JOURNAL OF THE SOUTHEAST ASIAN LINGUISTICS SOCIETY

Volume 6
2013

JSEALS

Journal of the Southeast Asian Linguistics Society

Volume 6, 2013

JSEALS is the peer-reviewed journal of the Southeast Asian Linguistics Society, and is devoted to publishing research on the languages of mainland and insular Southeast Asia.

JSEALS was formally established by decision of the SEALS 17 meeting, held at the University of Maryland in September 2007. It supersedes the Conference Proceedings, previously published by Arizona State University and later by Pacific Linguistics.

JSEALS welcomes articles that are topical, focused on linguistic (as opposed to cultural or anthropological) issues, and which further the lively debate that characterizes the annual SEALS conferences. Data papers, book reviews, and letters are welcome submissions, subject only to internal review.

Publication is continuous online, and consolidated annually in December. Papers should be submitted electronically, either to Editor-in-Chief Paul Sidwell (paulsidwell@yahoo.com) or to Managing Editor Peter Jenks (jenks@berkeley.edu).

Submission is open to all, although we expect that most JSEALS articles have been formally submitted for peer review and publication after having the opportunity to be presented and discussed at the SEALS conference. Note that papers are expected to be written in English.

As a service to the community, non-reviewed conference presentations are archived on our website (www.jseals.org). The SEALS On-Line Archives holds the texts of most of the yearly conference publications.

Asia-Pacific Linguistics: Open Access
THE AUSTRALIAN NATIONAL UNIVERSITY
ISSN 1836-6821

Contents

Volume 6, 2013

DELANCEY, Scott. 2013. *The History of Postverbal Agreement in Kuki-Chin.*
Journal of the Southeast Asian Linguistics Society (JSEALS) 5:1-17
Received 8/10/2012, text accepted 21/1/2013, published February 2013
ISSN: 1836-6821 | Website: http://jseals.org | Managing Editor Dr Paul Sidwell
URL: http://hdl.handle.net/1885/9718

www.jseals.org | *Volume 6* | *2012* | *Asia-Pacific Linguistics*

THE HISTORY OF POSTVERBAL AGREEMENT IN KUKI-CHIN

Scott DeLancey
University of Oregon
<delancey@uoregon.edu>

Abstract
In the Kuki-Chin branch of Tibeto-Burman we find both a widespread prefixal verb agreement paradigm and, in many languages, a distinct, competing postverbal agreement system. It is clear, and generally acknowledged, that the prefixal system is a KC innovation, while the postverbal system traces back to Proto-Tibeto-Burman. This paper assembles the evidence for the postverbal paradigm from the conservative Northern Chin, Old Kuki, and Southern Chin subbranches, and makes some suggestions toward a preliminary reconstruction of the paradigm in Proto-Kuki-Chin. The older paradigm has been lost in the Central Chin (e.g. Mizo) and Mara languages, but the older 2nd person index has been incorporated into the modern paradigms.
Key words: Tibeto-Burman, Kuki-Chin, verb agreement.
ISO 639-3 language codes: bap, cnh, csh, ctd, dao, hra, jya, kac, lus, mrh, mwq, nkd, puz, rab, suz, tro.

1. The Kuki-Chin branch

Kuki-Chin (or Mizo-Kuki-Chin) is a close-knit branch of Tibeto-Burman; the languages are spoken primarily in Chin State in Myanmar, Mizoram and Manipur in India, and the Chittagong Hills area of Bangladesh. KC is generally classified with the much more disparate "Naga" languages, but these two groups have little in common in terms of morphological structure beyond what is common to the broad Tibeto-Burman stereotype. Kuki-Chin is characterized by certain morphosyntactic innovations, in particular the morphologized stem alternation and the innovative prefixal agreement paradigm, in which the possessive proclitics are used with finite verbs as argument indices. The more conservative branches, Northern Chin, Old Kuki, and Southern Chin, also preserve an older postverbal conjugation, a few elements of which have been integrated into the newer paradigm in the languages of the Central and Maraic branches, where the older paradigm has otherwise disappeared.

1.1 The modern Kuki-Chin verb

Two morphosyntactic innovations in the finite verb starkly distinguish the Kuki-Chin languages from the rest of the family. The first, which at least for now has no evident relevance to the main purpose of this paper, is that the KC verb has two phonologically distinct stems, referred to as Stem I and II, whose distribution is morphosyntactically determined. The other distinguishing feature of KC morphosyntax, which is closely related to our central question, is the innovative preverbal agreement paradigm in which the pronominal possessives are prefixed or proclitic to the finite verb, agreeing with the subject and sometimes a 1st or 2nd person object (Tarao (Old Kuki) examples from C. Singh 2002):

Table 1: Proclitic possessives and subject indices in Tarao

	Possessive	Subject
1SG	*ki-pa*	*ki-sak*
	1-father	1-eat
	'my father'	'I eat'

2SG	*ni-pa*	*ni-sak*
	2-father	2-eat
	'your.sg father'	'you.sg. eat'
3SG	*ə-pa*	*ə-sak*
	3-father	3-eat
	'his/her father'	's/he eats'

I have argued (DeLancey 2010, 2011b) that this construction arose as a nominalized clause construction, which is why the finite verb governs possessive forms.

The forms used in Tarao are typical, except for the vowel; in most KC languages we find /a/ or /ə/ rather than /i/ in the 1st and 2nd person forms. But a few languages, show some variation:

Table 2: Agreement proclitics in Kuki-Chin languages

	Tarao	Mara	Mizo
1	*ki*	*ei*	*ka*
2	*ni*	*na*	*i*
3	*ə*	*a*	*a*

The heterogeneity of the prefixal paradigms is a comparative puzzle, but is explainable once we understand the history of the system. The anomalous Mara 1st and Mizo 2nd person prefixes both appear to reflect the PTB 1PL form **i*. As we will see, a version of the archaic finite agreement paradigm was retained in Proto-Kuki-Chin and beyond. Thus the nascent prefixal system coexisted for some time with a preexisting finite system. The use of a nominalized clause construction as a stylistic alternative to the more pedestrian finite construction is a very common phenomenon in Tibeto-Burman (Noonan 1997, Bickel 1999, DeLancey 2011b, Genetti 2011), and that was the situation in PKC. Thus any alternate means of pronominal reference could be used in the nominalized construction. The prefixal paradigm only finally grammaticalized into a fixed paradigmatic system with the decay of the older finite system. The differential preservation of the archaic system in the different KC branches implies that the preverbal agreement construction was still a nominalized construction in PKC, and its shift to a fully-finite system occurred independently in the various branches. Then it is not difficult to imagine how different forms could have finally grammaticalized in different branches and languages.

The story of the odd 1st and 2nd forms in Mara and Mizo, then, is that very far back – well before PKC – the language used the INC form for some kind of polite 2nd person reference. It is not that Mizo stopped using *na-* and started using *i-*, rather that both were in use, and the form which was not marked as polite eventually fell out of use in favor of the more polite one. The Mara data suggest that the Inclusive form also had a use where *i-* had 1SG reference. Both extensions of 1PL marking are very common across languages, and can easily co-occur in the same language. (In English both can be used in the same discourse by the same speaker – e.g. a nurse visiting a hospital patient: *How are we* (=2SG) *feeling this morning? How about we* (=1SG) *open the curtains and let some light in?*).

1.2 Classification

The KC languages fall into five groups: Northern Chin (Tedim, Paite, Sizang, Zahao, etc.), Old Kuki[1] (Aimol, Anal, Tarao, Koireng, etc.), Central Chin (Mizo or "Lushai", Lai, Bawm), Maraic (Mara or "Lakher"), and Southern Chin (Daai, Hyow, etc.). Scholars differ on the higher-order organization of these groups. In examining the postverbal agreement systems, we will see a clear distinction between the conservative Northern, Old Kuki, and Southern groups and the more innovative Central and Maraic languages. This is consistent with Peterson's (2000) suggested grouping of Northern and Southern Chin in a

[1] The Old Kuki languages are not well documented, and the exact constituency of this branch is not clear (see Mortensen 2011).

"Peripheral" branch of the family, also endorsed by VanBik (2009). Both Peterson and VanBik provide phonological evidence for this grouping, but the morphological correspondences among these and the Old Kuki languages which we will examine in this paper are shared retentions, not shared innovations, and thus do not constitute strong evidence for a Peripheral clade. The position of the Old Kuki languages is likewise unsettled: VanBik places this group close to Northern Chin and Peterson to Central Chin, while Bradley (1997) suggests a primary split between Old Kuki and the rest of KC. The data which we will examine here are most consistent with VanBik's classification, but, again, the close NC-OK correlations which we will see are shared retentions, and thus not necessarily probative of close genetic relationship.

In this paper I will treat each of the five groups separately. I will deal with Northern Chin, Old Kuki, and Southern Chin under the rubric "Conservative" languages, and the Central and Maraic languages as "Innovative". As we will see, morphological correspondences between Central and Maraic support the traditional classification in which these two groups form a genetic unit. With respect to the conservative branches, the forms and constructions which they share are shared retentions rather than innovations, and thus do not provide strong support for one higher-order classification scheme rather than another.

2. The postverbal paradigm in the conservative branches

The postverbal paradigm is best preserved in the Northern Chin, Old Kuki, and Southern Chin groups. Postverbal agreement paradigms in the Kuki-Chin branch were first reported by Henderson (1957, 1965) and Stern (1963), both of whom noted the significance of their data for the question of the provenance and antiquity of verb agreement, or "pronominalization", in Tibeto-Burman. Recently Peterson (2000, 2003a) has reported a clearly cognate paradigm from Southern Chin. In even more recent data from Old Kuki languages we see the most conservative version of the original paradigm yet attested.

2.1 Northern Chin

The argument for the PTB provenance of verb agreement begins with Northern Chin:

> As Konow has indicated … Kuki-Chin … share[s] the feature of postposed personal particles with Hodgson's "pronominalized" group within the Tibeto-Himalayan branch of Tibeto-Burman. The same feature is also to be found in such Burmese languages as Nung … while Gordon H. Luce … notes its presence in Trung, in Northwest Yünnan. The distribution of this feature should be carefully reconsidered and compared with other morphological and syntactic features. In this connection, it seems to me significant that while the **/khi˙/**–series in Sizang evinces phonetic resemblance to the personal nouns, the enclitic **/iŋ˙/**–series does not. (Stern 1963: 265)

> It appears not unlikely that improved knowledge of the Chin languages and of others equally remote geographically from the so-called pronominalized groups will bring further similarities to light. In this event linguists may be obliged to conclude that, contrary to what has often been supposed, pronominalization is after all a genuine Tibeto-Burman family trait. (Henderson 1957:327)

One could wish that in the half century since Henderson wrote our knowledge of the Chin languages might have improved more than it has, but nevertheless we do now have sufficient data to bring to light further morphological connections between KC and the well-known conservative branches, and, as I have argued elsewhere, we are indeed obliged to conclude that these features do trace back to Proto-Tibeto-Burman (DeLancey 2010, to appear).

Henderson (1965) and Stern (1963) describe a series of what I will call AGREEMENT WORDS, which index the subject but occur not as affixes but as phonologically independent words following the verb (Henderson 1965: 109):

Table 3: Tedim "general" conjugation

	Singular	Plural
1	Σ *ìŋ*	Σ *ùŋ* EXC
		Σ *hàŋ* INC
2	Σ *tɛʔ*	Σ *úʔ tɛʔ*
3	Σ	Σ *uʔ ~ úʔ*

These occur also with certain grammatical forms, in which case the 1SG morpheme combines with the other morpheme into a single syllable:

(1) *pài* *ní-ŋ*
go FUT-1SG
'I will go.'

(2) *pài* *ní* *tɛʔ*
go FUT 2SG
'You will go.'

The forms which take the postverbal agreement include various tense/aspect/modality markers, as well as a conditional morpheme which we will see again in Southern Chin (Henderson 1965: 111):

Table 4: Tedim conditional conjugation

	Singular	Plural
1	Σ *lè-**ŋ***	Σ *lè-**:ŋ***
2	Σ *lɛ **tɛʔ***	Σ *lɛ úʔ **tɛʔ***
	Σ *lɛ **cín***	Σ *lɛ úʔ **cín***
3	Σ *lɛʔ*	Σ *úʔ lɛʔ*

Henderson reports the use of *cín* as an alternant to 2nd person *tɛʔ* only in the conditional conjugation. As we will see below, 2nd person forms in both /c/ and /t/ are attested throughout KC, and the relation between them is a important problem.

2.2 Old Kuki

The postverbal paradigm is found in a number of the Old Kuki languages, always in the negative paradigm, and often other functions as well (DeLancey to appear). The phenomenon is not noted in the *Linguistic Survey of India*, but it shows up in the sample texts. The Parable of the Prodigal Son, which was used as a standard text for comparative purposes, has negative sentences with 1st, 2nd, and 3rd person subjects, and in several languages, such as Rangkhol, Hallam, and Purum, we see retentions of the PTB 1st person suffix (Grierson 1904):

Table 5: Person marking in negative forms from the *Linguistic Survey of India*

	Rāngkhōl	Hallām	Pūrūm
1st	Σ-*māü**ng***	Σ-*māi**ng***	Σ-*no**ng***
2nd	*nē*-Σ-*mā**k***	*nā*-Σ-*mā**k**-chei*	Σ-*no*
3rd	Σ-*mā**k***	Σ-*mā**k**-ngāi*	Σ-*ni-yo*

In Rangkhol and Hallam we see an alternation in the final consonant of the negative forms, with 1st person subject governing a final velar nasal while with 2nd and 3rd person the form is stop-final *māk.* In Pūrūm there is a newer negative morpheme *no*, but it also shows 1st person inflection in *-ŋ*.

2.2.1 Old Kuki paradigms

Recent work on Old Kuki languages by linguists at Manipur University has documented the phenomenon more extensively. Most authors record the postverbal forms as suffixes, rather than independent words, but the paradigm is virtually identical to what we have seen in Northern Chin. Consider the Koireng negative paradigm (C. Y. Singh 2010:113-4):

Table 6: Koireng realized negative paradigm

	Singular	Plural
1	Σ-*mək*-***iŋ***	Σ-*mək*-***uŋ***
2	Σ-*mək*-***ci***	Σ-*mək*-***ci-u***
3	Σ-*mək*-***e***	Σ-*mək*-***u***

Abstracting out the negative -*mək* we have:

Table7: Koireng agreement words

	Singular	Plural
1	*iŋ*	*uŋ*
2	*ci*	*ci-u*
3	*e*	*u*

This differs from the Northern Chin paradigm in the presence of an overt 3rd person form, the reversed order of person and number markers in the 2nd person plural, and the fact that we find only the palatalized *ci*, but no *t*- form, for 2nd person.

However, the *t*- form does occur in Old Kuki; it is clearly attested in Koireng (C. Y. Singh 2010) and Moyon (Kongkham 2010). In these two languages we find tense-based variation in the paradigm which permits some internal reconstruction. Both languages have a very conservative set of agreement indices in the unrealized or future negative paradigm (C. Y. Singh 2010:114-5):

Table 8: Koireng unrealized negative paradigm

	Singular	Plural
1	Σ-*no*-***ni-ŋ***	Σ-*no*-***mə-ni***
2	Σ-*no*-***ti-ni***[2]	Σ-*no*-***ti-ni-u***
3	Σ-*no*-***ni***	Σ-*no*-***ni-u***

The -*no* suffix is the negative morpheme, as in Purum. The other element present in all the forms of the unrealized paradigm, -*ni*, is identical to the copula *ni*, which is widespread in KC. In Koireng this conjugates regularly with the prefixal paradigm when it functions as a copula:

Table 9: Modern conjugation of Koireng independent copula *ni*

	Singular	Plural
1	*ki-ni*	*kin-ni*
2	*ni-ni*	*nin-ni*
3	*ə-ni*	*ən-ni*

But buried in the negative future paradigm we find a frozen version of an ancient agreement system quite different from the innovative proclitic paradigm:

[2] The *Koireng Grammar* has a misprint in example 24, p. 114, -*niti* should be -*tini*. The correct form is given in the text above on p. 114.

Table 10: Old conjugation of Koireng copula *ni*

	Singular	Plural
1	*ni-**ŋ***	***mə**-ni*
2	***ti**-ni*	***ti**-ni-**u***
3	*ni*	*ni-**u***

The personal indices 1st *-ŋ* and 2nd *tV-* are recognizable from the verbal systems in other TB branches (DeLancey 2010, 2011a, b). The *mə-* plural element is found also in Kiranti and Jinghpaw (DeLancey 2011c). So each of these, and thus the paradigm as a whole, represent inheritance from PTB into PKC. The plural *u* is found throughout KC, but has no evident corresponding forms outside the branch; it appears to be a KC innovation.[3]

2.2.2 Internal reconstruction in Old Kuki

Thus there are two somewhat different paradigms in Koireng and Moyon, as we see in the non-future and future (or realized and unrealized) negative paradigms. The two are evidently related, but correlating them requires some internal reconstruction. If we compare the forms in the realized negative paradigm (Table 6) with those in the unrealized negative paradigm based on *-ni-*, we see that, except for the 1PL, they appear to be formed by the same affixes added to a stem consisting of a front vowel:

Table 11: Koireng conjugations compared

	Singular	Plural
1	*ni-ŋ*	*mə-ni*
	i-ŋ	*u-ŋ*
2	*ti-ni*	*ti-ni-u*
	ci-	*ci-X-u*
3	*ni*	*ni-u*
	e	*-u*

On the basis of the Koireng data alone, it seems that the palatalization of the 2nd person form was conditioned by an organic following [i], but not by the harmonic [i] in *-ti-ni* (Kongkham reports *tə-* in Moyon). But the story of palatalization is more complex than this, as we will see (Section 5.2). I provisionally reconstruct the paradigm as follows:

Table 12: Reconstructed Koireng agreement words

	Singular	Plural
1	*i-ŋ*	*u-ŋ* < *i-u-ŋ*
2	*ci* < *t-i*	*ci-u* < *t-i-u*
3	*e*	*u* < *e/i-u*

The *i/e* element must, like *-ni*, be an old auxiliary, probably also a copula. Indeed, though they apparently no longer have that function in Koireng, in some Old Kuki languages the agreement words function as equational copulas, as in Tarao (Lakshmi 1992, C. Y. Singh 2002):

[3] Cp. Henderson on Tedim:

> The use of a pronominal suffix almost always precludes the use of a pronominal prefix … An exception is the pluralizing suffix 'uh', which must therefore be classed apart from the other pronominal suffixes. (1965: 109)

(3) *əma* *dəktər*
s/he doctor
'He is a doctor.'

4) *kəy* *dəktə* *əŋ*
I doctor 1SG
'I am a doctor.'

(5) *nəŋ* *dəktər* *ce*
you.sg doctor 2
'You sg. are a doctor.'

2.3 The nature and origin of agreement words

The Northern Chin and Old Kuki agreement word system is typologically odd, in that inflection for person and number does not occur directly on the verb, but either as in independent syllable, or combined with a tense/aspect or other verbal operator to form a distinct syllable. These phenomena are difficult to describe in existing terminology. Dai and Diehl (2003) use the label "sentence-final word" (a translation of Chinese 句尾词) for the equivalent phenomenon in Jinghpaw (see DeLancey 2011c), but this is not sufficiently specific. They are reminiscent of the "tense-marked pronouns" described by Anderson (2006:289-301), but since in the KC languages the agreement morphemes in these forms are not pronouns, the term is not appropriate. Arden adopts Dryer's "pronominal word" for the phenomenon in Mara. There it works, as we will see below, since the agreement morphemes are not bound to anything else, but this will not do for Northern Chin or Old Kuki, where the syllable often carries other information besides person indexation. For the time being I will call them agreement words. The same or very similar phenomena are found in Jinghpaw-Konyak (Dai and Xu 1992, Dai and Diehl 2003, DeLancey 2010, 2011c), and Meyor-Zakhring (Jaquesson 2001, Li and Jiang 2001, Landi 2005).

The obvious origin for this construction is in old conjugated auxiliaries. We have already seen that the simple agreement words are old inflected copulas. The other morphemes which combine with the agreement indices must have the same origin (cf. DeLancey 2011c). The most widely attested of these is a future/unrealized *n(i)-*, with likely cognates in Jinghpaw and elsewhere:

Table 13: 1st singular and plural forms of Koireng and Jinghpaw auxiliary *ni*

	Singular	Plural
Koireng	*ni-ŋ*	*mə-ni*
Jinghpaw	*niʔ ai*	*mə-niʔ ai*

The development of the Jinghpaw conjugation is outside of our topic here (see DeLancey 2010, 2011c), but the alternation in the 1st singular and plural forms confirms the pre-PKC provenience of one more element of the Koireng-Moyon relict paradigm. (Jinghpaw *niʔ* < **nik*, a paradigmatic alternant of *niŋ* (DeLancey 2011c)).

The two person indices in these paradigms, 1st *-ŋ* and 2nd *#tV-*, are elements of the Proto-Tibeto-Burman agreement paradigm (DeLancey 2010, 2011a, Jacques 2012); we find them in essentially this form in the rGyalrongic (J. Sun 2003, Jacques 2004) and Southern Kiranti (Ebert 2003, Rai 1985) languages, as in Bantawa (Rai 1985: 96):

(6) *ims-a-ŋ*
sleep-PST-1SG
'I slept'

(7) *tɯ-ims-a*
2-go-PST
'you.sg slept'

Comparing the 1sg and 2sg forms from the relict Old Kuki paradigm with the corresponding forms in rGyalrong and Southern Kiranti establishes the pre-PKC provenance of the conjugation:

Table 14: 1st and 2nd person forms in Kiranti, rGyalrongic, and Kuki-Chin

	1sg	2sg
Camling (Kiranti)	Σ-*uŋa*	*ta*-Σ
Bantawa (Kiranti)	Σ-*ŋ(a)*	*tɯ*-Σ
Caodeng (rGyalrongic)	Σ-*aŋ*	*tə*-Σ
lCogtse (rGyalrongic)	Σ-*ŋ*	*tə*-Σ
Koireng (Old Kuki)	AUX-*ŋ*	*tV*-AUX

The origin of agreement words is the same as the origin of new verbal suffixes in other branches, except that in KC the grammaticalized AUX+AGR bundle remains phonologically separate rather than agglutinating to the verb stem. Compare the innovative nonpast conjugation in the Kiranti language Sunwar (DeLancey 1992) with the future tense agreement word construction in Tedim:

Sunwar
(8) *pîi-n-uŋ*
come-NONPAST-1SG
'I am coming'

Tedim
(9) *pài n-íŋ*
go FUT-1SG
'I will go'

Sunwar nonpast *n-* is a grammaticalization of the inflected copula *nɔ*, as the Tedim future *n-* is a grammaticalization of the inflected copula *ni*. These look very much like independent grammaticalizations of the same source construction; the difference is that in Sunwar the erstwhile auxiliary has become phonologically bound to the stem, while in Tedim it has not.

2.4 Southern Chin

Peterson (2000, 2003a) describes a postverbal agreement paradigm in Southern Chin (see also Jordan 1969) and notes its close correspondence to that of Northern Chin. For Hyow (Cho, Sho), Peterson gives the following negative paradigm (Peterson 2003a):

Table 15: Negative paradigm in Hyow Chin

	Singular	Dual	Plural
1	Σ-*ŋa*	Σ-*hniʔ-ŋa*	Σ-*ʔu-ŋa*
2	Σ-*ti*	Σ-*hniʔ-ti*	Σ-*cu*
3	Σ-*aʔ*	Σ-*huʔy*	Σ-*ʔu*

The 1st and 2nd singular forms look very much like the reconstructed PTB affixes, but, on paradigmatic grounds, they cannot be direct reflexes of them. 2nd person #t*V*- was a prefix (DeLancey 2011a, Jacques 2012), while 1st person *-ŋ(a)* was a suffix, as we see in the conservative Koireng paradigm. So the only way to get both into the same paradigmatic slot is as paradigmatic forms of a postverbal auxiliary, as we have

seen in Northern Chin and Old Kuki. One peculiarity of this paradigm is that the 2nd person form retains the alveolar in the singular, but the palatalized *c-* in the plural; we will return to this problem in Section 5.2.

These forms also occur, without the negative morpheme, in different-subject non-final clauses (Jordan 1969:48-51). Final or independent finite clauses always index subject with the proclitic paradigm. In same-subject clause chains, a non-final clause may include a number index following the verb:

(10)	*ngaw*	*u*	*neh*	*kah*	*mi*	*bi*	*ci*
	sit	PL	NF.SIMUL	1	PL	work	FINAL
	'(We) sitting, we do (our) work'						

Plural *u* and dual *ni* are used in all three persons, but in same-subject non-final clauses there is no indexation for person (which is recoverable from the proclitic inflection of the final verb).

In the non-final clause of a different-subject clause chain, the lexical verb is followed by number and person agreement, prefixed to the NonFinal marker *tah*:

(11)	*chü*	*u*	*ah-tah,*	*law*	*ci*
	call	PL	1-NF	come	FIN
	'We having called, he came.'				

The paradigm is (Jordan 1969):

Table 16: Mindat Chin different-subject non-final agreement

	singular	dual	plural
1st	*(ng)ah-tah*	*ni (ng)ah-tah*	*u (ng)ah-tah*
2nd	*ci-tah*	*ni ci-tah*	*u ci-tah*
3rd	*khü-tah*	*ni khü-tah*	*u khü-tah*

The variation in the 1st person forms is dialectal; Jordan reports *ngah-*, *nah-*, and *ah-* dialects. In the 1st and 2nd person this is identical to the Hyow paradigm, except that where Hyow has the *t-* 2nd person form in the singular and *c-* in the plural, Mindat has the palatalized form throughout. Note also that the order of person and number in the 2PL form is reversed in the two languages, just as we have seen in the comparison of Northern Chin and Old Kuki: Mindat *u ci*, Hyow-*cu* < *-ci-ʔu*.

In Daai Chin (So-Hartmann 2009) we find the 1st person form in negated main clauses (So-Hartmann 2009:244-52):

(12)	*am*	*sit*	*be*	*khoh=ngü*
	NEG	go	back	able=1
	'I am/was not able to go back.' (So-Hartmann 2009:252)			

(13)	*am*	*pye:n*	*vaai*	*xa=ngü*
	NEG	speak	go	definitely-1
	'I definitely will not go and speak.'			

In contemporary Daai, this occurs only in negative sentences, but in an older form of the language preserved in oral texts, it can occur in some assertive clauses as well:

(14)	*tuh=ngooi:*	*ta*	*ni:ng*	*hmu*	*lo*	*veeng(=ngü)*
	today	FOCUS	2SG.OBJ	see	ASPECT	certainly(=1)
	'Today I will certainly see you.'					

As in Hyow and Mindat, this can be preceded by number indices:

(15) *am* *hmuh=ni=ngü*
NEG see=DUAL=1
'We2 did not see [him].'

The postverbal 2nd person form appears to have disappeared.

2.5 The postverbal paradigm in the conservative languages

The simple agreement words (i.e. not attached to other grammatical marking) in the three conservative branches of KC are summarized below (see DeLancey to appear for additional data supporting the reconstructed NC and OK forms):

Table 17: Postverbal indices in the conservative languages

	1SG	1PL	2SG	2PL	3SG	3PL
NC	**i(-)ŋ*	**u-ŋ*	**teʔ*	**u teʔ*	*(*a)*	**u*
OK	**i(-)ŋ*	**u-ŋ*	**ce*	**ce-u*	*(*ə)*	**u ~ *əi*
SC	*-ŋa* - *ʔu-ŋa*		*-ti*	*-cu/-u-ci*	*-aʔ*	*-ʔu*

At present I have no explanation for the difference in syllable shape of the 1st person form between NC/OK on the one hand and SC on the other. The alternation in the order of number and person in the 2nd person forms between NC and OK, and within SC, suggests that the ordering of these two morphemes was not fixed in PKC. Unlike all the rest of the morphological material in this table, the plural **-(ʔ)u* has no evident cognates outside of KC, and can thus be identified as a KC innovation (pace an erroneous suggestion in DeLancey to appear). It occurs in several other constructions in various KC languages.

Aside from these differences, and the problematic variation between 2nd person forms in *t-* and *c-*, the paradigms are identical, and must reflect common inheritance from PKC.

3. Remnants of the postverbal paradigm in Central and Maraic languages

Except for a striking archaic retention in the hortative paradigm in Hakha Lai, the Central and Maraic languages for which we have descriptions have not preserved the full postverbal paradigm. There is an apparent remnant 1st person form in one Mara (Lakher) paradigm, but it tells us nothing that we do not already know from the peripheral branches. But in both Mizo and Mara we find retentions of the 2nd person form in paradigms which do help to expand our understanding of the PKC situation and how the subbranches have developed from there.

3.1 Hakha Lai

Hakha Lai preserves particularly ancient forms in the "cohortative" ('Let's!') paradigm (Peterson 2003: 414-5); compare the corresponding 1→2 request forms in Trung (Dulong, Sun 1982: 108-10):

Table 18: Hakha Lai and Trung hortative paradigms

	singular	dual	plural
Hakha cohortative	*-niŋ*	*-ʔu-siʔ*	*-hnaa-ʔu-siʔ*
Trung 1→2 request	*-níŋ*	*-ɕìn*	*-nɯ*

This is the only evidence I have found for the survival into Kuki-Chin of the PTB dual **si* (LaPolla 2003:30). This does not bear directly on our main project, except insofar as it demonstrates the preservation of much archaic morphological material into PKC, at least in relict paradigms such as this one.

3.2 Mizo

In Mizo (Central Chin) the secondary prefixal paradigm has completely won out, and is the only verbal paradigm. The archaic 1st person marking has completely disappeared, but several reflexes of the old 2nd person form are found in the transitive paradigm, certain imperatives, and certain other relict constructions.

Chhangte (1993) and Peterson (2000) have noted the evident connection between the Mizo familiar imperative particle *teʔ* and the Northern Chin 2nd person form:

(16) *thû-**teʔ***
sit-**IMPERATIVE**
'Sit!'

But this is only the beginning of the story. This form, which is only one of several different Mizo imperative constructions, also occurs in a "less formal hortative":

(17) *i-kâl-**teʔ**-áŋ*
DU-go-**IMPERATIVE**-FUT
'Let's go (it's time)!'

There is also a weak imperative *-ta=cèè*, which is similar in form to the *c-* 2nd person forms which we have seen in Old Kuki and Southern Chin.

Even more interesting is another 2nd person particle, *cê*, which matches the Old Kuki 2nd person index in form, and both the OK and the Northern Chin in syntagmatic position. This occurs productively as an index of 2nd person object, which has been incorporated into the innovative, otherwise prefixal transitive paradigm (Chhangte 1993: 91-2):

Table 19: Agreement indices with singular arguments in Mizo

	Subject		
Object \	1	2	3
1		*ka*-Σ *cê*	*ka*-Σ
2	*mi*-Σ		*i*-Σ
3	*mi*-Σ	*a*-Σ *cê*	*a*-Σ

This form has no connection with the possessive proclitic system which forms the core of the prefixal paradigm, so it must have a different origin. After we have seen the development of 2nd person *tɛʔ* to *-ce* in Old Kuki, the source of the Mizo suffix is obvious. From its syntagmatic behavior we can see its origin in the Northern Chin pattern. The *cê* morpheme, as in Northern Chin, is a distinct word, as it does not trigger vowel reduction in a preceding stem. It follows all verbal suffixes, consistent with its origin as an inflected copular auxiliary. Thus it remains outside the phonological scope of continuative reduplication:

(18) *mî-sik* *mî-sik*
1OBJ-pinch 1OBJ-pinch
'[S/he] pinches me again and again.'

(19) *â-sik* *â-sik cê* *mòò*
3sg-pinch 3sg-pinch-2OBJ INTERROGATIVE
'Does [s/he] pinch you again and again?'

The order of person and number marking in Mizo is *cê û* (Chhangte 1993:90, Bedell 2004:53), just as in Old Kuki.

The 2nd person *cê* also occurs as a subject index in one formal, archaic request construction, when the 1st person object prefix is present:

(20) *mín-chââg* *ag* *cê*
1OBJ-answer HORT 2
'you [please] answer me'

(21) *mín-rhê-reŋ-áŋ* *cê*
1OBJ-know-always-FUT 2
'Please remember me!'

3.3 Mara

Our data come from several different reports on Mara, in earlier literature often called "Lakher" (Savidge 1908, Lorrain 1951, Weidert 1985, Arden 2010). Mara shares the innovative prefixal agreement paradigm characteristic of Kuki-Chin, except that its 1st person index is *ei* where other branches have *ka*. The agreement indices show considerable phonological independence from the verb; Arden (2010) is reluctant even to call them clitics.

In Mara, as in Mizo, the 2nd person agreement word is preserved in several constructions. Unlike Mizo, Mara also retains the 1st person form, but only marginally. (And perhaps no longer, since this construction is attested in Lorrain and Savidge's work of a century ago, but not reported in more recent work). There is a reflex of the 1st person agreement word **iŋ*, in a subjunctive conjugation recorded by Lorrain (1951: 18):

Table 20: Mara subjunctive conjugation

	Singular	Plural
1	*shi aw sha-la*	*shi i sha-la*
2	*shi la*	*shi u-la*
3	*shi sha-la*	*shi sha-la*

The *aw* which occurs uniquely in the 1sg construction, and alternates with 1PL *i*, is a regular reflex of PKC **iŋ* (VanBik 2009: 371). The *-la* element can be compared with the conditional *le-* which we have seen in other branches, where it also requires the postverbal conjugation, but it is not clear why it follows the agreement words in Mara, but precedes them in Northern Chin. (I also have no explanation for the *sha* which occurs in the 1st and 3rd but not 2nd person forms). Aside from this there is no apparent trace of the 1st person form in the available data on Mara.

The 2nd person agreement word occurs in three different forms in Mara. Like Mizo, Mara has *te* as an imperative marker. As in Mizo, the more interesting pattern is the distribution of the palatalized form in the transitive paradigm. It does not occur in the intransitive conjugation, but in the transitive paradigm it occurs preverbally as a 2OBJ index *cha*, and postverbally in the 2→1 form as *chi* (after Arden 2010; cp. Weidert 1985:929):

Table 21: Mara person agreement

	Object			
Subject \	1	2	3	Intr
1		*ei cha* Σ	*ei* Σ	*ei* Σ
2	*ei na* Σ *chi*	*na cha* Σ	*na* Σ	*na* Σ
3	*e na* Σ	*a cha* Σ	*a* Σ	*a* Σ

The 2→1SG configuration is the only construction in Mara in which a personal index (as opposed to a plural morpheme) follows the verb (Arden 2010).

There are many interesting aspects to this paradigm. The hierarchical distribution of 1st person *ei* and the apparent anomalous spread of 2nd person *na* into the 3→1 form are of great interest, but not for our present pursuit (see DeLancey 1989, 2010). Our concern is with the two other 2nd person indices, 2OBJ *cha* and 2SUBJ *chi*. Together, these two morphemes have the same distribution as Mizo postverbal 2nd person *cê*, which occurs everywhere that Mara has *cha-*, and in archaic forms also in the 2→1 form where Mara has *chi*.

The *cha-* is the only part of the prefixal paradigm which does not have an evident synchronic origin. The others are all possessive proclitics, reflecting the origin of the innovative prefixal paradigm in a clausal nominalization (DeLancey 2010, and below Sec. 4.1). The simplest hypothesis is that the 2nd person postverbal particle *#ce* was shifted to preverbal position as the proclitic paradigm began to crystallize. (The KC proclitics are unstressed and subject to phonological reduction, hence the reduction of the vowel in the preverbal but not the postverbal form is to be expected).

It cannot be that this is a retention of the *#tV-* prefix from PTB. There is no other evidence of *#tV-* prefixation on the main verb in PKC; as in Mara, so throughout the branch agreement proclitics and prefixes have evident and relatively shallow origins. Mara 2OBJ *cha* seems to be the sole exception,[4] and this points to its exceptional origin. The other preverbal indices originate as possessives, *cha* as an originally postverbal particle secondarily incorporated into the preverbal pronominal word construction in the new paradigm.

3.3 Reorganization of the paradigm in Mizo and Mara

The shared pattern of retention of the 2nd person agreement word in Mara and Mizo reflects a period of common development after their separation from PKC; these two groups are related to one another as a Central-Maraic subbranch within KC. If the 1SG subjunctive *aw* in Mara is in fact **iŋ*, then the 1SG agreement word also survived into the common ancestor of Mara and Mizo, but from its complete disappearance in Mizo and very marginal existence in Mara we can infer that already by that stage it had little remaining function, presumably because the prefixal conjugation had largely taken over the finite domain.

Proto-Central-Maraic had lost the postverbal paradigm as such, but had incorporated the 2nd person agreement word into the new finite verb, where it occurred in all SAP object forms with a 2nd person argument, i.e. 1→2, 3→2, 2→1:

Mara

(22) eina ly chi
1SG.OBJ thank 2
'You sg. thank me.' (Arden 2010:115)

Mizo

(23) *mín-rhê-reŋ-áŋ* *cê*
1OBJ-know-always-FUT 2
'Please remember me!' (Chhangte 1993: 92)

In Mizo, it is retained in its original postverbal position in all functions:

Mizo

(24) *kâ-ron-tlhààk-zéél-áŋ* *cê*
1-come-drop-keep.on-fut 2OBJ
'I will keep on dropping them to you' (Chhangte 1993: 180)

But Mara separated the 2SU and 2OBJ functions by moving the agreement word to a preverbal position consistent with both the preverbal position of the new agreement indices and the SOV constituent order of the clause:

[4] In DeLancey 1989 I suggested that Mara 1st person *ei* was another. As noted above, I now interpret it as an extension of 1pl **i*, which is well attested in KC.

Mara

(25) *ei* *cha* *ly*
1 2OBJ thank
'I thank you sg.' (Arden 2010: 114)

Both languages retain the 2nd person agreement word as an imperative form. This is not reported in the available data on the conservative branches, so it too appears to be a shared Central-Maraic innovation.

