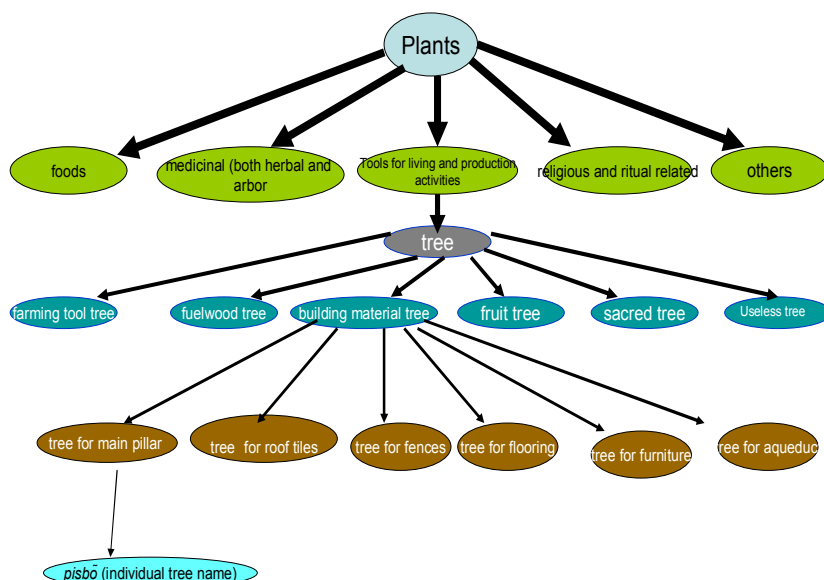


Figure 3. Function and use based Classification of Plants in Pumi Ethno-taxonomy

5.2 Ways of Naming Common Wildlife in the Southern Pumi Region

Corresponding to the richness of plant species, a similar situation is found with wildlife species in the Pumi region where this research is conducted. A functionality and practicality principle also applies in classifying and naming the wildlife species. There are indeed cases which correspond to the scientific ways of classifying and naming these species, for the most part, however, another system specific to local situations of the Pumi exists. In their language, they classify the most commonly seen wildlife into the following 4 primary categories: *useful* (*tehyi*⁵⁵) animals, *harmful* (*dzē*¹³) animals, *sacred* (*hi*⁵⁵) animals and collectively, *others* (*di*¹³). Each of these categories, in turn, is subdivided into several subcategories. Classification and naming the wildlife this way is closely related to livelihood and production activities of the Pumi people. The *useful animal* category consists of two sub-categories: species useful for crop production and species providing subsidiary food sources for the Pumi people. Species useful for crop production include some birdlife (*guē*³⁵*te*³⁵), frogs (*fa*⁵⁵*to*⁵⁵) and honey bees (*dzə*¹³*fi*⁵⁵). Some species of birdlife are classified based on the different seasons they appear, signalling sowing, planting and harvesting seasons. Species providing subsidiary food sources are some big game animals such as black bear (*uā*⁵⁵), musk deer (*sfə*⁵⁵), snub-

nosed monkey (*diu⁵⁵ diu⁵⁵*), wild boar (*pzi⁵⁵*) and pheasant (*luə⁵⁵ ɛō⁵⁵*). Traditionally, some Pumi were involved in hunting, farming and livestock grazing activities. Hunters could still be found among the older generations above 60 years old and they have a rich vocabulary in their language in describing the different species of wildlife they used to encounter in the forests near their villages. Only within the last two decades, hunting ceased to be a major daily activity as a result of the whole scale logging ban of the government and the implementation of some conservation programmes. If documentation of this part of the knowledge is not done soon, it is likely that the vocabulary describing these species will be totally gone. It is worth noticing that the names of the useful animals are mostly marked by certain features with references to crops, to farmlands location and habitats the species appear.

The same principle of functionality and practicality applies to the classification and naming of the *harmful animals*, which, as the name suggested, are the animals not good for their production and living activities for the Pumi people. As is suggested previously, the Pumi society is essentially agriculture based, and crop production is important for them. Farm production related activities occupy much of their time. However, most farmlands are found near forests which are normally the territory and habitats for a large number of wildlife species. Quite often, it is likely that the wild animals will stray into the farmlands to find food and, for this reason, it is understandable that the Pumi would separately assign a name for these wild animals. The most common ones are black bear (*uə*), wild boar (*pzi¹³ ɬɛ⁵⁵*), sparrow (*gue¹³ tsi⁵⁵*) and some birdlife (*gu³⁵ sa¹³*). It should be pointed out that there is overlapping among several species of this category and the species regarded as *useful animals* (i.e. black bear *uə*, wild boar *pzi¹³ ɬɛ⁵⁵* and etc. when treated as game animals).

Dog (*the¹³*) and goat (*tshi⁵⁵*), although not entirely treated as wildlife animals, are two important species playing important ritual and religious functions in the Pumi culture. The Pumi hold reverence for both animals. Dog is believed to have saved the life of the Pumi ancestor at one time and goat accompanies all the way of the Pumi migration from the Pumi homeland in northwest China to the present region. To trace the changes of the names of these two animals could possibly help to draw a clear line of Pumi language development over the thousand-year odd long migration.

6. Discussion

It is predictable, based on preliminary observations, that there exists a wildlife naming system in the Southern Pumi and the mechanisms governing such a system is different what is a generally acknowledged way of classification in the scientific community. This system incorporates a large but

finite set of taxa special in the Pumi region and specific to its members in their cognition and perception of biological reality. The classification and names the Pumi assigned to things they encounter in life are precisely the names of common speech.

To be specific, the Pumi plant and wildlife naming system is characterized by function and practicality oriented considerations, supplemented by folk beliefs and religious connotations. These are the dominant paradigms of Pumi classification. It is likely that the Pumi do not make structural distinctions as is specified by Berlin, they do have a semantic system and lexicon which could serve the daily necessities. And this, there can be a complementary and overlapping categorizations. It is possible that a functional categorization is distinct from, and co-exists with, a species categorization. A close study in the future of the system would reveal more secrets of the semantic features in naming the plant and wildlife in the Pumi region.

7. Conclusion

The Pumi people, like most other people in traditional societies, have a very large vocabulary of names for plants. However, the distinctions between the creatures so named are based on the pragmatic principles of function and practicality developed in hundreds of years' life adaptation to the new environment. This naming system may have connections to other TB language with generic affiliations with the Southern Pumi. With data from other areas, say, cultural, anthropological, ecological and linguistic, the connections seemingly vague now may become clearer to researchers in the field.

However, due to data availability from the linguistic perspective (data being collected from only two Pumi speaking villages in the southwestern-most part of the Southern Pumi Region), the research is too brief to be conclusive in presenting a whole picture, or even part of the picture, of the plant and wildlife naming system of the Southern Pumi. Further study is needed to describe the entire semantic domain of Southern Pumi experiences. This study just catch a glimpse of the cosmos biological knowledge of the Southern Pumi and with further and in-depth study, the picture of the ethno-semantic characteristics of the Southern Pumi can become clearer to the research community. In this sense, this research is not intended as an exhaustive study of the taxonomy mechanism of wildlife in the Pumi language, but a prelude to a series of researches dealing with the language and as a whole, worldview of the Pumi people from the ethno-semantic perspective, and thus can be regarded as a work in progress.

REFERENCES

- Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.
- Berlin, B., Breedlove, D. E. and Raven, P. H. (1974). *Principles of Tzeltal plant classification: An introduction to the botanical ethnography of Mayan-speaking people of Highland Chiapas*. New York, Academic Press.
- Mburu, D. N., Ocheing, J. W., Kuria, S. G., Jianlin, H., Kaufmann, B., Rege, J. E. O. and Hanotte, O. (2003). Genetic diversity and relationship of indigenous Kenyan camels (*Camelus droedarius*) populations: Implications for their classification. *Animal Genetics* 34(1), 26-32.
- Fillmore, C. (1982). *"Frame" semantics in linguistics of Morning Calm*. Seoul: Hanshin Publishing Co.
- Gordon, R. G. J. (ed.). (2005). *Ethnologue: Languages of the world. 15th ed.* Dallas, Texas. SIL International.
- Fei, X. (1988). *Fei Xiaotong Minzu Yanjiu Lunwenji* [Collected Research Papers on Ethnic Studies by Fei Xiaotong]. Beijing: Minzu chubanshe.
- Johnson, M. and Lakoff, G. (2003). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: the University of Chicago Press.
- Liu, Z.-H, Lin, J., Pan, D., Chen, W., Tao, H. and Xu, A. (2002), HLA-DPB1 allelic frequency of the Pumi ethnic group in south-west China and evolutionary relationship of Pumi with other populations. *European Journal of Immunogenetics* 29(3), 259-261.
- Lu, S. (1983). *An Outline of Pumi (Pumi Yu Jian Zhi)*. Beijing: Nationality Press.
- Li, H. (2008). Design and application of Latinization writing for Pumi language in Lanping County. 2nd National Minority Youth Seminar on Natural Language Processing, Hefei, China.
- Matisoff, J. (1997). Dayang Pumi phonology and adumbrations of comparative Qiangic. *Mon-Khmer Studies* 27, 171-213.
- Ottenheimer, H. J. (2006). *The anthropology of language: An introduction to linguistics anthropology*. Belmont, CA: Wadsworth Publishing.
- Reason, E. (1979). *Classification in their social context*. London: Academic Press.
- Thurgood, G. and LaPolla, R. (eds.). (2003). *The Sino-Tibetan languages*. London: Routledge.
- Wu, Z. (2004). *Yunnan Flora*. Beijing: Science Press.
- Xu, J. and Wilkes, A. (2003). Biodiversity impact analysis in Northwest Yunnan China biodiversity and conservation. Yunnan Press of Sciences and Technology.

- Yan, R. and Chen, J. (1986). *Pumi ethnic minority*. Beijing: Nationality Publishing Company.
- Yang, Z. (1999). Panoramic view on Pumi culture (Pumi zu wenhua daguan). Kunming: Yunnan Nationality Press.
- Yin, H. (1993). *Culture and customs of Pumi (Pumi Zu Fengsu Zhi)*. Beijing: China Nationality University Press.

Received: 30 August 2010

Southwest Forestry University
Kunming, PR China
E-mail address: hh_kunming@hotmail.com

A phonetic description and phonemic analysis of Jowai-Pnar¹

Hiram RING
NTU Singapore

Abstract

Pnar, a Mon-Khmer language of Meghalaya, India, has 28 phonemes: 7 vowels and 21 consonants. The language also exhibits onset consonant clusters that go against traditional views of sonority sequencing principles. I consulted two Jowai-Pnar male speakers of 26 and 25 years of age for this analysis, recording around 680 words which each speaker uttered twice. This resulted in over 1,400 tokens per speaker (some words were said more than twice). The resulting sound files were segmented using Praat for detailed acoustic analysis including duration, intensity, and formant measurements, allowing for quantitative as well as qualitative analysis of individual sounds.

Keywords: Pnar, phonetics, phonology

1 Introduction

Pnar is a Mon-Khmer language spoken in northeast India, primarily in the Jaintia Hills District of Meghalaya State. While a few books and articles have been written about the culture, society and history of the Pnar people (also called Jaintia or Synteng), much less research exists on their language. The reason for this may be that Pnar is often grouped with Khasi, a more dominant language of commerce which has a written standard. Pnar and Khasi do share a significant portion of their culture and certain lexical items, but Pnar is distinct linguistically and there is reportedly low mutual intelligibility between the languages. Determining innate intelligibility levels is problematic due to the fact that Khasi is used along with English as a medium of educational instruction in the Jaintia Hills, meaning that educated Pnar speakers can largely communicate in Khasi as well.

The form of the paper is as follows: the rest of section 1 gives a brief account of the language, historical and cultural factors, as well as a review of the meagre literature available for Pnar. Section 2 discusses the methodology used in gathering the data as well as the procedure used in sound segmentation.

¹ I would like to thank Ying Ying Tan and František Kratochvíl for assistance with initial phonetic data, as well as Alec Coupe, Paul Sidwell, and an anonymous reviewer for insightful comments and recommendations on earlier drafts of this paper.

The remaining sections deal with syllable structure, segmental phonemes, allophonic realisations, and challenges from phonotactics. Sound diagrams² and tables are included to elucidate claims regarding segments. A brief discussion of the phonological typology of Pnar concludes the paper, and a glossed wordlist in both close phonetic and broad phonemic transcription is included as an appendix.

1.1 Language

The Pnar language is described as belonging to the Austro-Asiatic family, Mon-Khmer branch, and is sub-categorized as Northern Mon-Khmer, Khasian, according to the Ethnologue (Lewis, 2009). Standard Khasi is described as AVO/SV (Rabel, 1961; Nagaraja, 1985) in which the predicate follows the A-argument and the O-argument follows the predicate, and the assumption in the literature has been that Pnar follows the same pattern (Choudhary, 2004; Tensen, 2011), though my current analysis (in process) suggests a VAO/VS basic word order, which patterns more with Bhoi, another Khasian variety (Nagaraja, 1997). There is no lexical tone, with intonation being used as a prosodic feature across utterances to encode pragmatic rather than lexical information. There are also indications from my initial field research that Pnar is more of a topic-prominent language than a subject-prominent language (see Li and Thompson, 1976), though there is no discussion of this possibility in the literature.

According to the Census of India, in 2001 there were 243,000 speakers of Pnar. Recent reports from 2006 (Daladier, 2011) place the number at around 700,000. Daladier (2010) counts fourteen dialect areas within Jaintia Hills District (Nartiang, Nonjngi, Nongbah, Mynso, Shilliang Myntang, Shangpung, Ralliang, Jowai, Rymbai, Sutnga, Nongkhlich, Lakadong, Narpuh, and Saipung); the prestige dialect is Jowai. Pnar is bordered by Khasi on the west, Assamese on the north, Tibeto-Burman (Kuki-Chin) languages on the east, and War and Garo on the south. The War-Jaintia are often grouped with Pnar, as speakers of War share cultural similarities, but the language is divergent and mutual intelligibility is low. As mentioned briefly above, Khasi is written using a roman orthography and taught in schools in Meghalaya, including in the Jaintia Hills.

An orthography for Pnar is also in some use within Jaintia Hills, but very little has been written in Pnar and thus there is lack of standardization. Further, it is primarily based on the Khasi orthography, and there is some confusion regarding how to represent all the phonemic Pnar sounds that Khasi

² The phrase “sound diagrams” here refers to visual representations of wave forms and spectrographs (images of the latter are called spectrograms), both of which are essential tools for segmenting words into individual sounds.

lacks using the Khasi alphabet. Khasi is listed in UNESCO's Atlas of the World's Languages in Danger³ as vulnerable, but due to its consistent use as a language of instruction, this classification is somewhat problematic. Pnar and other regional varieties that lack the hegemonic status accorded to Khasi are more likely to be in danger, particularly since they are not standardized or used as educational languages of instruction.

At the moment, these languages are in constant use by those who speak them, and children continue to learn Pnar and other regional varieties at home, despite Khasi being acknowledged as the dominant variety. This combined with the unity of cultural forms, practices, and stories among Khasian language speakers (while showing acknowledged variation) may mean that these languages remain strong enough to last another generation. What becomes a concern, then, is the increasing use of English and Hindi by certain sectors of Khasian society (J. War, personal communication) which poses a danger to the languages' continued maintenance. One potential buttress against such danger for Pnar is comprehensive documentation and description, combined with a grassroots effort for education and development of this variety.