4. The History of Agreement in Kuki-Chin

The PTB origins of the postverbal agreement indices in KC is clear (see also DeLancey 2010, to appear). The postverbal agreement words are old auxiliaries inflected for person, in a paradigm which included the PTB affixes 1SG *-ŋ*, 1PL *-i*, and 2nd *tV-*. In this section we will conclude the discussion of what we can infer about the history of this construction from PKC on. We have briefly looked at the incorporation of the 2nd person postverbal agreement word into the new paradigm in the Central-Maraic languages, but have not directly addressed the question of how and why the new paradigm arose and replaced the older one. We will consider this question in Section 5.1. The other problem which has been deferred until this point is the split of the original 2nd person form into *t-* and *c-* versions; I will offer some speculation on this question in Section 5.2.

4.1 Ancient and innovative paradigms in PKC

The preverbal agreement paradigms have shallow, transparent origins. The postverbal paradigm, in contrast, is opaque and mysterious in KC context, but immediately recognizable in a broader comparative Tibeto-Burman framework. Since the basic elements of the preverbal paradigm are shared across the branch, while the postverbal paradigm is still active in Northern Chin and Old Kuki, both must have been in productive use in PKC[5]. We must therefore reconstruct PKC with competing main clause constructions, an older finite form derived from the original PTB inflected verb, and a newer prefixal construction derived from a recent nominalized clause construction (DeLancey 2011d). This raises three questions: why were there two competing paradigms? What was the function of the innovative construction? And what led to the fading and ultimate extinction of the original paradigm?

Two characteristics of the innovative preverbal paradigm in the KC languages argue for its origin in a clausal nominalization. The first is the fact that the preverbal agreement morphemes are, for the most part, simply the possessive proclitics or prefixes pressed into service as verb agreement. The second is the syntax of the sentence final particle in the Northern languages. Such "indicative" or "affirmative" final particles are widespread among KC languages, as elsewhere in the family. I have suggested that such final particle constructions in Tibeto-Burman usually originate in the copula or nominalizer in a clausal nominalization construction (DeLancey 2010, 2011b). In Northern Chin, where the two paradigms are still in competition, the evidence is particularly clear. The final particle, *hî:*, is identical to the copula, and is obligatory with the preverbal construction, and impossible after postverbal agreement (Henderson 1965: 109-11):

(26) *pài:* ***ke****-ŋ*
go NEG-1SG
'I don't go.'

(27) *ká* *pài:* ***kei*** *hî:*
1st go NEG FINAL
'idem.'

[5] Cp. Peterson 2002: 99-100, who suggests that the prefixal paradigm was less well-developed in PKC, and that the paradigms in the various subbranches developed more independently, than I am suggesting here.

Thus the two features which suggest nominalization – possessive clitics as argument indices and a final particle which is transparently a grammaticalized copula – obligatorily occur together in one construction, and are both impossible in the other.

The nominalized paradigm must have still been formally a nominalization until after the divergence of the branches, since each language uses its own possessive proclitics, with Mizo substituting 1PL **i* for 2nd person, and Mara extending it to 1SG. If the prefixal paradigm were already completely reanalyzed as a finite verb construction by PKC, it should presumably no longer responsive to changes in the nominal paradigm. As long as the verbal paradigm remains in lockstep with the nominal, the verb is still being treated as a noun, as Konow long ago remarked (Konow 1902, Grierson 1904).

In most of KC the postverbal paradigm has mostly disappeared, as in Mizo and Mara, or been confined to specific constructions, as seems to be the case in Old Kuki and Southern Chin. But in the Northern languages Tedim and Sizang the two paradigms are still to some degree in competition. Henderson (1965) describes this as a register distinction, with the prefixal paradigm used in more formal, and the postverbal more informal or colloquial contexts. Sarangthem (2010) reports the same for Sizang.

4.2 Palatalization of #tV-

On phonological grounds the identification of the *cV* 2nd person forms with the demonstrably ancient *#tV* forms is shaky, as there is no regular phonological process which can explain the alternation. But on morphological grounds the identification is inescapable, as the two forms occur in exactly the same paradigmatic and syntagmatic slots.

In Northern Chin and Old Kuki, the OK palatal and NC alveolar forms correspond perfectly in meaning and paradigmatic and syntagmatic position. In Moyon and Koireng, they are in complementary distribution, apparently phonologically conditioned. In Hyow they are apparently arbitrarily distributed, nonpalatalized in singular and palatalized in plural. Finally, in Mizo, the two imperative forms are in functional contrast in (at least roughly) the same syntagmatic position. It is perhaps possible to imagine two different etyma becoming entwined in a network like this, but much easier to imagine an irregular phonological process producing two allomorphs which can later split.

Sources on Northern Chin are very limited, so claims about what may be absent there should be taken as provisional. In the available data, we find only *te*, and only as a postverbal agreement word. In Old Kuki we find *ce* as a postverbal agreement word. In Koireng and Moyon we also find the original #t*V*- prefix on *ni* in the negative paradigm, and thus an alternation between *ti* in the negative future and *ci* in the negative non-future. In Southern Chin we find the 2nd person agreement word as *ti* in the singular but *ci* in the plural. In Central-Maraic both forms occur as imperatives, but the palatalized version in indicative verb constructions.

The change *t* > *c* is not regular anywhere in KC,[6] so there is no purely phonological explanation for the variation. It may be that the KC 2nd person agreement word originated in a j-initial form with a prefix *tə-*, probably **tə-jik* > **tjik*, and that the palatalization was conditioned by the following glide. (This would be a unique development, since we have no other etyma with that initial cluster to show what the "regular" development would be). But since the palatalization is sporadic even with a single language, and is not consistently associated with a particular morphological position or syntactic function, the explanation cannot be simply phonological – we need to identify another factor. I suggest that the explanation is sociolinguistic, having to do with the marked character of 2nd person utterances in general. Note particularly the opposition in Mizo between the "familiar" imperative *te?* and the "weak" imperative *cèè*, where the opposition codes affective content.

References

Anderson, Gregory. 2006. *Auxiliary Verb Constructions*. Oxford and New York: Oxford University Press.

Arden, Michelle. 2010. A phonetic, phonological, and morphosyntactic analysis of the Mara language. MA thesis, San Diego State University. (accessed 22/01/2012 at http://scholarworks.sjsu.edu/etd_theses/3744).

[6] There is some alternation between /t/ and /ts/ in Tedim (VanBik 2009: 27-30), but nothing that could explain the distribution of #t*V* reflexes across the branch.

Bedell, George. 2004. "Agreement in Mizo." In: Burusphat, Somsonge (ed.). *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society*. Tempe: Program for Southeast Asian Studies, Arizona State University. Pp. 51-70.

Bickel, Balthasar. 1999. "Nominalization and focus constructions in some Kiranti languages." In: Yogendra Yadava and Warren Glover (eds.). *Topics in Nepalese Linguistics*. Kathmandu: Royal Nepal Academy. Pp. 271-296.

Bradley, David. 1997. "Tibeto-Burman languages and classification." In: Bradley, David (ed.) *Tibeto-Burman languages of the Himalayas*. [Pacific Linguistics Series A, no.86; Papers in Southeast Asian Linguistics, 14]. Canberra: Australian National University. Pp. 1-72.

Dài, Qìngxià 戴庆厦 and Xiījiān Xú 徐悉艰.1992. *Jǐngpō yǔ yǔfǎ* 景颇语语法. Beijing: Minzu Chu Ban She.

Dài, Qìngxià, and Diehl, Lon. 2003. "Jinghpo." In: Thurgood and LaPolla (eds.). Pp. 401-8.

DeLancey, Scott. 1989. "Verb agreement in Proto-Tibeto-Burman." *Bulletin of the School of Oriental and African Studies* 52: 315-33.

DeLancey, Scott. 1992. "Sunwar copulas." *Linguistics of the Tibeto-Burman Area* 15.1: 31-38.

DeLancey, Scott. 2010. "Towards a history of verb agreement in Tibeto-Burman." *Himalayan Linguistics* 9.1: 1-38.

DeLancey, Scott. 2011a. "Agreement prefixes in Tibeto-Burman." *Himalayan Linguistics* 10.1: 1-29.

DeLancey, Scott. 2011b. "Finite structures from clausal nominalization in Tibeto-Burman languages." In: Yap et. al. (eds.). Pp. 343-62.

DeLancey, Scott. 2011c. "Nocte and Jinghpaw: Morphological correspondences." In: Hyslop, Gwendolyn, S. Morey, and M. Post (eds.). *North East Indian Linguistics 3*. New Delhi: Foundation/Cambridge University Press India. Pp. 61-75.

DeLancey, Scott. to appear. "Postverbal agreement in Mizo-Kuki-Chin." In: Hyslop, Gwendolyn, S. Morey, and M. Post (eds.), *North East Indian Linguistics 5*. New Delhi: Foundation/Cambridge University Press India.

Driem, George van. 1993. "The Proto-Tibeto-Burman verbal agreement system." *Bulletin of the School of Oriental and African Studies* 56: 292-334.

Ebert, Karen. 2003. "Camling." In: Thurgood and LaPolla (eds.). Pp. 533-545.

Genetti, Carol. 2011. "Nominalization in Tibeto-Burman languages of the Himalayan area: a typological perspective." In: Yap et. al. (eds.). Pp. 163-94.

Grierson, G. A. (ed.). 1904. *Linguistic Survey of India, vol. III: Tibeto-Burman Family, part III: Specimens of the Kuki-Chin and Burma Groups*. Calcutta: Office of the Superintendant of Government Printing.

Henderson, E.J.A. 1957. "Colloquial Chin as a pronominalized language". *Bulletin of the School of Oriental and African Studies* 20: 323-7.

Henderson, E.J.A. 1965. *Tiddim Chin: A descriptive analysis of two texts*. London: Oxford University Press.

Jacques, Guillaume. 2004. *Phonologie et morphologie du japhug (Rgyalrong)*. PhD thesis, Université Paris VII - Denis Diderot.

Jacques, Guillaume. 2010a. "The inverse in Japhug Rgyalrong." *Language and Linguistics* 11.1: 127-157.

Jacques, Guillaume. 2010b. "The origin of the reflexive prefix in Rgyalrong languages." *Bulletin of the School of Oriental and African Studies* 73: 261-8.

Jacques, Guillaume. 2010c. "A possible trace of verb agreement in Tibetan." *Himalayan Linguistic*s 9.1: 41-50.

Jacques, Guillaume. 2012. "Agreement morphology: the case of Rgyalrong and Kiranti." *Language and Linguistics* 13.1: 83-116.

Jaquesson, François. 2001. "Person-marking in TB languages, Northeastern India." *Linguistics of the Tibeto-Burman Area* 24.1: 113-44.

Jordan, M. 1969. "Grammar of the Chin language." Appendix to *Chin Dictionary and Grammar*. Paris.

Kongkham, Hemabati. 2010. *A descriptive grammar of Moyon*. Imphal, India: Manipur University Ph.D. dissertation.

Konow, Sten. 1902. "Zur Kenntniss der Kuki-Chinsprachen." *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 56: 486-517.

Krishan, Shree. 1980. *Thadou: A Grammatical Sketch*. Kolkata: Anthropological Survey of India.

Landi, Victor. 2005. *The Meyors and their language*. Itanagar.

Lǐ, Dàqín 李大勤, and Dí Jiāng 江荻. 2001. "Zhá huà gàikuàng" 扎话概况. *Minzu Yuwen* 民族语文 2001.6:61-75.

Lorrain, Reginald. 1951. *Grammar and Dictionary of the Lakher or Mara Language*. Guwahati, India: Department of Historical and Antiquarian Studies, Government of Assam.

Mortensen, David, and Jennifer Keogh. 2011. "Sorbung, an undocumented language of Manipur: Its phonology and place in Tibeto-Burman." *JSEALS* 4.1: 64-114.

Noonan, Michael. 1997. "Versatile nominalizations." In: Joan Bybee, John Haiman, and Sandra Thompson (eds.), *Essays on language function and language type: In honor of T.Givón*. Amsterdam and Philadelphia: John Benjamins. Pp. 374-94.

Peterson, David. 2000. "On the status of the Southern Chin subgroup." XXXth ICSTLL, Bangkok.

Peterson, David. 2002. "On Khumi verbal pronominal morphology." *BLS* 28S: 99-110.

Peterson, David. 2003a. "Agreement and grammatical relations in Hyow." In: Bradley, David, et. al., *Language Variation: Papers on Variation and Change in the Sinosphere in honour of James A. Matisoff*. Canberra: Department of Linguistics, Australian National University. Pp.173-183

Peterson, David. 2003b. "Hakha Lai." In: Thurgood and LaPolla (eds.). Pp. 409-26.

Rai, Novel Kishore. 1985. *A Descriptive Study of Bantawa*. Ph.D. dissertation, Deccan College Post-Gradute & Research Institute.

Sarangthem, Bobita. 2010. *Sizang (Siyin) Grammar*. Ph.D. dissertation, Department of Linguistics, Manipur University, Imphal.

Savidge, Fred. 1908. *A Grammar and Dictionary of the Lakher Language*. Allahabad: Pioneer Press.

Singh, Chungkham Yashawanta. 2002. *Tarao grammar*. New Delhi: Akansha.

Singh, Chungkham Yashawanta. 2010. *Koireng grammar*. New Delhi: Akansha.

So-Hartmann, Helga. 2009. *A Descriptive Grammar of Daai Chin*. (STEDT Monograph 7). Berkeley: Sino-Tibetan Etymological Dictionary and Thesaurus Project.

Stern, Theodore. 1963. "A provisional sketch of Sizang (Siyin) Chin." *Asia Major* (n.s.) 10:222-78.

Sūn, Hóngkāi 孙宏开. 1982. *Dúlóng yǔ jiǎnzhì* 独龙语简志. Beijing: Minzu Chu Ban She.

Sūn, Hóngkāi. 1983. Wǒguó Zàng-Miǎn yǔ dòngcí de rénchēng fanchou" 我国藏缅语动词的人称范畴. *Minzu Yuwen* 民族语文1983.2: 17-29.

Sūn, Hóngkāi. 1988. "Shìlùn wǒguó jìngnèi Zàng-Miǎn yǔ de pǔxì fēnlèi 试论我国境内藏缅语的谱系分类. (A classification of Tibeto-Burman languages in China). In: Nishida, Tatsuo, and Paul Kazuhisa Eguchi (eds.). *Languages and history in East Asia*: *festschrift for Tatsuo Nishida on the occasion of his 60th birthday*. Kyoto: Shokado. Pp. 61-73.

Sūn, Hóngkāi. 1995. "A further discussion on verb agreement in Tibeto-Burman languages." Nishi et. al. (eds.). Pp. 17-29.

Sun, Jackson. 2003. "Caodeng rGyalrong." In: Thurgood and LaPolla (eds.). Pp. 490-502.

Thurgood, Graham, James Matisoff, and David Bradley (eds.). 1985. *Linguistics of the Sino-Tibetan area: The state of the art*. Canberra: Department of Linguistics, Australian National University. [Pacific Linguistics Series C, no. 87].

Thurgood, Graham, and LaPolla, Randy (eds.). 2003. *The Sino-Tibetan language*s. London and New York: Routledge.

VanBik, Kenneth 2009. *Proto-Kuki-Chin: A Reconstructed Ancestor of the Kuki-Chin Languages.* (STEDT Monograph 89). Berkeley: Sino-Tibetan Etymological Dictionary and Thesaurus Project, University of California.

Yap, Foong Ha, Karen Grunow-Hårsta, and Janick Wrona (eds.), *Nominalization in Asian Languages: Diachronic and Typological Perspectives*. Amsterdam: Benjamins.

SOMSONGE Burusphat. 2013. *The Language Shift in Progress of Thai Song.*
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:18-34
Received 30/6/2012, revised text accepted 24/9/2012, published February 2013
ISSN: 1836-6821 | Website: http://jseals.org | Managing Editor Dr Paul Sidwell
URL: http://hdl.handle.net/1885/9719

www.jseals.org | Volume 6 | 2012 | Asia-Pacific Linguistics

THE LANGUAGE SHIFT IN PROGRESS OF THAI SONG

Somsonge Burusphat
Research Institute for Languages and Cultures of Asia
Mahidol University, Thailand
<sburusphat@gmail.com >

Abstract
The research question of this paper is: To what extent has the Thai Song language been influenced by the Thai language? The finding of this question is obtained by a sociolinguistic study of phonological variation and lexical replacement among Thai Song speakers of three age groups: old generation (over 60), middle generation (35-55), and young generation (18-30). Thai Song speakers residing at two locations in Nakhon Pathom province are selected as a case study of Thai Song in the western region of Thailand. The study of phonological variation reveals that Thai Song has progressively shifted to Thai in the speech of middle and young generations. The study of lexical replacement shows that young generation speakers use fewer Thai Song words than the older generations.
Key words: language contact, historical phonology, language shift, Thai Song, Thai Song Dam, Tai Dam, Lao Song
ISO 639-3 language codes: soa, blt

1. Introduction

Previous studies have shown that in the seven provinces of the western region of Thailand, namely, Kanchanaburi, Ratchaburi, Phetchaburi, Suphanburi, Nakhon Pathom, Samut Sakhon, and Samut Songkhram, there are six major Lao ethnic groups, which are, Lao Song/Thai Song, Lao Yuan, Lao Phuan, Lao Khrang, Lao Vieng and Lao Tay. All of these Lao groups, apart from Lao Yuan, migrated from Laos more than 200 years ago. A study of the language use and attitude of these Lao ethnic groups shows that, among all such groups in these provinces, Thai Song are the most strongly united and thus have the strongest language vitality across all provinces, except for Samut Songkhram province where few Thai Song speakers reside. A survey of areas inhabited by Lao ethnic groups reveals that, in the western region of Thailand, Thai Song villages number the most at 394 villages (Burusphat et al. 2011). This number reveals that Thai Song is the dominant Lao ethnic group in this region.

Despite their strong language vitality and preservation of some cultural traits such as ritual practices and costumes, Thai Song people have assimilated well into the Thai community. The Thai Song language has been influenced by Thai as a standard language. Because of this dominant national language, it is possible that a language shift from Thai Song to Thai may take place in the future. Therefore, the purpose of this paper[1] is to explore how much the Thai Song language has been influenced by the Thai language. It is

[1] This paper is part of the sub-project "Ethnic language processes in progress" of the cluster research "Ethnicity: New paradigm in language and cultural transmission" sponsored by the Research-Team Promotion Grant 2010-2013, Thailand Research Fund (TRF). It is a revised version of the paper "The language maintenance and shift of Thai Song Dam in the western region of Thailand" presented at the 2011 Joint Conference of the Association for Asian Studies (AAS) & International Convention of Asia Scholars (ICAS) in Honolulu, Hawaii from March 31 to April 3, 2011. The author thanks Carol Compton for being a discussant of the original paper. Her comments and suggestions contribute a great deal to the revision of this paper. My special thanks go to Weera Ostapirat for sharing his expertise on Comparative Tai with me and Varisa Kamalanavin for being a commentator of this paper presented at the second annual meeting of the research project on August 18th, 2012 at Mahidol University.

hoped that the results of this study will help Thai Song community leaders become aware of the endangerment situation of Thai Song language and consider the implementation of a language revitalization program.

2. Origin and language affiliation

Thai Song people have been addressed by various names such as Thai Song, Thai Song Dam, Lao Song, Song, Tai Dam or Black Tai. The term "Tai" is distinguished from "Thai" in so far as: "Conventionally, Southeast Asianists use the term "Tai" in referring to any speakers of the Tai language family, reserving the aspirated "Thai" to designate only those citizens of the Kingdom of Thailand as a Siamese Tai state" (Sams 1988: 116). The term "Thai Song" is used in this paper.

The name "Thai Song Dam" refers to people dressed in black costumes. In the past, men wore black *Song* or trousers which became their ethnic name. Thai Song people migrated from Muang Thaeng (Myang Teng or Muoi)[2], Sipsongchutai (Sipsong Chao Tai) in the northern part of Laos. This place used to be under the Luang Prabang government (Sribusara 1987).

Because of common migration patterns and linguistic similarities between Thai Song and other Lao groups, later generations of Thai Song believed they were the same group as the Laos, and called themselves *Phu Laao* 'Lao people' (Chakshuraksha 2003). Linguistically, the Thai Song language is distinguished from the Lao language and other Lao ethnic group languages as a member of the Tai language group (Brown 1965). All languages belong to the Southwestern Branch of the Tai-Kadai language family (Li 1960).

The original settlement of Thai Song people in Thailand was in Phetchaburi province. Later on, Thai Song people moved to nearby provinces such as Kanchanaburi, Ratchaburi, Suphanburi, Nakhon Pathom, Samut Sakhon, and Samut Songkhram. In Nakhon Pathom province, there are many Thai Song communities which are strongly united, therefore this province has been chosen for the case study of this research.

In addition to their migration to Thailand, Tai Dam people also migrated from Son La (Muang La) in Vietnam to Laos. An interview with Bakam (2010), 78 years old, who lives in Vientiane, discloses that most Tai Dam people in Laos migrated from Son La more than 50 years ago.

3. Methodology

The research question of this paper is: To what extent has the Thai Song language been influenced by the Thai language? This study proposes the following hypotheses:

a) Young generation speakers mix Thai sounds with Thai Song sounds more than older generations.
b) Young generation speakers use fewer Thai Song words than older generations

The findings for the research question were obtained by means of a sociolinguistic study of phonological variation and lexical replacement among Thai Song speakers of three age-groups: old generation (over 60), middle generation (35-55), and young generation (18-30).

Variation Theory is used as a framework for this study. This theory holds that linguistic forms have variants that are alternatively used but their meanings remain the same. Variation Theory is an important part of sociolinguistics which asserts that there is no free variation. Linguistic variation is conditioned by social factors such as region, social class, educational background, or style (Snyder 1995). This study focuses on social variation. The primary concern is on variation as differences in pronunciation and word replacement. Age is chosen as an independent variable as it has been found in most Thai Song studies that age group is an important independent variable that conditions linguistic variation (Buranasing 1988, Liamprawat and Wattanaprasert 1996, and Saeng-ngam 2006). The phonological comparison is based on the work of L. Thongkum (2002).

The phonological variation of Thai Song was further compared with Tai Dam (Black Tai) dialect spoken in Muang La (L. Thongkum 2002). Supplementary Tai Dam data compiled by the researcher were

My heartfelt thanks are extended to Richard Hiam for his time and effort devoted to the English edition of this paper. I thank my research assistant, Iyaret Boonyarit for preparing figure 1 and figure 2.

[2] Muang Thaeng is presently Muang Dien Bien Phu in northwestern Vietnam.

also used. The Tai Dam informant was from Lae village, Muang La. She was female and 66 years old. She came to visit her relatives in Vientiane where the data were collected. The Tai Dam data were included in this study because some old forms which once shared by Tai Dam and Thai Song are still retained in Tai Dam while they are lost in Thai Song due to the influence of the Thai language.

This study used Fippinger and Fippinger's (1974) Tai Dam word list for a preliminary survey of phonological and lexical variation by age-group. Two hundred and twenty-two words were selected from this word list. The words were elicited in isolation, not in connected speech.

After the preliminary survey, a wordlist of eighty words and thirty-nine words was prepared for a study of consonants and vowels respectively. This wordlist was taken from *Old Tai Dam (Black Tai) Lexicon* (Manuscript) in which Old Tai Dam 1 (OTD1) and Old Tai Dam 2 (OTD2) were reconstructed by L. Thongkum (2002) Using this wordlist, the data were collected from two locations, village number six (Huathanon village-HTN), Donphutsa sub-district, Dontoom district and village number nine (Sakaeray village-SKR), Donyayhom sub-district, Muang district, Nakhon Pathom province. These two locations were chosen for this case study for four reasons. First, no previous studies of Thai Song language in these locations have been found. Second, these two Thai Song communities are strongly united. Third, the older generation, the middle generation and some young generation speakers still use Thai Song language in their villages. Finally, the community leaders and villagers are very cooperative. The gender and ages of the six informants are as follows:

Table 1: Genders and ages of the six informants

Age groups	Huathanon		Sakaeray	
	Age	Gender	Age	Gender
Old generation (G1)	66	Male	74	Male
Middle generation (G2)	40	Female	39	Male
Young generation (G3)	22	Female	21	Male

The words collected from speakers of all generations were compared to find how far pronunciation varied according to age-group. This linguistic variation was further compared with Thai and the Tai Dam language spoken in Vietnam. The comparison is limited to segmental features, i.e. initial consonants and vowels, excluding tones. Old Tai Dam initials constructed by L. Thongkum (2002) were provided to see how the Old Tai Dam initials were reflected in Thai Song, Tai Dam, and Thai.

The study of lexical replacement was carried out by identifying the typical Thai Song words from the wordlists. These words are unambiguously not Thai loans. The words that were cognates and showed a sound correspondence between Tai Dam, Thai Song and Thai were excluded. For example, the word [tɕa:ŋ^{31}] (Tai Dam) ~ [tɕa:ŋ^{441ʔ}] (Thai Song) ~ [tɕ^{h}a:ŋ^{55}] (Thai) 'elephant' was left out because all languages share the same cognates. On the other hand, words such as [hoŋ^{35}] (Tai Dam) ~ [maʔ^{45}hoŋ^{44}] (Thai Song) ~ [ma^{55}la^{55}kɔɔ^{33}] (Thai) 'papaya' were kept for lexical analysis as the Tai Dam and Thai Song words are not cognate with Thai. The word selection resulted in 45 Tai Dam/ Thai Song words. Out of these 45 words, the researcher counted the Thai words and Thai Song words each informant used. The final step was to calculate the frequency percentage of Thai and Thai Song words used by each informant and display the result in graphs.

4. Thai Song and Tai Dam Phonology

Most studies of Thai Song in Thailand report that there are nineteen single initial consonants (Ananthrawan 1978, Panka 1979, Wattanaprasert and Liamprawat 1988, Maneewong 1987, Unakornsawat 1993, and

Suesorsit 1992) as presented in table 2 compared with the consonant phonemes of Standard Thai in table 3. The consonants followed by a dash occur only in the initial position of words.

Table 2: The consonant phoneme inventory of Thai Song

p		t		k	ʔ
ph -		th -		kh -	
b -					
			c -		
	f -	s -			h-
m		n	ɲ -	ŋ	
		l -			
w			j		

Table 3: Consonant phoneme inventory of Standard Thai

p		t		k	ʔ
ph -		th -		kh -	
b -		d-			
			c -		
			ch-		
	f -	s -			h-
m		n		ŋ	
		r-			
		l -			
w			j		

Below is a list of lexical examples where the two languages have different phonemes in the same etymon.

Thai Song	**Standard Thai**	**Gloss**
nɛw^{33}	jiaw^{42}	'urine, to urinate'
caan^{41}	k^{h}laan^{33}	'to crawl'
ɲaa^{31}	jaa^{42}	'grass'
kɛw^{41}	k^{h}iaw^{55}	'to chew'
kwan^{55}	k^{h}wan^{33}	'smoke'
saj^{25}	k^{h}aj^{22}	'egg'
haʔ^{33}	raak^{42}	'root'

Gedney (1964) also found nineteen single initial consonants in Tai Dam spoken in Son La but some consonants are different. The consonants /ph-/ and /w-/ are absent and the consonants /d-/ and /v-/ are present in his work. The study of Tai Dam language in Son La by Fippinger and Fippinger (1974) is similar to that of Gedney (1964) except that the consonant /kh-/ is represented as /x-/ in Fippinger and Fippinger's work. The consonant [d-] is absent in Thai Song because it occurs in free variation with [l-] so it is analyzed as an allophone of /l-/. On the other hand, Daecha (1986) treats [d-] ~ [l-] as allophones of /d-/. The consonant /v-/ in Tai Dam has become /w-/ in Thai Song. This is evident in the work of Panka (1979) which found a free variation of [w-] and [v-] in only one word, i.e. [viaʔ^{44}] 'work'.

Most studies of Tai Dam and Thai Song have found the same initial clusters /kw-, khw-, ŋw-/ except for the work of Fippinger and Fippinger (1974) which found /xw-/ instead of /khw-/. Few words with /ŋw-/ have been found.

Tai Dam and Thai Song languages have the same final consonants /-p, -t, -k, -ʔ, -m, -n, -ŋ/. The final /-ʔ/ corresponds to /-k/ preceded by long vowels or diphthongs in Thai as in [piʔ^{45}] (Thai Song) and [pi:k^{22}] (Thai) 'wing'. Some studies report that there are nine final consonants because the final vowels /-u/ and /-i/ are treated as the finals /-w/ and /-j/ respectively (Ananthrawan 1978, Daecha 1986, Panka 1979, Wattanaprasert and Liamprawat 1988, Suesorsit 1992, and Unakornsawat 1993).

All Thai Song studies report that there are nine short single vowels /i, e, ɛ, ɨ, ɤ, a, u, o, ɔ/ with their long vowel counterparts, whereas the contrastive vowel length is found only in /a/ and /a:/ in Tai Dam (Gedney 1964 and Fippinger and Fippinger 1974).

The number of diphthongs found in Thai Song studies varies according to different analyses of final vowels. Most studies have four diphthongs /ia, ɨa, ua, aɨ/ whereas Tai Dam has three /ia, ɨa, ua/ because the final vowel /ɨ/ is treated as the final consonant /ɣ/ (Gedney 1964 and Fippinger and Fippinger 1974). Panka (1979) does not analyse /u, i, ɪ/ as final consonants so there are fifteen diphthongs and three triphthongs in her study.

The tonal analysis of Huathanon dialect is compared with Tai Dam (Fippinger and Fippinger 1974) as follows:

Table 4: Tai Dam and Thai Song tones

Tones No.	Tai Dam	Thai Song (Huathanon dialect)	Tai Dam sample words
1	lower-mid level [22]	low rising [213]	pi:[22] 'year'
2	high rising [45]	low rising to high [215]	si:[45] 'four'
3	low level or falling and glottalized [21ʔ]	low falling and glottalized [21ʔ]	xaw[21ʔ] 'rice'
4	high level [55]	high rising falling [452]	ma:[55] 'to come'
5	higher-mid level [44]	mid level [33]	pi:[44] 'older sibling'
6	mid falling and glottalized [31ʔ]	mid falling and glottalized [31ʔ]	ma:[31ʔ] 'horse'

5. Phonological variation by age-group

5.1. Initial consonants

The variation of initial consonants is presented in tables 5-10. The first column consists of Old Tai Dam (OTD) initials at the first stage with sample Thai words having the initials. The second column is composed of Old Tai Dam initials at the second stage. Both stages of Old Tai Dam initials were reconstructed by L. Thongkum (2002). The third column includes Modern Tai Dam (MTD) initials found in the work of L. Thongkum (2002). The Thai Song column is split into six columns, i.e. the first generation (G1), the second generation (G2), and the third generation (G3) of Sakaeray village and Huathanon village. The consonant variation is discussed as follows:

5.1.1 The d - ~ l- ~ n- variation

Table 5: the d - ~ l- ~ n- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ʔd- da:w[33] 'star'	*d-	d-/l-	d-	d-/l-	d-	d-	d-	d-
*ʔd- dom[33] 'to smell'	*d-	d-/l-	n-	d-	n-	n-	n-	d-

Thai Song has three variants of /d/, that is, [d - ~ l- ~ n-]. The original reflex of OTD1 *ʔd is [d-] as pointed out by Gedney (1964) that the change from [d-] to [l-] took place in Tai Dam dialects sporadically so this should be a recent change. So Thai Song speakers must have used [l-] before moving to Thailand. This inference is affirmed by the statement made by Maneewong (1987) that her main informant did not use /d-/ whereas young generation speakers used it. Using /d-/ is wrong because Thai Song ancestors did not use /d-/. In other words such as [laŋ[214]] 'nose' which is no longer used in Thai with the same meaning, all Thai Song speakers use [l-] with no variation with [d-]. In other cognate words such as [daaw[214]], [l-] has changed back to [d-] with the influence of the Thai sound [d-]. The middle-age speaker of SKR dialect still uses [d-] in free variation with [l-] while other speakers have completely replaced [l-] with [d-]. In particular words such as [dom[214]] 'to smell', the OTD1 *ʔd- is represented by [n-] in the speech of G1, G2 in HTN dialect and G1, G3 in SKR dialect. The change from [d-] to [n-] might be caused by the final nasal [-m] as remarked by Weera Ostapirat (p.c.) that Sui also uses [n-] when it is followed by nasal finals. The initial [n-] has changed back to [d-] in the speech of G2 in SKR dialect and G3 in HTN dialect because of the influence of the Thai consonant [d-].

5.1.2 The c - ~ k- ~ k^h- variation

Table 6: the c - ~ k- ~ k^h- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ɟ- k^hra:ŋ[33] 'to moan'	*c-	c-	c-	c-	k-	c-	c-	c-
*ɟ- k^hra:m[33] 'indigo'	*c-	c-	c-	c-	k-	k^h -	k^h -	k^h -

Thai Song reflects the OTD1 *ɟ- as [c-] such as in the word [caaŋ[41]] 'elephant'. However, in particular words such as [caam[452]] 'indigo', [c-] is replaced with the Thai sound [k^h-] as in the speech of all speakers in HTN dialect. As for the word [caaŋ[452]] 'moan', [c-] is kept by all speakers except G3 speaker of SKR dialect. This speaker replaces [c-] with the Thai sound [k^h-] which is further adapted to [k-] because [k^h-] in Thai corresponds to [k] in Thai Song.

5.1.3 The ŋw- ~ w- ~ ŋ- ~ h- variation

Table 7: the ŋw- ~ w- ~ ŋ- ~ h- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ŋw- (mɯa[42]) wa:n[33] 'yesterday'	*ŋw-	ŋw-	ŋw-	ŋw-	ŋw-	ŋw-	w-	ŋw-
*ŋw- ŋap[55]/ hap[22] 'to close (a door)'	*ŋw-	ŋw-	ŋw-	ŋw-	h-	ŋw-	ŋ-	ŋ-

The initial [ŋw-] is used only in a few words. L. Thongkum (2002: 93) lists only three words in her work (KhawYoy dialect), that is, [ŋwaa[55]] 'yesterday', [ŋwap[33]] 'to close (a door)', and [ŋwaaj[41]] 'to turn the face (in order to look)'. All Thai Song speakers of SKR and HTN dialects do not know the last word so it is excluded from this study. All speakers, except the G2 speaker of HTN dialect, still retain the initial [ŋw-] in the word [ŋwaa[55]] 'yesterday'. The G2 speaker of HTN dialect replaces [ŋw-] with the Thai sound [w-]. As for the word [ŋwap[33]] 'to close (a door)', the G1 and G2 speakers of SKR dialect and the G1 speaker of HTN dialect keep [ŋw-] while the G3 speaker of SKR dialect and G2 and G3 speakers of HTN dialect replace [ŋw-] with the Thai sounds [h-] and [ŋ-] respectively.

5.1.4 The h- ~ ŋ- variation

Table 8: the h- ~ ŋ- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*h- ŋa:j[24] 'to turn something up, to lie face up, to open up'	*h-	h-	h-	h-	ŋ-	h	ŋ-	ŋ-
*h- ŋɔ:n[24] 'comb (of fowls)'	*h-	h-	ŋ-	ŋ-	ŋ-	h-	ŋ-	ŋ-

Thai Song reflects the OTD1 *h- as [h-] as seen in the word [hak[35]] 'to break, broken off'. In this case, [h-] shows no variation by age-group but for the words whose initials were reconstructed by Li (1977) for Proto-Tai initials as [*hŋ-], there is a variation of h- ~ ŋ-. The G1 and G2 speakers of SKR dialect and G1 speaker of HTN dialect retain [h-] in the word [haaj[214]] whereas the G3 speaker of SKR dialect and G2 and G3 speakers of HTN dialect replace [h-] with the Thai consonant [ŋ-]. The G1 speaker of HTN dialect is the only one who keeps [h-] in the word [hɔn[214]] 'comb (of fowls)'. Others replace [h-] with the Thai consonant [ŋ-].

5.1.5 The h- ~ l- variation

Table 9: the h- ~ l- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ɦ- ruŋ[55] 'rainbow'	*h-	h-	h-	l-	l-	h-	l-	l-

Thai Song reflects the OTD1 *ɦ- as [h-] such as in the word [ham[452]] 'bran'. But in some words such as [huŋ[452]] 'rainbow', the G2 and G3 speakers of both locations replace [h-] with the Thai sound [l-], a simplified form of /r-/.

5.1.6 The m- ~ l- variation

Table 10: the m- ~ l- variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ml lɯ:m[33] (ta:[33]) 'to open (one's eyes)'	*m-	m-	m-	m-	m-/l-	m-	m-	m-

Thai Song reflects the OTD1 *ml as [m-] in the word [mɯn[452]] 'open (one's eyes)'. Most speakers at both locations retain [m-] while the G3 speaker of SKR dialect fluctuates between [m-] and the Thai sound [l-].

5.2. Vowels

While the initial consonants in Tai Dam and Thai Song are mostly consistent with those in Thai, vowel correlation appears sporadic, occurring randomly in individual words as seen in table 11.