1.2 History, culture, and society

According to historical records and Pnar oral tradition (Gurdon, 1914; Gait, 1963 [1906]), the Jaintia kingdom existed from before the 1500s and included not only the current Jaintia Hills district but also parts of the Khasi Hills, the Cachar Hills, parts of Assam, and the northern plains of Bangladesh, with its capital at Jaintiapur in the Bangladesh plains. This was the kingdom which the British East India Trading Company annexed in 1835, after creating an administrative centre for the area a few years earlier, in the Khasi town of Cherrapunji (Sohra). The British deposed the Jaintia king while maintaining the traditional representative system of locally elected officials to govern the Pnar people. Subsequent developments saw the British move their administrative centre of this region from Cherrapunji (Sohra) near Bangladesh to Shillong, further north, during which time the Shillong/Sohra variety of Khasi became the dominant language of communication and commerce for the hills of Meghalaya.

The Pnar people have a vibrant culture with many stories and legends regarding their land and environment. In most respects they share this cosmology with the Khasi people (Sen, 2004; Simon, 1966; Ruffy, 1920). Many Pnar have also converted to Christianity, the largest groups being Catholic and Presbyterian (Barooah, 2007). Agriculture is the main means of livelihood for

³ <http://www.unesco.org/culture/languages-atlas/> (4 January, 2012)

the Pnar, with rice as the staple food. Paddy field cultivation in the river valleys is combined with the raising of chickens, pigs, goats and cows, fishing and other agricultural and horticultural endeavors. The climate of the Jaintia Hills is generally cool and humid, with a regular warm rainy season and cold dry season. Trade and commerce is increasing with the discovery of coal and other raw materials.

Pnar society is based around a matrilineal structure, where the wealth of the clan passes from youngest daughter to youngest daughter, administered by her eldest brother. Some Pnar report that traditional roles are changing due to forces of modernisation (Barooah, 2007). While inheritance and ancestry is matrilineal, men traditionally serve as the spiritual and political leaders, being elected to office by all the men of a village, and performing the rituals and sacrifices appropriate to the circumstances. This traditional system exists alongside local, state and national institutions of India, with protective provisions being written into the Indian constitution under the Sixth Schedule (Pylee, 1994; Ghurye, 1980).

1.3 Previous Research

There is very little literature available on the Pnar language. The first linguistic description of any sort is a small volume by S.J. Grignard ([no date] 1992) based on the Mawkyndeng/Raliang dialect of Pnar and published by the Catholic Church. Though published in 1992, historical evidence that Grignard worked on another grammar published in 1924 (of the Oraon/Kurukh language) suggests that “The Pnar Grammar Book”, as it is called, was written in the early to mid 1900s. This fifty-nine page document includes a list of sounds, some words common to both Pnar and Khasi, a short list of articles, nominal declension, adjectives, pronouns, verbs, adverbs, conjunctions, prepositions, and interjections. It concludes with a brief appendix of conversations and vocabulary.

Most mentions of Pnar in the literature until recently has described it as a dialect or variety of Khasi, following Grierson (1904) and Bareh (1977). One of the first scholarly acknowledgements that Pnar may be a language in its own right comes from Gruessner in his brief article “Khasi: a minority language of northeast India; From an unwritten to a written language”, presented at the 18th European Conference on Modern South Asian Studies in 2004.

Several scholarly works by Indian linguists exist. The most comprehensive descriptive work to date on Pnar proper is unpublished: a PhD thesis of 268 pages by C. Bareh in 2007, himself a Pnar speaker of the Rymbai variety, titled “Descriptive analysis of the Jowai and Rymbai dialects of Khasi” and submitted to the linguistics department at NorthEastern Hill University

(NEHU) in Shillong. Choudhary's (2004) master's thesis from Jawaharlal Nehru University (JNU) is titled "Word order in Pnar", and gives an analysis of Pnar word order in relation to Greenberg's word-order universals. An MA dissertation by Sutradhar (2005) submitted to JNU titled "Relativization and Participialization in Pnar: A Mon-Khmer Language of the Austro-Asiatic Family" could not be accessed, however Sutradhar's ten page excerpt titled "Nominal modification in Pnar: a relativization strategy" (2006) was obtained - it describes relativisation processes and suggests that Pnar does not have a separate class of adjectives. Koshy's (2009) article on "Indefinite pronouns in Pnar" (JNU) is sixteen pages long and provides an analysis of Pnar indefinite pronouns within the framework of Haspelmath's 1997 work on indefinites, with numerous glossed examples.

Other recent sources of information on Pnar are Daladier's (2010) paper "A preliminary classification of the Mon-Khmer languages of India and Bangladesh", presented at the 28th South East Asian Linguistics Society (SEALS 28), and her (2011) paper titled "A multi-purpose project for the preservation of War oral literature", presented at the 4th North East India Language Symposium (NIELS 4) conference. While the latter paper of thirty pages focuses primarily on War, which is another Mon-Khmer variety spoken in this area, it provides some important comparisons between War, Pnar, and Khasi, particularly regarding phonetic differences. Finally, Temsen's (2011) "Variation in the Distribution and Position of Functional Categories in Varieties of Khasi" presented at the 29th South Asian Languages Analysis Roundtable, briefly looks at differences in word order among Khasian languages.

2 Data and Methodology

While in Jowai during May-June, 2011, I consulted two Jowai-Pnar male speakers of 26 and 25 years of age. Using both a Canon Legria HD video camera and a Zoom H4N1 solid state recorder with a Shure Beta 53 headworn microphone, we recorded over 680 words which each speaker uttered twice, resulting in more than 1,400 tokens per speaker (including some duplicates). In the interest of time and clarity, and so as to give a more detailed analysis of a single speech pattern, only the diagrams and data from Speaker 1 (26 yrs.) are given as examples here; in most cases they were corroborated by analysis of Speaker 2's recordings, with slight variation. Using Praat, words were segmented (via textgrid) into individual sounds, resulting in over 7,000 segments per speaker (2,100+ vowels, 4,500+ consonants, 400+ diphthongs). Each segment was analyzed automatically using a script that calculated duration (in milliseconds, based on the start and endpoints of each segment), maximum intensity (in HZ), and formant values (F1, F2, and F3, at the midpoint of the segment). Initial findings were double-checked with speakers in Jowai on a second fieldwork trip in February, 2012.

2.1 Segmenting sounds

The procedure used to segment words into individual sounds is similar to the procedure for measuring voice onset time (VOT) for word-initial stops (as discussed in Ladefoged 2003; 2005, and Singh and Singh, 1982). In the typical procedure for measuring VOT, a measurement (in milliseconds) is taken from 1) where energy is first evident in both the spectrogram and sound waveform to 2) where voicing is clearly visible in the spectrogram. This effectively gets a measurement for voiceless consonant duration from the point of initial release of oral closure to the onset of voicing for the following consonant or vowel.

Notably, this kind of measurement results in negative VOT for voiced consonants, which often exhibit voicing before release (during oral closure). The negative number signifies the duration from onset of voicing to the release spike (“negative” because it occurs before release), but this does not give adequate information about the length of the segment, as VOT does for voiceless consonants⁴. VOT measurements can also be confounded by environment, making it difficult to measure sounds that occur within words. Due to these factors, and desiring to find a means of consistently segmenting all sounds in all words, it seemed best to measure the closure period for voiced sounds separately, using the release spike and the onset of clear formants to identify start and endpoints of each consonant, and for semi-vocalic sounds to negotiate transitions between adjacent sounds by combining auditory evidence with visual evidence from both the spectrogram and sound waveform to identify the point at which transitions occurred, more like the process used in developing digital transcription software (Grabianowski, 2006). The procedure for segmenting sounds within a word using Praat is discussed below and illustrated in Figure 1.

The point where a sound begins at the start of a word is evident from increased energy in the waveform. For voiced sounds this is accompanied by voicing during closure, while for voiceless sounds voicing occurs only after release and is associated with the following sound. Vowels are distinguished from consonants in the spectrogram by having clear voicing and formants, with extremely dark horizontal bands representing intense vocal fold vibration. Voicing on consonants is not as intense as the voicing on vowels, and consonants tend to have more inconsistent high-frequency noise, unlike the clear formant bands visible for vowels. Where transitions are not clear from formants, the waveform view and the intensity contour aid in determining the point of transition between sounds, as the periodic amplitude of the visible

⁴ For other concerns about using VOT as a measure for consonants, see Rothenberg, 2009.

sound wave is lowest at transitional points, and intensity peaks (sharp changes in intensity) can be found near the beginning of a sound.

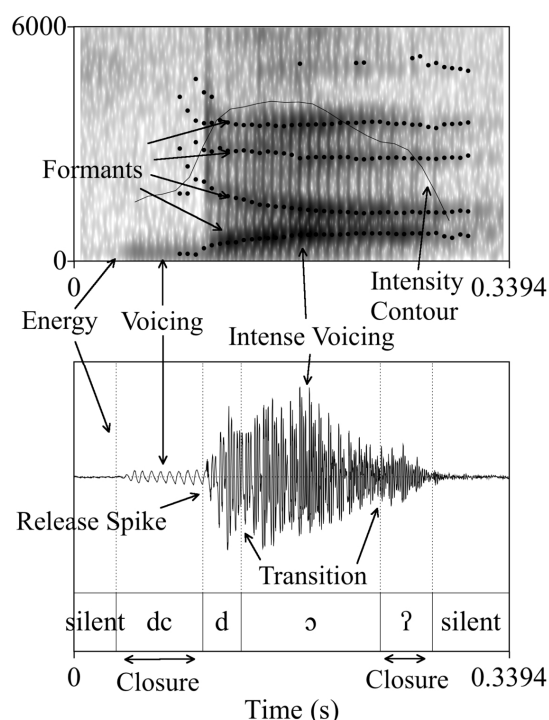


Figure 1. Illustration of sound segmentation, /dɔʔ/ ‘meat’

There were some difficulties in determining the length of consonants, in that they were often followed or preceded by a very short schwa-like sound ([ə] or [ɨ]). Voiced consonants particularly exhibited this feature when followed by other voiced sounds, whether consonants or vowels. For this analysis, short vocalic schwa-like sounds with intense voicing but non-steady formants were treated as transitional elements, with criteria such as intensity, amplitude, and formant location used to determine the point of transition. When they occurred between voiced consonants they were considered to be realisations of post-release voicing, and where they followed voiceless consonants or occurred before a vowel they were treated as constituents of the following voiced sound. The particular properties of schwa-like vowels are considered more explicitly in sections 3.2 and 4.3.

Using the cues of voicing, formants, and waveform amplitude, as well as auditory feedback, the words were segmented into distinct sounds, allowing for automatic measurements of particular constituents in various environments. For this study, the phonetic transcription of each sound was also entered into a

textgrid for labeling purposes, which allows for easy search and quantitative analysis of the resulting tab-delimited output. Segmenting a large database of words from the same speaker increases the likelihood that findings are valid, and comparing more than one such database can suggest that findings are representative.

Figure 1 also demonstrates the means by which sounds like [ʔ] can be identified. This sound is described as closure of the glottis, which in waveforms was represented by reduced amplitude and was represented in spectrograms by inconsistent formants, with the waveform's amplitude deteriorating until there was complete cessation of sound. As segmentation progressed, other sound segments became increasingly recognizable from their waveforms and spectrogram representations, which were quite similar to descriptions in Ladefoged (2003). Nasals showed continuous voicing of the kind associated with voiced stops and no noticeable bursts. Affricates and fricatives were distinguished by increased high-level noise and turbulent waveforms.

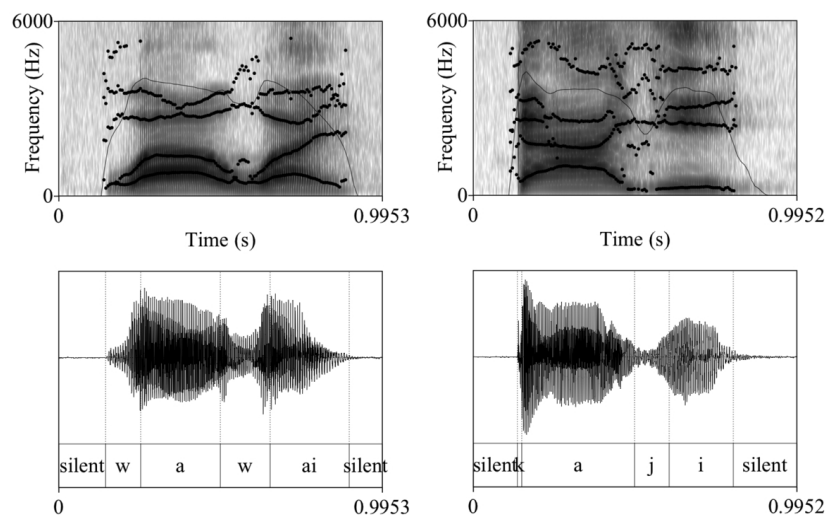


Figure 2. Segmentation of /w/ and /j/

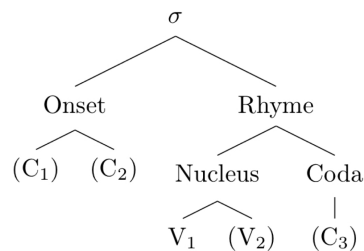
The most difficult sounds to segment were approximants or semivowels /w/ and /j/ (Figure 2) - these were differentiated from vowels and consonants using formant data and waveform amplitude. When the two sounds occurred between other sounds (as in multisyllabic words or compound words in fast speech), each had a lower amplitude in the centre of the segment than at the transition between it and other sounds. Formants for the two sounds were inconsistent throughout their duration, unlike vowels, and palatalisation/labialisation was clear at transitional points between these

sounds and the surrounding vowels. Each of these criteria assisted in helping to determine the exact place at which the transition between approximants and other kinds of sounds occurred.

3 Syllables

The syllable structure of Pnar has significant bearing on the discussion of phonemes: sounds in Pnar can be considered allophones of the same phoneme when they occur in complementary distribution within the syllable. The framework used here follows earlier approaches (from Pike and Pike, 1947, among others, and neatly summarized in Blevins, 1995), where the syllable is composed of two major components, an onset and a rhyme, the latter of which can be further broken down into a nucleus and a coda. In Pnar, onset and coda are optional, with the primary unit being represented by one or more vowels in nucleic position. Pnar poses the additional challenge of exhibiting complex onset clusters that don't conform to the Sonority Sequencing Principle proposed by Clements (1990; 2009), a challenge which will be dealt with further in section 5.1.

3.1 Linear Syllable Structure



Linear Syllable Structure

Minimally, syllables in Pnar can consist of a single nucleic vowel. Maximally, they can include a complex onset of two consonants, a diphthong nucleus, and a coda consonant. There are a few restrictions regarding which constituents can occur in the different slots, though Pnar patterns similarly to Khasi (Rabel, 1961) in having very few restrictions regarding onset consonant clusters. A more specific restriction is that final consonants tend to be unreleased, so that aspirated stops are not found in C_3 position, and neither are fricatives or most approximants (the exception being /w/). This makes it difficult to distinguish between voiced and voiceless syllable-final stops, suggesting a neutralization principle at work within the coda. For vowels, it seems that the nature of the vowel determines its relative intensity (measured in HZ) within the syllable, with the close front vowel [i] having a lower maximum intensity than most other vowels, with only the phonetic schwa [ə]

having a lower average maximum intensity. This, along with the tendency for the final syllable to take primary stress, is a major factor in creating the intonation patterns of individual Pnar words. With these restrictions taken into account, iterations of the structure above can be combined to form multi-syllabic words.

3.2 *Minor Syllables*

There is a second kind of syllable in Pnar, where a nasal/trill/lateral occurs in a semi-nucleic position, carrying the main weight of the first syllable in a disyllabic word. This kind of syllable is similar to the “minor syllable” type reported for Khmer (Henderson, 1952) in which vowels are not realized clearly as syllable nuclei. Several words in the wordlist of this type are given below in Examples 1-4.