Table 11: Vowel variation by age-group

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*mitD met^{55} ‘seed’	*mitD2	mit^{33}	i	e	e	i	e	e
*ɲipD yep^{55} ‘to sew’	*ɲipD2	ɲip^{44}	i	i	i	i	i	e
*khiwA (me:n^{24}) k^{h}iaw^{24} ‘smelly’	*khiwA1	khiw^{11}	i	i	i	i	iə	iə
*_khepD (ta^{33}) k^{h}a:p^{22} ‘centipede’	*_khepD1	cap^{24} khep^{24}	e	e	e	e	e	a
*khɛwA k^{h}iaw^{24} ‘green’	*khɛwA1	khɛw^{22} khiəw^{11}	ɛ	ɛ	ɛ	ɛ	iə	ɛ
*ʔɛwA ʔe:w^{33} ‘waist’	ʔɛwA1	ʔɛw^{11}	ɛ	ɛ	ɛ	ɛ	ɛ	e
*pɛwA ple:w^{33} (faj^{33}) ‘flame’	*pɛwA1	pɛw^{11} piəw^{22}	ɛ	ɛ	ɛ	ɛ	e	e
*hlɛwA le:w^{24} ‘fluid, liquid’	*lɛwA1	liəw^{11} dew^{11} lɛw^{22}	ɛ	ɛ	ɛ	ɛ	e	ɛ
*sa-ʔɤkD sa^{33}-ʔɯk^{22} ‘hiccup’	*sa-ʔɤkD1	sa-ʔɤk^{24}	ɤ	ɤ	ɤ	ɯ	ɯ	ɯ
*lɤkD lɯk^{55} ‘deep’	*lɤkD2	lɤk^{33} dɤk^{33}	ɤ	ɤ	ɤ	ɤ	ɯ	ɯ
*ʔɤkD ʔok^{22} ‘chest’	*ʔɤkD1	ʔɤk^{24}	ɤ	ɤ	ɤ	ɤ	ɤ	o
*hɤŋA hɯŋ^{24} ‘jealous’	*hɤŋA1	hɤŋ^{11}	ɤ	ɤ	ɤ	ɤ	ɤ	ɯ
*ʔotD ʔut^{22} ‘to stop, to plug (a hole)’	*ʔotD1	ʔot^{24}	o	o	u	u	u	u
*ʔoŋC ʔuŋ^{42}(mɯ:^{33}) ‘palm’	*ʔoŋC1	ʔoŋ^{21}	o	u	u	u	u	u

Table 11 shows that the vowels [i, e, ɛ, ɤ, o] in particular words are varied by age-group. All speakers of SKR dialect retain most of the typical Thai Song vowels except for the words [mit^{33}] ‘seed’ for which the G2 and G3 speakers use the Thai vowel [e]; [ʔot^{24}] ‘to stop, to plug (a hole)’ for which the G3 speaker uses the Thai vowel [u]; and [ʔoŋ^{21}] ‘palm’ for which the G2 and G3 speakers also use the Thai vowel [u].

At Huathanon village, the use of Thai vowels by all three generations of speakers increases. The G1 speaker keeps Thai Song vowels in most of the words except [sa-ʔɤk^{24}] ‘hiccup’, [ʔot^{24}] ‘to stop, to plug (a hole)’, and [ʔoŋ^{21}] ‘palm’. The G2 speaker uses fewer Thai Song vowels than the G1 speaker. She keeps the Thai Song vowels in five words, namely, [ɲip^{44}] ‘to sew’, [(cap^{24}) khep^{24}] ‘centipede’, [ʔɛw^{11}] ‘waist’, [ʔɤk^{24}] ‘chest’, and [hɤŋ^{11}] ‘jealous’. The G3 speaker replaces all of the Thai Song vowels with the Thai vowels except in the two words [khɛw^{22}] ‘green’ and [lɛw^{22}] ‘fluid, liquid’.

It should be noted that, in some words, no vowel variation is found because the speakers of all generations have replaced the Thai Song vowels with Thai vowels as seen in table 12.

Table 12: Thai song vowels with no variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*pɯtD pə:t[22] 'to open'	*pɯtD1	pɯt[24]	ə	ə	ə	ə	ə	ə
*mɤkD mɯk[22] 'ink'	*mɤkD2	mɤk[33]	ɯ	ɯ	ɯ	ɯ	ɯ	ɯ

The consonants and vowels which have been varied by age-group have been discussed. As mentioned earlier, the analysis of phonological variation is based on two wordlists, that is, a 222 wordlist and a wordlist of eighty words and thirty-nine words, so more phonological alternations may be found if more words are further studied. The next section will present the lexical replacement of the three generations of speakers.

6. Lexical replacement

A wordlist of 45 words (see the appendix) which are different in Thai Song and Thai forms was used for the study of lexical replacement. The lexical data were analyzed by counting the occurrences of Thai Song and Thai words used by each generation speaker at both locations. The frequency percentage of lexical items in each language was calculated. The results are shown in figures 1 and 2 respectively.

There are three words that all generations at both locations replace with Thai words. They are, /nam[6] be[3]/ 'sea', /thɔ[3]/ 'rabbit', /siŋ[2] saaw[1]/ 'spider' which are replaced with the Thai words [thə[32] le:[243]], [kə[32] ta:j[15]], [mæ:ŋ[243] mum[243]] respectively. Note also that the Thai Song tones are used in Thai words. As for the word /piəŋ[2] faa[3]/ 'cloud', only the G1 speaker of SKR dialect uses the Thai Song word [k[h]i:[22ʔ]fa:[31ʔ]] which is similar to /piəŋ[2] faa[3]/. The G2 and G3 speakers tend to replace some words with Thai words. For example, the G2 and G3 speakers of SKR dialect replace [kaɯ[214]] 'swollen' with the Thai word [buam[214]]. Both of them use more Thai words than the G2 and G3 speakers of HTN dialect.

Figures 1 and 2 show that the G1 speakers at both locations use Thai Song words the most. The G2 and G3 speakers at both locations use Thai Song words more than Thai words. However, the G2 and G3 speakers of SKR use fewer Thai Song words than those of HTN. That is, the former replaced more Thai Song words with Thai words than the latter.

This analysis of lexical replacement reveals that middle-aged and young speakers borrowed Thai words the most and it is expected that more Thai words will be borrowed into Thai Song. There are some words that are seldom used, for example, [ŋwaʔ[35]] 'to turn one's face to the right or left' and [khwɛʔ[33]] 'first time'. It can be predicted that these words will disappear from Thai Song in the near future.

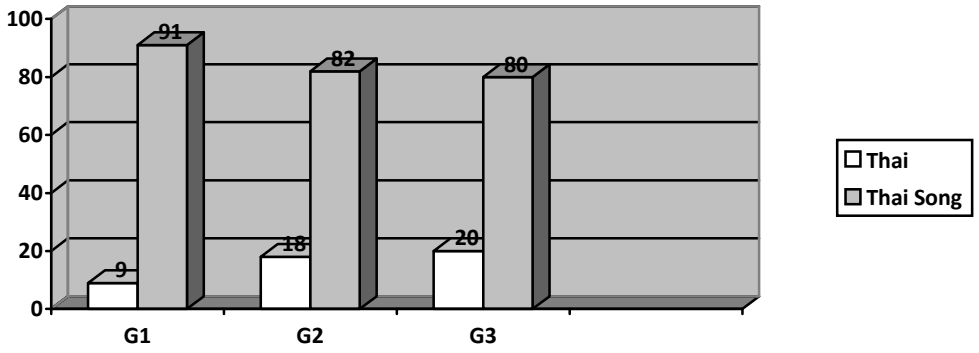


Figure 1: the percentage of Thai and Thai Song (HTN) words used by each generation

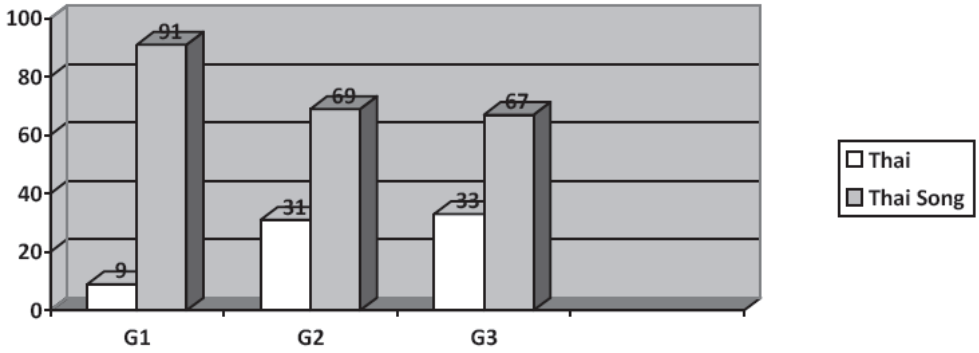


Figure 2: the percentage of Thai and Thai Song (SKR) words used by each generation

It should be noted that most speakers of HTN dialect use more Thai Song words than those of SKR dialect. On the contrary, the use of Thai vowels by the speakers of HTN dialect exceeds the use of Thai vowels by the speakers of SKR dialect. More studies should be done to find out whether phonological change occurs faster than lexical change or vice versa. The study of lexical replacement is based on forty-five words. If further studies use more words, some interesting pattern of lexical replacement, such as semantic fields of lexical borrowing, may be disclosed.

7. Conclusion and discussion

Phonological variation and lexical replacement in this study has been examined from a sociolinguistic point of view. The focus of this study is on age as a social variable controlling linguistic variation and lexical replacement. This work is a case study of differences in pronunciation and word replacement in only two locations, Huathanon village, Donphutsa sub-district, Dontoom district and Sakaeray village, Donyayhom sub-district, Muang district, Nakhon Pathom province. Further quantitative research should be done at larger locations so that the subjects might represent the whole Thai Song community.

The purpose of this study was to find out to what degree the Thai Song language in these two locations has been influenced by the Thai language. It has been found that Thai Song has progressively shifted to Thai in the speech of G2 and G3 generations as evidenced in the study of phonological variation and lexical replacement. The phonological variation shows the variation of initial consonants by age-group as listed in table 13. It is clear that most middle and young generation speakers tend to replace the typical initial consonants with Thai consonants as hypothesized.

Table 13: Variation of initial consonants by age-group

OTD 1	OTD 2	MTD	Initial consonant variation by age-group
*ʔd-	*d-	d-/l-	d - ~ l- ~ n-
*ɟ-	*c-	c-	c- ~ k- ~ k^h-
*ŋw-	*ŋw-	ŋw-	ŋw- ~ w- ~ ŋ- ~ h-
*h-	*h-	h-	h- ~ ŋ-
*ɦ-	*h-	h-	h- ~ l-
*ml	*m-	m-	m- ~ l

It is also interesting to note that some initial consonants are not varied by age-group as the Thai Song speakers of all generations at both locations use the same consonants which agree with the OTD2 consonants but some consonants are different from MTD as seen in table 14. It is questionable whether the Thai Song speakers used the same consonants as MTD speakers before migrating to Thailand and then replaced them with the Thai consonants afterwards, or always preserved these consonants as Thais do. According to the Wave Theory (Bailey 1973), an innovation takes place at the point of origin and spreads to neighboring areas like waves. The innovation barely reaches the outermost ring of the waves. Consequently, the farther a group

of people moves away from its homeland, the more likely the language will preserve older forms. The innovation does not affect the group whose language is spoken at the outer ring of the wave. As the Tai Dam area in Vietnam is the point of origin, any linguistic change taken place there might not reach Thai Song speakers in Thailand. So, the change from OTD2 *ph to [f-]; *b to [b-/v-]; and *w to [b-/v-] should have occurred after Thai Song speakers settled in Thailand since the Thai Song there continue to keep the original forms. Furthermore, it was pointed out by a Vietnamese researcher that the consonant [ph-] does not exist in Vietnamese. As Tai Dam speakers also speak Vietnamese, they have replaced [ph-] with [f-]. So, this evidence confirms the fact that the Thai Song consonants in table 14 are not from Thai but are original consonants of Thai Song.

Another notable point is the OTD *kh which is reflected in the speech of all Thai Song speakers at both locations as [kh-]. It seems undoubtedly that this consonant is an original Thai Song consonant, which is the same as the Thai consonant. But further data from a 94 year old speaker at Bangkung village, Muang District, Suphanburi Province shows the reflex [x-] which agrees with Fippinger and Fippinger's (1974) finding as mentioned in section 4. Consequently, it might be possible that Thai Song speakers in the western region of Thailand used to have [x-] but replaced it with the Thai consonant [kh-] or internally modified it as [kh-]. Another consonant which is also lost is [v-] which has become [w-]. As mentioned in section 4, Panka (1979) found a free variation of [w-] and [v-] in the word [viaʔ^{44}] 'work'. In this study, only [w-] is found. Contrary to the loss of consonants, Jirananthanaporn et al (2003) and Daecha (1986) found /ch-/ in some words such as /chut^{35}/ 'pull, drag'. These words are Thai loanwords. The initial /ch-/ is evidently a Thai consonant because the voiceless unaspirated initials are a typical feature of Thai Song and Tai Dam. This feature distinguishes Thai Song or Tai Dam from Thai and Lao which fall into the voiceless aspirated initial group (Chamberlain 1975). Jirananthanaporn et al. (2003) and Daecha (1986) also include /d-/ in their phonological inventory as the typical consonant [l-] has been completely replaced by the Thai sound [d-].

Table 14: the initial consonants with no variation

OTD 1	OTD 2	MTD	Thai Song					
			Sakaeray			Huathanon		
			G1	G2	G3	G1	G2	G3
*ph- p^{h}i$^{:24}$ 'ghost'	*ph-	f-	ph-	ph-	ph-	ph-	ph-	ph-
*ʔb- bin$^{:33}$ 'to fly'	*b-	b-/v	b-	b-	b-	b-	b-	b-
*ʔb- dɯan$^{:33}$ 'moon'	*b-	b-/v	b-	b-	b-	b-	b-	b-
*hw- wi$^{:24}$ 'comb'	*w-	b-/v-	w-	w-	w-	w-	w-	w-
*w- wa$^{:33}$ 'measure of length (2 meters)'	*w-	b-/v-	w-	w-	w-	w-	w-	w-

The study of vowel variation in the thirty-nine words shows that the vowels [i, e, ɛ, ɤ, o] in particular words are varied by age-group. In some words, there is no vowel variation by age-group because the speakers of all generations have replaced the Thai Song vowels with the Thai vowels.

Despite the gradual change of Thai Song consonants and vowels to Thai, it has been found that Thai Song speakers of all generations at both locations still preserve some typical features of Thai Song phonology. First, the palatal nasal /ɲ-/ such as in the word [ɲa:^{22ʔ}] 'grass' is preserved by speakers of all generations in both Thai Song communities. In particular words such as [ɲin^{452}] 'to hear', the Proto-initial *ɲ- is represented as /ŋ-/ in Tai Dam and in the speech of the G1 speaker (HTN). Other generation speakers at both locations use /ɲ-/. Second, the dropping of liquids in the Proto-labial and velar clusters *pl-, *bl/r-, *ʔbl-, *ml-, *kl-, *kr-, *khr-, *gl- remains, as seen in /p-, p-, b-, m-, k-, k-, s-, c-/ respectively. Third, the final glottal /-ʔ/, which corresponds to /-k/ preceded by long vowels or diphthongs in Thai, is kept, for example, in the word [luʔ^{32}] 'child'. And finally, the diphthong [aɬ] is a typical feature of Thai Song

phonology which has been kept at all locations. This diphthong has merged with [ai] in Thai but is still marked by a special symbol [may^{55} muan^{55}].

The study of lexical replacement supports the second hypotheses that "Young generation speakers use fewer Thai Song words than the older generations". This finding agrees with the work of Liamprawat and Wattanaprasert (1996) who used 200 words for their study of lexical usage among three generations of Thai Song speakers in Nongsonghong sub-district, Banphaew district, Samut Sakhon province. They found that the G1 speakers used 90 percent of Thai Song words whereas the G2 and G3 speakers used 68.5 and 23.5 percents respectively. In addition, the G1 and G2 speakers used Thai Song words more than Thai words whereas the G3 speakers used Thai words more than Thai Song words. This is different from the present study in that speakers of all generations at HTN and SKR use Thai Song words more than Thai words as seen in figures 1-2. Saeng-ngam (2006) also found the same result as this study. She used 70 words for her study of lexical usage among three generations of Thai Song speakers in Nongprong sub-district of Khao Yoy district in Phetchaburi province. The lexical study reveals that all three generations of Thai Song speakers used Thai Song words more than Thai words and young generation speakers used fewer Thai Song words than the older generations. She concluded that change in lexical usage has slowly begun to occur in the Thai Song spoken at this location and lexical change is occurring faster than tonal change. Buranasing (1988) also found that lexical changes are mostly evident in second and third generation Thai Song speakers in Ban Don sub-district of U-Thong district in Suphanburi province because of the influence of standard Thai.

Despite the positive attitudes towards the Thai Song language among the middle-aged and elderly, it can be concluded from the phonological variation and lexical replacement that language shift in progress has occurred. The major cause of language shift is the mass media which encourages the use of the Thai language in Thai Song communities. Other factors involved in language shift include limited domain of language use, change in social life, and negative attitudes towards the Thai Song language by the young generation. This is supported by Chakshuraksha's (2003) study in which she found factors that promote shift. These include the practice of intermarriage, formal schooling in schools far from the community and the widespread presence of electronic media in the village. Charoenchai (2008) has also found that technology influences change in the lexical usage of Tai languages, including Thai Song, in Lopburi province. The widespread use of Bangkok Thai lexical items in Lopburi is due to the arrival of modern technology. Buranasing (1988) suggests that the reason Thai loanwords can be found in the Thai Song language is because young generation speakers have become assimilated into Thai society and thus, their native language has also been influenced by Thai. Besides, they do not want to appear different from the Thai majority so they have adjusted their native pronunciation to that of Thai, such as the use of liquid clusters which are absent in the Thai Song language but have been introduced into the Thai Song phonology of Suphanburi dialect.

Consequently, it is anticipated that, in future, the Thai Song language will be used less and less. Over the next 30-40 years, the linguistic vitality of the language may decline.

References

Ananthrawan, Chakhrit. 1978. The phonology of Tai Dam, Panat village, Kawkaew sub-district, Chiangkhan district, Loei province. MA thesis. Srinakharinwirot University. (In Thai)

Bailey, Charles-James N. 1973. "The patterning of language variation." In Bailey, Richard W. and Jay L. Robinson (eds.). *Varieties of Present-day English*. New York: Macmillan Company. Pp. 156-186.

Brown, Marvin. 1965. *From Ancient Thai to Modern Dialects*. Bangkok: White Lotus Co. Ltd.

Buranasing, Anchulee. 1988. An analysis of lexical change among three generations in Thai Song dialect. MA thesis. Mahidol University. (In Thai)

Burusphat, Somsonge, Sujaritlak Deepadung, Sumittra Suraratdecha, Patama Patpong, Narong Ardsmiti, and Pichet Setapong. 2011. *Language Uses and Attitudes towards Languages and Ethnic Tourism of Ethnic Groups in the Western Region of Thailand.* Bangkok: Phimdee Ltd. (In Thai)

Chakshuraksha, Natthida. 2003. Language maintenance and shift in a displaced Tai language: a case study of Black Tai at Nongkhe village, Thailand. Ph.D. dissertation. Mahidol University.

Chamberlain, James R. 1975. "A new look at the history and classification of the Tai languages". In Harris, Jimmy G. and James R. Chamberlain (eds.). *Studies in Tai Linguistics in honor of William J. Gedney*. Bangkok: Central Institute of English Language Office of State Universities. Pp. 49-66.

Charoenchai, Payau. 2008. *A Lexical Study of Tai Languages in Lopburi Province and Influences of Technology on Changes of Lexical Usage*. Bangkok: Department of Oriental Languages, Faculty of Archeology, Silpakorn University, Thailand. (In Thai)

Daecha, Wilailuck. 1986. A comparative study of phonology of six Tai dialects spoken in Amphoe Tha Tako, Changwat Nakhon Sawan. MA thesis. Chulalongkorn University. (In Thai)

Fippinger, Jay and Dorothy Fippinger. 1974. "Black Tai phoneme, with reference to White Tai." *Anthropological Linguistics* 12.3: 83-95.

Gedney, William. 1964. "A comparative sketch of White, Black and Red Tai." *The Social Review*. Special number 14:1-47.

Jirananthanaporn, Supatra, Unchalee Singnoi, and Siriporn Maneechukate. 2003. *Grammatical System in Thai Song*. Linguistic Department, Faculty of Humanities, Naresuan University. (In Thai)

Li, Fang Kuei. 1960. "A tentative classifications of Tai dialects." In Diamond, Stanley (ed.). *Culture in History Essay in Honor of Paul Radin*. New York: Columbia University Press. Pp. 951-959.

Li, Fang Kuei. 1977. *A Handbook of Comparative Tai*. Hawaii: The university press of Hawaii.

Liamprawat, Suwattana and Kantima Wattanaprasert. 1996. *An Analysis of Lexical Use and Variation among Three Generations in Lao Language Communities of Thaicin River Basin*. Nakhon. Pathom: Silpakorn University Printing House. (In Thai)

L. Thongkum, Theraphan. 2002. *Old Tai Dam (Black Tai) Lexicon*. (Manuscript)

Maneewong, Orapin. 1987. A comparative phonological study of Lao Song in Petchaburi and Nakhon Pathom Provinces. MA thesis. Mahidol University.

Panka, Kanchana. 1979. The phonological characteristics of Lao dialects in Amphoe Muang, Nakhon Pathom. MA thesis, Chulalongkorn University. (In Thai)

Saeng-ngam, Suntharat. 2006. Lexical and tonal variation by age group and language attitude in the Song language (Black Tai) of Ampher Khao Yoy, Changwat Phetchaburi. MA thesis. Chulalongkorn University. (In Thai)

Sams, Bert F. 1988. "Black Tai and Lao Song Dam: the divergence of ethnocultural identities." *Journal of the Siam Society* 76: 100-116.

Snyder, Donna Marie. 1995. Variation in Bouyei. MA thesis. The University of Texas at Arlington.

Sribusara, M. (pen-name). 1987. *Thai Dam Wailing*. Bangkok: Bannakij Press. (In Thai)

Suesorsit, Rosarin. 1992. The phonology of Lao Song, Dontaw-it village, Rangway district, Ampher Phanomthuan, Kanchanaburi Province. Independent Study Paper. Silpakorn University. (In Thai).

Unakornsawat, Orapan. 1993. A comparative phonological study of Phu Tai and Lao Song. MA thesis. Silpakorn University. (In Thai)

Wattanaprasert, Kantima and Suwattana Liamprawat. 1988. *The Phonology of Lao Dialects in Thachin River Basin*. Nakhon Pathom: Silpakorn University Printing House. (In Thai)

Interview

Bakam, Khaming. 2010. Interview. November 21, 2010. Nongbuathongtay village, Sikhottabong district, Vientiane, Lao PDR.

Appendix

The Thai Song wordlist used for the study of lexical replacement

In the column of Huathanon and Sakaeray village, the first line consists of words used by the G1, G2, G3 speakers of Huathanon and the second line includes those of Sakaeray village. It should be noted that most of the Thai Song words in this wordlist are typical Thai Song words and some words are shared by Thai Song and Lao.

No.	Gloss	Thai	Black Tai	Huathanon and Sakaeray villages		
				G1	G2	G3
1.	cloud	[me:k^{42}]	/piəŋ^{2} faa^{3}/	[me:k^{32}]	[me:k^{32}]	[me:k^{32}]
				[k^{h}i:^{22ʔ} fa:^{441ʔ}]	[me:k^{32}]	[me:k^{32}]
2.	sea	[t^{h}a^{33} le:^{33}]	/nam^{6} be^{3}/	[t^{h}ə^{32} le:^{452}]	[t^{h}a^{32} le:^{45}]	[t^{h}a^{33} le:^{45}]
				[t^{h}ə^{32} le:^{452}]	[t^{h}a^{22} le:^{45}]	[t^{h}a^{33} le:^{45}]
3.	papaya	[ma^{55}la^{55} kɔ:^{33}]	/hoŋ^{2}/	[maʔ^{45} hoŋ^{215}]	[maʔ^{32} ho:ŋ^{214}]	[maʔ^{45} hoŋ^{215}]
				[maʔ^{45} hoŋ^{215}]	[maʔ^{45} hoŋ^{214}]	[maʔ^{45} ho:ŋ^{215}]
4.	rabbit	[ka^{33} ta:j^{22}]	/thɔ^{3}/	[kə^{32} ta:j^{215}]	[ka^{21} ta:j^{215}]	[kə^{33} ta:j^{215}]
				[kə^{32} ta:j^{215}]	[ka^{21} ta:j^{215}]	[ka^{33} ta:j^{215}]
5.	spider	[mæ:ŋ^{33} mum^{33}]	/siŋ^{2} saaw^{1}/	[mæ:ŋ^{452} mum^{452}]	[mæ:ŋ^{452} mum^{452}]	[mæ:ŋ^{453} mum^{453}]
				[mæ:ŋ^{452} mum^{452}]	[mæ:ŋ^{452} mum^{452}]	[mæ:ŋ^{453} mum^{453}]
6.	mouth	[pa:k^{22}]	/sop^{2}/	[sop^{45}]	[sop^{45}]	[sop^{45}]
				[sop^{45}]	[sop^{45}]	[sop^{45}]
7.	tooth	[fan^{33}]	/xɛw^{3}/	[k^{h}æw^{22ʔ}]	[k^{h}æw^{21}]	[k^{h}æw^{22ʔ}]
				[k^{h}æw^{22ʔ}]	[k^{h}æw^{21ʔ}]	[k^{h}æw^{22ʔ}]
8.	abdomen	[p^{h}uŋ^{33}]	/pum^{1}/	[pum^{214}]	[pum^{214}]	[pum^{215}]
				[pum^{214}]	[puŋ^{452}]	[pum^{213}]
9.	thigh	[ton^{42} k^{h}a:^{24}]	/paan^{3} xaa^{1}/	[kok^{45} k^{h}a:^{214}]	[kok^{45} k^{h}a:^{214}]	[kok^{45} k^{h}a^{213}]
				[kok^{45} k^{h}a^{214}]	[ton^{21} k^{h}a:^{214}]	[ko:n^{453} k^{h}a:^{213}]
10.	fat	[ʔuan^{42}]	/pi^{4}/	[pi:^{452}]	[pi:^{452}]	[pi:^{453}]
				[pi:^{452}]	[ʔuan^{21ʔ}]	[pi:^{453}]
11.	speak	[p^{h}u:t^{42}]	/paʔ^{2}/	[paʔ^{45}]	[paʔ^{45}]	[paʔ^{45}]
				[paʔ^{45}]	[paʔ^{45}]	[paʔ^{45}]
12.	think	[k^{h}it^{44}]	/ŋam^{2}/	[ŋam^{215}]	[ŋam^{215}]	[ŋam^{215}]
				[ŋam^{215}]	[k^{h}it^{45}]	[ŋam^{215}]
13.	swim	[wa:j^{42} na:m^{55}]	/lɔy^{4}/	[lɔ:j^{452} na:m^{441ʔ}]	[wa:j^{21} na:m^{441}]	[wa:j^{22} na:m^{31}]
				[wa:j^{32} na:m^{441ʔ}]	[wa:j^{32}na:m^{441}]	[wa:j^{22} na:m^{31}]
14.	throw	[k^{h}wa:ŋ^{42}]	/thim^{3}/	[bæ:n^{215}]	[bæn^{215}]	[bæ:n^{215}]
		[wiaŋ^{22}]		[bæ:n^{215}]	[bæ:n^{214}]	[wiaŋ^{215}]
15.	visit	[jiam^{42}]	/yaam^{1}/	[ja:m^{214}]	[ja:m^{215}]	[ja:m^{213}]
				[ja:m^{214}]	[ja:m^{214}]	[ɲa:m^{213}]
16.	tell a lie	[ko:^{33} hok^{22}]	/yam^{1}/	[k^{h}i:^{22ʔ} bæw^{22ʔ}]	[k^{h}i:^{22} bæ:w^{21}]	[k^{h}i:^{22ʔ} bæw^{22ʔ}]
				[k^{h}i:^{22ʔ} bæw^{22ʔ}]	[bæ:w^{21ʔ}]	[bæw^{32}]
17.	look at	[du:^{33}]	/bɤŋ^{2}/	[bɤŋ^{215}]	[bɤŋ^{215}]	[bɤŋ^{215}]
				[bɤŋ^{215}]	[bɤŋ^{215}]	[bɤŋ^{215}]
18.	father	[p^{h}ɔ:]	/aay^{3}/	[ʔa:j^{22ʔ}]	[ʔa:j^{21ʔ}]	[ʔa:j^{22ʔ}]
				[ʔa:j^{22ʔ}]	[ʔa:j^{21ʔ}]	[ʔa:j^{21ʔ}]

No.	Gloss	Thai	Black Tai	Huathanon and Sakaeray villages		
				G1	G2	G3
19.	mother	[mæ:^{42}]	/em^{4}/	[ʔem^{452}]	[ʔem^{452}]	[ʔem^{453}]
				[ʔem^{452}]	[ʔem^{214}]	[ʔem^{453}]
20.	corn	[k^{h}a:w^{42} p^{h}o:t^{42}]	/xaw^{3} li^{4}/	[k^{h}aw^{22ʔ} san^{214} li:^{452}]	[k^{h}aw^{21} sa^{214} li:^{452}]	[k^{h}aw^{22ʔ} sa^{213} li:^{453}]
				[k^{h}aw^{22ʔ} sa^{214} li:^{452}]	[k^{h}aw^{21} sa^{214} li:^{452}]	[k^{h}aw^{21} sa^{32} li:^{453}]
21.	chopstick	[ta^{33} kiap^{22}]	/thu^{2}/	[t^{h}u:^{215}]	[t^{h}u:^{215}]	[t^{h}u:^{215}]
				[t^{h}u:^{215}]	[taʔ^{21} kiap^{215}]	[ta^{21} kiap^{215}]
22.	firewood	[fɯ:n^{33}]	/luə^{1}/	[lua^{452}]	[lua^{215}]	[lua^{213}]
				[lua^{452}]	[lua^{214}]	[fɯ:n^{453}]
23.	trousers	[ka: ŋ^{33} ke:ŋ^{33}]	/suəŋ^{3}/	[suaŋ^{22ʔ}]	[suaŋ^{21}]	[suaŋ^{32}]
				[suaŋ^{22ʔ}]	[suaŋ^{21}]	[suaŋ^{32}]
24.	work	[t^{h}am^{33}ŋa:n^{33}]	/yet^{5}/	[ʔet^{45} wiaʔ^{32}]	[ʔet^{45} ŋa:n^{452}]	[ʔet^{45} wiaʔ^{22}]
				[ʔet^{45} wiaʔ^{32}]	[ʔet^{45} wiaʔ^{32}]	[ʔet^{45} wiaʔ^{22}]
25.	play	[le:n^{42}]	/in^{3}/	[ʔin^{22ʔ}]	[ʔin^{21}]	[ʔin^{22}]
				[ʔin^{22ʔ}]	[ʔin^{21ʔ}]	[ʔin^{21}]
26.	language	[p^{h}a:^{33}sa:^{24}]	/kwaam^{4}/	[kam^{452}]	[p^{h}a:^{22} sa:^{214}]	[p^{h}a:^{453} sa:^{213}]
				[kam^{452}]	[p^{h}a:^{22} sa:^{214}]	[p^{h}a:^{45} sa:^{213}]
27.	twenty	[ji: ^{42}sip^{22}]	/saaw^{4}/	[sa:w^{452}]	[sa:w^{452}]	[sa:w^{453}]
				[sa:w^{452}]	[sa:w^{452}]	[sa:w^{453}]
28.	many	[la:j^{24}]	/laay^{1}/	[ʔe:^{215}]	[ʔe:^{214}]	[la:j^{213}]
				[ʔe:^{215}]	[ʔe:^{214}]	[ʔe:^{215}]
29.	long	[ja:w^{33}]	/hi^{4}/	[hi:^{452}]	[hi:^{452}]	[hi:^{453}]
				[hi:^{452}]	[hi:^{452}]	[hi:^{453}]
30.	white	[k^{h}a:w^{24}]	/lɔn^{2}/	[lɔ:n^{214}]	[k^{h}a:w^{214}]	[k^{h}a:w^{213}]
				[lɔ:n^{214}]	[k^{h}a:w^{214}]	[k^{h}a:w^{213}]
31.	swell	[buam^{33}]	/kəw^{1}/	[kaɯ^{214}]	[kaw^{22}]	[kaw^{32}]
				[kaɯ^{214}]	[buam^{214}]	[buam^{213}]
32.	dirty	[pɯan^{42}]	/uəy^{2}/	[ʔuaj^{215}]	[ʔuaj^{215}]	[pɯan^{22}]
				[ʔauj^{215}]	[pɯan^{21}]	[pɯan^{21ʔ}]
33.	straight	[troŋ^{33}]	/sɨ^{5}/	[sɯ:^{32ʔ}]	[toŋ^{33}]	[sɯ:^{32}]
				[tɕɤŋ^{32}]	[sɯ:^{22}]	[tɕɤŋ^{32}]
34.	far	[klaj^{33}]	/lak^{2}/	[lak^{45}]	[lak^{45}]	[lak^{45}]
				[lak^{45}]	[lak^{45}]	[lak^{45}]
35.	near	[klaj^{42}]	/cham^{1}/	[kæ:m^{452}]	[kæ:m^{452}]	[kæ:m^{453}]
				[kæ:m^{452}]	[kæ:m^{452}]	[kæ:m^{453}]
36.	same	[mɯan^{24}]	/pɛʔ^{2}/	[k^{h}ɯ:^{452}]	[k^{h}ɯ:^{452}]	[k^{h}ɯ:^{453}]
				[k^{h}ɯ:^{452}]	[kɯ:^{452}]	[k^{h}ɯ:^{453}]
37.	where?	[thi:^{42}naj^{24}]	/kaa^{4} ləw^{1}/	[ka^{452} laɯ^{22ʔ}]	[ka^{22} law^{22}]	[ka^{33} law^{45}]
				[ka^{452} laɯ^{22ʔ}]	[ka^{22} law^{22}]	[kaʔ^{45} law^{33}]
38.	who?	[k^{h}raj^{33}]	/fəw^{3}/	[p^{h}aɯ^{452}]	[bɔ:^{22} p^{h}aw^{22}]	[p^{h}a:w^{453}, p^{h}aj^{45}]
				[p^{h}aɯ^{452}]	[p^{h}u^{21} law^{22}]	[bɔʔ^{32} p^{h}aj^{33}/ p^{h}a:w^{33}]
39.	what?	[ʔa^{33}raj^{33}]	/saŋ^{1}/	[ta^{32} haŋ^{452}]	[ta^{21} haŋ^{22}]	[taʔ^{33} haŋ^{34}]
				[to:^{32} haŋ^{452}]	[ta^{21} haŋ^{22}]	[taʔ^{33} haŋ^{33}]
40.	in	[naj^{33}]	/cuəŋ^{1}/	[kuaŋ^{214}]	[kuaŋ^{214}]	[taŋ^{45} kuaŋ^{213}]
				[kuaŋ^{214}]	[k^{w}uaŋ^{214}]	[kuaŋ^{213}]

No.	Gloss	Thai	Black Tai	Huathanon and Sakaeray villages		
				G1	G2	G3
41.	not	[maj^{42}]	/baw^{2}/	[bɔ:^{452}]	[baʔ^{45}]	[baʔ^{45}]
				[bɔ:^{452}]	[bɔ:^{22}]	[bɔʔ^{45}]
42.	pretty	[suaj^{24}]	/can^{1}/	[mæ:^{441ʔ} di:^{214}]	[mæ:^{22} di:^{214}]	[mæ:^{32} di:^{213}]
				[mæ:^{441ʔ} di:^{214}]	[mæ:^{22} di:^{214}, tɕan^{22}]	[mæ:^{32}suaj^{213}/ ŋa:m^{453}]
43.	short (length)	[san^{42}]	/ten^{2}/	[ten^{22ʔ}]	[ten^{441}]	[ten^{22ʔ}]
				[ten^{22ʔ}]	[ten^{21ʔ}]	[ten^{21}]
44.	like	[tɕ^{h}ɔ:p^{42}]	/mak^{5}/	[mak^{32}]	[mak^{45}]	[tɕɔ:p^{32}]
				[mak^{32}]	[mak^{32}]	[tɕ^{h}ɔ:p^{22}]
45.	with (accompaniment)	[duaj^{42}]	/naŋ^{3}/	[nam^{452}]	[nam^{452}]	[nam^{453}]
				[nam^{452}]	[nam^{452}]	[nam^{453}]

WICHAYA Bovonwiwat. 2013. Aspects in Fengshun Hakka Spoken in Thailand: Perfective, Experiential, and Inchoative.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:35-53
Received 26/5/2012, text accepted 24/1/2013, published June 2013
ISSN: 1836-6821 | Website: http://jseals.org | Managing Editor Dr Paul Sidwell
URL: http://hdl.handle.net/1885/10092

www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics

ASPECTS IN FENGSHUN HAKKA SPOKEN IN THAILAND: PERFECTIVE, EXPERIENTIAL, AND INCHOATIVE

Wichaya Bovonwiwat
Mahidol University
<bwichaya@hotmail.com>

Abstract
This research is part of my dissertation 'A Study of Hakka Aspectual System' for Mahidol University and Thailand Research Fund[1]. It aims to explain the syntactic and semantic structures of Perfective, Experiential, and Inchoative aspects of Fengshun Hakka spoken in Thailand. The Hakka aspectual system generally can be divided into two major categories: bounded and unbounded. According to Chappell (1989a, b), the bounded aspect refers to an event containing either the beginning or the end point, while the unbounded aspect refers to an event without a time limit. The bounded situations can be subcategorized into Perfective, Experiential, and Inchoative aspects. To add an interesting view to this study, the Miaoli Hakka dialect spoken in Taiwan, Jieyang Chaozhou, and Mandarin have been compared with the Fengshun dialect to point out real characteristics of the Hakka aspectual system. The comparison with the Miaoli dialect spoken in Taiwan demonstrates how the three bounded aspects of the two Hakka dialects are expressed. The language contact with Chaozhou in Fengshun Hakka is probably pointed out as one factor by which the bounded aspects in Fengshun Hakka are distinguished from those in the Miaoli dialect.
Keywords: aspect, Hakka, Perfective, Experiential, Inchoative
ISO 639-3 language codes: hak, cmn, tha

1. Introduction

Languages in the Sinitic group have rich aspectual systems and are different from Indo-European languages in which tense plays a more outstanding role than aspect to determine temporal relations. Xiao and McEnery (2004:2) broadly classify languages as tense and aspect languages. In tense languages, such as English and French, tense is morphologically combined with aspects such as English Past Simple and French Imparfait. The English Past Simple and the French Imparfait provide Perfective sense and relate the time of an action to the speech moment. On the contrary, tense in aspect languages, such as Mandarin Chinese, is presented by content lexicons such as adverbs of time, while aspect is denoted by aspect markers. A number of aspect markers have evolved from lexicons. In terms of diachronic study, Bybee et al. (1994:55) state that most aspect markers denoting past actions or relevant senses have evolved from verbs: stative verbs (especially 'have,' 'remain,' and 'wait') and dynamic verbs (especially verbs of movement and 'finish'). The markers that express ongoing events, repeated actions, or present situations are usually derived from locative words, such as postpositions or prepositions, verbs of posture, 'be at,' 'stay,' 'live,' or 'reside.'

[1] I would like to express my gratitude to the Royal Golden Jubilee Scholarship granted by Thailand Research Fund who financially supports my research in Hakka aspectual system.
Also, please note that while vernacular forms are generally rendered here in IPA, Mandarin examples are transcribed with Pinyin.

This aspectual study will provide an overview of Perfective, Experiential, and Inchoative aspects of Fengshun Hakka spoken in Thailand. To add a clearer view of the Hakka aspectual system, and to clarify the various features of Hakka dialects, another Hakka dialect, which is spoken in Miaoli in Taiwan, is compared with the Fengshun Hakka. Since Fengshun Hakka has experienced language contact with Chaozhou and Thai, while Miaoli Hakka has had language contact with Min Nan and Mandarin, the bounded aspects of both dialects are likely to have interestingly different characteristics due to those contacts. This study primarily discusses the aspect markers used to signal the three bounded aspects in Fengshun Hakka and their co-occurrence with verb classes, as well as a comparison with the aspectual system of the Miaoli dialect. Examples of Chaozhou dialect and Mandarin are also compared to show the shared aspectual features among Chaozhou, Mandarin, and Fengshun Hakka.

2. Hakka People

Norman (1988:182-183) proposes that there are three major groups of Chinese languages. The first group, referred to as the Northern group, contains Beijing, Xian, and Kunming. The second group, or Southern group, is composed of Meixian, Guangzhou, Fuzhou, and Jianou. This group includes three languages: Kejia, Yue, and Min. Lastly, Suzhou, Wenzhou, Changsha, Shuangfeng, and Nanchang are categorized as the Central group and include Wu, Gan, and Xiang. Among these three groups, the Northern group is spoken by the majority in China. The languages in this Northern group share similar linguistic characteristics and are less diverse than the languages in the Southern and Central groups, particularly in terms of phonology and lexicon. The most diverse group appears to be the Southern group.

Hakka or Kejia is considered to belong to the Southern dialect group because many of Hakka's linguistic characteristics are similar to other Southern dialects (Norman 1998:222). The most crowded region where Hakka people live is in Meixian. Hakka people also live in other provinces, including Fujian, Jiangxi, Guangxi, Guangdong, Sichuan, Hainan, and Taiwan. Ramsey (1987:111) states 'the dialect spoken around Meixian is considered to be standard Hakka.' Hakka is widely spoken in China, Taiwan, and in many countries in Southeast Asia. Chappell (2001:15) suggests that the Hakka population accounts for 3.7 percent of China's population. The Hakka population in mainland China seems relatively low compared to other groups, but a significant number resides outside China.

A number of Hakka people have migrated to Thailand. According to Smalley (1994: 207-213), the Hakka population accounts for the second largest Chinese group in Thailand, following the Chaozhou population. Chinese people migrated to Thailand a long time ago, and it has been recorded that Chinese have lived in the Tai kingdom since the Sukhothai period. A massive influx of Chinese people occurred in the nineteenth century, when mine laborers were needed. Nowadays, most Chinese Thais live in metropolitan areas. The total number of native Chinese speakers in cities and towns is estimated to 3,621,000 people: 2,200,000 speak Teochew, Swatow, or Chaozhou; 580,000 speak Hakka; 379,000 speak Hainanese; 275,000 speak Cantonese; 150,000 speak Hokkian; and 37,000 speak Taiwanese. Mandarin speakers constitute a much smaller number, less than 1% in cities and towns. The largest Chinese group, which represents some 60% of the Sino-Thai population, is Chaozhou people, whose language is the most widely spoken Chinese dialect in Thailand. Interestingly, Mandarin is the second most important language, rather than Hakka, which is spoken by more people. This is because many educated Chinese descendants are encouraged to study Mandarin. The third most important languages are Hakka and Cantonese, followed by Hainanese, Hokkien and Taiwanese.

The Hakka people in Thailand have migrated mostly from Fujian and Guangdong and live in every part of the country, but particularly in Bangkok and in the southern provinces of Songkla and Yala. Hakka people in Thailand speak various dialects that reflect the speakers' original hometowns. These dialects may be classified into two dominant groups: the first Hakka group, or Chim Khak, speaks dialects from Meixian, Dabu, Xingning, Wuhua, and Pingyuan; the second group, or Pan San Khak, speaks dialects such as Hongsun and Jieyang, which are areas where the Chaozhou people live in China (Ungsitipoonporn 2007:2-4). The Chim Khak group lives in its own areas and communities and experiences less contact with other languages, whereas the Pan San Khak group lives together with the Chaozhou people. Due to the influence of the Chaozhou language, the Pan San Khak group, which includes Jieyang and Fengshun, contains unique linguistic features that distinguish it from the Chim Khak group in terms of pronunciation and lexicon.

3. Methodology

This study focuses on the Fengshun Hakka dialect spoken in Thailand; it is considered a Pan San Khak dialect. The Fengshun Hakka dialect was originally spoken in Meizhou, Guangdong province in China. Fengshun is the name of a county in Meizhou, which is situated near Jieyang and Chaozhou municipalities. In this study, the Fengshun Hakka data was collected first-hand from a questionnaire as well as from natural speaking occurrences of three Hakka speakers who live in Narathiwat province in southern Thailand.

The questionnaire contains basic sentences plus needed aspectual expressions. Most sample sentences were developed and adapted from Li and Thompson (1981), Yue-Hashimoto (1993), and Iwasaki and Ingkaphirom (2005). The sentences retrieved from those sources were chosen from the examples in aspect chapters to make sure that they are all relevant to aspects. The first informant was interviewed with the questionnaire along with the explanation of each sentence situation so that the informants could understand the context of each sentence. Due to the fact that the selected sentences might influence the outcome of the research, the natural speaking data were collected. For the collection of natural speaking data, sample topics for conversation and narration (such as a trip to Laos, a money transfer to China, and a Chinese festival) were suggested to the second and third informants. However, some conversations naturally occurred because the two informants live in the same house and often speak Hakka at home.

After collecting the Fengshun Hakka data in Thailand, Miaoli Hakka data was collected in Taiwan. The same questionnaire was used in the interview with the two Miaoli Hakka informants. The contents of conversations and narrations with the Thai informants were explained to the two Miaoli informants before staging interviews. Both Miaoli Hakka informants were asked to speak in their own Hakka dialect to present the same meanings.

4. Aspect and Situation Types

Comrie (1976), Bybee et al. (1994) and Smith (1991) agree that two major aspectual categories are comprised of Perfective and Imperfective, which are distinguished by an event boundary. Perfective is considered as a completed action that presents the beginning, middle, and end of a situation as a single whole, but Imperfective views a situation which is not completed or is in progress. However, there is another way to classify aspects, which is introduced by Chappell (1989a:96, 1989b:117). According to Chappell's aspect categorization, the situation is divided into bounded and unbounded aspects. The bounded aspect, which has a close meaning to Perfective aspect, refers to a situation defined with either a beginning or an endpoint; the unbounded aspect refers to a situation with no time limit or ongoing event. The Perfective versus Imperfective framework cannot neatly fit in some aspectual categories, such as Inchoative (the beginning of a new situation). In the bounded versus unbounded aspectual classification, the Inchoative aspect can be put into the bounded aspect group that not only encodes the end of an event or state but also the beginning of an event or state. Under this kind of aspectual categorization, the Inchoative aspect is classified as a bounded aspect.

It is inevitable that aspect is associated with situation types that depend on verbs and have a great influence on the occurrence of aspect markers. Vendler (1967:97-121) points out that verbs and time are closely related. In addition to tenses, verbs involve more subtle temporal meanings, since verbs imply time schemata that elaborate on each type of verb. He describes the time schemata of English verbs that process verbs and non-process verbs play an important role in determining four main time schemata: *activity*, *accomplishment*, *achievement*, and *state*. The contrast between telic and atelic, and between process and non-process, distinguishes those verb classes. The termination and completion of an action in telic situations distinguish *accomplishment* and *achievement* from *activity* and *state*; process and non-process distinguish *activity* and *accomplishment* from *achievement* and *state*.

In addition to the classic work on verb classes by Vendler (1967), Smith (1991:55) and Van Valin (2005:33) elaborate more on situation types by adding a new situation type, *semelfactive,* which refers to a punctual or instantaneous event without any result state, such as 'blink,' 'knock,' and 'sneeze.' According to Smith (1991:28), the situation types are categorized by the semantic properties, static versus dynamic, telic versus atelic, and durative versus instantaneous. States are viewed as a single static event while dynamics are relevant to the processes of an action. The telic and atelic situations are distinguished by the event termination which yields the outcome or goal of the event. Telic situations contain a natural endpoint, but

atelic situations do not indicate any terminal point of the event. The distinction between durative and instantaneous depends on whether or not the event lasts momentarily.

According to Smith's (1994:108-109) classification, five groups of situation types are classified by the above mentioned semantic contrasts: static versus dynamic, telic versus atelic, and durative versus instantaneous.

Activity:	dynamic, atelic, durative
	zǒu 'walk', *tīng* 'listen'
Accomplishment:	dynamic, telic, durative
	gài yīzuò qiáo 'build a bridge'
Semelfactive:	dynamic, atelic, instantaneous
	tī 'kick', *qiāo mén* 'knock at the door'
Achievement:	dynamic, telic, instantaneous
	dǎ-pò 'break', *shuì-zháo* 'fall asleep'
State:	static, durative
	cúnzài 'exist', *zhīdào* 'know'

The distinction of aspect and situations types have been discussed for many years. Comrie (1976:3) defines that 'aspects are different ways of viewing the internal temporal constituency of a situation.' The situation types here concern about the inherit meanings of verbs indicating situation characteristics—dynamic or stative, telic or atelic, durative or punctual (See also Smith 1991:28-33, Xiao and McEnery 2004:14). The situation types play an important role in possible aspectual interpretations. Vendler (1967:99) gives some examples of English verbs that are not allowed in continuous tenses; for example, 'I am knowing/loving/recognizing' is not an acceptable answer for the question 'What are you doing?', but 'I am running/writing/working' is acceptable. Such stative verbs—'know, love, or recognize'—are non-process or non-dynamic verbs which are not allowed to co-occur with continuous tense focusing on an ongoing process. Thus, in general, the stative verbs are not used in Progressive sense referring to a continuing situation. On the other hand, 'run, write, or work' have an inherit meaning of a process, so they are naturally found in Progressive situations.

The relationship between aspect and situation types is also found in Chinese languages. In Mandarin, the Perfective marker *le* regularly tells a completed situation (such as *chī le* = 'already ate'); however, the appearance of *le* following a state verb or an adjective introduces the beginning of a new state or Inchoative aspect (such as *piàoliàng le* = 'become beautiful'). It can be concluded that different types of verb allow different aspectual interpretations and different uses of aspect marker.

In this Hakka aspectual study, the concept of bounded and unbounded aspects by Chappell (1989a, b) and Smith's (1991, 1994) situation types are applied in the aspect classification and the co-occurrence of aspect markers with each verb class.

5. Perfective

The Perfective aspect refers to an event completed at some point of time. Under this aspectual category, there are some relevant terms, such as Resultative Verb Compound (RVC) and Current Relevant State (CRS). RVCs are adjectives or verbs that show the state resulted from the thoroughly complete action in the preceding verb. The completive and terminative notion is implied in RVCs. CRS is the term used by Li and Thompson (1981:240) whereby the sentence final particle *le* in Mandarin shows the current state relevant to a past action. In other words, CRS is similar to Perfect in English.

In Fengshun Hakka, the most common Perfective markers are $liau^{42}$ and $lɔ^{42}$ (5.1), Resultative Verb Compound (RVC) (5.2), and jiu^{33} (5.3).

5.1 $liau^{42}$ and $lɔ^{42}$

The Perfective marker $liau^{42}$ is assumed to have grammaticalized from $liau^{42}$ 'finish,' which is an RVC showing a result state after the completeness of an activity, such as sit^{55} $liau^{42}$ 'ate up' and mai^{11} $liau^{42}$ 'sold up.' The Perfective $liau^{42}$ can be variantly pronounced as $lɔ^{42}$. When $liau^{42}$ appears at the post-verbal

position, especially after *activity* verbs, it can function as either a Perfective marker or an RVC depending on the context. When *liau*^{42} functions as an RVC, it is never pronounced *lɔ*^{42}; only the Perfective *liau*^{42} has a variant *lɔ*^{42}.

(1) a: *sit*^{55} *liau*^{42} *maŋ*^{24}
eat RVC NEG
'Did you eat up?'
b: *sit*^{55} *liau*^{42}
eat RVC
'(I) ate up.'

(2) a: *sit*^{55} *maŋ*^{24}
eat NEG
'Did you eat?'
b: *sit*^{55} *liau*^{42}/*lɔ*^{42}
eat PFV
'(I) already ate.'

From the above examples, it should be noted that one informant tried to explain the difference between *liau*^{42} and *lɔ*^{42} by specifying that *liau*^{42} should be used only to refer to the completeness of an action or as an RVC while *lɔ*^{42} should be used as a Perfective marker denoting that the activity took place at some point in the past but did not emphasize the result of the complete action. In example (1), the consumption has already been completed thoroughly. All of the food had been eaten and no food had been left. In example (2), the food was eaten, but it does not mean that all of the food is gone. In fact, it was found in her natural speaking, and in other natural conversations among other informants, that *liau*^{42} is also used in Perfective notion, and at the same time *lɔ*^{42} can replace *liau*^{42} in Perfective sense. Either *liau*^{42} or *lɔ*^{42} yields Perfective aspect. Only full-form pronunciation *liau*^{42} can also serve as an RVC.

The Perfective *liau*^{42} or *lɔ*^{42} usually appears at the sentence-final position. In a single clause, the Perfective *liau*^{42} or *lɔ*^{42} is always found at the end of the sentence. In a complex sentence, the Perfective *liau*^{42} or *lɔ*^{42} appears at the end of the first clause. In summary, there are two common positions of *liau*^{42} or *lɔ*^{42} in the sentence: at the sentence-final in a single clause (5.1.1.) and at the clause-final in the independent clause in a complex sentence (5.1.2).

5.1.1 liau^{42}*/lɔ*^{42} *in a Single Clause*

The marker *liau*^{42}*/lɔ*^{42} always appears after a predicate. If a verb is followed by an object like in verb-object compounds, *liau*^{42}*/lɔ*^{42} is located after the object or the compounds. Li and Thompson (1981:73) indicate that the VO compound is the combination of a verb and its direct object and is considered as a bound morpheme. In Fengshun Hakka (also in Miaoli Hakka and Mandarin), there are a number of VO compounds, such as *wan*^{11} *tɕhian*^{24} ('pay' + 'money') 'make a payment,' *sɔi*^{11} *muk*^{11} ('sleep/lay' + 'eye') 'sleep' and *sit*^{55} *fan*^{11} ('eat' + 'rice') 'have a meal.' With the transitive verbs and VO compounds, the Perfective marker *liau*^{42}*/lɔ*^{42} mostly appears at the end of a sentence (see Examples (3) and (4)).

(3) *kɯ*^{24} *wan*^{11} *tɕhian*^{24} *liau*^{42}*/ lɔ*^{42}
3SG pay money PFV/CRS
'He/She made/has made a payment.'

(4) *kɯ*^{24} *sit*^{55} *fan*^{11} *liau*^{42}*/lɔ*^{42}
3SG eat rice PFV/CRS
'He/She had a meal/has had a meal.'

The sentence-final position of *liau*^{42} not only denotes Perfective notion but also Current Relevant State (CRS) at the same time. In those above examples (3) and (4), *liau*^{42}/*lɔ*^{42} presents a completed event and also expresses that the completed event yields some current relevance. In example (3), he/she already paid and now he/she does not have pay anymore. In example (4), he/she already had a meal and now he/she is not hungry. The CRS, similar to Anterior mentioned in Bybee et al (1994:54), is distinguished from Perfective. Perfective refers just to the completion of an event rather than implying any relevance of a past action and a current situation, whereas the CRS exhibits the persisting result of a complete action. Nonetheless, the Perfective and CRS naturally semantically overlap each other. According to the grammaticalization process mentioned in Bybee et al (1994:61), the Perfective is developed from Anterior or CRS, so there is an area that both still share. The loss of current relevance in CRS leads to the interpretation of a past action or Perfective event. Hence, it is not unusual that *liau*^{42} can mark both CRS and Perfective. In Fengshun Hakka, the Perfective and CRS are not visibly distinguished from each other, especially if no specific time of an event is mentioned.

It can be clearly seen that the position of Perfective *liau*^{42}/*lɔ*^{42} in examples (3) and (4) differs from that of Perfective *le* in Mandarin, which can be placed right after the verb and before the object (between verb and object) as shown in Mandarin examples (5) and (6).

(5) *tā* *fù* *le* *qián*
3SG pay PFV money
'He/She made a payment.'

(6) *tā* *chī* *le* *fàn*
3SG eat PFV rice
'He/She had a meal.'

In terms of co-occurrence with verb classes, the post-verbal *liau*^{42}/*lɔ*^{42} in a single clause, which illustrates a completed event, can occur with both telic and atelic situations. The co-occurrence with *state* verbs will be discussed later in the Inchoative aspect section. The following examples will show the co-occurrence of Perfective *liau*^{42}/*lɔ*^{42} with *activity*, *accomplishment*, *achievement,* and *semelfactive*.

When *activity* verbs such as 'write' and *accomplishment* verbs such as 'drink a tea,' which contain the semantic features of durative and dynamic situations, occur with Perfective *liau*^{42} or *lɔ*^{42}, they are sometimes followed by phrases denoting quantity and frequency or duration of the action. An expression denoting quantity and frequency or duration indicates how long an event lasted. In the examples (7) and (8), the number of object (7) and that of hour (8) limit the duration of the situations.

(7) *kɯ*^{24} *sia*^{42} *kai*^{11} *kai*^{42} *tɕhɯ*^{11} *tɕhɔ*^{42} *liau*^{42}
3SG write that CL word wrong PFV
'He/She wrote that word wrongly.'

(8) *sɯ*^{42} *tɕha*^{33} *jit*^{11} *tiam*^{42}*tɕuŋ*^{33} *lɔ*^{42}
drive car one hour PFV
'(I) drove for an hour.'

The *achievement* verb class is telic and instantaneous in itself, such as 'wake up' and 'die.' Perfective *liau*^{42}/*lɔ*^{42} is also compatible with *achievement* verbs to express that the event is completed.

(9) *ŋai*^{11} *thɛu*^{11}*khi*^{42} *lɔ*^{42}
1SG get up PFV
'I already got up.'

(10) *khiu*^{11} *nian*^{24} *kɯ*^{24} *sɯ*^{42} *khɯ*^{42} *lɔ*^{42}
last year 3SG die go PFV
'He/She died last year.'

Another instantaneous situation is the *semelfactive* situation, which is a punctual event without any result state, such as 'cough' and 'jump.' Mostly a phrase defining duration of a situation is added to present how long the action had been running. In the example (11) below, the speaker coughed all night and now she does not cough anymore

(11) *ŋai*^{11} *ʨhuk*^{11} *wan*^{24} *ya*^{11} *lɔ*^{42}
1SG cough whole night PFV
'I coughed a whole night.'

In summary, *liau*^{42} in the Fengshun dialect can function as an RVC 'finish' and a Perfective marker. The appearance of *liau*^{42} in a single clause can be considered as either Perfective or RVC. The reduced form of *liau*^{42}, which is *lɔ*^{42}, shows only Perfective or CRS but cannot refer to event completeness.

It is found that a Perfective marker in Chinese languages such Mandarin, Jieyang Chaozhou, and Miaoli Hakka is situated right after a verb when a phase denoting quantity, frequency, or duration of an action is shown up. However, the appearance of the Perfective *liau*^{42} at the final position of a predicate is usually found only in Fengshun Hakka, different from the position of Perfective markers in Mandarin, Jieyang Chaozhou and Miaoli Hakka.

(12) Fengshun Hakka
ŋai^{11} *suŋ*^{42} *liɔŋ*^{42} *sam*^{33} *fuŋ*^{33} *sin*^{42} *pun*^{33} *ŋi*^{11} *liau*^{42}
1SG send two three CL letter give 2SG PFV
'I sent you a couple of letters.'

(13) Miaoli Hakka
ŋai^{11} *ki*^{55} *e*^{11} *lioŋ*^{31} *sam*$^{24\text{-}11}$ *foŋ*^{55} *ɕin*^{55}*e*^{11} *pun*^{24} *ṇ*^{11}
1SG send PFV two three CL letter give 2SG
'I sent you a couple of letters.'

(14) Jieyang Chaozhou
ua^{53} *taŋ*$^{53\text{-}35}$ *liau*$^{53\text{-}35}$ *puã*$^{213\text{-}53}$ *kai*$^{55\text{-}11}$ *tseŋ*^{33} *t'au*^{55}
1SG wait PFV half CL hour
'I waited for half an hour.' (Xu 2007:126)

(15) Mandarin
wǒ *děng* *le* *bàn* *ge* *xǐaoshí*
1SG wait PFV half CL hour
'I waited for half an hour.'

The clause-final position of the Perfective marker *liau*^{42} in Fengshun Hakka such as (12) may be resulted from the language contact with Thai. In Thai word order, the Perfective marker *lέɛw* is allowed to appear at the clause-final position whether a quantifier phrase is added or not.

(16) Thai
kin *khâw* *sǎam* *caan* *lέɛw*
eat rice three CL PFV
'I ate three dishes of rice.'

5.1.2 liau^{42}/lɔ^{42} in a Complex Sentence

In complex sentences, *liau^{42}/lɔ^{42}* occurs at the end of the first clause in a complex sentence to signal the sequence of events. That is, the event in a clause followed by *liau^{42}/lɔ^{42}* has happened before another event. Boundedness of the first event is noted by the Perfective *liau^{42}/lɔ^{42}*, which tells the order of two successive events.

(17) *sam^{33}* *khu^{42}* *tɕhau^{33}* *liau^{42},* *tɕhiu^{11}* *na^{33}* *lɔi^{24}* *tɕap^{11}*
shirt pants dry PFV, then take come fold
'When the clothes have dried, then fold them.'

(18) *ku^{24}* *tɕhut^{11}* *khu^{42}* *liau^{42} ,* *ŋai^{11}* *ham^{42}* *ŋa^{33}* *tɕhit^{55}* *kiak^{11}kiak^{11}*
3SG out go PFV, 1SG call POSS niece quickly
sɔ^{42} mun^{11}
lock door
'When she got out, I called my niece to quickly lock the door.'

The Perfective *liau^{42}* in this Hakka dialect is similar to *liau^{53}* in Jieyang Chaozhou and *le* in Mandarin. For instance, the Perfective *liau^{53}* in Jieyang Chaozhou (Xu 2007:129) also appears at the clause final position in the first clause of a complex sentence to express the boundedness of the first event. In Mandarin (Li and Thompson 1981:198), the verbal *le* is used to express the sequence of the events.

(19) Jieyang Chaozhou
ua^{53} *tsok$^{5-2}$* *tsaʔ^{2}-ŋiap$^{5-2}$* *liau^{53} $_{sub}$* *tsiã$^{213-53}$* *k'ɯ$^{213-53}$* *haʔ$^{5-2}$hau^{35}*
1SG do homework PFV then go school
'I will go to school after doing my homework.' (Xu 2007:129)

(20) Mandarin
wǒ *chī* *wán* *le* *nǐ* *chī*
1SG eat finish PFV 2SG eat
'After I have finished eating, then you eat.' (Li and Thompson 1981:198)

Similarly, the Miaoli Hakka Perfective marker *e^{11}* or *le^{11}* appears at the end of the first clause in a complex sentence to show the sequence of events, as illustrated in (21).

(21) Miaoli Hakka
sui^{31} *kun^{31}* *e^{11},* *tsaŋ^{55}* *pioŋ^{55}* *tsu^{24}* *ŋiuk^{2}*
water boil PFV, then put pork meat
'The water is boiled, then add pork.'

5.1.3 liau^{42}/lɔ^{42} and Past Tense

Tense is used to specify 'the relationship between the event frame and a reference locus in time' (Chung and Timberlake 1985:213). Past tense refers to an event that occurred before the moment of speaking. Semantically, past tense and Perfective point out the completion of an action in the past and signal the sequence of events in narration. Moreover, they both share the same lexical sources and develop through the anteriority stage (Bybee et.al. 1994:82). With the close relationship between past tense and Perfective, it seems possible that a Perfective marker can denote the notion of past tense.

However, a Perfective marker in many Chinese languages should not be classified as a past tense marker because it can express completion in the future, and the presence of Perfective marker is not always needed to note past tense. Li and Thompson (1981:213) do not define the Perfective *le* in Mandarin as a past tense marker. They indicate that *le* is found in many sentences that are not associated with past tense, such as in imperative sentences, as in (22) and even in sentences referring to future (23). Also, many sentences

without *le* note past events, as in (24). The following Mandarin examples are retrieved from Li and Thompson (1981:213-214).

(22) *hē* *le* *tā*
drink PFV 3SG
'Drink it.'

(23) *míngtiān* *wǒ* *jiù* *kāichú* *le* *tā*
tomorrow 1SG then expel PFV 3SG
'I will expel him/her tomorrow!'

(24) *zuótiān* *tā* *tiào* *zài* *chuáng* *shàng*
yesterday 3SG jump at bed on
'Yesterday, he/she jumped onto the bed.'

Xu (2007:131) points out that the Perfective *liau*^{53} in Jieyang Chaozhou does not always encode past tense. In a negative sentence, the Perfective *liau*^{53} is not allowed due to its incompatibility with an event that has not occurred, as shown in (25). Furthermore, *liau*^{53} is not always needed when an adverb of time appears to encode a past action, as shown in (26).

(25) **ua*^{53} *bo*$^{55\text{-}11}$ *t'oi*$^{53\text{-}35}$ *liau*$^{53\text{-}35}$ *tsi*$^{53\text{-}35}$ *puŋ*$^{53\text{-}55}$ *tsɯ*^{33}
1SG NEG read PFV this CL book
'I did not read this book.'

(26) *ua*^{53} *hoʔ*^{2} *tsa*^{53} *tsu*$^{35\text{-}21}$ *ts'e*^{53}
1SG very early then wake up
'I woke up very early.'

Similar to Mandarin and Jieyang Chaozhou, the Perfective *liau*^{42}/*lɔ*^{42} in Fengshun Hakka is not always used to denote a past event. There are several times when the Perfective *liau*^{42}/*lɔ*^{42} appears in a sentence noting future completion. Neither does it occur in a negative sentence that an event has not happened. When a phrase defining the time of a past action is added, the Perfective *liau*^{42}/*lɔ*^{42} is not required.

(27) *thian*^{33}*kuɔŋ*^{33}*ŋit*^{11} *ŋai*^{11} *tɕɔn*^{42} *wuk*^{11}*kha*^{33} *liau*^{42}
tomorrow 1SG return house PFV
'I will arrive home tomorrow.'

(28) **ŋai*^{11} *mɔ*^{24} *ta*^{42} *mun*^{11} *liau*^{42}
1SG NEG hit door PFV
'I did not knock the door.'

(29) *tɕhian*^{11}*pun*^{24}*ŋit*^{11} *kɯ*^{24} *lɔi*^{24} *thɛt*^{11} *man*^{11}
yesterday 3SG come over late
'Yesterday, he/she came late.'

As shown in the examples above, the Perfective *liau*^{42}/*lɔ*^{42} cannot be considered a past tense marker since it is used in many situations that are not relevant to an event in the past. Temporal adverbials or phrases specifying the actual time an event occurred can provide the notion of past tense regardless of the presence of the Perfective marker *liau*^{42}/*lɔ*^{42}.

5.2 Resultative Verb Compound (RVC)

Resultative verb compounds, or RVCs, usually are adjectives or verbs following a verb and indicate the state resulting from the thoroughly complete action in the preceding verb. The fact that RVCs express the resulting state of the prior verb yields the completive and terminative meaning and leads to event boundedness. The Perfective aspect, thus, can be provided by RVCs.

Some researchers (e.g. Smith 1991:344; Chao 1968:446) classify Mandarin RVCs as morphemes or phrase markers relevant to Perfective aspect. For example, in Mandarin, RVCs such as *wán*, *hǎo* and *guò* are said to be optional morphemes that can infer Perfective viewpoint by showing completion and termination.

(30) Mandarin

wǒ	*zuótiān*	*xiě-wán-le*	*yī*	*fēng*	*xìn*
1SG	yesterday	write-RVC-*le*	one	CL	letter

'Yesterday I wrote a letter (and finished).' (Smith 1994:113)

In Jieyang Chaozhou (Xu 2007:132), RVCs such as *liau*^{53} 'finish,' *ho*^{53} 'good/finish,' *tek*^{5} 'straight/finish' and *uaŋ*^{55} 'complete/finish' indicate completive and terminative meanings, and express Perfective aspect. Each RVC has a specific inherent meaning that delimits its co-occurrence with some kinds of verbs.

(31) Jieyang Chaozhou

ko$^{213\text{-}53}$	*iam*^{55}	*eŋ*^{11}	*liau*^{53} / *uaŋ*^{55}	*lau*$^{53\text{-}213}$
CL	salt	use	RVC	CRS

'(We) have used up the salt.' (Xu 2007:133)

In this research, I also treat RVCs in Fengshun Hakka as morphemes relevant to Perfective aspect that denote completive and and terminative notions to an event. Verbs such as *lɔi*^{24} 'come,' *khɯ*^{42} 'go,' and *tɔ*^{42} 'arrive' as well as adjectives such as *hɔ*^{42} 'good' are used as RVCs situated at the post-verbal position. These RVCs indicate that the action of the verb is finished and results in the current state. For example, *ʨha*^{33} *sɯ*^{24} *li*^{11} *hɔ*^{42} 'car repair good' means the car has already been repaired and now it is in good condition; *thai*^{11} *lɔi*^{24} 'grown up come' means a child has grown up, and is not a child anymore; *mai*^{11} *khɯ*^{42} 'sell go' means something has been sold and it is not with the agent anymore; *ʨɔn*^{42} *tɔ*^{42} 'return arrive' means the action of returning is finished and the agent has arrived at the destination. As mentioned earlier, the RVC *liau*^{42} 'finish' demonstrates the completeness of an event. Although all these RVCs share the same meaning of completion or 'finish,' each of them is compatible with different verbs due to their inherent meanings.

The RVCs discussed in the following sections are *lɔi*^{24} 'come,' *khɯ*^{42} 'go,' *hɔ*^{42} 'good,' and *tɔ*^{42} 'arrive.'

5.2.1 lɔi^{24} *'come'*

The directional verb 'come' is a common lexical source of anteriors, which then can develop to code Perfective aspect (Bybee et al. 1994:56). For instance, the verb 'come' in some Wu dialects can express Perfective aspect at the post-verbal position (Lamarre 2001:109). Likewise, in Fengshun Hakka, the post-verbal *lɔi*^{24} 'come' denotes an event that has happened at some point in the past before the speech moment. In (32), the RVC *lɔi*^{24} indicates where the referent went. If the RVC *lɔi*^{24} is replaced by the Perfective *liau*^{42}/*lɔ*^{42}, the meaning in (32) changes to where the referent goes. That is, someone or something has disappeared, then another person asks where he/she/it has gone. The RVC *lɔi*^{24} in example (33) notes that the referent went to a competition and now he/she is back. The RVC *lɔi*^{24} in (34) is used to encode the sequence of events in a complex sentence like *liau*^{42}/*lɔ*^{42} in a complex sentence.

(32)

khɯ^{42}	*nai*^{42}	*lɔi*^{24}
go	where	RVC

'Where did you go?'

(33) *kɯ*^{24} *khɯ*^{42} *pi*^{42}*sɔi*^{42} *lɔi*^{24} *liau*^{42}
3SG go compete RVC CRS
'He/She has already joined a competition.'

(34) *kɯ*^{24} *thai*^{11} *lɔi*^{24} *wa*^{11} *m̩*^{11} *thɛn*^{33}
3SG grown up RVC say NEG listen
'He has grown up and he does not listen to me anymore.'

The verb 'come' denoting a completed event is not only found in Hakka and Wu, it is also widely used in some Southeast Asian languages such as Thai (35) and Mon (36). *maa* in Thai and *klɤŋ* in Mon which mean 'come' express the same aspectual meaning as the verb *lɔi*^{24} 'come' in Fengshun Hakka.

(35) Thai
pay *nǎy* *maa*
go where RVC
'Where did you go?'

(36) Mon
a *lɔ* *klɤŋ*
go where RVC
'Where did you go?'

5.2.2 khɯ^{42} *'go'*
The RVC *khɯ*^{42} 'go' in Fengshun Hakka also denotes completive notion and Perfective aspect. As an RVC, *khɯ*^{42} signals the disappearance of something after the action is completed. Its appearance with certain verbs, such as *activity* and *accomplishment*, shows that an action is done and the object disappears.

(37) *ŋai*^{11} *sit*^{55} *khɯ*^{42} *liɔŋ*^{42} *liap*^{55} *jɔk*^{55}
1SG eat RVC two CL medicine
'I have eaten two tablets of medicine.'

(38) *ŋai*^{11} *jim*^{42} *khɯ*^{42} *lɔ*^{42}
1SG drink RVC CRS
'I have drunk it.'

Note that the RVC *khɯ*^{42} 'go' emphasizes on the disappearance of object after an action is completed, but the RVC *lɔi*^{24} 'come' only marks a completed situation happened before the time of speaking, whether or not something is gone. In other words, the RVC *khɯ*^{42} 'go' will be spoken when something must be gone or changes its state after a completed action, so *khɯ*^{42} 'go' cannot replace *lɔi*^{24} 'come' in the examples (32), (33), and (34). However, *lɔi*^{24} 'come' can replace *khɯ*^{42} 'go' in the example (37) and (38) since *lɔi*^{24} and *khɯ*^{42} share the same meaning of event completion but *lɔi*^{24} does not focus on the disappearance of object.

5.2.3 tɔ^{42} *'arrive'*
The RVC *tɔ*^{42} 'arrive' points out that the process of doing something is completed and has reached the goal when it occurs with *activity* situation types. If it appears with perception verbs, such as *thɛn*^{33} *tɔ*^{42} 'listen arrive' and *khɔn*^{42} *tɔ*^{42} 'look/watch arrive,' it encodes the occurrence of perception state.

(39) *ŋai*^{11} *ʨhim*^{11} *tɔ*^{42} *ʨhian*^{24} *lɔ*^{42}
1SG search RVC money CRS
'I have found money.'

(40) *khɔn*^{42} *tɔ*^{42} *ʔɔi*^{42} *ʔɛw*^{42}
look/watch arrive want throw up
'When I saw it, I wanted to throw up'

The RVC *tɔ*^{42} also occurs with other action verbs such as 'walk' and 'return,' which denote that the goal of the action is completed.

(41) *kɯ*^{24} *haŋ*^{11} *tɔ*^{42} *hɔk*^{55}*thɔŋ*^{11} *liau*^{42}
3SG walk RVC school CRS
'He/She has arrived school by walk.'

(42) *kɯ*^{24} *tɕɔn*^{42} *tɔ*^{42} *wuk*^{11}*kha*^{33} *liau*^{42}
3SG return RVC home CRS
'He/She has returned home.'

5.2.4 *hɔ*^{42} 'good'

The RVC *hɔ*^{42} 'good' at the post-verbal position, normally following *activity* verbs, denotes both the completion and success of an action. The post-verbal *hɔ*^{42} is considered as an RVC expressing the completed action, while the pre-verbal *hɔ*^{42} is a modal 'should.' In general, *hɔ*^{42} is found as an adjective meaning 'good,' as in *hɔ*^{42} *ŋin*^{11} 'good person,' as well as an adverb meaning 'well, very,' such as *hɔ*^{42} *siak*^{11} 'very cute.' At the pre-verbal position, it sometimes serves as a modal indicating the possibility or ability to do something and give a suggestion.

hɔ^{42} *juŋ*^{11}
good use
'can be used'

hɔ^{42} *tɕhut*^{11} *khɯ*^{42}
good exit go
'should get out'

m̩^{11} *hɔ*^{42} *khɔi*^{33}
NEG good open
'should not open'

As the post-verbal position, *hɔ*^{42} is known as an RVC and indicates that an event is successfully done. In the following example (43), the RVC *hɔ*^{42} can be replaced by the RVC *liau*^{42} 'finish' since they both share the meaning of completion. Nevertheless,the RVC *liau*^{42} 'finish' also holds the meaning of 'something has completely gone.' The verbs that usually co-occur with the RVC *liau*^{42} are the *activity* verbs that require an object. When verbs such as *sit*^{55} 'eat,' *jim*^{42} 'drink,' and *juŋ*^{11} 'use' appear with *liau*^{42} 'finish,' they also imply 'eat up,' 'drink up,' and 'use up,' apart from the notion of complete action. After an action such as eating, drinking, and using, something completely disappears. That is, the whole objects disappear after the action. On the other hand, the occurrence of such verbs with *hɔ*^{42} signals only the termination of an action. In (44), the whole amount of food is already consumed, but in (45), the action of eating is terminated without the implication of the disappearance of food.

(43) *ta*^{42} *pan*^{42} *hɔ*^{42}/*liau*^{42} *ŋi*^{11} *hɔ*^{42} *tɕut*^{11} *khɯ*^{42} *liau*^{42}
dress RVC 2SG good out go CRS
'When you have already put on clothes, you should go out.'

(44) *sit*^{55} *liau*^{42} *lɔ*^{42}
eat RVC CRS
'(I) have eaten up.'

(45) *sit*^{55} *hɔ*^{42} *lɔ*^{42}
eat RVC CRS
'(I) have finished eating.'