1) k̄ɪ̄n.de	2) p̄ɪ̄n.hap	3) k̄ɪ̄l.li	4) p̄ɪ̄r.t ^h ai
nine	CAUS.fall	ask	earth
‘nine’	‘drop’	‘ask’	‘earth’

In the Khasi orthography which Pnar speakers sometimes use for their language, the nature of this syllable type is considered to be $C_1V_1C_3$, where the syllabic consonant is considered to be in C_3 (coda) position, preceded by an extremely short vowel. This situation is illustrated by the Pnar causative morpheme, represented orthographically as ‘pyn’, but transcribed phonetically as [p̄ɪ̄] or [p̄ɪ̄ɪ̄] (as above in Ex. 2). The traditional convention seen in the Pnar-Khasi orthography maintains the linear syllable structure given above, representing the syllabic nasal/trill/lateral sounds as being preceded by a short vowel. However, this does not reflect pronunciation in all cases.

There are two competing explanations for the phenomenon of syllabification of nasal/trill/lateral sounds (called sonants in Rabel, 1961), the first being that the vowel is influencing the realisation of the coda, combining with the coda consonant to form a nucleic unit in preparation for the following syllable, which is stressed. The alternate analysis is that the vowel is a transitional element, a realization of pre-voicing on the syllabic consonant. The first analysis patterns more closely with minor-syllable analyses of Khmeric languages, whereas the second is closer to analyses such as for Bruu (Thongkum, 1979) which recognizes syllabic nasals.

While vowels can occur as the sole constituent of the syllable, Pnar syllables tend to prefer consonant onsets. In disyllabic or multisyllabic words, therefore, CVCV is most often syllabified as $C_1V_1.C_1V_1$, and in monosyllables consisting of a single vowel a glottal stop can often (but not always) be discerned in C_1 position, where it is invisible to most Pnar speakers.

4 Phonemes

The phonemes of Jowai-Pnar are listed in Table 1 and discussed further in 4.1 and 4.3 below.

Table 1. Jowai-Pnar Phonemes

Consonants:					Vowels:		
p, p ^h	t, t ^h	c	k, k ^h	ʔ			
b	d	ɟ			i		u
	s			h	e		o
m	n	ɲ	ŋ		ɛ		ɔ
	r, l	j	w			a	

4.1 Consonants

Pnar consonants have four places of oral articulation and one laryngeal, and five manners of articulation (six if affricates are included as a separate manner) with a voiced/voiceless contrast only holding for pre-velar stops, and aspiration only evident on voiceless plosives. Duration measurements in Table 2 give clear evidence for contrast based on manner of articulation.

Table 2. Mean Duration of Consonants According to Features (ms)

Features	PositionC ₁		PositionC ₂		PositionC ₃
	Closure	Release	Closure	Release	Closure
plosives		28.05		29.10	37.41
vl plosives		16.71		19.73	36.36
vd plosives	109.96	17.85	68.92	15.42	38.49
vl aspirated		74.02		81.04	N/A
affricates		69.23		30.85	44.60
vl affricates		87.33		N/A	N/A
vd affricates	114.17	54.88	54.87	30.85	44.60
trill/lateral		117.81		66.46	103.33
fricatives		123.62		111.62	112.15
approximants		135.70		101.60	136.67
nasals		150.96		74.15	143.70

The (near-)minimal pairs below illustrate phonemic contrasts for aspiration and voicing of stops (except in the case of /ʔ/, which is included for comparison with the other plosives).

/p/ and /p ^h /		/p/ and /b/	
[pa]	‘father’	[pam]	‘cut off’
[p ^h a]	‘2sg.F’	[bam]	‘eat’
/t/ and /t ^h /		/t/ and /d/	
[tam]	‘exceed’	[tiʔ]	‘dig’
[t ^h am]	‘crab’	[diʔ]	‘drink’
/k/ and /k ^h /		/ʔ/	
[ka]	‘3sg.F’	[ʔaʔ]	‘cut/slice’
[k ^h a]	‘birth’	[ʔiʔ]	‘be ripe’

The phonemes /c/ and /j/ are realised as affricates [tʃ] and [dʒ] when occurring in the onset of a syllable, which is where they occur most often in the wordlist. The nature of these consonants will be dealt with further in section 4.2. The following near-minimal pair illustrates the voicing contrast for affricates.

/c/ and /j/	
[tʃan]	‘basket’
[dʒan]	‘near’

The following near-minimal set (minimal for /m/, /n/, and /ɲ/, and near-minimal for /ŋ/) illustrates the place-of-articulation contrasts for nasals.

/m/, /n/, /ɲ/, and /ŋ/	
[m ^h aw]	‘cat’
[n ^h aw]	‘count’
[ɲ ^h aw]	‘seven’
[ŋ ^h aʔ]	‘be tired’

The following (near-)minimal pairs illustrate the contrast between trill and lateral.

/r/ and /l/	
[reʔ]	‘hide’
[leʔ]	‘white’

To round out the survey of phonemes, the following (near-)minimal pairs illustrate the contrast between glottal fricative /h/⁵ and both its glottal counterpart /ʔ/ and the only other fricative present in Pnar words, /s/.

/h/ and /ʔ/		/h/ and /s/	
[hap]	‘father’	[ha]	‘at’
[ʔap]	‘weigh’	[saʔ]	‘choke’

4.2 Consonant Allophonic Realisations

As is clear from the minimal pairs in 4.1, all phonemic consonants show contrast with each other when they occur as the single constituent of the syllable onset. It should also be noted that alveolar fricative [ʃ] exists in the wordlist, as does alveolar approximant [ɹ]. The former is recognized by Pnar speakers as being a Khasi sound that is used in place of /c/. Since Khasi is a neighboring related language and is used as the language of instruction in primary school, Pnar speakers grow up recognizing and sometimes producing the fricative. The alveolar approximant occasionally occurs as a realization of /r/, though very infrequently. This may be a factor of fast speech or articulatory undershoot.

The phonemes /c/ and /j/ are realised most often as [tʃ] and [dʒ]. These sounds have been reported elsewhere as voiceless and voiced palatal stops (Bareh, 1977; Daladier, 2011; Choudhary, 2004; Grierson, 1904; Koshy, 2009) but in Jowai-Pnar their realisation as stops only occurs in the coda of a syllable⁶. Acoustic analysis reveals that in onset positions they are affricates. Consider the minimal pair: /tit/ ‘mushroom’, and /cit/ ‘boil’. Were /c/ a true palatal stop we would expect it to be of similar duration to other voiceless stops in C₁ position. Spectrograms of these words using the same time frame (Figure 3) show a difference in duration between [t] and [tʃ], where over the same time span of 270 ms, [t] has a duration of 18 ms, and [tʃ] has a duration of 100 ms, more than five times longer than [t]. The voiced affricate [dʒ] patterns similarly to [tʃ].

⁵ The term “fricative” is perhaps a misnomer here, as this sound is more accurately described as unobstructed air movement through the vocal cavity.

⁶ Henderson (1976:525) describes the final /j/ as representing “a checked voiceless palatal stop [c]” for Khasi (p.525). In acoustic analysis of the six words where this orthographic representation was used, the sound was heard along with glottal closure, after which glottal pulsing and the sound signal ended.

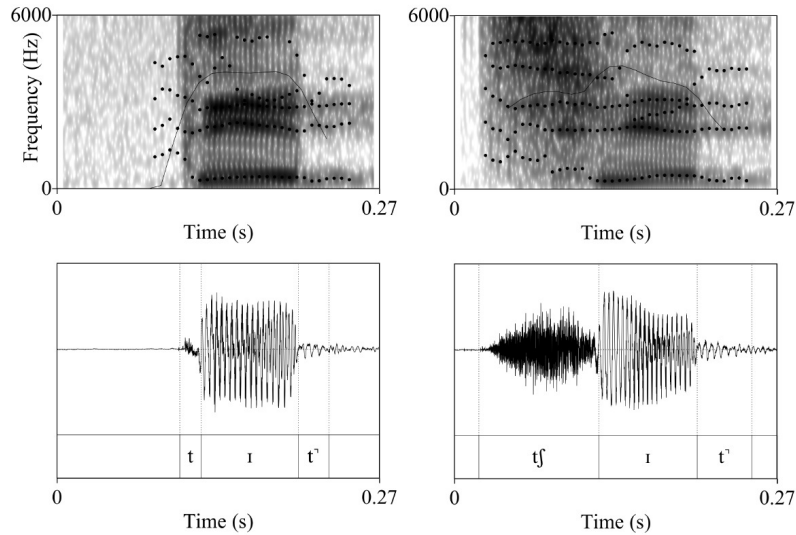


Figure 3. Sound Diagrams of /tit/ 'mushroom' and /cit/ 'boil'

Mean duration measurements, given above in Table 2 in milliseconds, show that the sounds identified in my transcription as affricates are significantly longer than the same speaker's unaspirated plosives in C_1 position, though in C_3 position the mean length is more similar. Furthermore, these sounds have high-frequency frication throughout their release, as illustrated by the spectrograms above. It is possible that in Jowai-Pnar (at least among my consultants) the sound has undergone lenition to an affricate, with length in C_3 position potentially governed by a tendency for coda consonants to be unreleased. In keeping with common practice in Mon-Khmer linguistics, I therefore consider /c/ and /ɟ/ part of the stop series while recognising an affricated manner of articulation for these sounds.

Two other consonantal sounds common to Pnar speakers are found to be allophones of existing phonemes. The Pnar alveolar sounds /t/ and /d/ are pronounced further forward in the mouth than alveolars in other languages, which becomes most clear when they occur word-finally. Rules governing these sounds and the affricates above are summarized in Table 3 below.

Table 3. Consonant Allophonic Realizations

C[alveolar, plosive] → C[dental] / _ #
 /t/ and /d/ realized as [t̚] and [d̚]
 word-finally

C[affricate] → C[unreleased] / _ #
 /tʃ/ and /dʒ/ realized as [t̚] or [d̚]
 word-finally

The rules here seem linked to specifics of allowable coda consonants in Pnar. The general observation is that coda consonants are unreleased, so that a consonant which would ordinarily be differentiated by voicing or frication must be distinguished through other means. It was quite difficult for me to tell the difference between oral voiced and voiceless stops in C₃ position - the source of differentiation may be tongue position, so as to prevent confusion with the glottal stop, or speakers may be listening for other cues, such as length on the preceding vowel. This deserves further investigation.

4.3 Vowels

Table 4. Jowai-Pnar Vowel Phonemes

	front	central	back
close	i		u
close-mid	e		o
open-mid	ɛ		ɔ
open		a	

Vowels in Pnar (Table 4, reproduced from table 1 for convenience) show a clear distinction between close and open positions, as well as front and back, but this is less clear for mid vowels. Mid vowels are the least represented in the wordlist, and rarely occur syllable-finally or syllable-initially, whereas close and open vowels are the most frequent and often occur both syllable-initially and syllable-finally. Several central vowels that occur between mid and close positions seem to be realizations of other vowels or of transitional vocoids. These vowels are analyzed here as phonetic segments but not as phonemic, an issue addressed further in section 4.3.1. A near-minimal set of words that illustrate vowel contrasts is given in Table 5.

Table 5. Vowel Near-Minimal Set

[ksi]	/ksi/	‘louse’	[ksɔʔ]	/ksɔh/	‘hold’
[kse]	/kse/	‘string’	[ksu]	/ksu/	‘grandchild’
[kseʔ]	/kseh/	‘pine’ (tree)	[so]	/so/	‘red’
[ksaŋ]	/ksaŋ/	‘bile’			

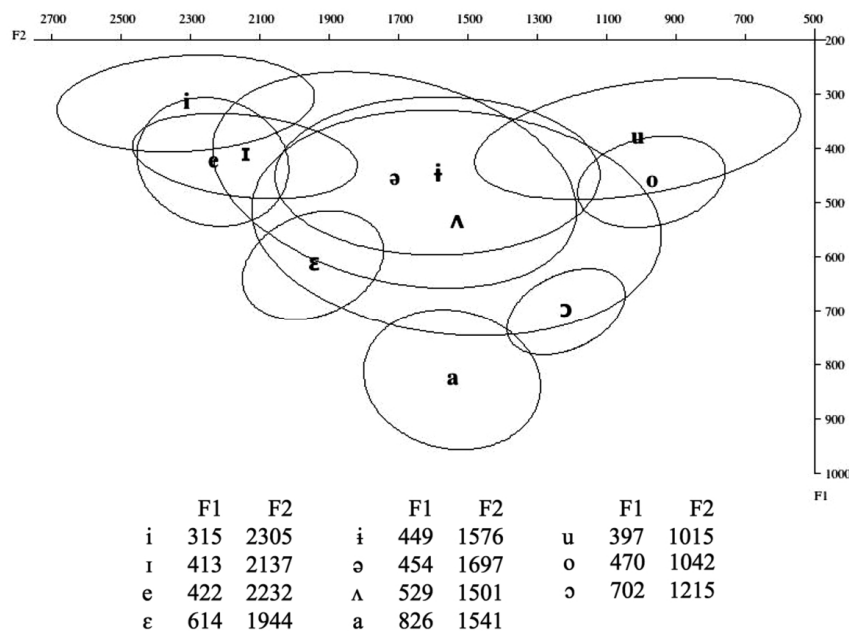


Figure 4. Jowai-Pnar Vowel Scatterplot

The seven vowels listed above are plotted in Figure 4, along with three phonetic mid central vowels (transcribed as [i̯, ə, and ʌ]). The formant data for each vowel is based on measurements of the first three formants that were automatically extracted at each midpoint. The resulting table of values were plotted on separate graphs using raw F1 and F2 values and Bark values for F1 and F2, as well as using the Bark Difference Metric (Syrdal and Gopal, 1986) which takes F3 into account as well. Since there were no significant differences in representation, the traditional manner of plotting F1 and F2 values is used here. The F1 and F2 values for each of the 2,116 vowels of Speaker 1 were first rounded to the nearest whole number in order to plot them using the computer program JPlotFormants (Billerey-Mosier, 2002). In Figure 4 each IPA character represents the mean value of the sound (given in the key below the chart) and ovals represent the general range of variation present in the data.

Each of the cardinal vowel targets in Pnar are evident as separate constituents in the formant plot, with [i, e, a, ɔ, o, and u] clearly occupying their own space, though not without some overlap. The phonetic mid central vowels [i̯, ə, and ʌ] can clearly be grouped together in terms of formants since they overlap so completely, yet it is important to distinguish between [ə, and i̯] and the slightly lower central [ʌ]. The distinction here is one of length, with the

former two vowels being much shorter than [ʌ]. A list of phonetic vowels according to length is given in Table 6.

Table 6. Average Phonetic Vowel Length (Speaker 1)

Vowel	Length (ms)	Vowel	Length (ms)
[ə]	27.47	[u]	123.62
[i]	36.65	[a]	145.63
[ʌ]	62.08	[i]	148.84
[ɪ]	84.85	[o]	184.09
[ɛ]	114.18	[e]	201.33
[ɔ]	120.44		

Phonetic length was initially observed on [o], but repeated tests of possible minimal pairs with speakers in Jowai demonstrated that length is only a significant factor for this vowel in differentiating it from [u] and [ɔ]. It seems that phonetic length contrasts were simply due to emphasis on the part of the speaker. Orthographic conventions confuse the issue, as the only means of differentiating between /o/ and /ɔ/ using the Khasi alphabet is to spell the former ‘oo’ and the latter ‘o’.

4.3.1 Arguments against [i] and [ə] as phonemic vowels

The sounds [i] and [ə] can be heard as preceding syllabic consonants in the minor syllables described above, where they are signified by the orthographic convention ‘y’. They can also be heard in onset consonant clusters or immediately following release of a voiced consonant, though when they occur in these positions their presence is not recognized by Pnar speakers. In the spectrogram, voicing and formants associated with vowels are only evident in these onset clusters when voicing is a feature of one of the constituents. Table 7 illustrates this principle at work.

Table 7. Onset Cluster Voicing

C ₁	‘y’	C ₂	Example	Gloss
[-voice]	NO	[-voice]	[tpai]	‘ashes’
[-voice]	YES	[+voice]	[t(ə)dəŋ]	‘tail’
[+voice]	YES	[-voice]	[b(ə)ha]	‘good’
[+voice]	YES	[+voice]	[b(ə)lai]	‘god’

This principle of consonant clusters seems related more closely to production processes than to phonemic inventory. The sounds [i] and [ə] are likely realizations of voicing within the cluster, and the orthographic convention ‘y’ which occurs before nasal/trill/lateral sounds in minor syllables is more likely to be a syllabification marker than representation of a separate vowel phoneme.

4.4 Vowel Allophonic Realizations

Table 8 displays allophonic realizations of phonemic vowels /i/ and /a/, which are due to environmental factors.

Table 8. Allophonic Vowel Realizations

/i/	→	[ɪ] / _ C [anterior]
		[i] / elsewhere
		<i>/i/ realized as [ɪ] before anterior consonants</i>
/a/	→	[ʌ] / C _ C [unstressed]
		[a] / elsewhere
		<i>/a/ realized as [ʌ] in unstressed syllables (as in fast speech)</i>

The first rule in Table 8 is illustrated by the transcriptions in Table 9, where /i/ takes a different form depending on the coda consonant. Burling (2004) reports a similar pattern for Mandi (Garó), a neighboring Tibeto-Burman language, where /i/ is realised as [ɪ] in closed syllables. Burling also reports /e/ to be realised as [ɛ] in closed syllables, which also occurred in the wordlist - however, the presence of clear minimal pairs for /e/ and /ɛ/ suggests that the confusion between the two vowels is likely due to the lack of a character to represent /ɛ/ in the current orthography. Pnar scholars are trying to address this by introducing the character ‘æ’ to represent the [ɛ] sound.

Table 9. Transcriptions of /i/

	/i/		/i/ → [ɪ]
[tiʔ]	‘dig’	[ʃim]	‘take’
[ʃi]	‘one/an, first’	[ʃiʔ]	‘hot’
		[tɪp]	‘know’

The last rule in Table 8 is illustrated by the transcriptions in Table 10, of words which Pnar speakers recognize as having the same vowel in both syllables. This was the case when the speakers slowed their speech. In fast speech, however, the vowel was heard as a mid vowel rather than an open vowel. As we see in the formant plot above, these transcriptions were corroborated by formant data.

Table 10. Transcriptions of /a/

[sʌ.laʔ] / [sa.laʔ]	‘potato’
[jʌ.liʔ] / [ja.liʔ]	‘smooth’
[saʔ]	‘choke, be stale’

It may be that the distinction for phonemes /i/ and /a/ with their allophones [ɪ] and [ʌ] is a tense/lax distinction. With regard to /a/, it is clear that stress plays a role in the different vowel qualities, where final syllables tend to be stressed, and their vowels to be realized as [a]. This also occurs in diphthongs, where /a/ is realized as [ʌ] when it occurs as the less-syllabic constituent of a diphthong. Further evidence is that /u/ covers such a large range in the formant plot above, suggesting that a more accurate transcription for some vowels represented by /u/ is [ɔ]. It may be that there is articulatory undershoot for unstressed syllables with normal vowels as the speaker psychologically prepares to hit the vowel target on the stressed syllable, which would mean that each of the primary vowels in Pnar should have an unstressed or lax counterpart. With /i/, however, articulatory movement in anticipation of the following anterior sound seems to cause the vowel change, which cannot be completely accounted for by a tense/lax distinction.

4.5 Diphthongs

Diphthongs can be viewed as having two articulatory targets within the same syllable, so that two vowels share the nucleus of a syllable as if they were a single vowel. They can also be viewed as a single vowel with an on- or off-glide which behaves as a secondary vowel constituent. Pnar exhibits both kinds of diphthongs, with the former occurring in open syllables and the latter occurring in closed syllables, where on- and off-glides are conditioned by environment and stress. In open syllables, diphthongs can be characterised as falling, with the first vowel in the nucleus being more syllabic than the second. In closed syllables, however, the syllabicity of each constituent in a diphthong is unpredictable, as speakers do not consistently stress either the first or second constituent. Palatal on-glides tend to be conditioned by alveolar and palatal onset consonants, and off-glides tend to be conditioned by alveolar and palatal coda consonants, but this was also found to vary in the wordlist data. Orthographic conventions treat each vocalic constituent as a separate vowel, which seems justified based on the unpredictability of diphthong stress and glide realisation. The eight diphthongs in the list are illustrated by the words in Table 11.

Table 11. Diphthongs

Diphthong	Open Syllable		Closed Syllable	
	Phonetic	Gloss	Phonetic	Gloss
/ia/	[bia]	‘marry’	[sʔiɬɾ]	‘chicken’
/ei/	[bei]	‘mother’	[deɲ]	‘tree, wood’
/ai/	[dai]	‘sell’	[daɪ̯]	‘bite’
/ie/	N/A		[bniɛɪ̯]	‘tusk’
/ui/	N/A		[ksuɪ̯]	‘evil’
/oi/	[ta.roi]	‘dirty’	[k ^h o:ɪ̯]	‘pure, clean’
/iɔ/	N/A		[ɲɔ̯m]	‘rough’
/io/	N/A		[sɲioʔ]	‘hair’

In relation to nearby languages, Grignard ([no date] 1992), shows how vowels and opening diphthongs in Khasi correspond to closing diphthongs in Pnar, and reports the same number of diphthongs for Pnar as noted here. Daladier (2011) does not report diphthongs for War, spoken south of Pnar, though she does report nasalisation of vowels and palatalisation, as well as possible phonemic length distinctions.

5 Challenges from Phonotactics

5.1 Sonority Sequencing

Several words in the list show complex syllable onsets that go against the Sonority Sequencing Principle (SSP; Clements, 1990), which claims a cross-linguistic tendency for the preference of onsets that maximally rise in sonority as they approach the nucleus of the syllable. A typical sonority sequence in the order of least sonorous to most sonorous is: plosive > affricate > fricative > liquid > nasal > glide > vowel. Voiced versions of each manner of articulation are generally viewed as more sonorant than the voiceless counterparts. The English word ‘stop’ does not follow this principle, and many words in both Pnar and Khasi also do not conform to the SSP (see Rabel, 1961 and Henderson, 1976 for further detail on Khasi initial clusters).

The first example in Table 12 illustrates the maximal syllable in Pnar. This example is allowed by the SSP since nasals are considered more sonorous than plosives. The final two examples in the table are problematic for the SSP, however, since [r] as a voiced trill is more sonorous than [k^h], a voiceless plosive, or [d], a voiced plosive.

Table 12. Three Pnar Onset Clusters

/k ^h ɲiaɲ/	‘insect’
/rk ^h ai/	‘laugh’
/rdəɲ/	‘neck’

When viewed in the spectrograph (Figure 5) there is slight evidence of a vowel between [r] and the following constituent, as noted above in Table 7, where a short central vowel can occur as a transitional element between consonants in a cluster, provided C₁ or C₂ is voiced. Due to the length of [r] and the necessity to voice this sound before the airflow is interrupted, it is difficult to know whether the visible formants are associated with a separate vowel or with the voiced trill [r]. This suggests that [r] is a potentially syllabic constituent that can occur as a minor syllable of the type identified for other Mon-Khmer languages. The possibility is elaborated below in section 5.2, with illustration in Figure 5 and discussion of how native speakers count syllables for these words.

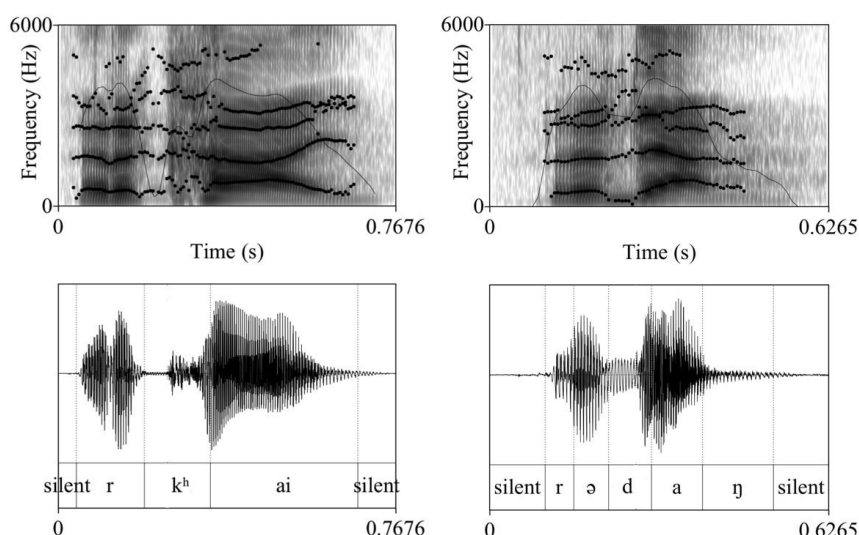


Figure 5. Sound Diagrams of /rk^hai/ and /rdaŋ/

5.2 Sesquisyllables

The term “sesquisyllable” is used for a kind of syllable that is somewhere between a monosyllable and a disyllable (Matisoff, 1989; Thomas, 1992; Bradley, 1980). These kinds of syllables have been reported for a number of Mon-Khmer languages (see Thomas, 1992, for a review of similar structures in about 10 Mon-Khmer languages). The general principle is that over time, a previously disyllabic word is stressed more on the final syllable, so that the penultimate syllable increasingly becomes attached to the final syllable and the word becomes a monosyllable with a complex onset.

In Pnar, final syllables are generally stressed, with non-final syllables tending to have less weight. Over time, the process of lenition could easily

cause the penultimate syllable to combine with the final syllable. This seems evident from the word /rdan/, which consultants said could also be given as [rən.dan] ‘neck’, depending on the dialect. The weakest kind of sesquisyllabic form is one “where the phonetic presyllable is only a predictable open transition between consonants. Phonemically it is a monosyllable.” (Thomas, 1992: 206) This seems to be the case for Pnar, as most consultants claimed a single syllable for both /rk^hai/ and /rdan/ above.

5.3 Feature Rules

Pnar sounds can be described in terms of feature rules, as in Figure 6 (see Chomsky and Halle, 1968, also Carr, 1993). The purpose of using distinctive features to describe consonants and vowels is to create rules with greater explanatory power regarding what exactly is happening in a native speaker's production of sounds.

	p	p ^h	b	t	t ^h	d	k	k ^h	ʔ	tʃ	dʒ	m	n	ɲ	ŋ	r	l	s	h	w	j	
cons	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
syllabic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
voice	-	-	+	-	-	+	-	-	-	-	+	+	+	+	+	+	+	+	-	-	+	+
cont	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
del rel	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
lateral	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
nasal	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-
anterior	-	-	+	+	+	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-
coronal	-	-	-	+	+	+	-	-	-	-	-	-	-	+	-	+	+	+	-	-	-	+
dorsal	-	-	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-
labial	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
strident	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	+	-	-	-
sonorant	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	+	+
approx	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+
spread	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 6. Jowai-Pnar Consonant Features

Using features, the Pnar syllable can also be illustrated by Table 13, though this has its difficulties, particularly in trying to represent the various constituents with feature rules. For elucidation of the difficulties, in the table the actual possible constituents for each slot or node (C₁, C₂, V₁, V₂, C₃) are given in the first set of brackets, and an attempt to combine them in terms of features is given in the second set of brackets.

Table 13. Maximal Syllable Structure with Features

Maximal Template:	(C ₁) (C ₂) V ₁ (V ₂) (C ₃)
where C ₁ =	[p, p ^h , b, t, t ^h , d, k, k ^h , ʔ, c, ʃ, m, n, ɲ, r, l, s, h, w, j] [-syllabic, +cons]
C ₂ =	[p, b, t, d, k, k ^h , ʔ, c, ʃ, m, n, ɲ, r, l, s, w] [-syllabic, +cons]
V ₁ =	[i, e, ε, u, o, ə, a] [+syllabic]
V ₂ =	[i, a] [-syllabic, -cons]
C ₃ =	[p, t, d, ʔ, ʃ, m, n, ɲ, l, r, w] [+cons, -spread, ±syllabic]

The table demonstrates that distinctive feature rules for consonants in C₁ and C₂ position are not particularly explanatory. In C₃ position, using the features [+cons, -spread] seems to be the only way to unite the collection of consonants that can occur (including the feature [±syllabic] allows for this model to incorporate the minor syllables identified above). The only place where feature rules seem to be predictive is in the vowel nodes, where V₂, if it occurs, is generally non-syllabic⁷, in that it shares the syllable with V₁, which retains the larger amount of stress. The words in Table 14 illustrate possible syllables in Pnar:

Table 14. Possible Syllables in Jowai-Pnar

Syllable Template	Phonemic	English Gloss
V1	/i/	‘we’
C1 V1	/le/	‘three’
V1 C3	/ar/	‘two’
C1 V1 V2	/lai/	‘go’
C1 V1 V2	/k ^h ia/	‘heavy’
V1 V2 C3	/eit/	‘dung’
C1 C2 V1	/kba/	‘rice’ (paddy)
C1 V1 C3	/hεʔ/	‘big/wide’
C1 V1 V2 C3	/dait/	‘bite’
C1 C2 V1 V2	/rk ^h ai/	‘laugh’
C1 C2 V1 C3	/pseɲ/	‘snake’
C1 C2 V1 V2 C3	/k ^h ɲiaɲ/	‘insect’

While it is difficult to find unifying features for each place node, there is some evidence from phonotactic analysis that not all consonants can follow

⁷ And possibly [-back], if [a] is analyzed as a mid vowel.

each other in onset clusters. Figure 7 shows the possible consonants and what consonants are found to follow them in the onset of a syllable.

Consonant clusters in onsets of related languages may clarify this issue. Specifically, Rabel (1961) describes Khasi as having consonant clusters in syllable onsets whose order does not seem to be constrained by sonority, suggesting that this is a characteristic of Khasian languages.

This column followed by:	Stops/Plosives							Fricatives/Aff				Liquids		Nasals				Approx			
	p	p ^h	b	t	t ^h	d	k	k ^h	ʔ	s	h	tʃ	dʒ	l	r	m	n	ɲ	ŋ	w	j
p									+	+					+			+			
p ^h														+	+				+		
b											+							+			
t	+		+			+		+						+				+		+	
t ^h														+	+	+	+				
d												+							+		
k	+		+	+	+	+			+	+			+	+	+	+	+	+		+	
k ^h														+	+	+	+	+			
ʔ																					
s	+			+		+	+		+					+		+	+	+	+		
h																					
tʃ			+			+	+		+					+	+		+				
dʒ											+			+					+		
l									+												
r			+			+		+												+	
m															+						
n																					+
ɲ																					+
ŋ																					

Figure 7. Jowai-Pnar Phonotactics

6. Typological Considerations and Conclusion

Attempts toward a typology of phonological systems are limited by the pool of data available, but have resulted in some important theories about what is possible and likely in languages of the world (Crothers, 1978; Lass, 1980; Lass, 1984; Ladefoged and Maddieson, 1996; Maddieson, 2011). The most current information is found in the World Atlas of Language Structures (WALS) online, a database on over 564 languages, with the chapters on phonological typology edited by Ian Maddieson (2011).