The RVC *liau*^{42} is unique in the Pan San Khak group. In Jieyang Hakka (Shiwaruangrote 2008:230), a Pan San Khak dialect, *liau*^{31} can present a resultative state, similar to *liau*^{42} in Fengshun Hakka. It is assumed that the RVC *liau*^{42} in the Pan San Khak group has evolved to a Perfective marker. In the Miaoli dialect spoken in Taiwan, the verb *thet*^{2} 'kick' is used as an RVC while *e*^{11} or *le*^{11} serves as a Perfective marker. Therefore, *sɪt*^{5} *thet*^{2} *e*^{11} in Miaoli dialect means '(I) have eaten up' and is equivalent to *sit*^{55} *liau*^{42} *lɔ*^{42} in Fengshun dialect. Since the Fengshun dialect is a Hakka dialect spoken in the Chaozhou area, it is possible that *liau*^{42} 'finish' as an RVC and a Perfective marker has been influenced by the Chaozhou language. As shown in the example of RVC in Jieyang Chaozhou in (31), the function of the RVC *liau*^{42} in Fengshun Hakka is similar to that of the RVC *liau*^{53} in Jieyang Chaozhou. The Perfective marker *liau*^{42} in Fengshun Hakka and *liau*^{53} in Jieyang Chaozhou were grammaticalized from their RVC 'finish'. The word *thɛt*^{11} in Fengshun Hakka also exists but only means 'kick' and cannot serve as an RVC to refer to the completeness of a situation.

All in all, it can be clearly seen that RVCs can express Perfective aspect without the appearance of a Perfective marker. The appearance of *liau*^{42} or *lɔ*^{42} at the end of sentence, instead, notes the current relevant state. Perfective meaning is already exhibited by RVCs.

5.3 jiu^{33} *'have'*

The verb *jiu*^{33} 'have,' a verb of possession and existence, can signal Perfective aspect. As an auxiliary, it is situated before the verb phrase and notes that the action is completed at some point prior to the time of speaking. The negator of *jiu*^{33}, *mɔ*^{24} 'not have' is used to denote the negation of a past event and also serves as a question marker in a Perfective situation.

(46) a: *ku*^{24} *jiu*^{33} *ta*^{42}*pan*^{42} *mɔ*^{24}
3SG PFV dress NEG
'Did she get dressed?'
b: *jiu*^{33}
PFV
'Yes, she did.'

Although *liau*^{42} and *jiu*^{33} are similar in terms of Perfective illustration, *liau*^{42} and *jiu*^{33} are not exactly the same in all functions. The first and most noticeable difference is that *liau*^{42} is post-verbal whereas *jiu*^{33} is pre-verbal. Second, the appearance of *jiu*^{33} not only expresses Perfective aspect, but also highlights that an action really took place or affirmative. The affirmative function of *jiu*^{33} distinguishes *jiu*^{33} from the Perfective marker *liau*^{42} because the affirmative sense is not implied in *liau*^{42}. Another interesting fact is that the Perfective marker *liau*^{42} is not allowed in negative sentences. Only the negator *mɔ*^{24} 'not have' is used to negate a Perfective event.

Like Southern Min spoken in Taiwan, the verb 'have' is also used to express a completed action. Chappell (1989b:119) states that Southern Min uses the verb 'have' to show affirmative notion in a future context and also Perfect notion. Apart from these notions, it also expresses event completion or a past event. Kubler (1982:162) also mentions that the appearance of the verb 'have' in Taiwanese Mandarin is the result of the influence of Southern Min. In this paper, only examples of aspectual Perfect situation and completed action are identified and discussed.

(47) Southern Min

û	*k'ì*	*Pāk'kiã*
have	go	Beijing

'(They)'ve gone to Beijing.' (Chappell 1989b:119)

(48) Southern Min

lì	*û*	*k'uã:kĩ:*	*ī*	*buê?*
2SG	PFV	see	3sg	NEG

'Did you see her?' (Kubler 1982:162)

The verb *û* in Southern Min in (47) and (48) presents a past action that was already completed, so it provides the Perfective notion. In Taiwanese Hakka, such as the Miaoli dialect, the verb iu^{24} 'have' preceding VP also provides Perfective meaning as illustrated in (49).

(49) Miaoli Hakka

a:	ki^{11}	$toŋ^{24\text{-}11}tsu^{55}$	$iu^{24\text{-}11}$	$soi^{55}muk^{2}$	mo^{11}
	3SG	at noon	PFV	sleep	NEG

'Did he take a nap at noon?'

b:	ki^{11}	$iu^{24\text{-}11}$	$soi^{55}muk^{2}$ /	ki^{11}	mo^{11}	$soi^{55}muk^{2}$
	he/she	PFV	sleep /	he/she	NEG	sleep

'Yes, he did/ No, he didn't.'

Sometimes the verb iu^{24} is used to express Experiential aspect when it co-occurs with Experiential marker ko^{33}. This is a difference between Fengshun Hakka and the Miaoli dialect. The verb jiu^{33} 'have' in Fengshun Hakka can only express Perfective aspect, but the verb iu^{24} in the Miaoli dialect can denote both Perfective and Experiential aspects. This difference will be discussed later in the following section.

6. Experiential

Experiential aspect, which is called Experiential Perfect in Comrie (1976:58), refers to an experience or action which was done in the past and has become an experience of the agent. Experiential aspect in Fengshun Hakka can be presented by two markers, $kuɔ^{42}$ 'pass' and sit^{11} 'know,' in three syntactic structures:

verb + $kuɔ^{42}$
sit^{11} + verb
sit^{11} + verb + $kuɔ^{42}$

6.1 $kuɔ^{42}$ 'pass'

The marker $kuɔ^{42}$ 'pass or cross' is a movement verb that is used as an Experiential aspect marker indicating an action or event has been experienced at least once. The Experiential marker $kuɔ^{42}$ is situated at the post-verbal position, like *guò* in Mandarin.

(50) Fengshun Hakka

$ŋai^{11}$	$khɯ^{42}$	$kuɔ^{42}$	$thɔŋ^{11}san^{33}$
1SG	go	EXP	China

'I have been to China.'

(51) Mandarin

wǒ	*qù*	*guò*	*zhōngguó*
1SG	go	EXP	China

'I have been to China.'

In terms of co-occurrence with verbs, the marker *kuɔ*^{42} in Fengshun Hakka mostly appears with dynamic verbs such as *activity*, *semelfactive*, and *accomplishment* situation types, but it will not follow *state* and *achievement* verbs and RVCs as shown in (52) and (53).

(52) **ŋai*^{11} *ʔɔi*^{42} *kuɔ*^{42} *kɯ*^{24}
1SG love EXP 3SG
'I used to love him/her.'

(53) **ŋai*^{11} *thɛu*^{11}*khi*^{42} *kuɔ*^{42}
1SG get up EXP
'I used to get up.'

6.2 sit*^{11} *'know'

The marker *sit*^{11} 'know' also denotes an action or event that has been done in the past. Different from the post-verbal *kuɔ*^{42}, *sit*^{11} is placed at pre-verbal position. The marker *sit*^{11} expresses the same Experiential meaning as *kuɔ*^{42} does, despite their different position.

(54) *ŋai*^{11} *sit*^{11} *khɯ*^{42} *thɔŋ*^{11}*san*^{33}
1SG EXP go China
'I have been to China.'

In contrast to *kuɔ*^{42}, the marker *sit*^{11} can be used with *state* and *achievement* verbs (55) and then yields the meaning of 'used to.'

(55) *ŋai*^{11} *sit*^{11} *ʔɔi*^{42} *kɯ*^{24}
1SG EXP love him/her
'I used to love him/her.'

6.3 sit*^{11} *'know' and kuɔ*^{42} *'pass'

Both *sit*^{11} and *kuɔ*^{42} can appear together in the same sentence to denote Experiential aspect.

(56) *ŋai*^{11} *sit*^{11} *khɯ*^{42} *kuɔ*^{42} *thɔŋ*^{11}*san*^{33}
1SG EXP go EXP China
'I have been to China.'

All three structures express the same meaning of having done something in the past. In brief, the Experiential aspect in the Fengshun dialect can be expressed by either *sit*^{11} or *kuɔ*^{42}. The presence of both markers is also acceptable and still indicates a past experience.

As mentioned earlier concerning another Hakka dialect spoken in Taiwan, the Experiential aspect in Miaoli can be expressed by the existential verb *iu*^{24} 'have,' yet in Fengshun Hakka, only *kuɔ*^{42} and *sit*^{11} can encode the Experiential aspect. In the Miaoli dialect, three markers denote Experiential: *ko*^{55}, *sɪt*^{5} and *iu*^{24}. The markers *ko*^{55} and *sɪt*^{5} are used in the same way as the markers *kuɔ*^{42} and *sit*^{11} in the Fengshun dialect. The verb *iu*^{24} in Miaoli Hakka can denote the Experiential aspect if the Experiential marker *ko*^{55} follows a verb, see (57). Without the appearance of *ko*^{33}, the verb *iu*^{33} contributes Perfective meaning, as shown in (58).

(57) Miaoli Hakka
n̩^{11} *iu*$^{24\text{-}11}$ *hi*^{55} *ko*^{55} *pet*^{2}*kin*^{24} *mo*^{11}
2SG EXP go EXP Beijing NEG
'Have you been to Beijing?'

(58) Miaoli Hakka

ṇ^{11}	*iu*$^{24\text{-}11}$	*hi*^{55}	*pet*^{2}*kin*^{24}	*mo*^{11}
2SG	PFV	go	Beijing	NEG

'Did you go to Beijing?'

Kubler (1982:162) and Chappell (1989b:119) indicate that due to the influence of Southern Min, the verb 'have' usually is found in Taiwanese Mandarin to express a completed action. Thus, the use of the verb 'have' to note Experiential aspect in Miaoli Hakka might also have been influenced by Southern Min, which is the mother language of the Taiwanese majority and also uses the verb 'have' to refer to a past event. That illustrates the difference in the Experiential aspect between Fengshun Hakka and Miaoli Hakka.

7. Inchoative

Inchoative is a term used to explain the change of state and the continuation of a new state. Comrie (1976:19) calls it Ingressive, which refers to the inception of a new state. In Fengshun dialect, Inchoative aspect is shown when adjectives or *state* verbs are followed by *liau*^{42} 'finish.' The Perfective marker *liau*^{42} or *lɔ*^{42} following *activity* verbs usually refers to bounded event or Perfective aspect, but those following *state* verbs or adjectives indicate that there is a complete change of state and a new state has been suggested. In (59), the agent was unclear about something but now understands it. In (60), the agent's look has changed from plain to beautiful, and *lɔ*^{42} at the sentence-final position is considered a sentence-final particle to express some annoyance of the speaker who did not want to wait anymore.

(59)

ŋai^{11}	*min*^{11}*phak*^{55}	*liau*^{42}
1SG	understand	INC

'I have understood.'

(60)

kiau^{42}*liaŋ*^{42}	*lɔ*^{42},	*ṃ*	*sɯ*^{42}	*ta*^{42}*pan*^{42}	*lɔ*^{42}
beautiful	INC,	NEG	have to	dress	SFP

'You have become beautiful, so you don't have to dress up anymore.'

Like *liau*^{42}/*lɔ*^{42} in Fengshun dialect, *e*^{11}/*le*^{11} in Miaoli dialect also allows Inchoative aspect at the post-verbal position of *state* verbs or adjectives. As shown in (61), a new state of a person has been achieved: the color of of his/her face has changed.

(61) Miaoli Hakka

kia^{11}	*mian*^{55}	*fuŋ*^{11}	*e*^{11}/*le*^{11}
POSS	face	red	INC

'His/Her face has turned red.'

8. Conclusion and Discussion

This preliminary analysis of the bounded aspects in Fengshun Hakka spoken in Thailand has exhibited both syntactic and semantic functions of Perfective, Experiential, and Inchoative aspects. In some constructions, the comparison among the Miaoli Hakka dialect, Mandarin, and Jieyang Chaozhou provides an overview of those three bounded aspects in Chinese languages. It has also been shown that Perfective, Experiential, and Inchoative aspects in Fengshun Hakka share some features with those in Chaozhou and in Miaoli Hakka. The following table summarizes the comparisons of the three bounded aspects in the Fengshun dialect and the Miaoli Hakka.

Bounded Aspects	**Fengshun Hakka**	**Miaoli Hakka**
Perfective	*liau*^{42}/*lɔ*^{42} 'finish'	*e*^{11}/*le*^{11} 'finish'
	jiu^{33} 'have'	*iu*^{24} 'have'
	Resultative Verb Compounds: *lɔi*^{24} 'come,' *khɯ*^{42} 'go,' *tɔ*^{42} 'arrive,' *hɔ*^{42} 'good,' *liau*^{42} 'finish'	Resultative Verb Compounds: *loi*^{55} 'come,' *hi*^{55} 'go,' *to*^{55} 'arrive,' *ho*^{55} 'good,' *thet*^{2} 'finish'
Experiential	*kuɔ*^{42} 'pass'	*ko*^{55} 'pass'
	sit^{11} 'know'	*sit*^{2} 'know'
	sit^{11} + Verb + *kuɔ*^{42}	*sit*^{2} / *iu*^{24} + Verb + *ko*^{55}
Inchoative	*liau*^{42}/*lɔ*^{42} 'finish'	*e*^{11}/*le*^{11} 'finish'

Although many of the bounded aspect markers in both Fengshun Hakka and Miaoli Hakka are used similarly, differences between the two Hakka dialects can be noted.

First, *thet*^{2}, which means 'finish' in Miaoli, serves as an RVC to indicate event completeness, yet in Fengshun Hakka, *liau*^{42}, a Perfective marker, is also used as an RVC showing event completeness just as *thet*^{2} in Miaoli does. Because of the close contact with the Chaozhou language, e.g. the Jieyang Chaozhou, which also uses *liau*^{53} as an RVC, Fengshun Hakka uses *liau*^{42} to code the completeness of a situation rather than *thet*^{2}.

Second, in the presence of a quantifier phrase, the Perfective marker *e*^{11}/*le*^{11} in Miaoli is situated in post-verbal position, but the Perfective marker *liau*^{42}/*lɔ*^{42} in Fengshun Hakka appears after the quantifier phrase. The occurrence of *liau*^{42}/*lɔ*^{42} in Fengshun Hakka at clause-final position could possibly be influenced by the Thai Perfective Marker *lέεw*, which can locate after a quantifier phrase.

Another interesting feature of *e*^{11} in Miaoli Hakka is the fact that *e*^{11} is used as a sentence-final particle (SFP) more often than *liau*^{42}/*lɔ*^{42} in Fengshun Hakka. In addition to denoting the current relevant state, the SFP *e*^{11} signals the closure of a sentence. That is, the appearance of the SFP *e*^{11} makes a sentence sound complete and the speaker does not want to say anything further.

(62) Miaoli Hakka

sui^{31}	*kun*^{31}	*ten*^{55}	*e*^{11}
water	boil	PROG	SFP

'The water is boiling.'

From the above example, the SFP *e*^{11} can appear at the end of the sentence containing the Progressive marker *ten*^{55} which actually means 'wait' because it just signals the end of the sentence not the end of the event. This particular use of SFP *e*^{11} distinguishes the characteristics of *e*^{11} in Miaoli Hakka from those of *liau*^{42} in Fengshun Hakka. The sentence-final *liau*^{42}/*lɔ*^{42} is only used to note the current relevant state and it does not appear with Progressive markers. It can be seen that the grammaticalization of particle *e*^{11} in Miaoli Hakka is more advanced than that of *liau*^{42} in Fengshun Hakka.

Finally, *jiu*^{33} 'have' in Fengshun Hakka can express only Perfective aspect but *iu*^{24} 'have' in Miaoli can be used as both Perfective marker and Experiential marker. The function of *iu*^{24} 'have' as an Experiential marker in Miaoli is most likely influenced by contact with Southern Min and Taiwanese Mandarin, which use the verb 'have' to encode an experience.

As a Pan San Khak Hakka dialect, Fengshun Hakka typically has several features shared with Chaozhou. In terms of an aspectual system, it is also found that several aspectual characteristics in Fengshun Hakka are more similar to Chaozhou than Mandarin, and Thai. Although Fengshun Hakka is spoken in Thailand where Thai language is the most predominant language, the language contact between these two languages does not have much effect on the Fengshun Hakka aspectual system. Due to the fact that Hakka is usually spoken in a family domain and Chaozhou is preferable in a business domain, the contact with Thai mostly generates language borrowing in the lexical level rather than the grammatical level in Fengshun

Hakka. However, it is found in the collected natural occurrence data that a grammatical unit borrowed from Thai is *kɔ̂* 'then, so, too'. It can serve many functions and express many meanings: it can be used as a conjunction, topic marker and interjection. In fact, Fengshun Hakka has the equivalent word *ja*^{33} which is used similarly to the Thai word *kɔ̂*, but in natural conversations the word *ja*^{33} is sometimes substituted by *kɔ̂*.

(63) *ʨhian*^{24} *kɔ̂* *mɔ*^{24}, *mak*^{11}*kai*^{42} *kɔ̂* *mɔ*^{24}
money too NEG, what too NEG
'She does not have money or anything.'

Thai and Fengshun Hakka share some aspectual features in common. For example, the Perfective marker *lɛ́ɛw* in Thai and *liau*^{42} in Fengshun Hakka are possibly from the same word. They both follow the VO compound, not in between the verb and the object. Semantically, they can express the completion of an event and the current relevant state in the Perfective aspect, and they can refer to the beginning of a new state in the Inchoative aspect.

(64) Fengshun Hakka
kiau^{42}*liaŋ*^{42} *lɔ*^{42}
beautiful INC
'You have become beautiful.'

(65) Thai
sŭay *lɛ́ɛw*
beautiful INC
'You have become beautiful.'

Moreover, the directional verbs 'come' and 'go' are also used as Resultative Verb Compounds in Thai to imply the resulting state from a thoroughly complete action. The directional verb 'come' in Thai and Fengshun Hakka refers to a completed action (as shown in 5.2.1). The directional verb 'go' of both languages is used to show the disappearance of something after the action is completed (as shown in 5.2.2).

In conclusion, the aspectual system of Fengshun Hakka shares several aspectual features with Chaozhou and some with Thai. The influence of Chaozhou and Thai of Fengshun Hakka spoken in Thailand distinguishes the aspectual features of Fengshun Hakka from those of Miaoli Hakka spoken in Taiwan.

References

Bybee, J., Perkins R., & Pagliuca W. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angles: University of California Press.

Chappell, H. (1989a). On so-called 'empty particles' in Yue-Guangzhou: a study of aspect. *La Trobe University Working Papers in Linguistics, 2*, 91-112.

Chappell, H. (1989b). The grammatical category of aspect in Southern Min: preverbal coding of aspect in Min-Xiamen. *La Trobe University Working Papers in Linguistics, 2*, 113- 125.

Chappell, H. (2001). Synchrony and Diachrony of Sinitic Lanugages: A Brief History of Chinese Dialects. In Hillary Chappell (Ed.), *Sinitic Grammar: Synchronic and Diachronic Perspectives* (pp.3-28). New York: Oxford University Press.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge Univeristy Press.

Chung, S., & Timberlake, A. (1985). Tense, Aspect, and Mood. In Timothy Shopen (Ed). *Language Typology and Sysntactic Description: Grammatical Categories and The Lexicon* (Vol.3). Cambridge: Cambridge University Press.

Hashimoto, M. J. (1973). The Hakka Dialect: A Linguistic Study of Its Phonology Syntax and Lexicon. Cambridge: Cambridge University Press.

Iwasaki, S., & Ingkaphirom, P. (2005). *A Reference Grammar of Thai*. Cambridge: Cambridge University Press.

Kubler, C. C. (1982). The influence of Southern Min on Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156-176.

Lamarre, C. (2001). Verb Complement Constructions in Chinese Dialects: Types and Markers. In Hillary Chappell (Ed.), *Sinitic Grammar: Synchronic and Diachronic Perspectives* (pp.85-120). New York: Oxford University Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Norman, J. (1988). *Chinese.* Cambridge: Cambridge University Press.

Ramsey, S. R. (1987). *The Languages of China*. Princeton: Princeton University Press.

Shiwaruangrote, N. (2008). *A Phonological Description of Jieyang Hakka Dialect As Spoken in Thailand*. Unpublished doctoral dissertation. Nakhon Pathom: Mahidol University.

Smalley, W. A. (1994). *Linguistic Diversity and National Unity: Language Ecology in Thailand*. Chicago/London: The University of Chicago Press.

Smith, C. S. (1991). *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publisher.

Smith, C. S. (1994). Aspectual viewpoint and situation type in Mandarin Chinese. *Journal of East Asian Linguistics, 3*, 107-146.

Ungsitipoonporn, S. (2007). *Phonological and Acoustic Analyses of the Tone System of Hakka as Spoken in Bangkok, Thailand*. Unpublished doctoral dissertation. Nakhon Pathom: Mahidol University.

Van Valin, R. D., Jr. (2005). *Exploring the Syntax-Simantics Interface*. New York: Cambridge University Press.

Vendler, Z. (1967). Verbs and Times. *Linguistics in Philosophy*. New York: Cornell University Press.

Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A Corpus-based Study*. Amsterdam: John Benjamins.

Xu, H. L. (2007). Aspect of Chaozhou grammar: a synchronic description of the Jieyang variety. *Journal of Chinese Linguistics, 22*.

Yue-Hashimoto, A. (1993). Comparative Chinese Dialectal Grammar: *Handbook for Investigators (Collection des Cahiers de Linguistique d'Asie Orientale 1)*. Paris: Ecole Des Hautes Etudes En Sciences Sociales.

NGUYỄN, Anh-Thư T. 2013. *Acoustic and perceptual correlates of Vietnamese folk poetry rhythmic structure. Journal of the Southeast Asian Linguistics Society* (JSEALS) 6:54-77
Received 26/8/2012, text accepted 13/3/2013, published August 2013
URL: http://hdl.handle.net/1885/10278
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

SOUTHEAST ASIAN LINGUISTICS SOCIETY SEALS
www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics
Australian National University

ACOUSTIC AND PERCEPTUAL CORRELATES OF VIETNAMESE FOLK POETRY RHYTHMIC STRUCTURE

Anh-Thư T. Nguyễn
Mountain Creek, Sunshine coast, Australia
<thunguyen1972@gmail.com>

Abstract
This paper reports a study on the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poetry. Ten speakers of Sài Gòn dialect recite four folk poems that were made up of three-word, five-word, six-word, seven-word, and eight-word lines. The acoustic analysis showed that the duration and intensity results mirror each other in indicating a strong iambic pattern of prominence, supporting the literature that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn, 1985). The perception results showed that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used. Also, majority of listeners were not finely tuned to these acoustic cues and only a few listeners could detect them in parsing the poetic lines into detail bi-syllabic iambic units.
Key words: folk poetry, rhythmic structure, acoustics, perception.
ISO 639-3 language codes: vie

1. Introduction

This paper reports a study on the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poetry. Rhythm can generally be characterized as the repetition of patterned sequences of elements, often varying in prominence (Fraisse 1974, 1982). This characterization applies most strictly to certain forms of music and poetry; however, normal speech also exhibits tendencies toward rhythmic patterning — for example, in the alternation of stressed and unstressed syllables. In addition, listeners may perceptually exaggerate the rhythmicity of utterances. For example, English-speaking listeners hear interstress intervals as more evenly spaced than they actually are (Darwin & Donovan 1980; Donovan & Darwin 1979; Lehiste 1977). Although the claim that speech is characterised by a rhythmical component is universally accepted, there is far less consensus as to what should be identified as the rhythmic unit and as the carrier of rhythmic be at in spoken language.

To identify rhythm in speech, we have to search for something that is either structured or recurrent within a limited time range. This is the same as that stands at the base of the conception of metre in poetry. Metres fulfil the requirements in that they are simple structures composed of regular successions of stresses and syllables, resulting in alternations of stressed and unstressed syllables. Furthermore, stresses and syllables fall precisely within the relevant time range. In effect, the study of speech rhythm has long been associated with metrics and poetry. Aristotle, in The Art of the Rhetoric made an early attempt to describe the rhythm of language, using metric concepts to describe the different speech styles of the people (e.g. the iambic metre was said to be the rhythm of the common people, the trochaic that of rhetoricians, etc.).

Although the question of metre is still of interest today, many phoneticians of the last century have shifted the focus of research on two other issues. One consisted in finding an acoustic correlate of the perception of the rhythm beat: results in this field are not utterly uniform, but there seems to be general

agreement on the importance of the vowel onset (see Allen 1972 and 1975). The second issue concerns the alleged regular occurrence of syllables and/or stresses, that retained most of the attention through the years.

The two categories of stress-timed and syllable-timed languages have been introduced by Pike (1945) referring to the impression that stresses seem to occur at regular temporal intervals in English, while syllables seem to have similar durations in Spanish. Abercrombie (1967) drew on the distinction, also on the basis that the different rhythmic structure of these languages seems to be reflected by the metrical units adopted in poetry: Germanic languages count the length of verses in feet, while Romance languages use the syllable as the basic metric unit. This view had been of great influence in the following years, but various instrumental experiments failed to give evidence of isochrony at the foot or at the syllable level.

After these failures, some linguists (Bertinetto 1977 and Dauer 1983, among others) attributed the impression of stress-timing or syllable-timing to structural properties of languages, such as the absence vs. presence of vocalic reduction and a complex syllabic structure. Relying on these theories, Ramus, Nespor & Mehler (1999) and Grabe & Low (2002) proposed acoustic correlates of these phonological properties based on vocalic and consonantal durations. The authors claimed that their measures allowed for a scalar characterisation of languages on the basis of rhythm properties. The validity and the stability of the acoustic correlates (soon re-baptised rhythm metrics) were soon tested in other studies, which gradually introduced new variables, such as different speech rates (e.g. Dellwo & Wagner 2003). Some authors have proposed modifications of the formulae (e.g. Dellwo 2006, and Benton 2010), or have applied the formulae to different durations, such as voiced and unvoiced intervals (e.g. Galves et al. 2002) or feet and syllables (e.g. Wagner & Dellwo 2004, and Asu & Nolan 2006), or have proposed new metrics that are based on different rationales (e.g. Bertinetto & Bertini 2008). Despite some criticisms and a few failures (e.g. Barry & Russo 2003, and Arvaniti 2009), these measures and the perspectives they offer have raised (and are still raising) a growing interest within the scientific community: various authors have used them with the aim or in the attempt of categorising different languages, different language varieties (e.g. Giordano & D'Anna 2010) and even to detect the interference of the rhythm properties of L1 on productions by L2 speakers (e.g. White & Mattys 2007). The latest developments in this field include some attempts to merge the two aspects of speech rhythm (namely, the segmental and the accentual levels) into multi-layer models (Bertinetto & Bertini 2010, and O'Dell et al. 2010).

This study examined the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poetry on the basic of Lehiste's assumption (1984, 1986, 1990a) that "the suprasegmental system of a language is crystallized in the metric structure of its traditional poetry". Poetic meter is also considered to be a strong indicator for the metrical pattern of the language in which the poetry is written (Cutler 1994, cf. Wagner 2010): "poetic speech maximizes the language specific rhythmical constraints which are probably often violated in spontaneous speech. Thus, poetic speech can be called to be characterized by a maximal level of rhythmical harmony" (Wagner, forthcoming). Therefore, it was found by Lehiste (1990a) that within a poetic line, all metric feet have approximately the same duration in English and Latvian poetry even though no evidence of isochrony at the foot or at the syllable level has been found in the languages by instrumental experiments. A number of early experiments measuring feet in English have shown that foot duration is proportional to the number of syllables they contain (Shen & Peterson 1962; Bolinger 1965; Uldall 1971; see Lehiste 1977). This applied also to reiterant speech, used by Nakatani, O'Connor & Aston (1981) to test isochrony.

In many languages, a poetic line is structured in terms of number and type of syllables. For example, the epic folksongs of Serbia, the so-called deseterac, have ten syllable lines. The classic Balto-Finnic folk song meter, the Kalevala meter, employs an eight-syllable line. In other languages, rhythm may be based on a succession of stressed syllables. Germanic languages developed a metric of that kind. In stress-timed poetry in contemporary English, metric feet consist of a stressed syllable and a flexible number of unstressed syllables, whose duration is subordinated to the requirement that within a poetic line, all metric feet have approximately the same duration (Lehiste 1990a). Lehiste (1990a) also found that Latvian tended to equalize the duration of the metric feet while in Finnish the duration is determined by the quality of the syllables constituting the foot.

Vietnamese folk poetry meter is conditioned by the length of lines and arrangement of tones within the line. Vietnamese is a tonal language consisting of six tones which are designated even or flat (bằng) and

uneven or sharp (trắc). Uneven tones are also referred to as "oblique". The two even/flat tones are: high level (ngang) and low falling (huyền). The four sharp/oblique tones are: high rising (sắc), curve (hỏi), broken (ngã) and low falling (nặng). The number of tones, as well as their realisation, varies according to dialect: the Northern dialect has 6 tones while Central and Southern dialects have 1 tone less due to the historical merging of the curve (hoi) and broken (ngã) tones. The rule of distribution of tones in poetry — the aim of which is to give music to a poem — specifies the slots in a line to be filled with even/flat tones as opposed to those to be filled with oblique/sharp tones (Công Huyền Tôn Nữ 2001).

Vietnamese folk poetry (ca dao), particularly the two common folk meters: six-eight (lục bát) and "double-seven-six-eight" (song thất lục bát) was impressionistically noted to "reproduce the common cadence of Vietnamese speech, a rising rhythm – the iamb (a group of two syllables with the stress on the second syllable), and the anapaest (a group of three syllables, with the primary stress on the last syllable and sometimes a secondary accent on the first" (Durand and Nguyễn 1985). It is also observed that "in literary and also in modern spoken Vietnamese, there has been a strong tendency toward using a two-two syllable rhythm. Vietnamese syllables are usually grouped in two-two rhythmic groups. This tendency has been crystallized into poetic norms of four-syllable lines" (Liêm 1970, page 1). Nevertheless, no empirical acoustical studies have been done to verify the stress and rhythmic pattern of the language or its folk poetry.

Vietnamese has no system of culminative word stress; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson 1965; Nguyễn Đăng Liêm 1970). Duration, intensity, full tonal realisation of accented syllables have been observed to be important parameters for describing stress in Vietnamese (Đỗ 1986; Chaudhary 1983; Hoàng & Hoàng 1975; Gsell 1980). Regarding the stress patterning in utterances, it is generally agreed by some researchers that there is an alternating pattern of strong and weak syllables. Thompson (1965) stated that the majority of the syllables have medium stress. In a sequence of syllables, alternating ones are slightly louder (but not in a distinctive manner): "each pause group has at least one heavy stress and weak stresses are fairly frequent in rapid passages, rarer in carefully speech" (p. 50). Jones and Huỳnh (1960) stated that "normally the stresses in a Vietnamese utterance are conditioned by the junctures," and regarded the fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress. Consistent with Jones and Huỳnh's observation, it is remarked by Cao (2003) that due to the demarcative function of stress/accent in Vietnamese, native listeners tended to hear a juncture after a stressed syllable even though there is no such pause in reality as examined by spectrograms.

In recent studies on more carefully phonetically controlled and specialized sets of Vietnamese disyllabic compounds and reduplications, Nguyen and Ingram (2007a, b) found that there was at least a phonetic tendency for the right hand element of a disyllabic compound word to be more prosodically prominent by a number of relevant phonetic measures: greater tonal f0 range, higher intensity, greater duration of the second syllable, and formant measurements indicative of more centralized vowel nuclei (vowel reduction) on the first syllable. Nguyen (2010) investigated the rhythmic patterns in Vietnamese polysyllabic words by examining the rhythmic patterns and their acoustic correlates in polysyllabic reduplicative words (2-,3-,4-,5-,6- syllable pseudo-words). The results showed that there is a tendency of syllable coupling indicated mainly by syllable duration pattern and supported by the native listeners' perception results, suggesting that polysyllabic words in Vietnamese tend to be parsed into bi-syllabic iambic feet with a rightward or retrograde rhythmic pattern.

2. Cadao - Vietnamese folk poetry- metric structure

Ca dao takes a wide range of forms, from two-syllable to six- and eight-syllable verse (lục bát). The range of *ca dao* includes children's game songs, love songs, lullabies, riddles, work songs, and reveries about spiritual and social orders (Balaban 1980). The two most common folk meters are six-eight (lục bát) and "double-seven-six-eight"(song thất lục bát). It is estimated that 90% of folk poetry are in six-eight meter (Nguyễn 1992). Both verse forms can be traced to two basic meters: a rising rhythm, the iamb (a group of two syllables with the stress on the second syllable), and the anapaest (a group of three syllables, with the primary stress on the last syllable and sometimes a secondary accent on the first) (Durand and Nguyễn 1985). Vietnamese sayings and proverbs are often folk poems in miniature; they combine two iambs, two anapaests, one iamb, and one anapaest, or varying numbers of both. A line of folk verse tends to be made up

of an even number of syllables, a series of iambs in particular, when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest.

The constituent unit of six-eight (lục bát) verse is a couplet in which the first line has six syllables and the second line eight syllables. Usually, both lines consist of iambs, and the stress falls on each even-numbered syllable: in such a case, they can be described as iambic trimeter and iambic tetrameter, respectively. On occasion, the first line departs from the norm and comprises two anapaests, instead: two equal, balanced hemistichs divided by a medial caesura (three | three). The last (or sixth) syllable of the first line rhymes with the last-but-two (or sixth) syllable of the second line (or, as an archaism, with the fourth syllable). What stamps six-eight verse with individuality, however, is its pattern of euphonic requirements, an ingeniously plotted alternation of "flat" (bằng) and "sharp" (trắc) tones. The scheme, as observable in a typical couplet, is diagrammed in figure 1: the double-headed arrow indicates rhyming, an open space represents an unstressed syllable, and F and S stand for flats and sharps. It will be noted that the tonal rules apply only to accented (even-numbered) syllables; that the same type of tones is assigned to syllables taking up the same position in both lines (flats for I-2 and II-2, sharps for I-4 and II-4, and flats for I-6 and II-6); and that flats predominate: they flank a single sharp in each line and monopolize the rhyme-fellows (I-6 and II-6) as well as the last two stressed syllables of the second line (II-6 and II-8).

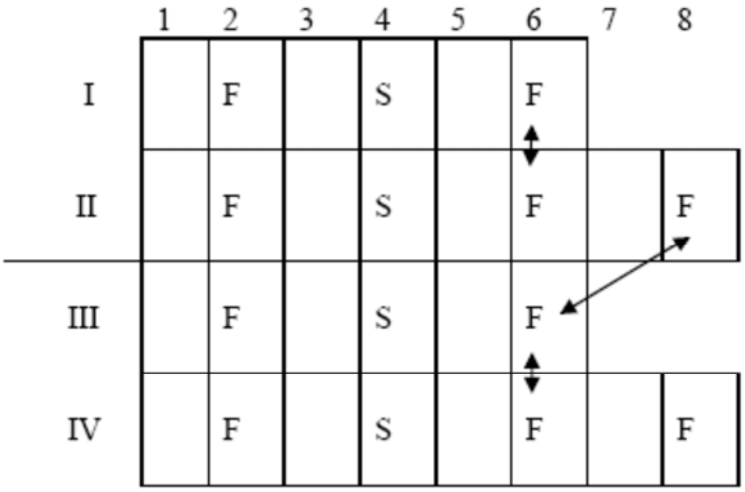


Figure 1. Six-eight verse diagram (adapted from Durand and Nguyễn 1985).
F: flat tone, S: sharp tone.

A six-eight couplet can stand by itself as a poem, and it often does in folklore. Six-eight verse amounts to a concatenation of such couplets. The needed linkage between couplets is that the last (or eight) syllable in the second line of one couplet rhymes with the last (or sixth) syllable in the first line of the next couplet. Each line rhymes with the next, but a new, different rhyme appears in every other line, in every couplet.

The constituent unit of double-seven-six-eight verse is a four-line stanza in which a double-seven doublet is followed by a six-eight couplet. Each line of the double seven doublet has seven syllables: an anapaest followed by two iambs (three | two | two). The disposition of rhyming syllables and prescribed tones is shown in figure 2 below. The double- seven doublet resembles the six-eight couplet in that the last syllable of the first line rhymes with the penultimate stressed syllable of the second line. In everything else, the double-seven is the opposite of the six-eight. Each line has an odd number of syllables. The rhyme-fellow carry sharp tones. Odd-numbered syllables are accented and embody a tonal scheme which, significantly, contradicts that of six-eight verse and no longer favours flat tones: three flats are balanced by three sharps. Figure 2 represents the double-seven-six-eight quatrain – the doublet and the couplet are joined by the rhyme-fellows II-7 and III-6. It should be noted that II-7 and III-6 carry flats; as a result, such tones still predominate within the rhyming scheme of the double-seven quatrain, although sharps more than hold their own in the first two lines: they equal flats in number and appropriate the rhyme syllables (I-7, II-5).

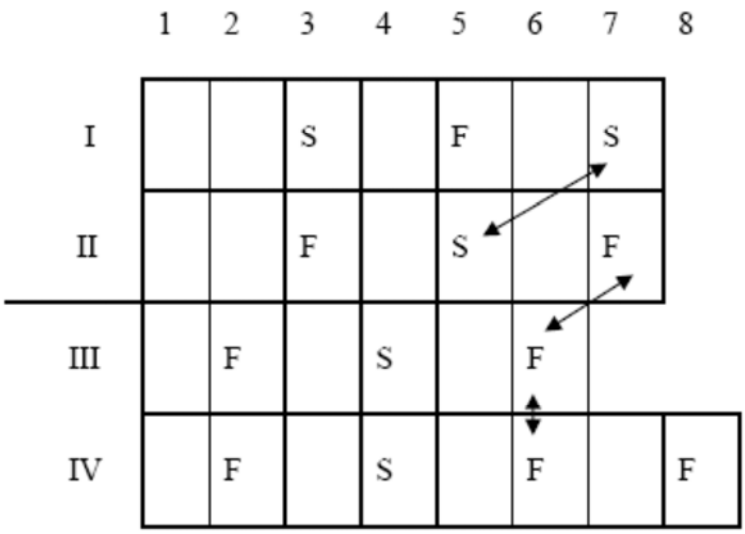


Figure 2. Double-seven-six-eight stanza diagram (adapted from Durand and Nguyễn 1985)
F: flat tone, S: sharp tone, arrows: rhyming.