Jowai-Pnar clearly exhibits contrast between 21 consonant phonemes and 7 vowel phonemes, placing it in the “average” category in terms of consonants (19-25 consonants), in the “large” category in terms of distinctive vowel quality (7-14 distinctions), and in the “average” category in terms of consonant/vowel ratio. What sets Pnar apart from other similar languages typologically is the complexity in onset consonant clusters that does not follow typical sonority constraints.

Within average sized consonant systems it is typical to have three to four place distinctions, a voicing contrast, and three or more manner contrasts. Pnar has four place distinctions for nasals. If the sounds which are currently affricates have undergone lenition, then historically there were four place distinctions for the plosive category as well. The majority of plosive sounds in Pnar (and the affricates) exhibit a voicing contrast, while voiceless members of this class (excluding affricates) also show an aspiration contrast, which may be an areal feature since neighboring language such as Bangla have an expanded aspiration system.

The lack of a voiced velar plosive can often be attributed to the physiological difficulty of voicing a velar or post-velar closure (Ohala, 1983), and is common to languages geographically distant from each other, but for Pnar it is likely that the loss of this sound is due to historical processes (Paul Sidwell, personal communication). Pnar also exhibits four place distinctions in the nasal continuant category, but only two fricatives (both voiceless), a single trill and lateral in the alveolar place of articulation, and two approximants with two places of articulation (labial-velar and palatal). There was little information to be found regarding the typological status of such sounds.

The contrast Jowai-Pnar exhibits between 21 consonant phonemes and 7 vowel phonemes⁸ means that Pnar patterns with a majority of the languages in the WALS database. However, the complexity in onset consonant clusters that do not follow typical sonority constraints is a unique feature. Further investigation into morphology and continued observation of phonetic realisation may give more insight into the phonemic structure of Pnar.

REFERENCES

- Bareh, Hamlet. 1977. *The language and literature of Meghalaya*. Simla: Indian Institute of Advanced Study.
- Barooah, Jeuti. 2007. *Customary laws of the Jaintias of Meghalaya: with special reference to their land holding system*. Shillong: Law Research Institute, Eastern Region, Gauhati High Court. North Eastern Council.
- Billerey-Mosier, Roger. 2002. Jplotformants v1.4 <<http://www.linguistics.ucla.edu/people/grads/billerey/PlotFrog.htm>>.

⁸ According to Maddieson, diphthongs can be considered phonologically as a sequence of vowels (as in this analysis) or as single vocalic units. Since Pnar speakers regard diphthongs as a combination of two separate vowels, it seems best not to posit a larger number of vowel phonemes than is clearly indicated by the formant plots above. Were diphthongs to be considered individual phonemes, this would raise the vowel inventory count to 14, keeping Jowai-Pnar in the “large” category typologically but changing the consonant/vowel ratio to “low”, a rarer type.

- Blevins, Juliette. 1995. The syllable in phonological theory. In *The handbook of phonological theory*, ed. John A. Goldsmith. Oxford: Blackwell Publishers.
- Bradley, David. 1980. Phonological convergence between languages in contact: Mon-Khmer structural borrowing in Burmese. In *Proceedings of the Sixth Annual Meeting of the Berkeley Linguistic Society*, 259–267.
- Burling, Robbins. 2004. *The language of the Modhupur Mandi (Garo)*, volume 1: Grammar. New Delhi: Bibliophile South Asia, in association with Promilla and Co., Publishers.
- Carr, Philip. 1993. *Phonology*. New York: St. Martin's Press.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Choudhary, Narayan K. 2004. *Word order in Pnar*. Master's thesis, Jawaharlal Nehru University.
- Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I*, ed. John Kingston and Mary Beckman, volume I, 283–333. Cambridge: Cambridge University Press.
- Clements, G. N. 2009. Does sonority have a phonetic basis? In *Contemporary views on architecture and representations in phonology*, ed. Raimy Eric and Charles E. Cairns. Boston: MIT Press.
- Crothers, J. 1978. Typology and universals of vowel systems. In *Universals of human language*, ed. Joseph H. Greenberg, volume 2, 93–152. Stanford University Press.
- Daladier, Anne. 2010. A preliminary classification of the Mon-Khmer languages of India and Bangladesh. In Paper presented at SEALS XX . Zurich, Switzerland.
- Daladier, Anne. 2011. A multi-purpose project for the preservation of War oral literature. In *NEILS 4*, ed. Gwendolyn Hyslop Stephen Morey, Mark Post. New-Delhi: Foundation, Cambridge University Press.
- Gait, Edward Albert. 1963 [1906]. *A history of Assam*. Calcutta: Thacker, Spink.
- Ghurye, G. S. 1980. *The scheduled tribes of India*. New Brunswick, N.J.: Transaction Books.
- Grabianowski, Ed. 2006. How speech recognition works. *HowStuffWorks.com* <<http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm>>.
- Grierson, G.A. 1904. *Linguistic survey of India*, volume II. New Delhi: Low Price Publications.
- Grignard, S. J. [no date] 1992. *The Pnar Grammar Book*. Catholic Church, Mawkyndeng, 793150 Raliang, Meghalaya.
- Gurdon, Philips Richards Thornhugh. 1914. *The Khasis*. London: Macmillan and Co.

- Henderson, Eugénie J. A. 1952. The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14:159–174.
- Henderson, Eugénie J. A. 1976. Khasi initial clusters. *Oceanic Linguistics Special Publications* 13, *Austroasiatic Studies Part I*:523–538.
- India, Government of. 2001. Census 2001, Statement 1: Abstract of speakers' strength of languages and mother tongues. Technical report, Bureau of Statistics, Online.
- Koshy, A. 2009. Indefinite pronouns in Pnar. *The Mon-Khmer Studies Journal* 39.
- Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.
- Ladefoged, Peter. 2005. *Vowels and consonants: An introduction to the sounds of languages*. Malden, MA: Blackwell.
- Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Wiley-Blackwell.
- Lass, Roger. 1980. *On explaining language change*. New York: Cambridge University Press.
- Lass, Roger. 1984. Vowel system universals and typology: prologue to theory. *Phonology* 1:75–111.
- Lewis, M. Paul, ed. 2009. *Ethnologue: Languages of the world*. SIL International, sixteenth edition edition.
- Li, Charles N., and Sandra A. Thompson. 1976. Subject and topic: A new typology of language, 475. New York: Academic Press.
- Maddieson, Ian. 2011. Consonant inventories, vowel quality inventories, consonant-vowel ratio, and syllable structure. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath, chapter 1-3, 12. Munich: Max Planck Digital Library.
- Matisoff, James A. 1989. The bulging syllables, or the mora the merrier: echo-vowel adverbialization in Lahu. In *Southeast Asian linguistics: essays in honour of Eugénie J.A. Henderson*, ed. J.H.C.S. Davidson, 163–97. London: School of Oriental and African Studies.
- Nagaraja, K.S. 1985. *Khasi, a descriptive analysis*. Doctoral Dissertation, Deccan College, Pune.
- Nagaraja, K.S. 1997. Word order in Khasi. In *Languages of tribal and indigenous peoples of India, the ethnic space*, ed. A. Abbi, 347–359. Delhi: Motilal Banarsidass.
- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. Peter F. MacNeilage, volume 22, 189–216. Heidelberg and Berlin: Springer Verlag.
- Pike, Kenneth, and Eleanor Pike. 1947. Immediate constituents of Mazatec syllables. *International Journal of Anthropological Linguistics* 13:78–91.
- Pylee, M. V. 1994. *India's constitution*. New Delhi: R. Chand.
- Rabel, L. 1961. *Khasi, a language of Assam*. Louisiana State University Press.

- Rafy, Mrs. 1920. *Folk-tales of the Khasis*. London, Macmillan and Co., Limited.
- Rothenberg, Martin. 2009. Voice onset time vs. articulatory modeling for stop consonants. *Logopedics Phoniatrics Vocology* 34:171–180.
- Sen, Soumen. 2004. *Khasi-Jaintia folklore: context, discourse, and history*. Chennai: National Folklore Support Centre.
- Simon, I. M. 1966. *Khasi and Jaintia tales and beliefs*. Gauhati: Department of Tribal Culture and Folklore Research, Gauhati University.
- Singh, Sadanand, and Kala S. Singh. 1982. *Phonetics: principles and practices*. Baltimore, Md: University Park Press.
- Sutradhar, Joyshree. 2005. *Relativization and participialization in Pnar, a MonKhmer language of the Austro-Asiatic family*. Master's thesis, Jawaharlal Nehru University, New Delhi.
- Sutradhar, Joyshree. 2006. *Nominal modification in Pnar: a relativization strategy*.
- Syrdal, A. K., and H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79:1086–100.
- Temsen, G.M. 2011. *Variation in the distribution and position of functional categories in varieties of Khasi*. South Asian Languages Analysis Roundtable 29, Mysore: Central Institute of Indian Languages.
- Thomas, David. 1992. On sesquisyllabic structure. *Mon-Khmer Studies Journal* 21:206–210.
- Thongkum, Theraphan L. 1979. The distribution of the sounds of Bruu. *Mon-Khmer Studies Journal* 8:221–93.

Received: 8 December 2011

Nanyang Technological University
Singapore
E-mail address: HIRAM1@e.ntu.edu.sg

Appendix A: Wordlist, Speaker 1

	<i>Phonetic</i>	<i>Phonemic</i>	<i>Gloss</i>
1.	bneɲ	bneɲ	sky
2.	ləʔəʔ	lʔəʔ	cloud
3.	rɨmp ^h um	rɨ.p ^h um	mist
4.	ləʔəʔ k ^h ɨndaw	lʔəʔ k ^h ɨ.daw	fog
5.	sʔem sɲi	sʔem sɲi	sun
6.	sɲi	sɲi	day
7.	bnai	bnai	moon
8.	k ^h lor, k ^h lur	k ^h lur	star
9.	lʔer	lʔer	wind
10.	slap	slap	rain
11.	seɲt ^h ɨlleɪt	seɲ.t ^h ɨ.leɪt	rainbow
12.	meɪt	meɪt	night
13.	snem	snem	year
14.	p ^h ria ^h	p ^h ria	hail/ice
15.	t ^h aʔ	t ^h aʔ	freeze/ice
16.	ʔum	um	water
17.	waʔ	waʔ	river
18.	liɹ	liɹ	stream
19.	leɲ	leɲ	boat
20.	nan	nan	lake
21.	duɹiɹɿw	du.riɹiaw	sea
22.	k ^h ɨndaw pɨɹt ^h ai	k ^h ɨ.daw pɨ.t ^h ai	earth
23.	mo:	mo	stone
24.	ʃʔɛɹp	cʔɛɹp	sand
25.	kteʔ	kteʔ	mud
26.	dʒɨlp ^h oʔ	ʒɨ.p ^h oʔ	dust
27.	ksiɹ	ksiɹ	gold
28.	rupa	ru.pa	silver
29.	lo:m	lom	mountain
30.	pɨnt ^h ɔɹ	pɨ.t ^h ɔɹ	valley
31.	ʃɹɨɲ, ʃɹaiɲ	ca.riɲ	slope
32.	ro:dwaʔ	rod.waʔ	bank (river)
33.	ri	ri	country
34.	rɔŋ	rɔŋ	colour/paint
35.	ʃɹɹɹɹmer	ca.la.mer	shade
36.	pɨt ^h at	pɨ.t ^h at	thunder
37.	ʃɹɹɹɹslap	ca.li.slap	lightning
38.	sɹmoi kɔʒam	sa.moi kɔʒam	cold season
39.	kɨŋrɔŋ I pɨɹt ^h ai	kɨ.rɔŋ i pɨ.t ^h ai	dawn
40.	step	step	morning
41.	ʃɹisɲi	ci.sɲi	day (24 hrs)
42.	rɛm rɛm	rɛm.rɛm	twilight
43.	dʒan meɪt	ʒan meɪt	dusk

44.	dejn	dejn	tree
45.	k ^h lo:	k ^h lo	forest
46.	dejn ksɛ?	dejn ksɛ?	pine tree
47.	pɪr ^h ap	pɪr ^h ap	plank
48.	skum rɪnda	skum rɪ.da	shavings
49.	sdai	sdai	axe
50.	ʃfe?	ce?	thorn/splinter
51.	tɪn do: dejn	tɪ.do dejn	stump
52.	tnaɬ	tnat	branch
53.	ʔum dejn	um dejn	sap
54.	sla	sla	leaf
55.	snei? dejn	snei? dejn	bark (tree)
56.	sɪntu	sɪ.tu	flower
57.	ʃɪlluid	cɪ.luid	bud
58.	t ^h eid	t ^h eid	root
59.	sɔ?	sɔ?	fruit
60.	ʃʔejn	cʔejn	seed/grain
61.	p ^h laŋ	p ^h laŋ	grass
62.	lɔdaw	la.daw	banana
63.	sɔkɪmp ^h ɔr	sa.kɪ.p ^h ɔr	papaya
64.	sni.kɔr	sni.kɔr	coconut
65.	bam	bam	food/eat
66.	ʔumŋap	um.ŋap	honey
67.	k ^h o:	kho	rice
68.	k ^h o: wa duŋ	kho wa duŋ	pounded rice
69.	dʒa	ʒa	cooked rice
70.	sɔru	sa.ru	corn
71.	sɪmbe kba	sɪm.be kba	seed of paddy
72.	kba	kba	rice (paddy)
73.	nai	nai	sesame
74.	blo:ʔ	bloh	salt
75.	muɾɪɬ	mu.rit	red pepper
76.	kwai, pɔt ^h i	kwai, pa.t ^h i	betel
77.	ʔumsɔ?	um.sɔ?	juice(fruit)
78.	dɔ?	dɔ?	meat
79.	sɔla?	sa.la?	potato
80.	tɔ?	tɔ?	beans
81.	ʃru	cru	yam
82.	sɔdhɪɔr	sa.dhier	rice beer
83.	p ^h niɔŋ	p ^h niɔŋ	oil
84.	dud	dud	milk
85.	tit	tit	mushroom
86.	bi?	bi?	poison
87.	skum	skum	husks/nest
88.	sɪŋrai	sɪ.rai	pestle
89.	t ^h lɔŋ	t ^h lɔŋ	mortar

90.	ʃēt	cet	to steam
91.	dejn t ^h aj dij	dejn t ^h aj dij	firewood
92.	dij	dij	fire
93.	tpai	tpai	ashes
94.	tdem	tdem	smoke
95.	t ^h ille? dij	t ^h l.le? dij	flame
96.	k ^h nap	k ^h nap	tongs (chimta)
97.	juj	juj	house
98.	tno:m	tnom	roof
99.	surək	su.rək	road
100.	lūti	lu.ti	path
101.	kse	kse	string
102.	t ^h ria	t ^h r.nia	needle
103.	sula	su.la	clothing
104.	juslejn	ju.slejn	loincloth
105.	kdə?	kdə?	knot/bind
106.	k ^h riwə?	k ^h r.wə?	bind
107.	kɨmp ^h ad	kɨ.p ^h ad	cotton
108.	sɨnk ^h ein	sɨ.k ^h ein	pillow
109.	dejn paləŋ	dejn pa.ləŋ	bedstead
110.	t ^h aw ʃəŋ	t ^h aw cəŋ	seat
111.	wawai	wa.wai	opening
112.	ʃken	ckən	bamboo
113.	rimet	ri.met	cane
114.	ʃdo:?	cdo?	walking stick
115.	sej?	sei?	bamboo spike
116.	t ^h ri	t ^h ri	bamboo thong
117.	ta.dəŋ sej?	ta.dəŋ sei?	bamboo cup
118.	muna	mu.na	bag
119.	ʃaŋ	caŋ	basket (sq)
120.	k ^h nam	k ^h nam	arrow
121.	som	som	spear
122.	waīt	wait	sword
123.	stei?	stei?	shield
124.	ʃama	ta.ma	bronze
125.	nar	nar	iron
126.	rīntej?	rɨ.tei?	bow
127.	k ^h a	k ^h a	wound (n)
128.	metjap	met.jap	carcass/corpse
129.	jap	jap	death/die
130.	dziŋim	ʃiŋ.im	life
131.	k ^h a	k ^h a	birth
132.	blai	blai	god
133.	ksūd	ksuid	evil
134.	t ^h ma	t ^h ma	war
135.	pa.rəm	pa.rəm	tale,story/tell

136.	ʃe	ce	light
137.	sɪŋu	sɪ.ŋu	shadow
138.	mɪŋseɪn	mɪ.ŋ.sein	soul/breath
139.	tiɛŋ	tiɛŋ	fear
140.	kɔʒuɕ	kɔʒut	pain/sick/ill
141.	pʰoːʔ lɔŋtɪɕ	pʰoːʔ lɔŋ.tit	mould
142.	sapsaraʔ	sap.sa.raʔ	rust
143.	kɾem	kɾem	cave
144.	madan	madan	ground/flat
145.	kteɪʔ	kteɪʔ	dirt/filth
146.	tʰlu	tʰlu	hole/pit
147.	eɪɕ	eit	dung/faeces
148.	mɾaɕ	mɾad	animal
149.	ksaw, ksau	ksaw	dog
150.	sniɔŋ	sniɔŋ	pig
151.	sʔiɔɾ	sʔiɔɾ	chicken
152.	rapasa	ra.pa.sa	duck
153.	dʒiŋri	ʒiŋ.ri	cattle
154.	muid	muid	buffalo
155.	maɕi	ma.si	cow (fem)
156.	maɕi kɪɾtɔŋ	ma.si kɪ.ɾ.tɔŋ	bull (male)
157.	kʰɔn maɕi	kʰɔn ma.si	calf (animal)
158.	kule	ku.le	horse
159.	blaŋ	blaŋ	goat
160.	skai	skai	deer (barking)
161.	tʰam	tʰam	crab
162.	da.kʰa	da.kʰa	fish
163.	pseɪŋ	pseɪŋ	snake
164.	rapmanʃeɪʔ	rap.man.ceiʔ	lizard
165.	kʰɾɔʔ sa.kɪɾduk	kʰɾɔʔ sa.kɪ.ɾ.duk	toad
166.	kʰɾɔʔ	kʰɾɔʔ	frog
167.	ɾasaŋ	ra.saŋ	squirrel
168.	kʰne	kʰne	rat
169.	ɾabit	ra.bit	rabbit
170.	ʃɾeɪʔ	ɾeiʔ	monkey
171.	mɪɾsiaŋ	mɪ.ɾ.siaŋ	fox
172.	miow	miow	cat
173.	siŋ	siŋ	lion
174.	krɔŋ	krɔŋ	leopard
175.	dŋeɪm	dŋeɪm	bear
176.	kʰla	kʰla	tiger
177.	haɕi	ha.ti	elephant
178.	bnɪɕɕ	bniet	(elephant) tusk
179.	sim	sim	bird
180.	dhɔʔ	dhɔʔ	owl
181.	rapati	ra.pa.ti	pigeon

182.	nirai	ni.rai	fowl
183.	radha	ra.dha	goose
184.	sʔiar k ^h lo:	sʔiar k ^h lo	junglefowl
185.	k ^h lejn	k ^h lejn	eagle
186.	klaw	klaw	peacock
187.	laʔa:p, laʔa:p	la.ʔap	crow (bird)
188.	t ^h awanɛr	t ^h a.wa.nɛr	wing/feather
189.	p ^h llejn	p ^h l.lejn	egg
190.	tdɔŋ	tdɔŋ	tail
191.	t ^h ɪsim	t ^h ɪ.sim	claw/nail
192.	rejn	rejn	horn
193.	k ^h muɗ	k ^h muɗ	beak
194.	l ^h mbɪt	l ^h m.biɪt	bat
195.	ʔum ɲap	um.ɲap	honey bee
196.	weiʔ	weiʔ	worm
197.	daʔ ^h am sɪŋja	da.t ^h am sɪ.ŋja	scorpion
198.	ɲjaŋ ^h laʔwa	ɲjaŋ.t ^h a.la.wa	spider
199.	ksi	ksi	louse
200.	maɪt	mait	mosquito
201.	p ^h are	p ^h a.re	a fly
202.	k ^h ɲiaŋ	k ^h ɲiaŋ	insect
203.	puid ^s ru	puid.sa.ru	locust
204.	tk ^h u	tk ^h u	ant
205.	l ^h mbnaŋ	l ^h m.bnəŋ	leech
206.	samaɪt	sa.mat	bed bug
207.	t ^h ɪllaʔ	t ^h ɪ.laʔ	flea
208.	wi	wi	one
209.	ʔar	ar	two
210.	le	le	three
211.	so:	so	four
212.	san	san	five
213.	ɲru	ɲ.ru	six
214.	ɲɲaw	ɲ.ɲiaw	seven
215.	p ^h ra	p ^h ra	eight
216.	k ^h ɪnde	k ^h ɪ.de	nine
217.	ʔip ^h aw	ci.p ^h aw	ten
218.	k ^h at wi	k ^h at wi	eleven
219.	k ^h at ar	k ^h at ar	twelve
220.	k ^h at le	k ^h at le	thirteen
221.	k ^h at so:	k ^h at so	fourteen
222.	k ^h at san	k ^h at san	fifteen
223.	k ^h at ɲru	k ^h at ɲ.ru	sixteen
224.	k ^h at ɲɲaw	k ^h at ɲ.ɲiaw	seventeen
225.	k ^h at p ^h ra	k ^h at p ^h ra	eighteen
226.	k ^h at k ^h ɪnde	k ^h at k ^h ɪ.de	nineteen
227.	ʔarp ^h aw	ar.p ^h aw	twenty

228.	lep ^h aw	le.p ^h aw	thirty
229.	so:p ^h aw	so.p ^h aw	forty
230.	sanp ^h aw	san.p ^h aw	fifty
231.	ɲrup ^h aw	ɲ.ru.p ^h aw	sixty
232.	ɲɲjawp ^h aw	ɲ.ɲiaw.p ^h aw	seventy
233.	p ^h rap ^h aw	p ^h ra.p ^h aw	eighty
234.	k ^h iɲdep ^h aw	k ^h ɲ.de.p ^h aw	ninety
235.	ʃispaʔ	cis.paʔ	hundred
236.	sanspaʔ	san.spaʔ	five hundred
237.	ʃiɦʌdʒar	ci.ɦa.ʒar	thousand
238.	wʌrɔʔ	wa.rɔʔ	all
239.	ʃiɓo:n	ci.bun	many
240.	ʃiteɲ	ci.teɲ	half
241.	k ^h ʌdʒiak	k ^h a.ʒiak	some/few
242.	dʌbiaɲ	da.biaɲ	enough
243.	ʔidʒo:ʔ	i.ʒoʔ	equal
244.	ʔidɔr	i.dɔr	price
245.	ɦɛʔ	ɦɛʔ	big/wide
246.	k ^h eɲ, k ^h im	k ^h eɲ, k ^h im	small/narrow
247.	dʒrɔɲ	ʒrɔɲ	long/tall
248.	tbien	tbien	short/low
249.	p̄illun	p̄l.lun	round
250.	laɲdɔɲ	laɲ.dɔɲ	circular
251.	ʒʌliʔ	ja.liʔ	smooth
252.	rəben	rben	thick
253.	staɲ	staɲ	thin
254.	jɔɲ	jɔɲ	black
255.	so:	so	red
256.	leiʔ	leiʔ	white
257.	ʒiɲɲam	ʒɲ.ɲam	green
258.	stem	stem	yellow
259.	rau, raw	raw	dry
260.	dʒheɪʔ	ʒheɪʔ	wet
261.	ʔaɖ	ad	swollen
262.	dap	dap	full (container)
263.	tɲit, tʌroi	tɲit, ta.roi	dirty
264.	ɲɛp	ɲɛp	sharp (blade)
265.	len	len	dull
266.	t ^h iɲme	t ^h ɲ.me	new
267.	ʃiɖ	ciɖ	hot
268.	kɖʒam	kɲam	cold
269.	k ^h ia	k ^h ia	heavy
270.	beɖ	beit	straight
271.	tɔʔ	tɔʔ	right (correct)
272.	bha	b ^h a	good
273.	siʔ	siʔ	bad

274.	rim	rim	old (object)
275.	t̪im̪m̪en	t̪im̪.m̪en	old (person)
276.	d̪ɔŋai	ɟŋai	far
277.	d̪ɔan	ɟan	near
278.	wa	wa	and/with
279.	ha	ha	in
280.	hei	hei	at
281.	ʃapɔʔ	ca.pɔʔ	inside
282.	ha de'n	ha dein	behind
283.	ha k ^h mat	ha k ^h mat	in front
284.	p̪i̪rdi	p̪i̪.di	centre/middle
285.	hā p̪i̪rdi	ha p̪i̪.di	between
286.	kdɔŋ	kdɔŋ	corner
287.	ha k ^h mut	ha k ^h mut	edge
288.	hapɔʔ	ha.pɔʔ	below
289.	timun	ti.mun	rightside
290.	tidieŋ	ti.dieŋ	leftside
291.	miʔsɟi	miʔ.sɟi	east
292.	sɛpsɟi	sɛp.sɟi	west
293.	d̪ɔʔ	ɟoʔ	same
294.	ɟap ^h er	ja.p ^h er	different
295.	he'ni	hei.ni	here
296.	he'tai	hei.tai	there
297.	kani	ka.ni	this
298.	katai	ka.tai	that
299.	ɲnu	ɲ.nu	when?
300.	ʃewɔn	ce.wɔn	where?
301.	ʔu / ka ji	u/ka yi	who?
302.	i.leʔ	i.leʔ	what?
303.	kawɔn	ka.wɔn	which?
304.	i.leʔ	i.leʔ	why?
305.	neibhaʔ	nei.bhaʔ	because
306.	kammɔn	kam.mɔn	how
307.	lada	la.da	if
308.	mtɔʔ	m̪.tɔʔ	not
309.	mpuʔɛm	m̪.pu ɛm	not yet
310.	d̪adɛp	da.dɛp	already
311.	ʃɲŋraŋ	cɲ.raŋ	boy/male
312.	k̪i̪ŋ ^h ai	k̪i̪.t ^h ai	girl/female
313.	bru	bru	person
314.	pa	pa	father
315.	bei	bei	mother
316.	k ^h i̪ŋnaʔ	k ^h i̪.naʔ	child
317.	ʔu kurim	u ku.rim	husband
318.	ka kurim	ka ku.rim	wife
319.	paju, diaʔ	pa.ju, diaʔ	brother

320.	paju k ^h ien	pa.ju k ^h ien	younger sibling
321.	ki bei	ki bei	mother-in-law
322.	bei k ^h ien	bei k ^h ien	m's y. sister
323.	bei hɛ?	bei hɛ?	m's o. sister
324.	sam pun	sam pun	older sister
325.	ɲiasan	ɲia.san	f's older sister
326.	ɲiaroɰ	ɲia.roit	father's y. sister
327.	ki pa	ki.pa	father-in-law
328.	paroɰ pasan	pa.roit, pa.san	father's bro
329.	men	men	father's mother
330.	beipun	bei.pun	grandmother
331.	papun	pa.pun	grandfather
332.	ksu	ksu	grandchild
333.	paju wa hɛ?	pa.ju wa hɛ?	older brother
334.	kʌ wa bia	ka wa bia	bride
335.	ʔu wʌ bia	u wa bia	bridegroom
336.	pɨrsa kurim	pɨ.sa ku.rim	daughter-in-law
337.	k ^h ɔn kɨŋt ^h ai	k ^h ɔn kɨ.t ^h ai	daughter
338.	k ^h ɔn ʃɨŋraŋ	k ^h ɔn cɨ.raŋ	son
339.	paju ʔar bei	pa.ju ar bei	cousin
340.	wʌ ha juŋ	wa ha juŋ	relative
341.	kur	kur	clan
342.	ɲa	ɲa	self/I
343.	pɨrtuɰ	pɨ.tuid	name
344.	p ^h i	p ^h i	you
345.	p ^h i ʔar ɲut	p ^h i ar ɲut	you two
346.	p ^h i wʌ rɔ?	p ^h i wa rɔ?	you pl
347.	ʔu, ka	u, ka	he, she
348.	ʔi	i	we (incl)
349.	ʔi ʔar ɲut	i.ar.ɲut	we two
350.	wa rɔ?	wa rɔ?	we pl
351.	ki	ki	they
352.	ki ʔar ɲut	ki ar ɲut	they two
353.	nɔŋkɨŋdɔŋ	nɔŋ.kɨ.dɔŋ	village
354.	k ^h maɰ	k ^h mat	face/eye
355.	k ^h muɰ	k ^h mut	nose
356.	t ^h lu k ^h muɰ	t ^h lu k ^h mut	nostril
357.	ʃkɔr	ckɔr	ear
358.	k ^h lei?	k ^h lei?	head
359.	ktein	ktein	mouth
360.	snei? ktein	snei? ktein	lip
361.	lʌmeɲ, lameɲ	lameɲ	tooth
362.	ʃɨŋraɲ	cɨ.raɲ	gums
363.	t ^h llei?	t ^h l.lei?	tongue
364.	rɔdaŋ	rdaŋ	neck
365.	taŋbeɲ	taŋ.beɲ	chin

366.	t̄mp ^h a	t̄m.p ^h a	shoulder
367.	rupai	ru.pai	chest
368.	r̄iŋk ^h i	r̄ŋ.k ^h i	back
369.	kl̄oŋsnam	kl̄oŋ.snam	heart
370.	kp̄o?	kp̄o?	abdomen
371.	sner	sner	intestine
372.	nud	nud	liver
373.	kti	kti	arm/hand
374.	k ^h aŋbuid̄	k ^h aŋ.buid	elbow
375.	p̄o?tilt	p̄o?.ti.lit	armpit
376.	pru?	pru?	arm length
377.	sla kti	sla kti	palm
378.	pari ^l am kti	pa.riam kti	finger
379.	kd̄zət̄	k̄jat	leg
380.	sla kd̄zət̄	sla k̄jat	foot
381.	k ^h le? k̄i ^r su	k ^h le? k̄r.su	knee
382.	sk̄ep	sk̄ep	thigh
383.	k ^h awa?	k ^h a.wa?	calf (leg)
384.	k ^h iŋd̄oŋ	k ^h iŋ.d̄oŋ	sole, of foot
385.	pari ^l am kd̄zət̄	pa.riam k̄jat	toe
386.	k ^h awa?	k ^h a.wa?	shin
387.	snam	snam	blood
388.	ʃ?ej̄n	c?ej̄n	bone
389.	snei?	snei?	skin
390.	s̄ŋa ⁱ d̄	s̄ŋaid	fat
391.	?a ⁱ t̄ p ^h are	?ait p ^h a.re	mole on skin
392.	d̄o?	d̄o?	body
393.	lai k ^h lo:	lai k ^h lo	bowels
394.	t ^h l̄o?, pu ⁱ d̄	t ^h l̄o?	penis
395.	kdi, p̄en, t̄pu	kdi, p̄en, t̄pu	vulva
396.	t ^h e ⁱ d̄	t ^h eid	vein
397.	s ^l t ^h ar, sat ^h ar	sa.t ^h ar	sinew
398.	s ^l p̄e ^{t̄} , sap̄e ^{t̄}	sa.p̄et	navel
399.	t̄mp̄oŋ	t̄m.p̄oŋ	waistline
400.	j̄i ^m bu	j̄m.bu	breast
401.	pna	pna	buttocks
402.	ʃr̄o?	cr̄o?	side
403.	t ^h araw	t ^h a.raw	lungs
404.	sp̄u?	sp̄iu?	hair
405.	k̄iŋte ⁿ	k̄ŋ.tein	word
406.	sur	sur	voice
407.	klam	klam	speech/speak
408.	p̄iŋbi	p̄ŋ.bi	joint
409.	d̄z ^l be ^j n	ʃa.be ^j n	brain
410.	ksaŋ	ksaŋ	bile
411.	?a ⁱ t̄ s ^l k ^h mut̄	ait sa k ^h mut	snot