3. Production experiment

3.1. Linguistic materials

In order to pursue the aim of the present study, which examines the rhythmic patterns of three-word, five-word, six-word, seven-word, and eight-word lines, four folk poems were used. The investigated poems are in the appendix. The first is a popular six-eight poem titled "Tát nước đầu đình "(Fetching the water before the village communal house), which has 16 lines. Generally, the folk poem conforms to the prescribed metrical structure described in figure 1, except for the three syllables which varied from the prescribed tones (the prescribed tones should be even/flat ones but they are in uneven/sharp tones, *được*, in the third line, *mượn* in the eight line, and *lúc* in the tenth line). The second is the first four double-seven-six-eight stanzas of the "Chinh phụ ngâm" (The song of a soldier's wife) by Đặng Trần Côn, rendered in Chữ Nôm (Vietnamese scripts) and in double- seven- six -eight folk meter by Phan Huy Ích. The four four-line stanzas consisted of 16 lines in total. Generally, the tones of the poem conform to the prescribed metrical structure described in figure 2 above, except for two syllables which have flat/even tones instead of sharp/uneven tones (*thành* in the fifth line and *bình* in the ninth line). The third item is a five-word folk poem which has ten lines. The fourth is four three-word riddles and sayings of about two, three and four lines each, which have 11 lines in total.

3.2. Subjects

Ten speakers of the Sài Gòn dialect (5 males, 5 females) who came from Hồ Chí Minh city participated in the study. They were either visitors or newly arrived immigrants to Australia and had been in Australia from 2 weeks to 4 years. Their age ranged from 38 to 70 years. Their education levels ranged from high school to higher degrees.

3.3. Procedure

Subjects were given the poems in print to practice reading before the recording. They were asked to read the poems in a natural manner. In fact, since these are popular poems, most of them have either read, heard or even learnt about them in schools. Recordings were made in a quiet room using sound recording and editing computer software PRAAT (Boersma and Weenink 2007) at 22050 Hz sampling rate.

3.4. Measurement

The acoustic parameters measured included syllable duration, line duration, duration of pause following each poetic line, intensity, vowel formant and vowel duration. Temporal variations of F0 and tonal shapes are obvious components but will not be treated here due to the nature of the linguistic material. The test items were segmented via the Emu Speech Tools, (Cassidy 1999). First, the Emu Labeller was used to mark the

edges of the target syllables and vowels, relying primarily on the spectrographic display in the Labeller. Then the Emu-R statistical software was used to extract the target duration (ms). Peak intensity (dB) in syllables and vowel first and second formants (Hz) and vowel duration were measured manually via Praat (Boersma and Weenink 2007).

3.5. Analysis

There were in total 530 lines of poem (16 lines of six-eight poem + 16 lines of double-seven-six-eight poem + 10 lines of five-word poem + 11 lines of three-word riddles and sayings read by 10 speakers).

A mixed (fixed and random) effects analysis of variance (ANOVA) model, using the restricted maximum likelihood method to estimate variance components was used to statistically analyse the data for each poem. The dependent variables were syllable duration, line duration, pause duration, tentative foot duration (grouping of two syllables), and intensity. The fixed effects were syllable or foot positions (in a line), lines (number of words in a line type), and speakers. The random effect was syllables. Tukey post-hoc tests were carried out to determine the significant differences among levels of the main factor of syllable or foot positions and lines when necessary.

3.6. Results

3.6.1. Duration

The three-way mixed effect ANOVA (syllable positions x lines x speakers) on syllable duration of the six-eight poem showed a significant effect for main factors syllable positions: $F(7, 1120) = 80.03$, $p < .0001$, lines: $F(1, 1120) = 8.93$, $p < 0.001$, speakers $F(9, 1120) = 44.52$, $p < .0001$, and the interaction syllable positions x lines: $F(5, 1120) = 7.09$, $p < .0001$, while other interaction effects were not significant (positions x speakers: $F(63, 1120) = 1.26$, $p = 0.08$ ns., speakers x lines: $F(9, 1120) = 0.8$, $p = 0.5$ ns., syllable positions x speakers x lines: $F(45, 1120) = 0.9$, $p = 0.5$ ns.)

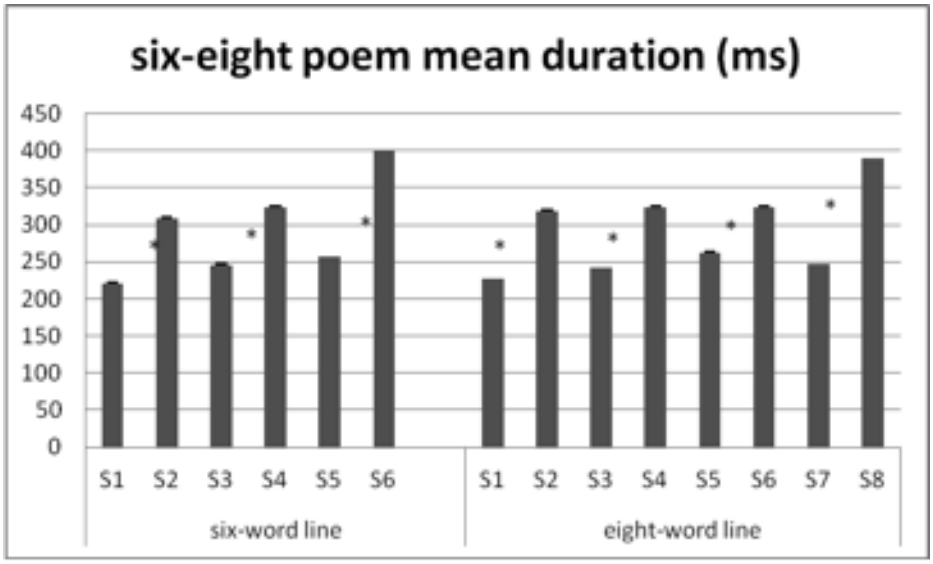


Figure 3. Mean syllable duration (ms) of the six-eight poem. S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

A post-hoc Tukey test for the interaction syllable positions x lines (figure 3 above) showed that the even syllables were significantly longer than its preceding odd syllables for both six and eight lines of the poem, suggesting a tendency of syllable coupling headed by the even syllables. The last syllables of the lines were longest (significantly longer than all other syllables in the lines), possibly indicating both a final lengthening effect and a right –headed accented effect if the same result is found for intensity.

The three-way mixed effect ANOVA results for the double-seven-six-eight poem showed a significant effect for main factors syllable positions: $F(7, 1120) = 17.37$, $p < .0001$ and speakers $F(9, 1120) = 26.98$, $p <$

.0001, but non-significance for lines: F(1, 1120) = 2.14, p = 0.1ns., and the interaction syllable positions x lines: F(12, 1120) = 0.37, p=0.8 ns. Other interaction effects were not significant (positions x speakers: F(63, 1120) = 1.64, p = 0.5 ns., speakers x lines: F(18, 1120) = 1.67, p = 0.3 ns., syllable positions x speakers x lines: F(100, 1120) = 0. 98, p = 0.7 ns.)

The post-hoc results of the interaction syllable positions x lines (figure 4 below) showed that the timing pattern of the six- and eight-word couplets mirrors that of the six-eight poem: even syllables are significantly longer than their preceding odd counterparts. In the seven-word lines, the odd syllables (S3, S5, S7) were significantly longer than their preceding even syllables, supporting the literature claim that the odd-numbered syllables are accented and the line has an anapaest followed by two iambs (three | two | two). The last syllable of the lines also had the longest duration as found in the six-eight poem. Also, there seems to be some type of line-medial prominence in double-seven-six-eight (figure 4), but not in six-eight poem (figure 3); particularly, in the eight-word lines, the fourth syllable was as long as the last syllable.

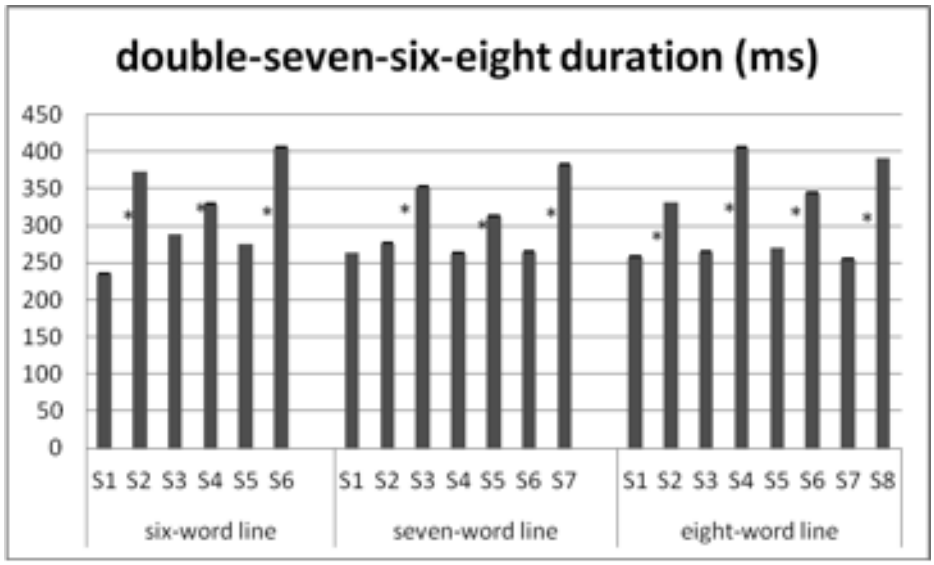


Figure 4. Mean syllable duration (ms) of the double-seven-six-eight poem.
S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line.
The symbol * indicates significance at p < .001.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable duration of the five-word poem showed a significant effect for main factors syllable positions: F (4, 500) = 7.64, p < .0001, speakers F(9, 500) = 23.55, p < .0001, but the interaction syllable positions x speakers was not significant: F(36, 500) = 1.15, p = 0.2 ns.

A Tukey post-hoc test showed that the second syllable was significantly longer than the first syllable. The third syllable was significantly longer than the first and the second syllable. The fifth syllable was significantly longer than the fourth syllable and also the longest, while the fourth syllable was the shortest. This result suggests a rhythmic pattern of an anapaest followed by an iamb.

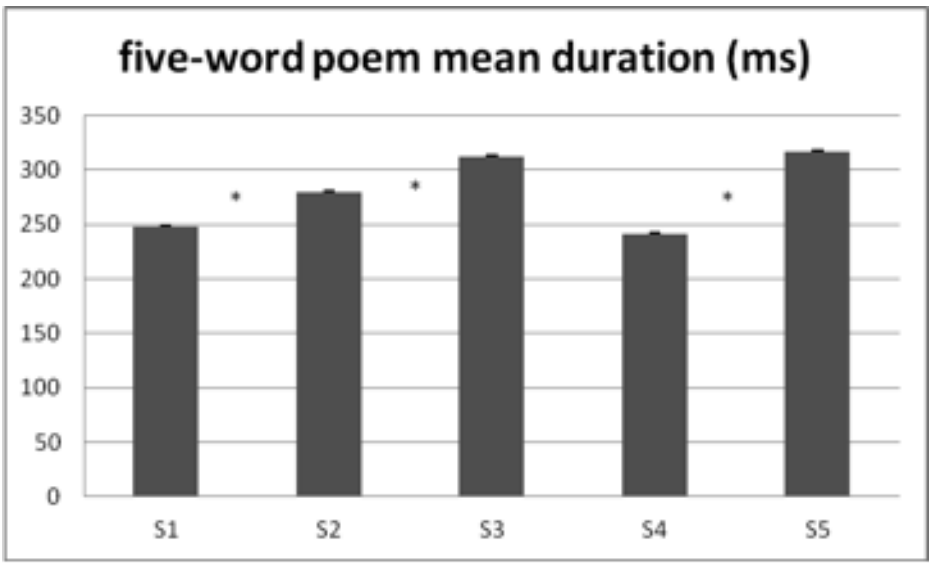


Figure 5. Mean syllable duration (ms) of the five-word poem. S1, S2, S3, S4, and S5 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable duration of the three-word riddles and sayings showed a significant effect for main factors syllable positions: $F(2, 330) = 13.09$, $p < .0001$, speakers $F(9, 330) = 16.50$, $p < .0001$, but the interaction syllable positions x speakers was not significant: $F(18, 330) = 1.27$, $p = 0.05$ ns.

A post-hoc test showed that the first syllable was significantly longer than the second syllable. The third syllable was the longest, significantly longer than the first and the second syllable. This can suggest either an anapaest pattern or a tentative two-foot pattern in which the first syllable stand by itself and was lengthened to fill the foot template and the other two syllables constitute an iamb. This needs to be verified in a perception study to see how listeners parse the rhythmic pattern of the line.

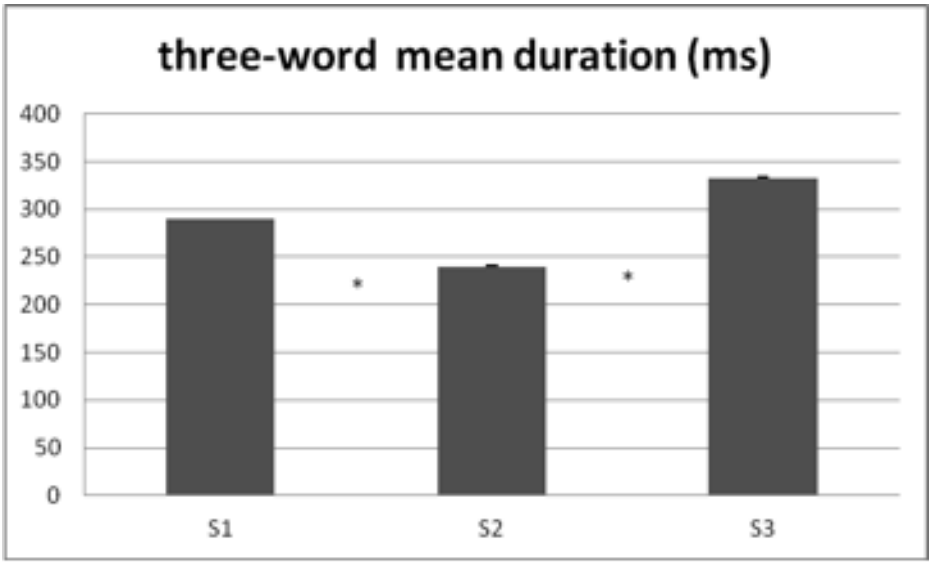


Figure 6. Mean syllable duration (ms) of the three-word riddles and sayings. S1, S2, and S3 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

3.6.2. Intensity

The three-way fixed effect ANOVA (syllable positions x lines x speakers) on syllable intensity of the six-eight poem showed a significant effect for main factors syllable positions: $F(7, 1120) = 10.26$, $p < .0001$, lines: $F(1, 1120) = 35.09$, $p < .0001$, speakers $F(9, 1120) = 195.93$, $p < .0001$, and the interaction syllable

positions x lines: F(5, 1120) = 2.81, p < .02, while other interaction effects were not significant (positions x speakers: F(63, 1120) = 1.44, p = 0.01 ns., speakers x lines: F(9, 1120) = 1.54, p = 0.1 ns., syllable positions x speakers x lines: F(45, 1120) = 0. 92, p = 0.6 ns.)

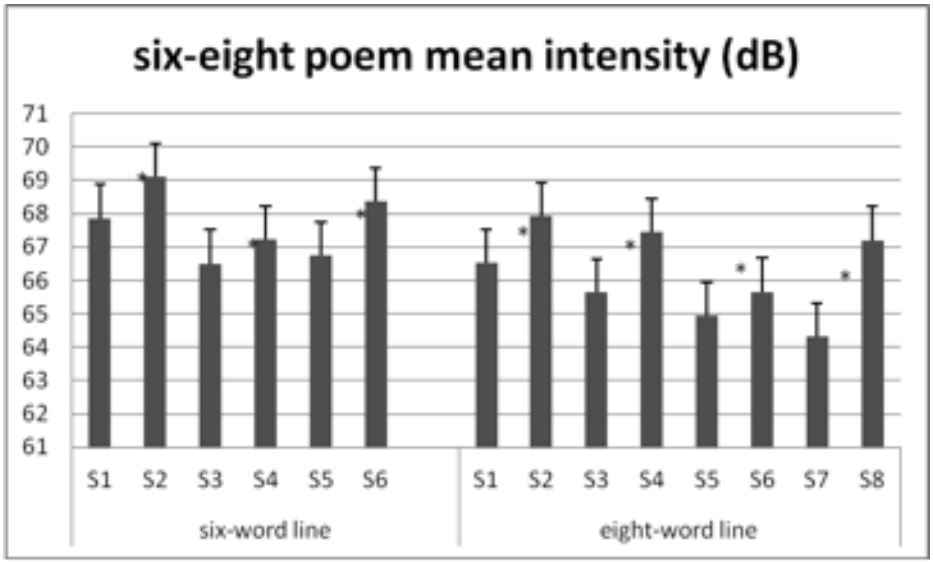


Figure 7. Mean syllable intensity (dB) of the six-eight poem. S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line. The symbol * indicates significance at p < .001.

A post-hoc Tukey test for the interaction syllable positions x lines (Figure 7 above) showed that the even syllables had significantly higher intensity than their preceding odd syllables for both sixth and eighth lines of the poem. This matches the duration results, suggesting a pattern of syllable coupling headed by the even syllables. The last syllables of the lines also had very strong intensity, together with the duration results, suggesting a right –headed accented effect.

The three-way mixed effect ANOVA results for the double-seven-six-eight poem showed a significant effect for main factors syllable positions: F(7, 1120) = 10.50, p < .0001, lines: F(1, 1120) = 5.85, p < .05, speakers F(9, 1120) = 101.45, p < .0001, and the interaction syllable positions x lines: F(12, 1120) = 3.17, p < .001, while other interaction effects were not significant (positions x speakers: F(63, 1120) = 1.29, p = 0.2 ns., speakers x lines: F(18, 1120) = 1.17, p = 0.3 ns., syllable positions x speakers x lines: F(100, 1120) = 0. 81, p = 0.8 ns.)

The post-hoc results of the interaction syllable positions x lines (figure 8 below) showed that for the six-word and eight-word lines, the even syllables had higher intensity than their preceding odd counterparts. Nevertheless, the significant patterns were less robust than that of the duration results. The reason the fourth syllable of the eight-word line had lower intensity is that these syllables in the double-seven-six-eight poem carry either the low falling tone (*nặng*) or the curve tone (*hỏi*) which tend to be creaky and thus had low intensity values. There was not a significant intensity pattern for the seven-word lines because many accented syllables (S5 and S7) carry the low-falling (*nặng*) and curve (*hỏi*) tones, which have low intensity due to creakiness.

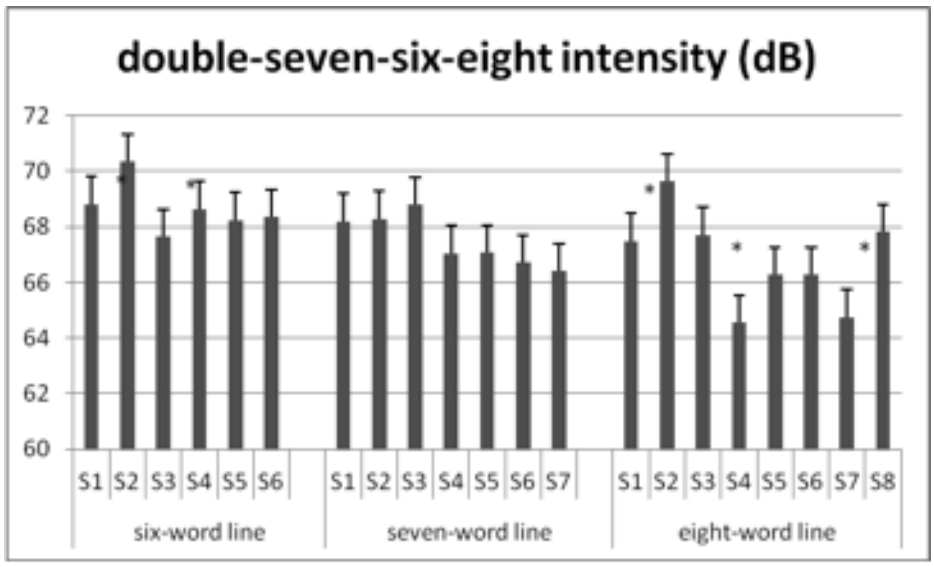


Figure 8. Mean syllable intensity (dB) of the double-seven-six-eight poem. S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable intensity of the five-word poem showed a significant effect for main factors syllable positions: $F(4, 500) = 5.60$, $p < .001$, speakers $F(9, 500) = 187.46$, $p < .0001$, but the interaction syllable positions x speakers was not significant: $F(36, 500) = 1.63$, $p = 0.24$ ns.

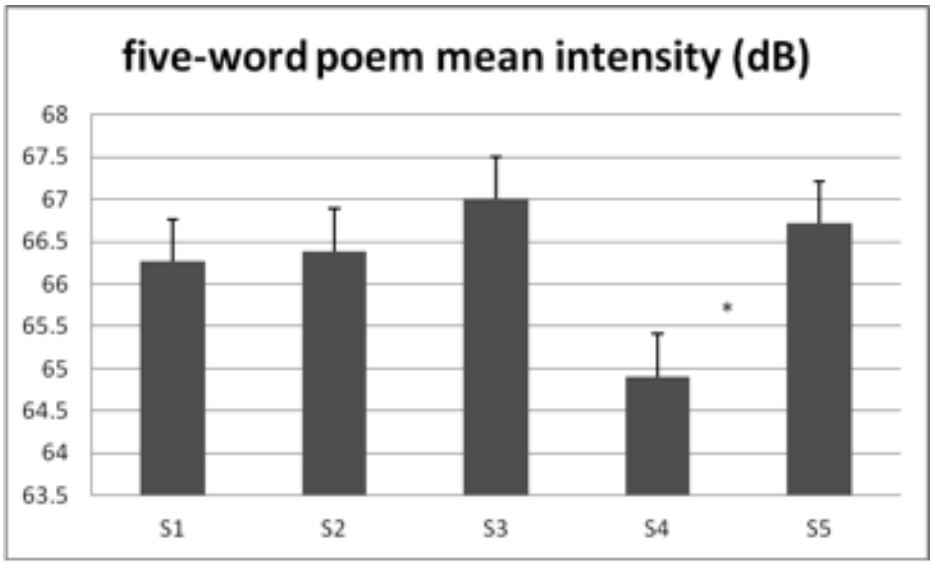


Figure 9. Mean syllable intensity (dB) of the five-word poem. S1, S2, S3, S4, and S5 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

A Tukey post-hoc test showed that the third syllables had higher mean intensity than the first and the second syllables, however it did not reach significance. The fifth syllable was significantly higher in intensity than the fourth syllable and the fourth syllable had the lowest intensity. This, together with the duration result, suggests a rhythmic pattern of an anapaest followed by an iamb.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable intensity of the three-word riddles and sayings showed a significant effect for main factors syllable positions: $F (2, 330) = 6.1$, $p < .01$, speakers $F(9, 330) = 9.87$, $p < .0001$, but the interaction syllable positions x speakers was not significant: $F(18, 330) = 0.16$, $p = 1$ ns.

A post-hoc test showed that the first syllable was significantly higher in intensity than the second syllable. The third syllable significantly had higher intensity values than the second syllable. Particularly, the

first syllable had the strongest intensity. This, together with the duration result, suggests a tentative two-foot pattern in which the first syllable was lengthened and strongly accented to fill the foot template and the other two syllables constitute an iamb. However, this needs to be verified in a perception study to see how listeners parse the rhythmic pattern of the line.

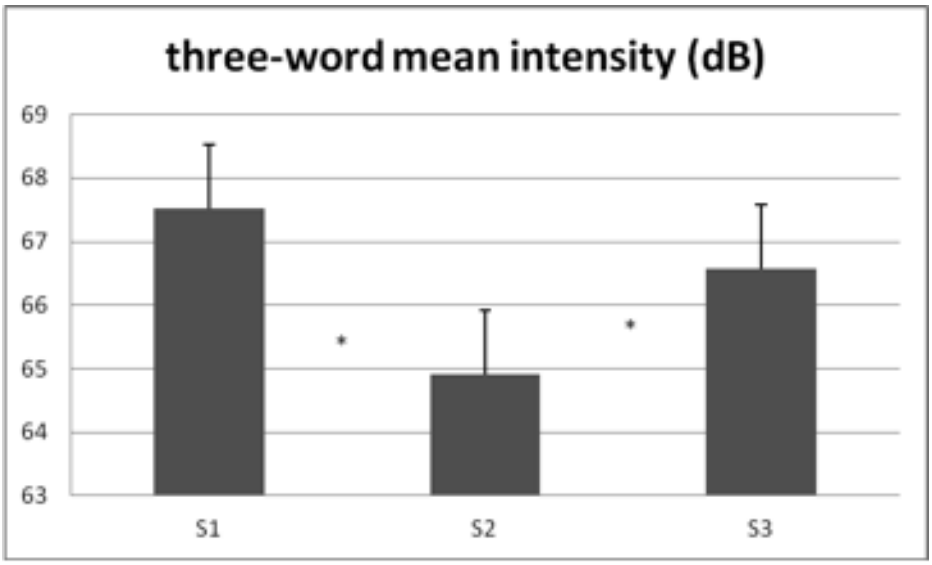


Figure 10. Mean syllable intensity (dB) of the three-word riddles and sayings. S1, S2, and S3 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

In summary, the duration and intensity results mirror each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs, and that when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). This suggests that syllables in Vietnamese folk poetry tend to be grouped into a tentative iambic bi-syllabic foot and when there is only one odd syllable, the syllable tends to be lengthened to fill the bi-syllabic foot template. If native listeners' parsing pattern matched this acoustic prominence pattern, it would suggest that a bi-syllabic foot is a prosodic unit above the syllable in Vietnamese folk poetry meter, which is the aim of the perception experiment reported in a later section of this paper.

3.6.3. Vowel reduction

In order to examine whether there is phonetic reduction associated with the examined acoustic pattern of asymmetric prominences, a small follow-up experiment was carried out. Vowels of the first syllable (unaccented syllable) and second syllable (accented syllable) of the tentative iambic foot in the six-eight poem were compared with matching vowels elicited in a control condition in which speakers said the target word in a sentence "Đọc lại từ ----- đi nhé " (Please say the word ----- again). Two of the ten speakers (one male and one female) were asked to record the control vowels which were then compared with their counterpart vowels in the poem. The vowels investigated included: /i/ /e/, /ɛ/, /a/, /ɐ/, /o/ and /ɔ/ embedded in the targets words as shown in Table 1 (below). The target tentative foot used for this vowel analysis were bolded in the six-eight poem in the appendix of stimulus material. First and second formants of the vowels under investigation were taken at vowel mid- point manually via Praat. Vowel duration was also measured. The result is shown in figures 11 and 12 (below).

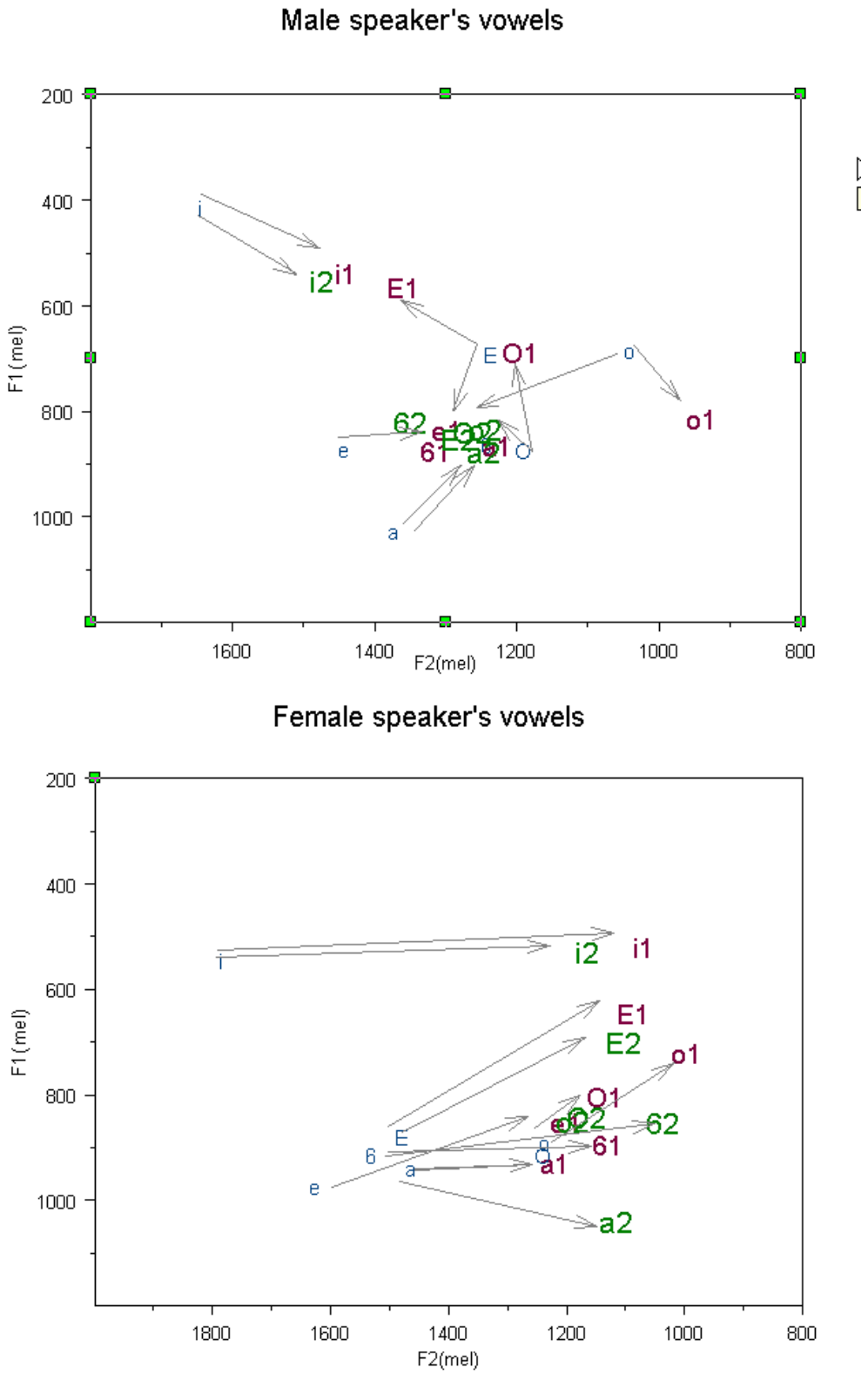


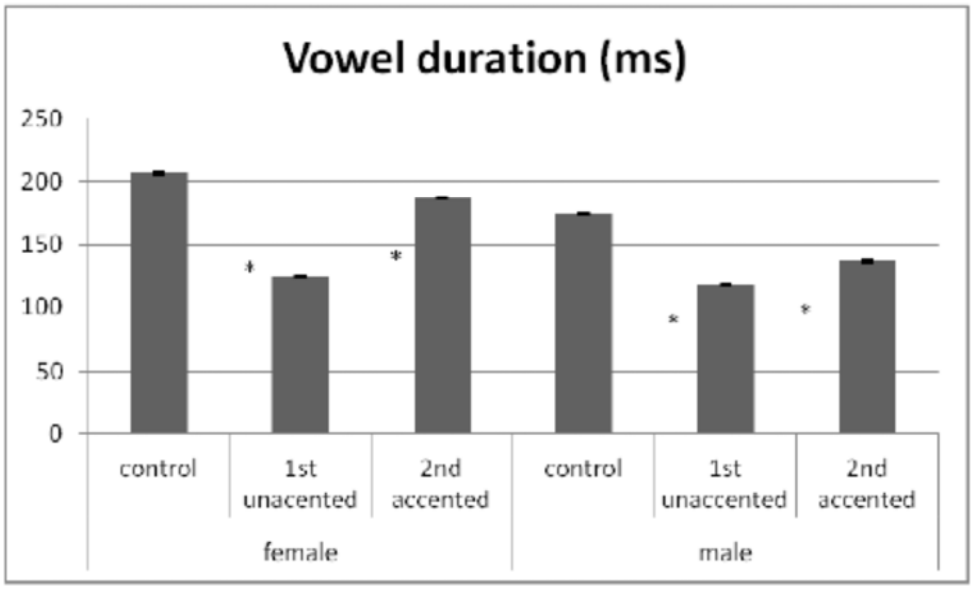
Figure 11. Vowel plots produced by the male (top fig.) and female (bottom fig.) speakers. Vowel SAMPA symbols without a number: control condition. Vowels with number 1: target word in the first unaccented position of a two-word foot in the 6- word or 8-word poetic line. Vowels numbered 2: target word in the second accented position of a two-word foot in the 6-word or 8-word line. SAMPA symbols: i: /i/, E: /e/, e:/ɛ/, a:/a/, 6:/ ɐ/, o: /o/, and O: /ɔ/.

Table 1. Vowels and target words

Vowels	Control words	1st unaccented words	2nd accented words
/i/	Thì /tʰi^{21}/	Thì /tʰi^{21}/ **(thì** cho)	chỉ /ci^{313}/ (sứt **chỉ**)
/e/	Quên /wen^{33}/	Trên/ tʂen^{33}/ (**trên** cành)	quên /w en^{33}/ (bỏ **quên**)
/ɛ/	mẹ /mɛ $^{3'2?}$ /	mẹ /mɛ$^{3'2?}$/ (**mẹ** già)	-
/a/	qua /wa ^{33}/	trả / tʂa^{313}/ (**trả** công)	qua /wa^{33}/ (hôm **qua**)
/ɐ/	cành /kɐn^{21}/	ăn /ɐn^{33}/ (**ăn** rau)	cành /kɐŋ^{21} / (trên **cành**)
/o/	hôm /hom^{33}/	hôm /hom^{33}/ (**hôm** qua)	công koŋ͡m^{33}/ (trả **công**)
/ɔ/	cho /cɔ^{33}/	cho /cɔ^{33}/ (**cho** cùng)	cho /cɔ^{33}/ (thì **cho**)

As shown in figure 11, there was not much distinction between the vowels of the unaccented first syllable and those of the accented second syllable of the foot. For the male speaker, both vowels were more centralised in comparison with the control vowels. For the female speaker, both vowels in the orally produced poem tended to be more back in comparison with their counterparts in the control condition. This result only indicated that vowels in poetry readings tend to be somewhat altered or reduced in acoustic quality compared with when they were produced in a more carefully speaking style, a phenomenon predicted to be more pervasive in spontaneous Vietnamese than already documented and needs to be more systematically studied.

A mixed effect two- way ANOVA (conditions x speakers) was conducted on vowel duration. The fixed effect was conditions (control vs. 1st unaccented syllable vs. 2nd accented syllable) and speakers. The random effect was vowels. The results showed significant effects for conditions ($F(2,41)= 5.4$, $p<0.02$), speakers ($F(1, 41) = 23.7$, $p<0.01$), and the interaction conditions x speakers ($F(2, 41)=4.2$, $p<0.05$). A post-hoc test showed that the 1st unaccented vowel was significantly shorter than the control vowel and the 2nd accented vowel for both speakers as indicated in figure 12. This result indicates that while the 2nd syllable in the tentative bi-syllabic foot was lengthened when accented, the first unaccented syllable was reduced, probably to fit in a foot timing template as investigated in the section on foot duration below.



Figures 12. Mean vowel duration (ms) . The symbol * means p < 001.

3.6.4. Foot duration

In order to examine whether the tentative iambic feet are isochronous, a three-way ANOVA (foot positions x lines x speakers) was conducted on the duration of bi-syllabic iambic feet (duration of a bi-syllabic foot = duration of 1st syllable + duration of 2nd syllable) of the six-eight poem. The result showed a significant effect

for main factors foot positions: F (3, 560) = 31.84, p < .0001, lines: F(1, 560) = 5.95, p < .02, speakers F(9, 560) = 25.06, p < .0001, and the interaction foot positions x lines: F(2, 560) = 9.5, p < .0001, while other interaction effects were not significant (positions x speakers: F(27, 560) = 0.79, p = 0.76 ns., speakers x lines: F(9, 560) = 0.48, p = 0.8 ns., foot positions x speakers x lines: F(18, 560) = 0. 9, p = 0.56 ns.)

A post-hoc Tukey test (figure 13) showed that only the last foot of the six- and eight- word lines was significantly longer than all the other preceding feet in the line respectively, while the non-final feet were not significantly different in duration from one another, indicating a pattern of foot isochrony. The non-final feet of the six-word line were also found to be equal (not significantly statistically different) to non-final counterparts of the eight-word line. This strongly suggests an isochronous bi-syllabic foot-timing template.

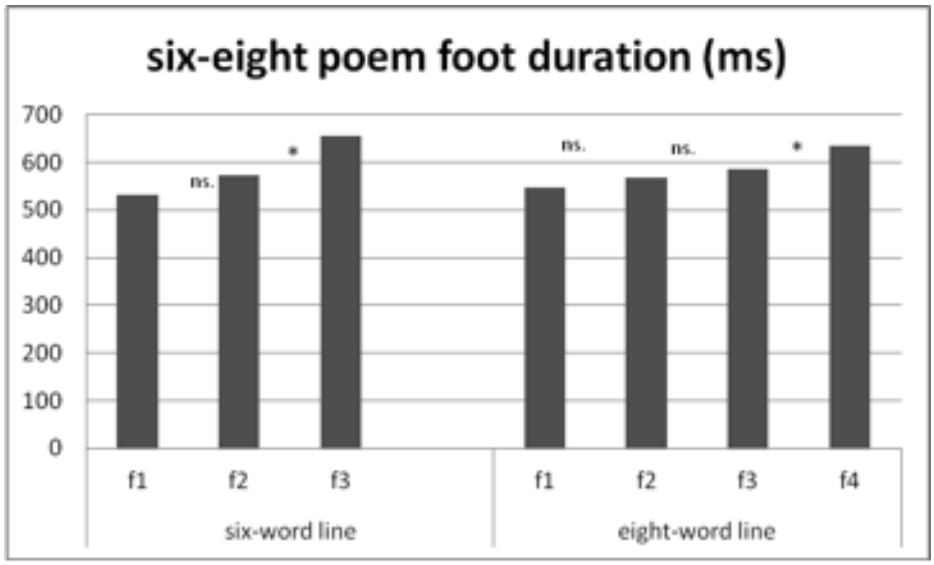


Figure 13. Mean foot duration (ms) of the six-eight poem; f1, f2, f3, and f4 are foot positions in a line. The symbol * indicates significance at p < .001; ns. means non-significance.