412.	k ^h ak	k ^h ak	phlegm
413.	dʒuŋ	ʒuŋ	urine
414.	ksu ^h ɔ̄	ksuid	pus
415.	p ^h ɪ ^h hier	p ^h ɪ.hier	diarrhoea
416.	ʔim	im	live
417.	reɪ ^h mɪ ^h seɪ ^h m	reɪ ^h mɪ ^h .seɪ ^h m	breathe
418.	sn ^h aw	sniaw	hear
419.	paɪ ^h	pait	see/look
420.	rk ^h ai	rk ^h ai	laugh
421.	jam	jam	weep/cry
422.	bu	bu	suck
423.	mɪ ^h t ^h u	mɪ ^h .t ^h u	spit
424.	slu	slu	blow
425.	daɪ ^h	dait	bite
426.	di ^h ʔ	di ^h ʔ	drink
427.	mu ^h ʔaɪ ^h	mu ^h .aɪ ^h	drunk
428.	prai	prai	vomit
429.	sma	sma	smell/stink
430.	pɪ ^h rk ^h aɪ ^h	pɪ ^h .k ^h at	think
431.	tɪp	tɪp	know
432.	n ^h aw	niaw	count
433.	teɪ ^h	teɪ ^h	afraid
434.	kwa ^h ʔ	kwa ^h ʔ	want
435.	t ^h ia ^h ʔ	t ^h ia ^h ʔ	sleep
436.	t ^h ia ^h ʔ	t ^h ia ^h ʔ	lie down
437.	jeɪ ^h	jeɪ ^h	stand
438.	ʃɔ̄ŋ	cɔ̄ŋ	sit
439.	lai kdʒaɪ ^h	lai kɪat	walk/go
440.	wan	wan	come
441.	psia ^h ʔ	psia ^h ʔ	enter
442.	liwan	li.wan	return
443.	dɔ̄ŋ	dɔ̄ŋ	turn
444.	ji ^h pa	ji ^h .pa	swim
445.	pɛr	pɛr	float
446.	tu ^h ɔ̄	tuid	flow
447.	ɲiaɪ ^h	ɲiat	push
448.	tan	tan	pull
449.	p ^h ak ^h ɔ̄t/pak ^h ɔ̄t	pa.k ^h ɔ̄t	throw
450.	hap	hap	fall
451.	ʔe	e	give
452.	ʃim	cim	take
453.	saɪ ^h	sait	wash
454.	saɪ ^h t ^h ad	sait t ^h ad	launder
455.	p ^h ia ^h ʔ	p ^h ia ^h ʔ	split/diverge
456.	kdɔ̄ʔ	kdɔ̄ʔ	tie
457.	n ^h am	niam	wipe

458.	kiꞥfuꞥ	kr.cut	rub
459.	daꞥ	dat	hit
460.	ʔaʔ	aʔ	cut/slice
461.	duꞥ	duꞥ	stab/pound
462.	tiʔ	tiʔ	dig
463.	ʃbɔꞥ	cbɔt	scratch
464.	ksiʔ	ksiʔ	squeeze
465.	k ^h iꞥweiꞥ	k ^h r.weiꞥ	wring
466.	leiʔkmen	leiʔ.kmen	enjoy
467.	jiꞥp ^h o	jiꞥ.p ^h o	dream
468.	ʃʌniaʔ ʃdeꞥn	ca.niaʔ cdein	lean back
469.	sʌpai	sa.pai	waste
470.	bɔꞥm, daꞥ	bɔꞥm, dat	beat
471.	da ʃe	da ce	be light
472.	tɪꞥmɔʔ	tɪꞥ.mɔʔ	cover
473.	kɛꞥrkuꞥ	kɛꞥr.kut	surround
474.	sɪꞥn ^h iaʔ	sɪꞥn ^h .iaʔ	be sleepy
475.	dai	dai	sell
476.	pɪꞥnsum	pɪꞥ.n.sum	bathe
477.	sɔꞥŋ	sɔꞥŋ	wrap
478.	da doꞥm	da dom	be dark
479.	lum	lum	sweep
480.	t ^h aꞥŋ	t ^h aꞥŋ	burn
481.	sɪꞥnʔaꞥ	sɪꞥn.aꞥ	roast
482.	soꞥr	sor	sew
483.	kreʔ	kreʔ	work
484.	kiꞥndeꞥʔ ke	kɪꞥ.deiʔ ke	play
485.	rwai	rwai	sing
486.	ʃaꞥ	cad	dance
487.	bɔꞥm, ʃɔʔ bɔꞥm	bɔꞥm, choʔ bɔꞥm	drum, to drum
488.	t ^h eꞥiꞥ	t ^h eid	buy
489.	sieꞥ	siet	shoot
490.	lai siet dɔʔ	lai siet dɔʔ	hunt
491.	pɪꞥnɟap	pɪꞥ.n.jap	kill
492.	jaꞥʃɔʔ	ja.cɔʔ	fight
493.	ŋiaʔ	ŋiaʔ	be tired
494.	wiꞥr	wier	disappear
495.	ksɔʔ	ksɔʔ	hold
496.	k ^h ur	k ^h ur	shave (verb)
497.	sɪꞥaw kdʒam	sɪꞥiaw kjam	feel cold
498.	k ^h ɔʔ sneiʔ	k ^h ɔʔ sneiʔ	strip off
499.	kʔi	kʔi	nod
500.	dʌ dɛp	da dɛp	finish
501.	pɪꞥnɟɔ	pɪꞥ.ɟo	drip
502.	baʔ (ha riꞥk ^h i)	baʔ (ha rɪꞥ.ki)	carry (back)
503.	raʔ (ha tɪꞥp ^h a)	raʔ (ha tɪꞥ.p ^h a)	carry (shlder)

504.	jaʔmən	jaʔ.mən	be spoiled
505.	laʔtlo:ɔ̄	lait.loid	be free
506.	sɲəm	sɲəm	be moist
507.	daŋ kʰi:mʲeɕ	daŋ kʰɾ.niet	be awake
508.	kʰut	kʰut	call
509.	wan ɕʌpɔʔ	wan ca.pəh	come down
510.	lai ɕʌpɔʔ	lai ca.pɔʔ	go down
511.	pam, ʔaʔ	pam, aʔ	cut off
512.	ja.ha:p, ja.tɔʔ	ja.hap, ja.tɔʔ	be fitting
513.	kɿlla blu	kɿ.la blu	be blue
514.	tʰaɲ, so:r	tʰaɲ	weave
515.	tɔ:ɔ̄	toid	sag
516.	hadʒrɔŋ	ha.ɟrɔŋ	be high/above
517.	ŋam	ŋam	sink/drown
518.	ɕɔŋŋiaʔ	cɔŋ.ŋiaʔ	rest
519.	sɲeʔ	sɲeʔ	stop
520.	rieʔ	rieʔ	hide (v.)
521.	kʰaraɕ	kʰa.rat	saw (wood)
522.	waʔ	waʔ	hang
523.	blad	blad	be faded
524.	dʒlɔp	ɟlɔp	be wilted
525.	klamkɿnde	klam.kɿ.de	jest
526.	tɿlliaɔ̄	tɿ.liad	grind
527.	peɪʔ	peiʔ	scoop
528.	ɕam	tam	exceed
529.	kʔiaɔ̄	kʔiad	scream
530.	kʰur	kʰur	scrape
531.	ja ʔawri	ja aw.ri	quarrel
532.	kɿrsu, ɕɔʔ	kɿ.su, cɔʔ	knock
533.	kɿlliaŋ	kɿ.lianŋ	exchange
534.	dʒəm	ɟəm	be soft
535.	pɿŋ ʔeʔ	pɿŋ eʔ	coerce
536.	tikʔkaliaw	tik.ka.liaw	tickle
537.	pɿŋɕkeh	pɿŋ.ckɛʔ	be tough
538.	ʔawri	aw.ri	be noisy
539.	kʰleʔ	kʰleʔ	mix
540.	kʰiɾwaitɕ	kʰɾ.wait	twist
541.	pɿŋŋad	pɿŋ.ŋad	be fresh
542.	jaɕəm	ja.cəm	meet
543.	ʔo:d	od	whine
544.	maham	ma.ham	threaten
545.	sdeiʔ	sdeiʔ	fry
546.	takʰo:ʔ	ta.kʰoʔ	hit (collide)
547.	janja	ja.nia	argue
548.	pʰaʔ	pʰaʔ	send
549.	kɿlli	kɿ.li	ask

550.	man ki w ^Λ miε̄t	man ki wa miet	be beautiful
551.	mure?	mu.re?	conceal
552.	brait	brait	scatter
553.	sa:m	sam	distribute
554.	jabha?	ja.bha?	divide
555.	sa?	sa?	choke, be stale
556.	p _h jar	p _h .jar	spread
557.	ra?	ra?	carry/bring
558.	?e?	e?	abandon
559.	k ^h əŋ	k ^h əŋ	block/close
560.	lai tʃ ^Λ dʒrəŋ	lai ca.ʒrəŋ	go up
561.	ku tʃ ^Λ dʒrəŋ	ku ca.ʒrəŋ	ascend (stairs)
562.	ku lo:m	ku lom	come up (hill)
563.	sumk ^h leɪ?	sum.k ^h leɪ?	wash hair
564.	sait t _h ər	sait tiar	wash utensils
565.	wai	wai	open
566.	maja	ma.ja	love
567.	tlep	tlep	bury
568.	beɪsian	beɪ.ɹ.sian	be stupid
569.	?ɪnk ^h aw	ɪŋ.k ^h aw	burned (rice)
570.	jada	ja.da	guard
571.	t ^h ε?	t ^h ε?	pour
572.	psia? tʃ ^Λ pə?	psia? ca.pə?	go in
573.	wan tʃ ^Λ pə?	wan ca.pə?	come in
574.	t ^h uŋ	t ^h uŋ	plant
575.	p _h ndap	p _h .dap	fill
576.	dʒo? sa?	ʒo? sa?	dwell
577.	p _h ɪbi, p _h ɪdaɪt	p _h .bi, p _h .dait	affix
578.	t ^h nam, t ^h ɪrta	t ^h nam, t ^h ɹ.ta	boil
579.	kupt ^h aɖ	kup.t ^h ad	wear shawl
580.	t _h ɪmpia?	t _h ɪ.pia?	break
581.	laɪt	lait	lick
582.	də?, kbiat	də?, kbiat	kiss
583.	ɲo:m	ɲo:m	be rough
584.	t ^h aw təŋ ?um	t ^h aw təŋ um	ladle water
585.	daŋ k ^h i?	daŋ k ^h i?	be in motion
586.	kbai	kbai	shout
587.	dʒɛm	ʒɛm	be tender
588.	jo: luɪtɪ	jo lu.ti	long for
589.	sma	sma	stink
590.	dʒaidʒai	ʒai.ʒai	be gentle
591.	sarə:m, sarə:m	sa.rə:m	be shy
592.	k ^h o:ɪɖ	k ^h oid	be pure/clean
593.	bo:ɪɖ	boid	itch
594.	saɪt kti	sait kti	wash hands
595.	p _h ɪŋtʃkɛh	p _h .ckɛ?	be strong

596.	bo:ɔ̄	bod	follow
597.	wan hadzan	wan ha.ɟan	adhere
598.	ŋo:ɔ̄	ŋoid	swallow
599.	t ^h aw	t ^h aw	measure
600.	ʔap kilo	ap ki.lo	weigh
601.	ŋ sɲ ^h aw ʃa:ɬ	ŋ sɲiaw cait	be nauseated
602.	duk ^h i	du.k ^h i	be sullen
603.	ʃɔʔ kba	cɔʔ kba	thresh
604.	pɪŋk ^h eɪŋ	pɪŋ.k ^h eɪŋ	compress
605.	ksiʔ	ksiʔ	press
606.	dɛmʃʌpɔʔ	dɛm.ca.pɔʔ	bend down
607.	jaʃo:n	ja.con	enmity with
608.	dʒaw	ɟaw	be sour
609.	kt ^h aŋ	kt ^h aŋ	be bitter
610.	ʃʌɾiŋ	ca.ɾiŋ	be steep
611.	pɪŋjo:	pɪŋ.jo	show
612.	ʔiʔ	iʔ	be ripe
613.	t ^h iʌŋ	t ^h iʌŋ	be sweet
614.	rɛm, da do:ʔ	rɛm, da doʔ	defeat/lose
615.	jo:beɪŋ	jo.beɪŋ	oppress
616.	pɪŋbna	pɪŋ.bna	announce
617.	ŋam jillu	ŋam jil.lu	be deep
618.	pɪŋʃitɔm	pɪŋ.ci.tɔm	punish/torture
619.	mɪllij	ml.lij	forget
620.	lai liŋba	lai liŋ.ba	go across
621.	wan liŋba	wan liŋ.ba	come across
622.	lai ʃalaɬ	lai ca.lait	go out
623.	wan ʃalaɬ	wan ca.lait	come out
624.	jeɪŋ	jeɪŋ	erect, v.
625.	pɪŋhap	pɪŋ.hap	drop (vtr.)
626.	ʔɛmdɔr	ɛm.dɔr	be valuable
627.	kɪŋt ^h aʔ	kɪŋ.t ^h aʔ	sting
628.	pɪŋbo:ɔ̄	pɪŋ.bod	imitate
629.	wʌtɔʔ, watɔʔ	wa.tɔʔ	be correct
630.	p ^h o:ʔ	p ^h oʔ	bloom
631.	miʔluŋ	miʔ.luŋ	sprout
632.	kɪŋt ^h iʔ	kɪŋ.t ^h iʔ	jump
633.	bɛɬ, t ^h huŋ	bɛt, t ^h uŋ	reap
634.	wa dɔŋ	wa dɔŋ	be crooked
635.	miʔʃiɬ	miʔ.cit	sweat
636.	k ^h ai kʔɔʔ	k ^h ai kʔɔʔ	cough
637.	ʔaɔ̄	ʔad	be swollen
638.	laŋan	la.tan	shrink
639.	wʌ suk, wa suk	wa suk	be easy
640.	daŋ ʔim	daŋ ʔim	be alive
641.	ro:ɔ̄	roid	draw

642.	pʔut	pʔut	decay/rotten
643.	ro:ɔ̄d lejn	roid lejn	draw a line
644.	rap bo:n	rap bon	be more
645.	tʰlɛ:r	tʰlɛr	slip/slippery
646.	da wiɛr	da wiar	extinguish
647.	kɨlli	kɨ.li	roll up
648.	ʃɔŋ kule, ja?	cɔŋ ku.le, ja?	ride
649.	ʃaŋ	caŋ	bark (of dog)
650.	tikna	ti.kna	be certain
651.	kʰai, mi?	kʰai, mi?	rise
652.	ʃlai	clai	overflow
653.	su kɨlliaŋ	su kɨ.li.aŋ	repay
654.	saiɬ kʰmaɬ	saiɬ kʰmat	wash face
655.	kɨrpaɬ	kɨr.paɬ	request
656.	naŋ dʒor	naŋ ɟor	worse, grow
657.	kʰi?, lɛ?	kʰi?, lɛ?	do
658.	lɛ? biaŋ	lɛ? biaŋ	repeat
659.	wa:je, waje	wa:je	be capable
660.	ʃna	cna	make
661.	dʒiɛɬ	ɟied	pick
662.	jut	jut	flee
663.	ja? luti	ja? lu.ti	lead
664.	pʰɛɬ	pʰɛt	run
665.	no?	no?	cast away
666.	ra:m	ram	lend
667.	paɬ	pat	fall over
668.	kne?	kne?	seize
669.	daiɬ lɛmeɨŋ	daiɬ la.meɨŋ	grind teeth
670.	ʃdɨɬ	cdit	pinch
671.	kɔ:p	kɔp	challenge
672.	pɨŋdʒɛɬ	pɨ.ɟɛt	tear
673.	sʔuɨŋ	sʔuɨŋ	revolve/spin
674.	kʰlei?	kʰlei?	knead
675.	pai?, dʒo:	pai?, ɟo	leak
676.	kɨŋri?	kɨ.ri?	winnow
677.	ʔu pɨŋher	u pɨ.her	to fly

NOTES

Origin of Tai numeral *ha*³ (five)

Xiaohang QIN

Minzu University of China, Beijing

Abstract

Opinions vary with regard to the origin of the Kam-Tai word *ha*³(meaning five) in Chinese academic circles. This paper examined that *ha* ‘five’ in Tai is not borrowed from the Chinese for several reasons. It’s inferred that *ha*³ ‘five’ of Tai originated from the word meaning ‘palm’ or ‘hand’ because the primitive man counted something with his palm or hand and some other reasons.

Keywords: *ha* ‘five’, origin

Opinions vary with regard to the origin of the Kam-Tai word *ha*³(meaning five) in Chinese academic circles. As early as 1990, Prof. Ni Dabai, an expert in Kam-Tai, put forward a new idea on the subject. He said in his article *an introduction to Kam-Tai languages* (1990),

There are two words for the meaning ‘five’, one is *ha*³ and another *ɲo*⁴ in Kam-Tai. And *ɲo*⁴ was probably borrowed from Chinese, but *ha*³ remains a mystery to this day. Fanggui Li reconstructed the primitive shape of *ha*³ as **hɲ* on the basis of *ha*³ (five) in Thai, *lɲa* (five) in Tibetan and **ɲag* (five) in ancient Chinese, so as to prove the cognation between *ha*³ of Kam-Tai and the word meaning ‘five’ in Chinese. But no words meaning ‘five’ prefixed *h* have been found in any Chinese dialects up to now. On the contrary, *p* and *m* correspond to *k*, *kh* and *h* in Kam-Tai. In addition, ‘five’ and ‘hand’ share *mei*³¹ in Gelao (Liangshui Xiang, Qing long county, Guizhou province). And it corresponds to *mu*² (hand) in Thai, *mja*² (hand) in Kam, *mi*² (hand) in Mo and so on.

From these arguments Ni inferred that *ha*³ (five) in Kam-Tai is related to *ma* from the Indonesian word *lima* (meaning five). In some Indonesian

languages such as Paiwan *lima* is used in both the word for *five* and the word for *hand*. From this Ni drew the conclusion that *ha*³(five) in Kam-Tai originated from *lima* meaning ‘hand’ in the Indonesian group.

15 years later, Ding Bangxin, an academician from Taiwan, China, published his paper *On Five and Six in Sino-Tibetan Languages*(in *Minzu Yuwen*,2005(3)).In this paper Ding states that no initial of ‘five’ had been pronounced ‘*h*’ whether in historical or modern Chinese dialects, and so it is unlikely that *h*- in Tai was borrowed from Chinese. There is a cognation between Tai and Chinese judging by the differences of both tone categories in Yin-Yang (Yin= odd – numbered tone such as in the Thai words *ta*:¹‘grandfather’, *ja*:³‘grandmother’ and *khun*³‘officer’; Yang=even-numbered tone such as in the Thai words *kep*² ‘collect’ and *sam*⁴ ‘again’).On this basis Ding believed that *ha*³ of Tai has evolved from **hyag* in his hypothetical ancient Chinese-Tai group.

Only a few months after Ding’s paper had been published, Prof. Wu Anqi also published his paper *Numerals in Sino-Tibetan Languages* (in *Minzu Yuwen*, 2006(2)), discussing the origin of ‘five’ in Kam-Tai. Wu argued that ‘five’ in modern Chinese could be reconstructed as **ho?* in middle ancient Chinese (middle ancient times ranging from the 3rd to the 9th century of Chinese history), and **sya(-?)* in ancient Chinese (ancient times being prior to the 3rd century of Chinese history). Wu drew the conclusion since ‘five’ in the age of Chinese pictophonetic characters had the prefix **s-*. Now ‘five’ in Be has three pronunciations, i.e., *ya*³, *yo*⁴ and *yu*⁴. From these, *ya*³ is just proof that ‘five’ in the proto Tai was borrowed from the ancient Chinese word **hya*^B ‘five’.”

Though Ding and Wu stick to their opposing views of the origin of the Kam-Tai word for ‘five’, they both relate it to the Chinese. Ni however, believes that no relationship exists between the Tai and Chinese words for ‘five’, on which we happen to hold the same view. On this point I agree with Ni, since I consider *ha*³ to have originated from the same word from which the Tai for ‘palm’ also had its origin. However I disagree with him regarding which language these words originated from, and what phonetic form it stemmed from.

In our opinions, *ha* ‘five’ in Tai is not borrowed from the Chinese for the following reasons:

First, the Tai word *ha*³ ‘five’ differs from **yu* ‘five’ in ancient Chinese in initial and there is no correspondence between them. More importantly, there is no correspondence between the tones of these two words. The tone of the Chinese **yu* ‘five’ is a voiced falling-rising tone and ought to correspond to the fourth tone of Kam-Tai. However the tone of the Kam-Tai word *ha*³

‘five’ is the third tone of the Kam-Tai tone system. These are inconsistent so the words cannot be related to Chinese **ɲu* ‘five’.

Second, *ɲa*³ ‘five’ of Be was not borrowed from Chinese but is a native word that has a cognate relation with *ha*³ ‘five’ in Kam-Tai. There are many reasons which support this.

One is that the ‘five’ of Chinese was a word composed with *Yimu* 疑母. *Yimu* is the name of an initial consonant in ancient Chinese. As there were no phonetic symbols in ancient times to record Chinese speech sounds, our ancestors had to use Chinese characters to indicate initial consonants. Thus *Yi* [i³⁵] is written as 疑 in the phonetic system of ancient Chinese, representing the consonant *ɲ*, and 母 [mu²¹⁴] is the name of a Chinese character indicating initial consonants. In addition to this, the initials of all words with *Yimu* were pronounced *ɲ* in Kam-Tai and there is no evidence that it ever changed into *h*, e.g.

Meaning	remote ancient Chinese	middle ancient Chinese	Zhuang	Buyi	Be	Dai
bud	疑鱼* <i>ɲɲea</i>	疑麻开二平假* <i>ɲa</i>	<i>ɲa</i> ²	<i>ɲa</i> ²	<i>lək</i> ⁸ <i>ɲou</i> ⁴	<i>ɲak</i> ⁸
tile	疑歌* <i>ɲɲoa</i>	疑马合二上假* <i>ɲwa</i>	<i>ɲva</i> ⁴	<i>ɲwa</i> ⁴	<i>ɲua</i> ⁴	
forehead	疑铎* <i>ɲɲeak</i>	疑陌开二入梗* <i>ɲ k</i>	<i>ɲek</i> ⁸		<i>ɲak</i> ⁸	
five	疑鱼* <i>ɲɲa</i>	疑姥合一上遇* <i>ɲu</i>	<i>ɲu</i> ⁴		<i>ɲo</i> ⁴	
endure	疑宵* <i>ɲɲau</i>	疑豪开一平效* <i>ɲau</i>		<i>ɲa:u</i> ⁴		
lotus root	疑侯* <i>ɲɲo</i>	疑厚开一上流* <i>ɲəu</i>	<i>ɲau</i> ⁴	<i>ɲəu</i> ³	<i>ɲou</i> ³	
money	疑文* <i>ɲɲiən</i>	疑真开三平臻* <i>ɲiən</i>	<i>ɲan</i> ²	<i>ɲan</i> ²	<i>ɲən</i> ²	<i>ɲ n</i> ²

Meaning	Remote ancient Chinese	Middle Ancient Chinese	Kam	Mulao	Shui	Ma Maonan	Li	Thai
bud	疑鱼* <i>ɲɲea</i>	疑麻开二平假* <i>ɲa</i>	<i>ɲe</i> ²	<i>ɲa</i> ²		<i>ɲa</i> ²		<i>ɲo:k</i> ³
tile	疑歌* <i>ɲɲoa</i>	疑马合二上假* <i>ɲwa</i>	<i>ɲwe</i> ⁴	<i>ɲwa</i> ⁴	<i>ɲwa</i> ⁴	<i>ɲwa</i> ⁴		
forehead	疑铎* <i>ɲɲeak</i>	疑陌开二入梗* <i>ɲ k</i>	<i>ɲek</i> ⁸		<i>ɲak</i> ⁸			
five	疑鱼* <i>ɲɲa</i>	疑姥合一上遇* <i>ɲu</i>	<i>ɲo</i> ⁴	<i>ɲo</i> ⁴	<i>ɲo</i> ⁴	<i>ɲo</i> ⁴		
endure	疑宵* <i>ɲɲau</i>	疑豪开一平效* <i>ɲau</i>		<i>ɲa:u</i> ⁶				
lotus root	疑侯* <i>ɲɲo</i>	疑厚开一上流* <i>ɲəu</i>	<i>ɲau</i> ⁴	<i>ɲau</i> ⁴			<i>ɲou</i> ³	<i>ɲau</i> ³
money	疑文* <i>ɲɲiən</i>	疑真开三平臻* <i>ɲiən</i>					<i>ɲən</i> ¹	

Wang Li, a famous linguist, said, “Speech sounds change according to a category.” Thus, it is not possible that the initial of the Chinese word ‘five’ could have been changed of its own accord into the *h* of Kam-Tai.

The other is that *ŋ* of Be corresponds to some initial *h* of other languages of Tai, e.g.

meaning	Be	Zhuang	Dai
paddy	ŋau ⁴	hau ⁴	—
ox	ŋuP ^{2P}	—	hoP ^{2P}
cheek	ŋakP ^{8P}	huwkP ^{7P}	—
five	ŋa ³	ha ³	ha ³

This shows that *ŋa³* and *ha³* are cognates within Kam-Tai and they have no relation to the Chinese ‘five’. Their similar primitive shapes can be reconstructed as **ŋha³*.

Third, there is a correspondence law, still applicable in today’s modern language, between the eight tones of Kam-Tai and the eight tones of Chinese (the Ping¹, Shang², Qu³ and Ru⁴ tones were divided into Yin⁵ and Yang⁶ categories). Some examples of the correspondence law between Zhuang tones and Chinese tones are as follows:

	Tone categories		Tone marks	Tone pitches	Example	Chinese tone categories
Stretchy tone	The 1th tone		No mark	24	Na (thick) daeng (lamp)	Yinping
	The 2nd tone		Final+z	42	Naz (field) daengz (arrive)	Yangping
	The 3ird tone		Final+j	55	Naj (face) daengj (erect)	Yinshang
	The 4th tone		Final+x	33	Ngvax (title) max (horse)	Yangshang
	The 5th tone		Final+q	44	Vaq (trousers) daengq (stool)	Yinqu
	The 6th tone		Final+h	21	Vah (word) caengh (steelyard)	Yangqu
Plosive tone	High-sounding groups (The 7th tone)	short vowel	The terminal sounds: p, t, k	55	Daep (liver) baet (sweep) baek (plant)	Yinru
		long vowel	Ditto	44	Dap (pagoda) bak (basin) bak (mouth)	Yinru
	Low-sounding groups (The 8th tone)	short vowel	The terminal sounds: b, d, g	21	Haeb (bite) moed (ant) baeg (tired)	Yangru
		long vowel	Ditto	21	Hab (box) mad (socks) bag (chop)	Yangru

From the table above we can identify when a Kam-Tai word has been borrowed from Chinese. Discerning this depends on recognizing whether or not the tone of the Kam-Tai word corresponds to the tone of a similar Chinese word. If there is no relationship between the tones of the two words, the word

has not been borrowed from the Chinese. This is the case even if the words contain the same or similar initials and *Yunmu* (simple or compound vowels in Kam-Tai, sometimes with a terminal *n* or *ŋ*). There is no correspondence between the tone of *ha*³ ‘five’ of Tai and that of **ŋu* ‘five’ of ancient Chinese because the tone of Chinese **ŋu* ‘five’ belongs to the voiced falling-rising one and ought to correspond to the fourth tone of Kam-Tai, but *ha*³ ‘five’ of Tai belongs to the third tone in the Kam-Tai tone system, which is inconsistent with that of Chinese **ŋu* ‘five’.

Therefore, *ha*³ and *ŋa*³ of the Kam-Tai language were not borrowed from the Chinese word for ‘five’. However the tone *ha*³ and *ŋa*³ of Kam-Tai were not borrowed from Chinese ‘five’. But the tone of another word meaning ‘five’—*ŋu*⁴ or *ŋo*⁴ corresponds to the Yangshang-tone of Chinese ‘five’ indicating it was borrowed from Chinese ‘five’.

NOTE ON TONE VALUES

1. Ping: level tone, one of the four tones in classical Chinese, which has evolved into the high and level tone and the rising tone in modern standard Chinese pronunciation
2. Shang: falling-rising tone, one of the four tones in classical Chinese and the third tone in modern standard Chinese pronunciation.
3. Qu: falling tone, one of the four tones in classical Chinese and the fourth tone in modern standard Chinese pronunciation.
4. Ru: entering tone, one of the four tones in classical Chinese pronunciation, still retained in certain dialects.
5. Yin: the tone of the word with voiceless initial.
6. Yang: the tone of the word with voiced initial.

REFERENCES

- Ding Bangxin. 2005. On Five and Six in Sino-Tibetan Languages. *Minzu Yuwen* 3: 1-6
- Wu Anqi. 2006. Numerals in Sino-Tibetan Languages. *Minzu Yuwen* 2: 3-13
- Somsong Burusphat and Qin Xiaohang. 2006. *Northern-Zhuang-Chinese-Thai -English- Dictionary*. Thailand: Ekphimthai Ltd.
- ILCRD and IKS. 1996. Languages and Cultures of the Kam-Tai Group: A Word List. Thailand: Sahadhammika Co., Ltd.
- Ni Dabai. 1990. *an introduction to Kam-Tai languages*. Beijing: Cun press.
- Zhang Yuansheng, Ma Jialin, Wen Mingying & Wei Xinglang. 1985. *Be Language in Hainan*. Nanning: Guangxi Minzu Press.
- Wang Li. 2004. *History of Chinese Language*. Beijing: Zhonghua Shuju (press).

Received: October 4th 2010

Department of Linguistics
Central University for Nationalities
Beijing
PR CHINA
E-mail address: qinxiaohang@2911.net