The three-way ANOVA (foot positions x lines x speakers) on the duration of tri-syllabic and bi-syllabic iambic foot of the double-seven-six-eight poem showed a significant effect for main factors foot positions: F (3, 501) = 66.3, p < .0001, lines: F(2, 501) = 50.41, p < .0001, speakers F(9, 501) = 19.48, p < .0001, and the interaction foot positions x lines: F(4, 501) = 85.2, p < .0001, while other interaction effects were either weakly or not significant (positions x speakers: F(27, 501) = 2.3, p <05., speakers x lines: F(18, 501) = 2.6, p < 05, foot positions x speakers x lines: F(36, 501) = 1.1, p = 0.28 ns.)

A post-hoc Tukey test (figure 14) showed a similar pattern of foot isochrony for the six-eight couplets as that of the six-eight poem above. That is, only the last foot of the six-word line was significantly longer than all the other preceding feet in the line respectively, while the non-final feet of both lines were not significantly different in duration from one another, strongly confirming a pattern of foot isochrony regardless of the segmental structure of the disyllabic sequence, consistent with Latvian and English metric feet (Lehiste 1990a). On the other hand, in the seven-word line, the first three-syllable foot was much longer than its following bi-syllabic feet, suggesting foot duration increases with the number of syllables, consistent with Shen & Peterson (1962), Bolinger (1965), Uldall (1971) and Nakatani et al (1981). It is wondered how listeners parse the three-syllable anapaest of the seven-word lines: as one unit of three words together or as two units one | two.

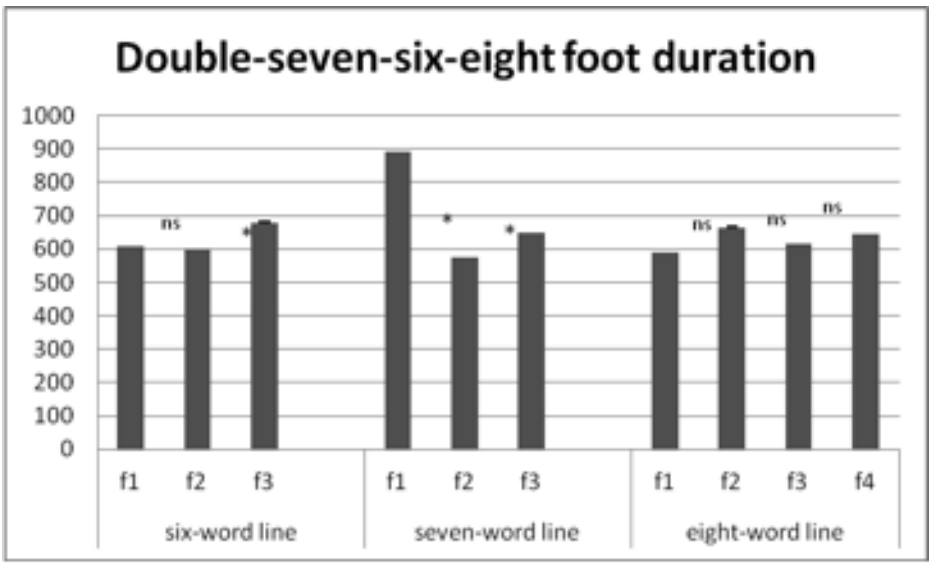


Figure 14. Mean foot duration (ms) of the double-seven-six-eight poem; f1, f2, f3, and f4 are foot positions in a line. The symbol * indicates significance at $p < .001$; ns. means non-significance.

3.6.5. Line and pause duration

In order to examine the higher-level timing pattern above the line level, two-way ANOVAs (lines x speakers) were conducted on the line and pause duration of the double-seven-six-eight poem. The results showed significant main effect of lines (6-word vs. 7-word vs. 8-word): line duration: $F (3, 150) = 41$, $p < .0001$, pause duration: $F (3, 150) = 10.14$, $p < .0001$) and speakers: $F (3, 150) = 7.5$, $p < .0001$ and $F (3, 150) = 2.83$, $p < .0001$ respectively. The interaction of lines x speakers were not significant: $F (27, 150) = 1.9$, $p = 0.26$ ns. for line duration data and $F (27, 150) = 0.8$, $p = 0.5$ ns. for pause duration data).

A post-hoc test (figure 15) showed that the six-word line was the shortest, the seven-word line was significantly longer and the eight-word line was the longest. In the seven-word doublet, the second seven-word line was significantly longer than the first line. This indicates that with isochrony at the foot level in Vietnamese six-eight poem, there was no isochrony at the line level as found in Lehiste (1990b)'s study in Lithuanian poems. In terms of pause duration, the pauses following the second seven-word line and the eight-word-line were significantly longer than those following their preceding counterpart: the first seven-word line and the six-word line respectively. Nevertheless, the pause following the eight-word line (i.e. that last line of the stanza) was not significantly larger than that of the second seven-word line. This result indicates that speakers tended to organise the double-seven-six-eight poems into units of doublets and couplets instead of four-line stanza units.

In summary, the duration and intensity results mirror each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). The results of this experiment also mirror the findings on acoustic and perceptual rhythmic pattern of polysyllabic word (Nguyen 2010) and supporting the iambic acoustic pattern in a subset of bi-syllabic compounds and reduplications investigated (Nguyen and Ingram 2007a, 2007b). This suggests that syllables in Vietnamese folk poetry tend to be grouped into an iambic bi-syllabic foot. This is strongly supported by the isochronous iambic foot pattern regardless of the segmental structure of the disyllabic sequence in the six-eight metric. It is likely that "poetic speech maximizes the language specific rhythmical constraints which are probably often violated in spontaneous speech. Thus, poetic speech can be said to be characterized by a maximal level of rhythmical harmony" (Wagner, forthcoming). The next section reports the results of the perception test, which aimed to examine whether the acoustic cues are employed in the perception of the rhythmic structure of the folk poetry by native listeners.

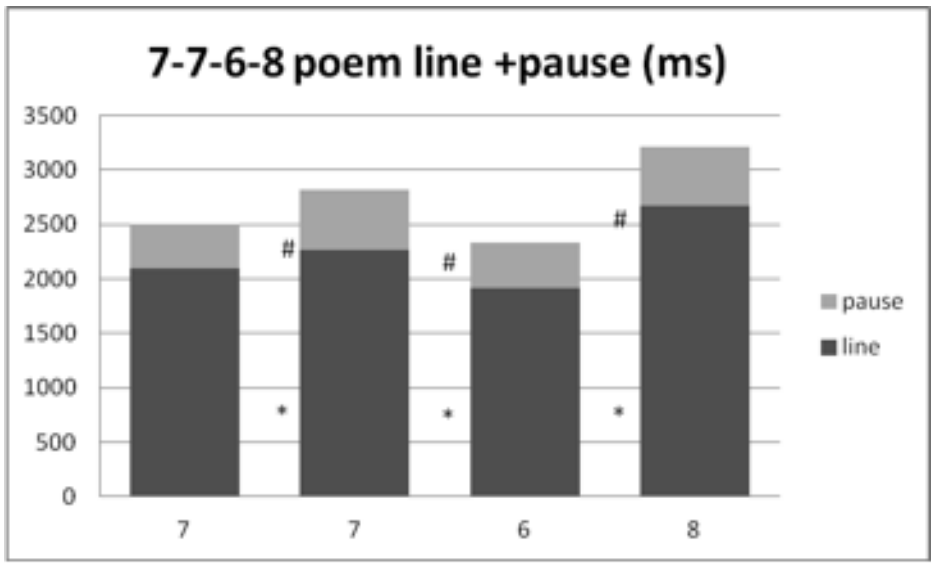


Figure 15. Mean line and pause duration (ms) of the double-seven-six-eight poem;
The symbol * indicates significance at $p < .001$ for the line duration data, # for pause duration data.

4. Perception experiment

4.1. Stimuli

Stimuli of two speakers (1 male and 1 female) from the production experiment were used for the perception study. The criteria used for the selection of stimuli were that the speakers had an easy-to-hear voice, which was judged to be clear, loud enough to be heard by the experimenter. There were 106 utterances to be judged ((16 six-eight lines + 16 double-seven-six-eight lines + 10 five-word lines + 11 three-word lines) x 2 speakers). The tokens were put in random order each with two immediate repetitions with a gap of about 6 seconds between each item.

4.2. Subjects

Twenty subjects of the Southern dialect (8 males, 12 females) with no known auditory deficiencies participated in the perception experiment. Listeners in this study were all tertiary students between 23 to 39 years of age.

4.3. Procedure

In the perception test, subjects listened to the test tokens twice each and were asked to judge the rhythmic pattern of the poetic line and put a slash after the syllable where they thought the rhythmic beat fell. A sample of the answer sheet is as follows:

Hôm qua / tát nước / đầu đình /

The experiment was carried out using a laptop computer with loud speakers in a quiet classroom at Can Tho University located in the delta to the southwest of Sài Gòn.

4.4. Analysis

There were five data sets of poetic lines (three-word lines, five-word lines, six-word lines, seven-word lines, and eight-word lines). Separate mixed effect ANOVAs were conducted for each data set. The dependent variable was the percentage number of responses (the number of listeners over the total listeners) for each stimulus token. The independent variables were rhythmic patterns and speakers. The random effect was poetic lines or stimulus tokens.

A second type of analysis involved the correlation between the acoustic parameters, namely duration and intensity, and the perception scores. Perception scores which were equal to the proportion of listeners who marked that syllable/word as having a stress beat were calculated for each syllable in the poetic lines,

namely, three-word, five-word, six-word, seven-word and eight-word lines. Then Pearson correlations were computed between the acoustic parameters and the perception scores for each set of the poetic line data. Results are presented in Table 2 at the end of the next result section.

4.5. Results

4.5.1 Three-word lines

The ANOVA result on three-word poetic line set showed no significant effect for the main factors : rhythmic patterns ($F(1,44) = 0.33$, $p = 0.62$, ns) and speakers ($F(1,44) = 0$, $p = 0.98$, ns), but the interaction effect patterns x speakers was significant: ($F(1, 44)= 12$, $p < 0.01$).

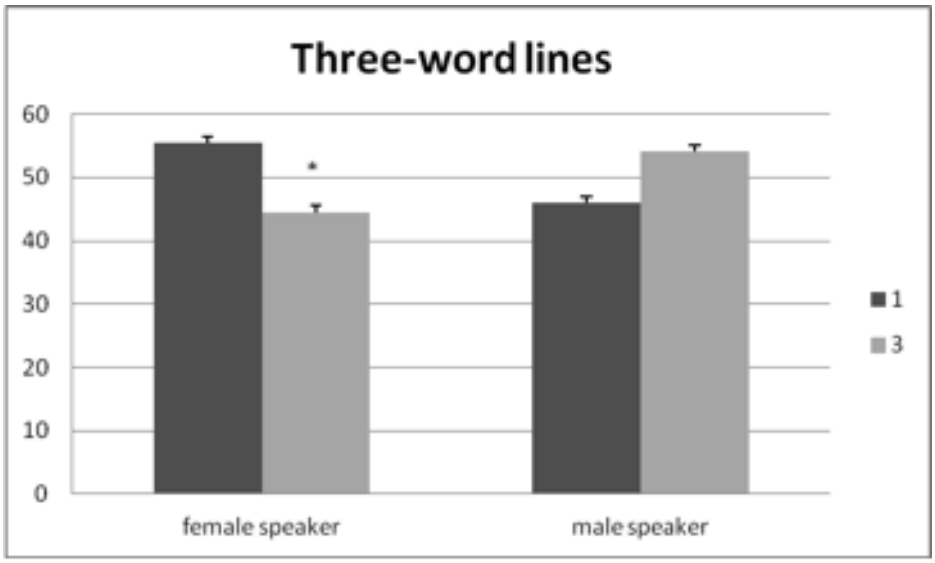


Figure 16. Mean response percentage of the three-word line data set. 1 and 3 are rhythmic patterns (1: the rhythmic beat judged to fall on the first syllable or 3: on the third syllable). The y-axis: mean response percentage. * indicates significance at $p < 0.01$.

As shown in figure 16, two rhythmic patterns were judged for the three-word poetic lines: the rhythmic beat fall on the first syllable or on the third syllable. Nevertheless, the agreement rate among listeners for each rhythmic pattern is not high: there was no significant difference between the two rhythmic patterns for the male speaker's stimuli, there was a significant effect for female speaker's stimuli but the difference is not large (56% on first syllables vs. 44% on third syllables). This result shows that the three-word lines can be heard as either two rhythmic units (one | two) by some listeners or one unit of three words together by other listeners, indicating that the intensity and duration cues found on the first syllable, though salient acoustically, can only be detectable in perception by some Vietnamese listeners but not all. On the other hand, this may also indicate that the three-word line is not long enough to carry at least one repetition of the basic unit of rhythm. It is found by neurophsyiological and psychoacoustic research that there exist "sliding windows of temporal integration" of approximately the size of 150-200ms and a window of roughly 400-600ms can be regarded as temporal present and the time span of the temporal present co-occurs with typical foot length across a variety of languages (Wagner, forthcoming). The mean duration of the three-word lines is 862.53ms, which is greater than the "temporal present."

4.5.2 Five-word lines

The ANOVA result on five-word poetic line set showed a significant effect for the main factor rhythmic patterns ($F(3,57) = 19.5$, $p < .0001$) but no significant effect for speakers ($F(1,57) = 0.02$, $p = 0.9$, ns) and the interaction effect patterns x speakers: ($F(3, 57)= 0.59$, $p =0.7$, ns). As shown in figure 17 below, four different rhythmic patterns were perceived for the five-word poetic line: on the second syllable (16%), on the third syllable (50%), on the fifth syllable (19%) and on the first, third, and fifth syllables (14%). Even though

the agreement rate was not very high, generally a majority of the five-word lines were parsed on the third syllable: a three | two pattern, consistent with the duration and intensity results and supporting the literature of an anapaest followed by an iamb pattern. Only the last four lines of the poem were parsed on the second syllable with a two | three pattern by some listeners. This could be due to the syntactic structure of the lines. Apparently few listeners could attend to the fine acoustic cues in parsing the lines at 1st 3rd and 5th syllables or into one | two | two pattern.

Figure 17. Mean response percentage of the five-word line data set. 2, 3, 5 and 135 are rhythmic patterns. The y-axis: mean response percentage. * indicates significance at $p < 0.01$.

4.5.3 Six-word lines

The ANOVA result on six-word poetic line set showed a significant effect for the main factor rhythmic patterns ($F(3, 62) = 7.8$, $p < .0001$) and the interaction effect patterns x speakers: ($F(2, 62) = 8.5$, $p < .0001$) but no significant effect for speakers ($F(1,62) = 0.7$, $p = 0.3$, ns). As shown in figure 18 below, only 46% listeners could detect the acoustic cues (duration and intensity) in parsing the six-word poetic lines in detail into two-syllable iambs (at 2nd, 4th, and 6th syllables or two | two | two pattern). By contrast, 42% listeners parsed the lines at the 2nd syllable and 12% parsed at the 4th syllable.

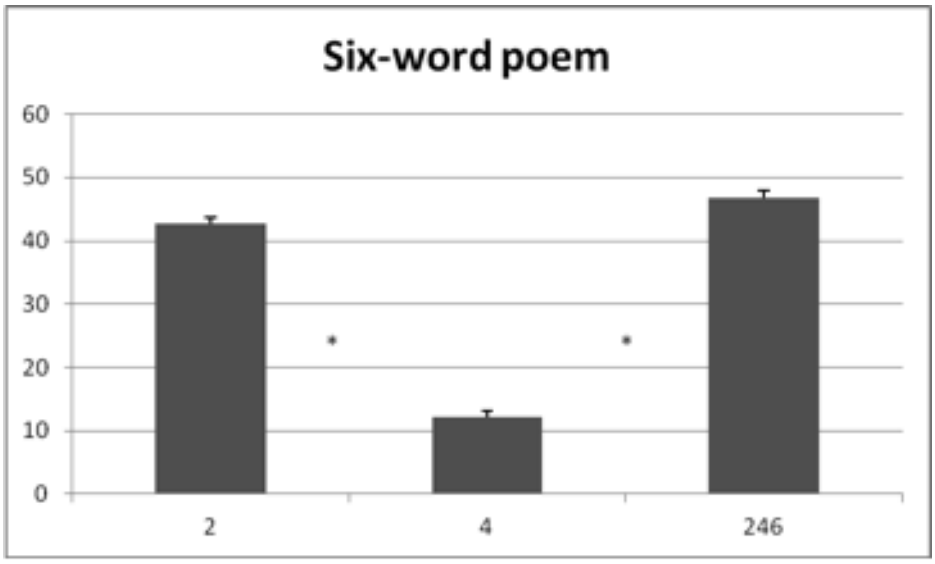


Figure 18. Mean response percentage of the six-word line data set. 2, 4 and 246 are rhythmic patterns. The y-axis: mean response percentage. * indicates significance at $p < 0.01$.

4.5.4 Seven-word lines

The ANOVA result on seven-word poetic line set showed a significant effect for the main factor rhythmic patterns ($F(4,42) = 95$, $p < .0001$) but no significant effect for speakers ($F(1,42) = 0.17$, $p = 0.7$, ns) and the interaction effect patterns x speakers: ($F(3, 42) = 0.11$, $p = 0.9$, ns.). As shown in figure 19, 81% of listeners parsed the seven-word lines at the 3rd syllable and into a three | four pattern, while only 7% of listeners parsed at the 3rd and 5th syllables into a three | two | two pattern, or as an anapaest followed by two iambs, as suggested by the acoustic results and as claimed by the literature (Durand and Nguyễn 1985).

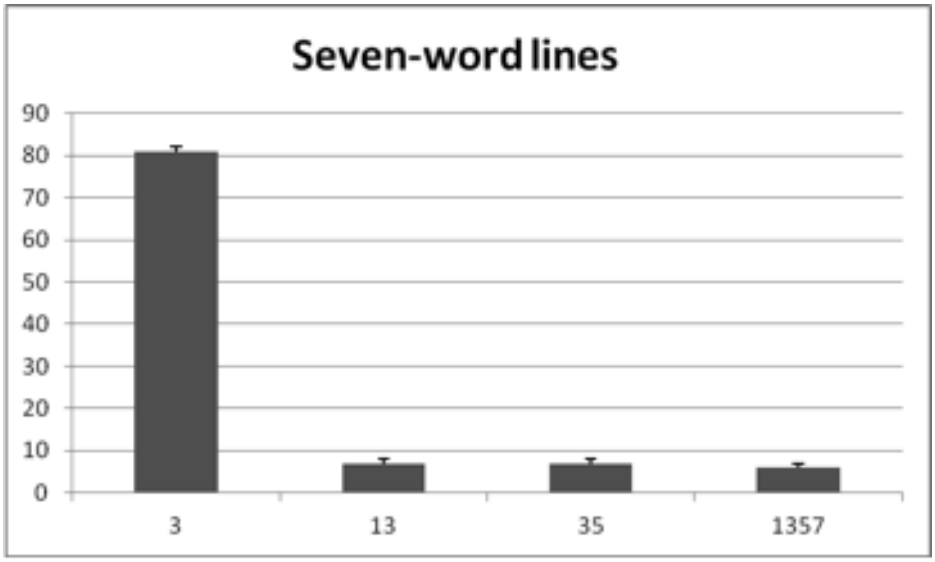


Figure 19. Mean response percentage of the seven-word line data set. 3, 13, 35 and 1357 are rhythmic patterns. The y-axis: mean response percentage. * indicates significance at $p < 0.01$.

4.5.5 Eight-word lines

The ANOVA result on eight-word poetic line set showed a significant effect for the main factor rhythmic patterns ($F(3,69) = 47$, $p < .0001$) but no significant effect for speakers ($F(1,69) = 0.27$, $p = 0.8$, ns) and the interaction effect patterns x speakers: ($F(3, 69) = 1.3$, $p = 0.3$, ns.). As shown in figure 20, 62% of listeners parsed the eight-word lines at the 4th syllable consistent with the line-medial prominence in the acoustic results, while only 15% of listeners could detect the acoustic cues in parsing the lines at 2nd, 4th, 6th, and 8th syllables or into two-syllable iambs as suggested by the acoustic result.

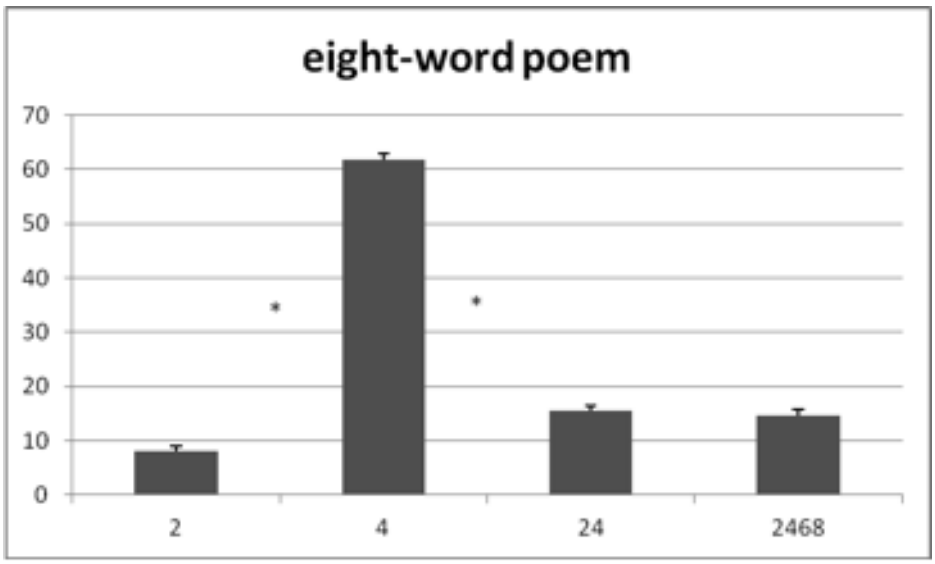


Figure 20. Mean response percentage of the eight-word line data set. 3, 13, 35 and 1357 are rhythmic patterns. The y-axis: mean response percentage. * means significance at $p < 0.01$.

4.5.6 Correlation results

As shown in table 2 below, there was a significant correlation between duration and perception scores for all poetic lines while the correlation between intensity and perception scores was not significant. This indicates that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used.

Table 2. Results of Pearson correlation between duration and intensity and perception scores, ns: non-significance.

Poetic lines	**Duration**	**Intensity**
Three-word lines	r = 0.45, p < .0001	r = 0.16, p = 0.19 ns.
Five-word lines	r = 0.37, p < .0001	r = 0.06, p = 0.51 ns.
Six-word lines	r = 0.27, p < .001	r = 0.11, p = 0.16 ns.
Seven-word lines	r = 0.33, p < .0001	r = 0.13, p = 0.17 ns.
Eight-word lines	r = 0.28, p < .0001	r = 0.05, p = 0.49 ns.

4.6. Discussion and conclusion

The perception results showed that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used. Even though the acoustic results suggest iambic patterns of prominence for the poetic lines, supporting the literature claim about the poetic rhythmic pattern (Durand and Nguyễn 1985) and strongly implying that syllables in Vietnamese folk poetry tend to be grouped into iambic bi-syllabic foot, the perception experiment showed that majority of listeners were not finely tuned to these acoustic cues and only a few listeners could detect them in parsing the poetic lines into detail bi-syllabic iambic units.

Why listeners in Nguyễn (2010) could parse the polysyllabic reduplicative pseudo-word into iambic patterns as reflected in the duration and intensity cues but most listeners of this study could not parse the poetic lines in detail is probably due to the structure of the stimuli. Stimuli in Nguyễn (2010) were polysyllabic words with constant segmental makeup and tones, while the syllables in poetic lines in this study were of different segmental composition and tones. This is consistent with the result of a perception study (Nguyễn and Ingram, submitted) in which Vietnamese listeners listened to disyllabic words (disyllabic compounds, phrases, full reduplications and tone sandhi reduplications) and judged which syllable of the sequence (the first or the second) was more prominent or if both syllables were of equal prominence. The results show that subjects performed at chance level for most of the dissyllabic word types and only the tone sandhi forms (with constant segmental makeup but with an alternation of tones) were judged to have more prominent on the second syllable. This contradicts results showing that listeners in other languages, such as English, can hear stress contrasts regardless of syllable segmental composition (Fry 1958). This suggests that prosodic asymmetry in Vietnamese is a 'sub-perceptual' threshold phonetic effect and can only be detected by some individuals with phonetically fine-tuned ears.

In the production study, the duration and intensity results mirrored each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). The results of this experiment also mirror the findings on acoustic and perceptual rhythmic pattern of polysyllabic word (Nguyen 2010) and support the iambic acoustic pattern in a subset of bi-syllabic compounds and reduplications investigated (Nguyen and Ingram 2007a, 2007b). This suggests that syllables in Vietnamese folk poetry tend to be grouped into an iambic bi-syllabic foot. This is supported by the isochronous iambic foot pattern regardless of the segmental structure of the disyllabic sequence in the six-eight metric, consistent with Latvian and English metric feet (Lehiste 1990a). On the other hand, in the seven-word line, the first three-syllable foot was much longer than its following bi-syllabic feet, suggesting foot duration increases with the number of syllables, consistent with Shen & Peterson (1962), Bolinger (1965), Uldall (1971), and Nakatani *et al*. (1981). It is likely, that "poetic speech maximizes the language specific rhythmical constraints which are probably often violated in

spontaneous speech. Thus, poetic speech can be said to be characterized by a maximal level of rhythmical harmony" (Wagner, forthcoming).

Acknowledgements
I would like to thank the subjects for their voluntary participation in the experiments. Thanks to Mrs. Tho Nguyen for administering the perception experiment. Thanks to Dr. John Ingram and the anonymous reviewers for their insightful feedbacks.

References

Abercrombie, D. 1967. Elements of General Phonetics, Edinburgh University Press.

Allen, G. D. 1973. Segmental timing control in speech production. Journal of Phonetics, 1, 219-222.

Allen, G. D. 1975. Speech Rhythm: its Relation to Performance Universals and Articulatory timing. Journal of Phonetics, 3, 75-86.

Arvaniti, A. 2009. Rhythm, Timing and the Timing of Rhythm. Phonetica, 66, 46-63

Asu, E. L. & Nolan, F. 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In: Proceedings of Speech Prosody 2006, Dresden, Germany, 2-5 May 2006.

Balaban, J. 1980. Ca Dao Việt Nam. *NhipSong Magazine*. Mosaic Press/Valley Editions, Canada.

Benton, M. 2010. A Preliminary Analysis of the Relationship of Speech Rate to Speech-Timing Metrics as applied to Large Corpora of Non-Laboratory Speech in English and Chinese Broadcast News. In: M. Hasegawa-Johnson eds.. Proceedings of Speech Prosody 2010 , Chicago, 11-14 May 2010.

Bertinetto, P. M. 1977. "Syllabic Blood", ovvero l'italiano come lingua ad isocronismo sillabico. *Studi di Grammatica Italiana*, 6, 69-96.

Bertinetto, P.M. & Bertini, C. 2008. On modeling the rhythm of natural languages. In: P.A. Barbosa, S. Madureira & C. Reis (eds.). Proceedings of Speech Prosody 2008, Campinas (Brazil), 6-9 May 2008, 427-430.

Bertinetto, P.M. & Bertini, C. 2010. Towards a unified predictive model of Natural Language Rhythm. In M. Russo (ed.) Prosodic Univers als. Comparative Studies in Rhythmic Modeling and Rhythm Typology. Rome: Aracne.

Boersma P. & Weenink D. 2007. Praat: doing phonetics by computer (Version 4.5.18) [Computer program]. Retrieved March 30, 2007, from http://www.praat.org/

Cao, X. H. 2003. Vietnamese: issues of phonology, grammar and semantics. Education press.

Cassidy, S. 1999. Compiling multi-tiered speech databases into the relational model: Experiments with the Emu system. In G. Gordos ed., *Proceedings of Eurospeech 99*, pp. 2239-2242.: (European Speech Communication Association, Budapest, 1999.

Chaudhary, C. C. 1983. Word stress in Vietnamese: A preliminary investigation. Indian Linguistics 44:1-10.

Cong Huyen Ton Nu, Nha Trang. Versification of Vietnamese Riddles. www.geocities.com/chtn_nhatrang/riddles.html

Cutler, A. 1994. The perception of rhythm in language. Cognition, 50, 79–81.

Darwin, C. J., & Donovan, A. 1980. Perceptual studies of speech rhythm: Isochrony and intonation. In J. C. Simon (Ed.), Spoken language generation and understanding (pp. 77-85). Dordrecht: Reidel.

Dauer, R. M. 1983. Stress-timing and Syllable-timing Reanalysed. Journal of Phonetics, 11, 51-62.

Dellwo, V. & Wagner, P. 2003. Relations between language rhythm and speech rate. In: Proceedings of the 15th International Congress of Phonetics Sciences , Barcelona (Spain), 3-9 August 2003, 471-474.

Dellwo, V. 2006. Rhythm and Speech Rate: a variation coefficient for C. In: Language and language Processing. Proceedings of the 38th Linguistic Colloquium Piliscsaba , 231-241.

Do, T. D. 1986. Elements pour une e'tude comparative de l' intonation en Francais et en vietnamien: L'accent de mots en vietnamien. Memoire de DEA (Universite' de Paries 3 ILPGA.

Donovan, A., & Darwin, C. J. 1979. The perceived rhythm of speech. In Proceedings of the Ninth International Congress of Phonetic Sciences (Vol. 2, pp. 268-274). Copenhagen: Institute of Phonetics.

Durand, M. Maurice, and Nguyen, Tran Huan 1985. *An Introduction to Vietnamese Literature*, translated from the French by D.M. Hawke. New York: Columbia University Press.

Fraisse, P. 1974. Psychologie du rythme. Paris: Presses Universitaires de France.

Fraisse, P. 1982. Rhythm and timing. In D. Deutsch (Ed.), The psychologyof music (pp. 149-180). New York: Academic Press.

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech 1*, 126–152.

Galves, A., Garcia, J.E., Duarte, D., Galves, C. 2002. Sonority as a Basis for Rhythmic Class Discrimination. In: Proceedings of Speech Prosody 2002 , Aix-en- Provence France), 11-13 April 2002

Giordano, R. & D'Anna, L. 2010. A comparison of rhythm metrics in different speaking styles and in fifteen regional varieties of Italian. In: M. Hasegawa-Johnson (eds.) Proceedings of Speech Prosody 2010, Chicago USA), 11-14 May 2010.

Grabe, E., & Low, E. L. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhover, C., Warner, N. (eds.), Papers in Laboratory Phonology 7, Berlin: Mouton de Gruyter, 515-546.

Gsell, R. 1980. Remarques sur la structure de l' espace tonal en Vietnamien du sud (Parler de Saigon). *Cahiers d'etudes Vietnamiennes, 4* (Universite' Paris 7).

Hoàng, T.; Hoàng, M. 1975. Remarques sur la structure phonologique du Vietnamien. Essais Linguistiques. *Etudes Vietnamiennes 40*.

Jones, R. B. and Huỳnh, S. T. 1960. *Introduction to Spoken Vietnamese*. Washington, D.C

Lehiste, I. 1977. Isochrony reconsidered. Journal of Phonetics, 5,253-263.

Lehiste, Ilse 1984. "The metric structure of a recited Finnish spell." Ural-Altaische Jahrbücher 4, pp. 83-89.

Lehiste, Ilse 1986. "Phonetic realization of metrical structure in some types of Estonian verse." Ural-Altaische Jahrbücher 6, pp. 11-20.

Lehiste, Ilse 1990a. Phonetic investigation of metrical structure in orally produced poetry. *Journal of Phonetics 18*, pp. 123-133.

Lehiste, Ilse 1990b. An acoustic analysis of the metrical structure of orally produced Lithuanian poetry. *Journal of Baltic Studies 21,* 2, 145-156.

Nguyễn Đ. L. 1970. *Four-syllable idiomatic expressions in Vietnamese* Pacific Linguistics, Series D – Special Publications 6. Canberra: Australian National University.

Nguyễn , T.A.T. and Ingram J.C. 2007a. Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America, 112* 3), 1746-1757.

Nguyễn , T.A.T. and Ingram J.C. 2007b. Stress and tone Sandhi in Vietnamese reduplications. *Mon-Khmer studies Journal, 37*, 15-40.

Nguyễn , D. L. 1970. *A contrastive phonological analysis of English and Vietnamese* Pacific Linguistics Series, No 8. Canberra: Australian National University.

Nguyễn , T.A.T. & Ingram, J.C. Submitted. Perception of prominence pattern in Vietnamese dissyllabic words.

Nguyễn, T.A. T. 2010. Rhythmic pattern and corrective focus in Vietnamese polysyllabic words. *The Mon-Khmer Studies Journal, 39*, 1-28.

Nguyễn, Xuân Kính. 2004. *Thi pháp ca dao* (Folk poetry metric). NXB. ĐHQG Hà Nội Vietnam National University Hanoi publisher.

O'Dell, M. L., Nieminen, T. & Mustanoja, L. 2010. Assessing Rhythmic Differences with Synchronous Speech. In: M. Hasegawa-Johnson (eds.) Proceedings of Speech Prosody 2010, Chicago USA), 11-14 May 2010.

Pike, K. L. 1945. The Intonation of American English . Ann Arbor. University of Michigan Press.

Ramus, F. & Mehler, J. 1999. Language identification with suprasegmental cues: a study based on speech resynthesis. Journal of the Acoustic Society of America, 105 (1), 512-521.

Ramus, F., Nespor, M. & Mehler, J. 1999. Correlates of Linguistic Rhythm in the Speech Signal. Cognition , 73/3, 265-292.

Thompson, L. 1965. *A Vietnamese Reference Grammar*. Honolulu: University of Hawaii Press.

Wagner, P. & Dellwo, V. 2004. Introducing YARD (Yet Another Rhythm Determinator) and Re-Introducing Isochrony to Rhythm Research. In: Proceedings of Speech Prosody 2004 , Nara Japan), 23-26 March 2004.

Wagner, P. 2010. Two Sides of the Same Coin? Investigating Iambic and Trochaic Timing and Prominence in German Poetry. Proceedings of Speech Prosody 2010, Chicago, P1.47.

Wagner, Petra. Forthcoming. The rhythm of language and speech: Constraints, Models, Metrics and Applications.

White, L., & Mattys, S.L. 2007. Calibrating rhythm: First language and second language studies. Journal of Phonetics, 35 , 501-522.

Appendix: Folk poems

NB: The bolded words were those used for the analysis of vowel formants.

(1) *Tat nuoc dau dinh poem*

Hôm qua tát nước đầu đình
Bỏ quên cái áo **trên cành** hoa sen
Em được **thì cho** anh xin
Hay là em để làm tin trong nhà?
Áo anh **sứt chỉ** đường tà
Vợ anh chưa có, **mẹ già** chưa khâu
áo anh sứt chỉ đã lâu
Mai mượn cô ấy về khâu **cho cùng**.
Khâu rồi anh sẽ **trả công**
Đến lúc có chồng anh lại giúp cho
Giúp em một thúng xôi vò
Một con lợn béo, một vò rượu tăm.
Giúp cho đôi chiếu em nằm
Đôi chăn em đắp, đôi tằm em đeo
Giúp em quan tám tiền cheo
Quan năm tiền cưới lại đèo buồng cau.

(2) *The first four stanzas of the Chinh Phu Ngam*

Thuở trời đất nổi cơn gió bụi
Khách má hồng nhiều nỗi truân chuyên
Xanh kia thăm thẳm tầng trên
Vì ai gây dựng cho nên nỗi này
Trống Trường Thành lung lay bóng nguyệt
Khói Cam Tuyền mờ mịt thức mây
Chín tầng gươm báu trao tay
Nửa đêm truyền hịch định ngày xuất chinh
Nước thanh bình ba trăm năm cũ
áo nhung trao quan vũ từ đây
Sứ trời sớm giục đường mây
Phép công là trọng, niềm tây sá nào
Đường giong ruổi lưng đeo cung tiễn
Buổi tiễn đưa lòng bận thê noa
Bóng cờ tiếng trống xa xa
Sầu lên ngọn ải, oán ra cửa phòng

(3) *Five-word folk poem*

Ai qua cầu rơi dép
Ai xuống cầu nhặt giùm
Nhặt ba lần không nản

Vẩn kính cận ung dung
Biết người có trí lớn
Là một đấng anh hùng
Trao cho sách Thái ất
Về nhà học thuộc lòng
Ngày sau ra giúp nước
Lập được đại kỳ công

(4) *Three-word riddles and sayings*

Một mẹ nằm
Trăm con bước.
bằng cái vung
Vùng xuống ao
Đào không thấy
Lấy không được.
Đói ăn rau
Đau uống thuốc.
Vào mồng ba
Ra mồng bẩy
Rẫy mồng tám.

ZAKARIA, Muhammad. 2013. *On the Origin of Hyow.*
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:78-86
Received 3/3/2013, text accepted 29/7/2013, published August 2013
URL: http://hdl.handle.net/1885/10277
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

SOUTHEAST ASIAN LINGUISTICS SOCIETY SEALS
www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics
Australian National University

ON THE ORIGIN OF HYOW

Muhammad Zakaria
East West University, Dhaka
<rehman.zakaria@gmail.com >

Abstract
The origin of Hyow, one of the Kuki-Chin languages, in the current study of the language family is not a conclusive one. This study aims to compare the previous findings and the research data of Hyow that is collected from Chittagong Hill Tracts in Bangladesh. Based on the shared innovations of the Kuki-Chin languages and the findings of my research, I have tried to draw a conclusive origin of Hyow in the Southern Kuki-Chin branch. The comparative study reveals that Hyow should be aligned with Khumi and Cho-Asho rather than be kept under Asho of the Southern-Plains-Chin group.
Key words: Proto Kuki-Chin, Southern-Plains-Chin, Hyow.
ISO 639-3 language codes: rmz, cfm, csh, cnk.

1. Introduction

This paper aims to give a description of the origin of Hyow. Hyow belongs to Kuki-Chin group of Tibeto-Burman branch. Kuki-Chin is divided into different sub-groups according to the shared proto features of the languages. We can find different sub groupings of Kuki-Chin, but the recent one by Kenneth VanBink (2009) is more convincing and correct. Due to the unavailability of enough data Kenneth could not provide with enough information on Hyow. In this paper, I will try to restate the position of David A. Peterson (2000) on Hyow in the group through detail examples and comparative discussion. The data used in this article are taken from my field notes that I have been intermittently collecting for last six years under the supervision of David A. Peterson. I have worked with three consultants, but my main consultant is Hla Kroy Prue who is the first woman to complete BSS among the Gungrupara Hyows. I have also consulted with Hla Kroy's husband, Nio Jai, Assistant Headmaster of a primary school in Banderban, for several occasions. I owe a great deal to this couple for making themselves available whenever I requested them to. I really want to thank David A. Peterson for his continuous support through the fund of NSF and by mentoring me to do my research work on Hyow for a long period of time.

2. The name Hyow

The speakers of this language call themselves 'Hyow' and are called 'Khyang' by other indigenous people and the Bangalees. The history indicates that 'Khyang' form was first used by the Burmese to call the Asho-Chin people. As the Marmas live in a great number around the Hyows, it can be assumed that the Burmese group started to call them 'Khyang' first and later on started to be used by others. According to Luce (1959:25) the name 'Khyang' is the archaic form of the term 'Chin' which means 'friends' in modern Burmese. Kenneth VanBik (2009:4) states that the origin of this name can be traced in the languages of Asho-hin. According to Joorman (1906:12), the word for 'person' in Asho-Chin is 'hklaung' (sometimes ***khlaaŋ*** or ***khloŋ***). VanBik (2009:4) says that when the Burmese met the Asho Chin people, they used to call them by this name. But, as the Burmese had already lost the ***khl-*** cluster and the closest approximation was ***khy-***, the word 'Khyang' denoted any Chin Group.

The word 'hyow' has cognates in other Kuki-Chin languages like 'zaw (Mizo) and 'zaw (Lai). Amy Campbell (2007:4) identifies 'zaw' as human constituent question marker in Falam as in example (1) and (2).

(1) *zaw* *ni* *zaw* *a-ŋam*
who ERG who 3sgS-kiss
'Who kissed whom?'

(2) *zaw* *ni* *ziang* *kha* *zaw* *si:* *a-pe:k*
who ERG what PAT who FOC 3sgS-give
'Who gave what to whom?'

So, I would say that 'zaw' is convincing basic form for all these cognates in Kuki-Chin languages. The PKC palatal 'y' is retained in Southern Plains languages with some exceptions in Hyow. It has free variations 'y' and 'z' in Northern Chin languages and in Central Chin languages it is 'z'. For the sonorant final words in Hyow this palatal 'y' has become voiceless. In some places PKC *y is retained in Hyow (see section *5.2.3*).

2.1 PKC *y > SPKC y > Hyow hy

From the following table it is quite evident that PKC 'y' is retained in SPKC languages other than Hyow. In Hyow it has become voiceless. So, Hyow has its own innovation of the voiceless palatal 'j' and evidently the name 'Hyow' means 'person' which has cognates 'zaw' or 'zo' in Central Kuki Chin languages.

Table 1: *PKC *y in SPKC and Hyow*

PKC	NKC	CKC	SPKC	Hyow	Gloss
*yuu	zûu (Thado Kuki)	zùu (H.Lai)		hyu	mouse
*yuun		zuun (H.Lai)	yún (Asho)	hyun	creeper
*yiing	khua zing (Tedim)			hyung	dark
*yaang-*X* *yaan-*Y*	zâang-*X* zaan-*Y*	zaang-*X* zaan-*Y* (H.Lai)	yang (Asho)	hyɔng	light (weight)
*yuul	zuuy-*X* zuy-*Y* (Thado Kuki)	zuul-*X* zulh-*Y* (Mizo)	yun (M.Cho)	hyul	follow
*yuum	zum (Tedim)	zum[2] (Mizo)	m-yum	hyum	Pointed (sharp)
*yaa-*X* *yaat-*Y*	zaa-*X* zaat-*Y*	za-*X* za-*Y*	ya	hyɔ	itch
*yun	zun[3] (Tedim)	zún (H.Lai)	pyun-dűi (Asho)	hyun	urine

3. Early observations

In Linguistic survey of India Sten Konow put Kuki-Chin under Naga-Kuki-Chin and Robert Shafer (1974) called it Kukish putting it in the second level of his schema. The schema of Paul K. Benedict shows Kuki-Chin under Kuki-Naga group. The recent schema of James A. Matisoff (2003) considers Kuki-Chin under Kamarupan of Tibeto-Burman branch. Kenneth VanBik (2009) followed Matisoff's schema and went on further sub grouping of the Kuki-Chin group. According to VanBik, "There are two shared innovations that separate Kuki-Chin languages from the rest of the Tibeto-Burman family: a thorough going verbal stem alternation and a phonological change of the PTB initials *s/sy to PKC *th." He also emphasizes the study of David A. Peterson (2001) who proposed the idea of peripheral subgroup. The schema of VanBik is shown below.

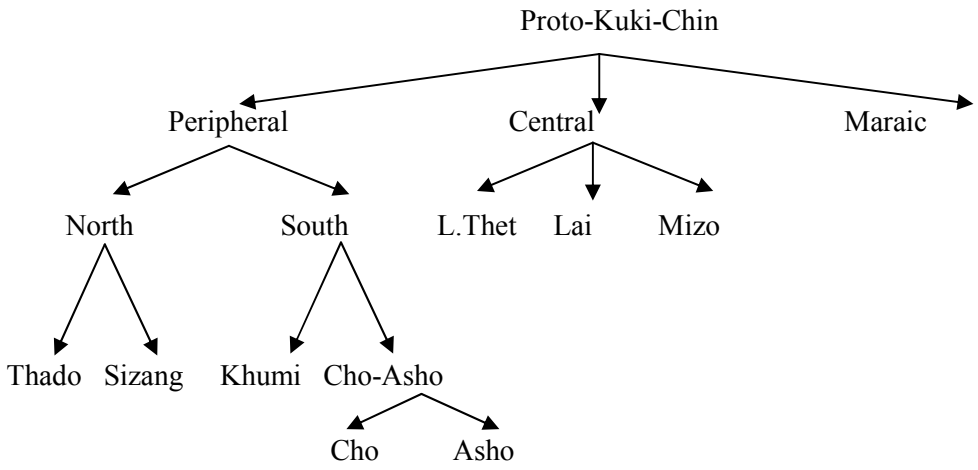


Figure 1: *Kenneth VanBik's Sub-grouping*

4. Status of Hyow in PKC

Other than the verb stem alternation (see ***4.1***), another phonological innovation PTB *s/sy > PKC *th makes the group separated from the others in the PTB branch.

Table 2: *PTB *s/sy > PKC *th > Hyow th*

PTB	PKC	Maraic (Mara)	Central (H.Lai)	Southern (Hyow)	Northern		Gloss
					(Tedim)	T.Kuki	
*m-sin (34) WB: sâŋ	*thin	pa-thī	thĭn	thin	sin3	thìn	liver
*sey (129)WB: sî	*thay	théy	thày	thɛy		thêy	fruit
*sat (136) WB: sat	*that	that-*X* tha?-*Y*		ʃɔt	that3	thàt-*X* thàa-*Y*	kill
*g-sum (36) WB: sûm	*thum	pa-thò	thûm		thum3	thúm	liver

5. Place of Hyow in the subgroup

The identical characteristics of Hyow to be a Southern-Kuki-Chin language will be discussed in this section. Further discussion will be done to determine the place of Hyow within the subgroup.

5.1. Stem alternation

One of the main common features of Kuki-Chin languages is that there are two forms of verb stems. This stem alternation is also found in Hyow. And there are many linguistic environments which show the verb stem alternations.

5.1.1. Affirmative-Negative

In Hyow verb shows Form *X* in Affirmatives and Form *Y* in Negatives as in the following examples.

(3) *l̥akrɔy la hɔ luak tuk kʰɔ* (Form Y)
Hlakroy ERG bird CLS[1] kill AUX
'Hlakroy killed a bird'

(4) *l̥akrɔy la hɔ luak tu a hɔ* (Form *X*)
Hlakroy ERG bird CLS kill NM TM

[1] 'luak' is glossed as CLS (numeral classifier).

'Hlakroy did not kill a bird'

(5) *kei naŋ koŋ kɔ l̥ɔy* (Form *Y*)
I you for VMP try
'I will try for you'

(6) *kei naŋ koŋ lɔ ŋa* (Form *X*)
I you for try NM
'I will not try for you'

(7) *pʰöl la döʔ hɔ* (Form *Y*)
snake ERG strike TM
'The snake is striking'

(8) *pʰöl la dö a hɔ* (Form *X*)
snake ERG strike NM TM
'The snake is not striking'

5.1.2. Clause

In main clause, Form *X* can be found in Hyow as in example (10).

(9) *naŋ ei kʰrak boi* (Form *Y*)
you him harm TM
'You will harm him'

(10) *kei kɔ mɔt kʰö naŋ ei na kʰra ai hjɔ* (Form *X*)
I VMP know TM you him that harm TM AUX
'I know that you will harm him'

5.1.3. Tense

Form *Y* can be found in present tense and Form *X* in future as in example (11) and (12).

(11) *ei kʰol ʃel le a ini ʧet tʰɔ* (Form *X*)
they (PL) cow buy DIR VMP go TM
'They are going to buy cow'

(12) *ei la ʃel hat e l̥ey ei hjɔ* (Form *Y*)
he ERG cow one VMP buy TM AUX
'He will buy a cow'

5.1.4. Transitivity

In Kuki-Chin languages except Hyow, valence changing operations take place at the final position of a stem as in example (13) and (14) from Falam Chin (Deborah King, 2010). On the other hand, Hyow valence increasing is done by devoicing the initial consonant of the verb as in the examples (15), (16), (17) and (18).

(13) *A zir* (Form *X*)
he learn
'He learned'

(14) *Ka* *lo* *zirh* (Form *Y*)
I you teach
'I teach you'

(15) *kei* *ke* *nek* (Form *X*)
I VMP sink
'I sink'

(16) *kei* *ani* *ke* *n̥ek* (Form *Y*)
I him VMP sink (transitivization)
'I make him sink'

(17) *kei* *kɔ* *lɔm* (Form *X*)
I VMP dance
'I dance'

(18) *kei* *ani* *kɔ* *l̥ɔm* (Form *Y*)
I him VMP dance (transitivization)
'I make him dance'

Table 3: *Stem alternation in Hyow*

Parameters	Form *X*	Form *Y*	Gloss
Affirmative-Negative	tu	tuk	kill
	l̥ɔ	l̥ɔy	try
	dö	döʔ	strike
Tense	l̥e	l̥ey	buy
Clause	kʰra	kʰrak	harm
Valence increasing (causative)	nek	n̥ek	sink
	lɔm	l̥ɔm	dance

5.2. Sound change

5.2.1. PKC Prefix

"The Southern-plain group comprise of languages of the Khumi-Cho-Asho type that have prefixes…" VanBik (2012)

Hyow does not have mentionable prefixes like other Khumi-Cho-Asho type languages possess. There might be some clitics (used as pronouns) in some sentences as in example (19).

(19) *ei* *i=* *tʰin* *ɔ=* *tʃɔp* *ei* *-ti* *-tʃɛ* *ɔ=*
that CLIT liver CLIT lung eat who TPM CLIT
nui *-hi* *-tʃɛ* *kʰɛi* *bo* *-ai* *-hjɔ* *a=* *miak*
laugh COND TPM flower fall TM AUX CLIT dream
bo *-ai* *-hjɔ*
fall TM AUX
'he (son of constable) saw in dream, he who will eat that (hen's) liver, (hen's) lung if he laughs, flower will fall on him.'

A further study is required to draw a conclusive statement regarding prefixes and clitic in Hyow. From the examples given in the following table, it can be stated that Khumi-Cho-Asho type languages' prefixes are not found in Hyow.

Table 4: *Prefixes in Southern-Plains-Chin Languages*

PKC	Southern Kuki Chin				Gloss
	Khumi	M.Cho	Asho	Hyow	
*looy	alawy			lo	return
*laa-*X* *laak-*Y*	laa	la-*X* laak-*Y*	lö	low	to take
*lay	plaáy	m-lei		ley	tongue
*li	plúue		m‘l í	l̥i	four

*5.2.2. PKC *khl- > PSPC *khl- > Hyow kh/l/khl*

PKC *khl- is retained in some places of Hyow and in some places they have lost either /l/ or /kh/.

Table 5: *PKC *khl- in SKC and Hyow*

PKC	PSPC	Southern Kuki Chin				Gloss
		Khumi	M.Cho	Asho	Hyow	
*khlaa	*khlaa	lo	cha	k‘ló	khɔ	moon
*khlaa	*khlaa				lɔy	wing
*khlii	*khlii	alii	chi	k‘lí	khli	wind
*khluu-*X*, *khluuk-*Y*	*khluu-*X*, *khluuk-*Y*		chuk		khlu	fall over
*khlu(u)ng	*khlu(u)ng	sliiwng	chung		klung	top
*khleng	*khleng	plúue		m‘l í	khlong	arrive
*khlan	*khlan			a k‘law̆	khlɔng	sweat

*5.2.3. PKC *y- > PSPC *y- > Hyow y*

PKC *y- is retained everywhere in PSPC and also in Hyow.

Table 6: *PKC *y- in SKC*

PKC	PSPC	Southern Kuki Chin				Gloss
		Khumi	M.Cho	Asho	Hyow	
*yan	*yan		yan	a-yan	yɔn	night
*yu	*yu		yu	a- ű	yu	rice beer
*yong	*yong		yawng	yón	yong	monkey
*yum	*yum		yum	yón-é	yum	believe
*yuuk	*yuuk			yük	yök	pull out
*yuar	*yuar	jó	yawi	yi´	yæ	sell

*5.2.4. PKC final *-r > PSPC final *-y > Hyow final ø*

Hyow retains the PKC initials of words ending with ‘r’ but drops the final ‘r’.

Table 7: PKC final *-r in SKC

PKC	PSPC	Southern Kuki Chin				Gloss
		Khumi	M.Cho	Asho	Hyow	
*paar	*pay	paw	pai	pa‘a	paa	flower
*ʔaar	*ʔaay		ai		a	fowl
*thar	*thay	kthaá	ak-thai	t‘a	tha	new
*baar	*bay		bai		ba	feed with palms of hand
*khur	*khuy	khoó	khui	konk‘ő	kʰö	hole
*tsaar	*caay	ajaá	cui	sá	tʃa	to dry

*5.2.5. PKC *r > PSPC *g > Hyow y*

In some cases PSPC *g became k^h in Hyow as in 'rain'.

Table 8: *PKC *r in SKC*

PKC	PSPC	Southern Kuki Chin				Gloss
		Khumi	M.Cho	Asho	Hyow	
*rii	*gii		ng-gu		yæ[1]	border
*rii	*gii		ng- gi		yæ[2]	banyan tree
*ruak	*guak	tlángveew	gawk	aŏ	yok	corpse
*rua	*go		gaw	yó	yo	bamboo
[1]*ruung-*X* *ruu^{ʔ}-*Y*					yöng	scoop up
*ruy	*guy	vuuy	gui		yöi	rope

*5.2.6. PKC *w > PSPC *w/v > Hyow h*

PKC *w becomes 'h' at the initial position of a word in Hyow. In some words PKC *w is retained in Hyow showing the conservative feature and some distinctive innovations of the language.

Table 9: *PKC *w in SKC*

PKC	Southern Kuki Chin				Gloss
	Khumi	M.Cho	Asho	Hyow	
*wut				hi	dust
*wom	tvóeeng			hom	bear
*waa	tvoo	va 'a	waw̆-k'o	hɔ	bird
*wua-*X* *wuak-*Y*		m-vawk	wó-é	wüt	strike
*wat	tvaw			wöt	leech (land)
*wok	ew	vawk	wŏ	wok	pig
*wun		vun	u-üń	wun	skin

5.3. Southern Plains Chin

VanBik (2012) mentions that his interview with U Ba Thaung Ti, an Asho Chin from Sandaway , indicates there are about six different dialects of Asho Chin, most of them mutually intelligible. Among the dialects, Laitu and Kaitu draw my attention as I have found two dialects called Laitu and Kantu of Hyow in CHT, Bangladesh. And, this almost convinces me that Hyow is nothing but a dialect of Asho. Before doing so, let's look at VanBik's schema of Proto Southern Plains Chin.

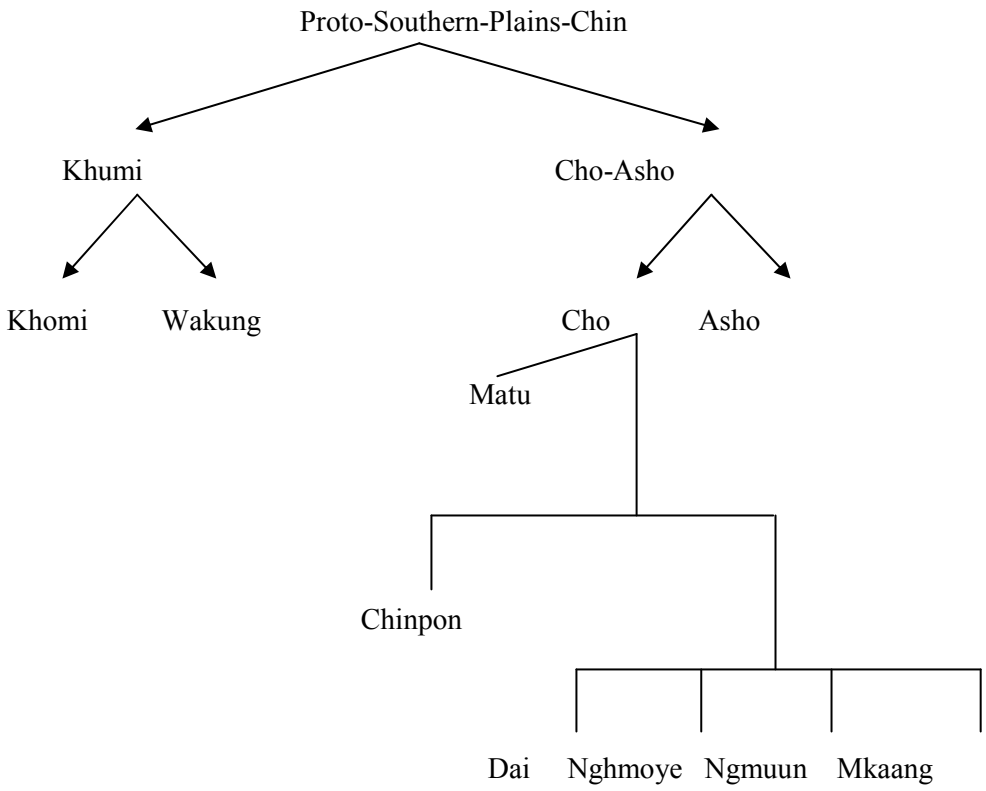


Figure 2: *Ken VanBik's schema for the Southern-Plains group*

Even though, data analysis in the above sections puts Hyow closer to Asho, for few of the features I would like to place Hyow distinctively rather than putting it under Cho-Asho. First of all, in **Table 5,** it is quite visible that other than Hyow in all the other Southern Chin languages there are prefixes such as consonant (m) or vowel (a) at the beginning of the Proto forms whereas Hyow has retained the Proto form. Secondly, PKC *khl- is retained in some places of Hyow and in some places they have lost either /l/ or /kh/ whereas other languages of PSPC have lost the Proto *khl. Thirdly, PSPC *y- is intact in Hyow even in the final nasals and vowels as in PSPC *yu> Hyow /yu/ (rice beer), PSPC *yum> Hyow /yum/ (believe), etc. Fourthly, for valance changing operations Hyow does have its own innovation of making the initial sound of the Form *X* devoiced. And, most interestingly, Hyow has its own innovation for the palatal 'y' which is also considered in this article to explain the name 'Hyow' itself. PKC palatal *y has become voiceless in words like 'hyul' (follow), hyum (pointed), etc. in Hyow. Fourthly, PKC *w has become 'h' in some Hyow words and in other words it is retained whereas in other SPKC languages this has become either 'v' or 'u'. And finally, Hyow has a very distinct way of forming its causatives. So, it is quite visible that Hyow is more conservative in its nature of retaining Proto features than the other languages in the subgroup and shows quite a number of own innovations.

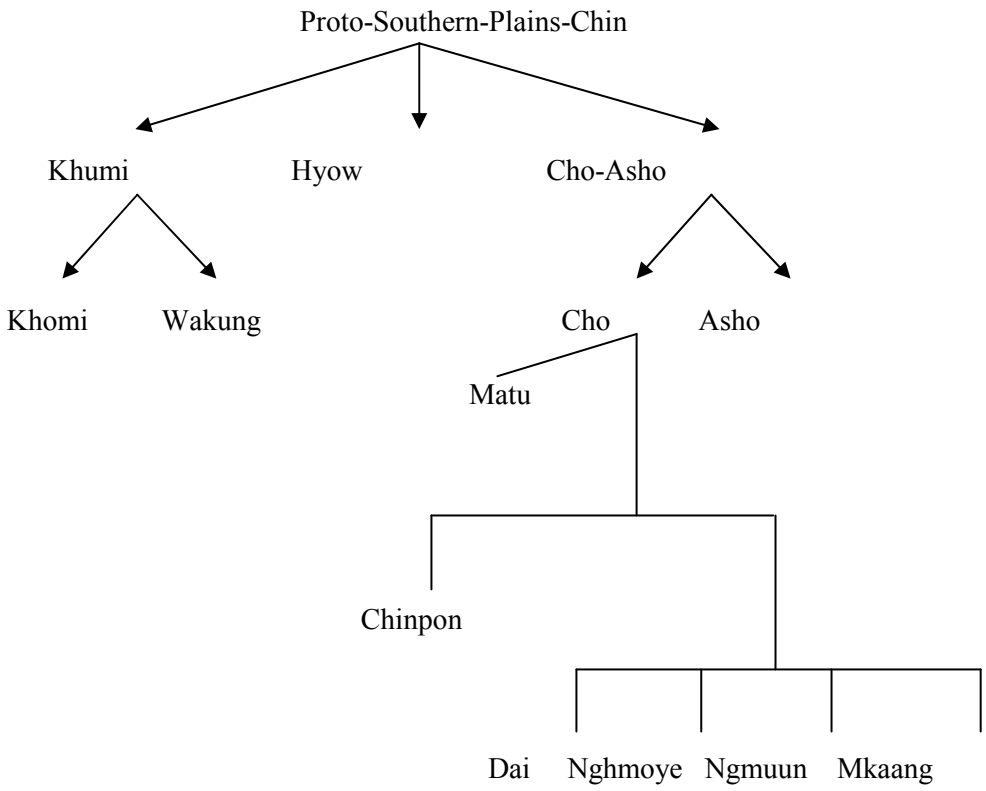


Figure 3: *Zakaria's schema for the Southern-Plains group*

6. Conclusion

In conclusion, it can be restated that Hyow has some own innovations which it doesn't share with other Southern Kuki-Chin languages. Hyow is not a dialect of Asho neither it can be put under Chin-Asho as it has quite a number of innovations which are distinctive in nature. Hence, Hyow fits best as a separate language within the Southern Plain languages.

References

Campbel, Amy. 2007. *Question and Answer in Falam*. Papers from Field Method Class, Department of Linguistics, UC Berkeley.

DeLancey, Scott. 2012. The History of Postverbal Agreement in Kuki-Chin. JSEALS.

Joorman, H. 1906. *Chin Grammar*. Rangoon: American Baptist Mission Press.

King, Deborah. 2010. Syntactic and Pragmatic Functions of Kuki-Chin Verbal Stem Alternations and the Agentive and Nonagentive Voice. JSEALS

Luce, Gordon. H. 1959. *Chin Hills-Linguistic Tour (Dec.1954)-University Project*. Journal of the Burma Research Society.

Matisoff, A. James. 2003. *Handbook of Proto-Tibeto-Burman*. University of California Press.

Payne, E. Thomas. 2006. *Describing morphosyntax*. Cambridge University Press.

Peterson, A. David. 2002. *On Khumi Verbal Pronominal Morphology*. Max Planck Institute for Evolutionary Anthropology.

Peterson, A. David. 2000. *On the Status of Southern Chin Subgroup*. The 33rd International Conference on Sino-Tibentan Languages and Linguistics, Bangkok, Thailand.

Thurgood, Graham & Lapolla, J. Randy. 2006. *The Sino-Tibetan Languages*. Routledge.

VanBik, Kenneth. 2009. *Proto-Kuki-Chin*. STEDT (http://stedt.berkeley.edu).

CHANG, Ya-Yin Melody, Paul LAW & Qing-Lian ZHAO. 2013. Plurality in Naxi and its Typological Implications.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:87-98
Received 5/3/2013, text accepted 20/7/2013, published September 2013
URL: http://hdl.handle.net/1885/10376
ISSN: 1836-6821 | Website: http://jseals.org |
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics

PLURALITY IN NAXI AND ITS TYPOLOGICAL IMPLICATIONS*

Ya-Yin Melody CHANG[1], Paul LAW[2] & Qing-Lian ZHAO[3]
[1, 2] City University of Hong Kong, [3] Yunnan Minority Languages Committee
<lilianzhao@aliyun.com>

Abstract
Plurality in Naxi can be explicitly expressed by suffixation or tone change on common nouns denoting human beings. This paper complements the published literature in providing a detailed description of its morpho-syntax and semantics of plurality. It shows that Naxi belongs to the typologically rare type of language in which the coding of plurality and definiteness is in one morpheme, the other three languages known to date to have this property are Chinese, Khmer and Maori. Evidence for the definiteness property of explicit plurality comes from it being excluded in syntactic environments in which definiteness noun phrases are ruled out. It is argued that the empirical basis of Greenberg's (1974) generalization regarding the relation between numeral classifiers and compulsory expression of nominal plurality is subject to the interpretation of explicit expression of plurality. If it is taken to embody in a morpheme (or a set of morphemes) specifically for expressing plurality, then Naxi is consistent with Greenberg's claim that languages with numeral classifiers do not have compulsory expression of plurality on nouns..
Key words: morphology, plurality, semantics
ISO 639-3 language codes: nxq

1. Introduction

Among the several ways a language indicates plurality on nouns by means of morphology, marking plurality by tone is relatively rare. In a survey of 986 languages coding nominal plurality in the World Atlas of Language Structures (WALS) (Haspelmath et al. 2005),[1] only four languages are listed as using tone marking plural on nouns. All of them are in Africa. It is in this context that Naxi, a language spoken in southwestern China, is especially interesting, for the language sometimes uses tone to mark nominal plurality, in addition to suffixation.

The purpose of this paper is three-fold. First, it provides a detailed description of the morpho-syntax of plurality as well as its semantic properties in Naxi. It complements the published literature in documenting the language. He (1987) and He and Jiang (1985) have only a few examples of nominal plurality with little description of the morphosyntactic and semantic properties of the coding of plurality in the language. Second, it compares the coding of plurality in Naxi with that of some other languages and examines the extent to which languages may differ with respect to expression of nominal plurality. It is shown that Naxi belongs to the typologically rare type of language in which plurality is expressed together with definiteness in one morpheme. Three other languages known to combine plurality and definiteness are Chinese, Khmer

* An earlier version of the paper was presented at SEALS 20. We would like to thank Lichuan He, Zhili He, Yunfang Yang, and Yongqiang Zhu for sharing their intuitions with us. All errors of fact or interpretation are our responsibilities. Our work is financially supported by grant 146010 of the General Research Fund of the Hong Kong Special Adminstrative Region government, for which we are very grateful.

1 The online map can be found at: http://wals.info/feature/33A?tg_format=map&v1=cd00&v2=c00d&v3=ca00&v4=c000&v5=cff0&v6=cccc&v7=cf60&v8=df60&v9=cfff

and Maori (cf. Dryer 2011). Third, we consider the relation between numeral classifiers and explicit coding of plurality. While the facts in Naxi appear to be consistent with the complementary distribution of numeral classifiers and (explicit) coding of plurality on nouns (cf. Chierchia 1998), we argue that the empirical basis of Greenberg's (1974) generalization regarding the relation between numeral classifiers and compulsory expression of nominal plurality is subject to the interpretation of what is meant by expression of plurality.

2. Expressing plurality in Naxi

Naxi is a Tibeto-Burman language spoken in Yunnan, southwestern China. According to Bradley (1975) Naxi belongs to the Lolo-Burmese branch of Tibeto-Burman, but the issue does not seem to be settled (Thurgood and LaPolla 2003). The variety investigated in this paper is the Western dialect, primarily in Dayanzhen in Lijiang, Yunnan, China. The data considered here are from the published literature as well as from the third author and other consultants.

Naxi is mostly a head-final language. The verb comes at the end of the VP, yielding the SOV order. PPs are postpositional but the noun comes first in noun phrases. It is an analytic language in that it has very little inflectional morphology on nouns and verbs. Nouns are not inflected for gender, case, or number, and verbs carry no morphology for tense, aspect or agreement with the subject. The focus of this paper is on nouns, primarily the coding of plurality and its morphosyntactic properties.

The **base nouns** without morphology indicating plurality can be understood to be either singular or plural; they can be either definite or indefinite, depending on the context. For example, in (1a), the noun *sɿ˩˧dzɿ˧* 'teacher' has no morphology indicating plurality. It can be understood to be singular or plural, definite or indefinite (PERF=perfective, SG=singular):[2]

(1) a. sɿ˩dzɿ˧ tsʰɿ˩ se˩
teacher come PERF
'A/The teacher has come.' or 'Teachers/The teachers have come.'

b. ŋɤ˩ çøˊ se˧ ndø˩
1SG student see
'I saw a student/students/the student/the students.'

However, plurality in Naxi can be explicitly expressed in two ways. Either the morpheme *hɤ˧* is suffixed to the base noun or by tone change from mid-level ˧ or mid-falling ˩ tone to low-rising tone ˊ (He 1987 and He and Jiang 1985):

(2) a. çi˧ 'person' → çi˧hɤ˧ 'people'
b. zy˥ 'child' → zy˥xɤ˧ 'children' (He & Jiang 1985: 44)

(3) a. bɚ˧ 'guest' → bɚˊ 'guests'
b. swe˧ 'official' → sweˊ 'officials'
c. dzɿ˧ 'friend' → dzɿˊ 'friends'

The nouns suffixed with the morpheme *hɤ˧* in (2) and those undergoing tone change from mid-level ˧ or mid-falling ˩ tone to low-rising tone ˊ in (3) are necessarily understood to be plural. The tone change in (2)-(3) can be taken to be the effect of the phonological rule in (4):

(4) *Coding plurality by tone in Naxi nouns*
N˧/N˩ → Nˊ

[2] As is well-known, there are substantial phonological differences amongst dialects. The transcriptions given here are according to the pronunciations of the third author, including those cited from the original articles, for uniformity's sake. See Pinson (2012) for the different pronunciations in different dialects. As far as we can tell, the morphosyntax and semantics of plurality has no bearing on the phonological differences.

Pluralization by suffixation or tone change may apply to multi-syllabic base nouns or even phrases, regardless of whether they are morphologically decomposable:

(5) a. çi˧-ndɯ˩ ‘adult’ → çi˧-ndɯ˩-hɤ˧ ‘adults’
person-big

b. zø˧-zy˥ ‘young man’ → zø˧-zy˥-hɤ˧ ‘young men’ (He 1987: 42)
male-young

c. pʰa˥ci˧me˧ ‘girl’ → pʰa˥ci˧me˧-hɤ˧ ‘girls’
girl

d. zy˥-zy˩ ‘small child’ → zy˥-zy˩-hɤ˧ ‘children’
child

e. mi˥-cʰy˧ ‘woman’ → mi˥-cʰy˧-hɤ˧ ‘women’
female-kind

f. sø˩-ndzɻ˧ ‘classmate’ → sø˩-ndzɻ˧-hɤ˧ ‘classmates’
study-mate

g. çi˧-mu˥ ‘elder’ → çi˧-mu˥-hɤ˧ ‘elders’
person-old

h. pɤ˩ci˧-me˧ ‘Beijing woman’ → pɤ˩ci˧-me˧-hɤ˧ ‘Beijing women’
Beijing-female

i. ʔɤ˥lø˧-ʔɤ˥na˧ ‘old people’ → ʔɤ˥lø˧-ʔɤ˥na˧-hɤ˧ ‘old people’
grandfather-grandmother (He 1987:60)

j. sɻ˩-dzɻ˧çø˩se˧ ‘teacher and student’ → sɻ˩-dzɻ˧çø˩se˧hɤ˧ ‘teachers and students’
teacher-student

(6) a. sɻ˩-dzɻ˧ ‘teacher’ → sɻ˩-dzɻ˩ ‘teachers’ (He & Jiang 1985: 44)
teacher

b. mi˥-cʰy˧ ‘woman’ → mi˥-cʰy˩ ‘women’ (He & Jiang 1985: 44)
female-kind

c. zø˧-cʰy˧ ‘man’ → zø˧-cʰy˩ ‘men’
male-kind

d. zy˥-zy˩ ‘small child’ → zy˥-zy˩ ‘small children’
child-child

f. çi˧-ndɯ˩ ‘adult’ → çi˧-ndɯ˩ ‘adults’
person-big

g. lɑ˧pɤ˩ ‘Mount Baoshan’ → lɑ˧pɤ˩ ‘Baoshan inhabitants’ (He 1987:61)
place name

h. pɯ˥ndzɻ˩ ‘carpenter’ → pɯ˥ndzɻ˩ ‘carpenters’ (He 1987:61)
carpenter

f. kø˩-ndy˩ ‘mountaineer’ → kø˩ndy˩ ‘mountaineers’ (He 1987:61)
mountain-place

Explicit expression of plurality in Naxi is subject to certain constraints. Human nouns may be suffixed with the morpheme *hɤ˧* or undergo tone change for plurality, but non-human nouns may not:

(7) a. ȵi˧ → *ȵi˧-hɤ˧ ‘fish’
fish

b. kʰɯ˧ → *kʰɯ˧-hɤ˧ ‘dogs’
dog

c. zwɑ˧ → *zwɑ˧-hɤ˧ ‘horses’
horse

	d.	ɯ˧	→	*ɯ˧-hɤ˧	‘cows’
		cow			
	e.	ndzɚ˩	→	*ndzɚ˩-hɤ˧	‘trees’
		tree			
	f.	t^{h}e˧ɯ˧	→	*t^{h}e˧ɯ˧-hɤ˧	‘books’
		book			
	g.	ɟi˩	→	*ɟi˩-hɤ˧	‘houses’
		house			
(8)	a.	ȵi˧	→	*ȵi˩	‘fish (pl.)’
		fish			
	b.	k^{h}ɯ˧	→	*k^{h}ɯ˩	‘dogs’
		dog			
	c.	zwɑ˧	→	*zwɑ˩	‘horses’
		horse			
	d.	ɯ˧	→	*ɯ˩	‘cows’
		cow			
	e.	ndzɚ˩	→	*ndzɚ˩	‘trees’
		tree			
	f.	t^{h}e˧ɯ˧	→	*t^{h}e˧ɯ˩	‘books’
		book			
	g.	ɟi˩	→	*ɟi˩	‘houses’
		house			

Naxi is thus similar to many languages that differentiate human from non-human nouns with respect to nominal plurality (Corbett 2000:56-58).

There are nevertheless some exceptions. Not all human nouns may be suffixed with *hɤ˧* or undergo tone change for plurality:

(9)	a.	zø˧ ‘male’	→	*zø˧-hɤ˧ ‘males’
	b.	mi˥ ‘female’	→	*mi˥-hɤ˧ ‘females’
	c.	nɑ˩ʂi˧ ‘Naxi person’	→	*nɑ˩ʂi˧-hɤ˧ ‘Naxi people’
	d.	ʂʅ˧kɑ˩ ‘butcher’	→	*ʂʅ˧kɑ˩-hɤ˧ ‘butchers’
(10)	a.	nɑ˩ʂi˧ ‘Naxi person’	→	*nɑ˩ʂi˩ ‘Naxi people’
	b.	ndzʅ˧ ‘mate, friend’	→	*ndzʅ˩ ‘mates, friends’
	c.	dze˧ɣɯ˧ ‘nephew’	→	*dze˧ɣɯ˩ ‘nephews’
	d.	dze˧me˧ ‘niece’	→	*dze˧me˩ ‘nieces’

The base nouns therefore must be marked in the lexicon that they may or may not be pluralized with the suffix *hɤ˧* or by tone change.

Given that the suffix *hɤ˧* can be attached to human nouns, it is unsurprising that pronouns referring to persons too can be suffixed with the morpheme *hɤ˧* (He 1987, He & Jiang 1985):[3]

[3] In addition to the suffix *hɤ˧* the suffix *ŋgɯ˩* can also be used to pluralize pronouns (He & Jiang1985 and He 1987). Some pronouns take only *ŋgɯ˩* while some others take either:

(i)	a.	ŋɤ˧ŋgɯ˩	‘we (exclusive)’
	b.	ŋɤ˥gɯ˩	‘we (inclusive)’
	c.	nɯ˧ŋgɯ˩	‘you (peer)’
	d.	thɯ˧ŋgɯ˩	‘they (peer)’
(ii)	a.	nɑ˩hɤ˧ or nɑ˩ngɯ˩	‘you (inferior)’
	b.	thɑ˩hɤ˧ or thɑ˩ngɯ˩	‘they (peer)’

(11)

a.	nɑ˨ ‘you (inferior)’	→	nɑ˨hɤ˧ ‘you pl. (inferior)’
b.	wu˧ ‘you (peer)’	→	wu˧hɤ˧ ‘you pl. (peer)’
c.	ŋv˥ ‘you (superior)’	→	ŋv˥hɤ˧ ‘you pl. (superior)’
d.	wɑ˥ ‘you (peer/respectful)’	→	wɑ˥hɤ˧ ‘you, pl (peer/respectful)’
e.	t^{h}ɑ˨ ‘he (peer)’	→	t^{h}ɑ˨hɤ˧ ‘they (peer)’
f.	t^{h}e˩gv˧mɑ˥ ‘he (superior)’	→	t^{h}e˩gv˧mɑ˥hɤ˧ ‘they (superior)’

But there is no coding for plurality by **tone on pronouns**:

(12)

wu˧	→	*wu˨
‘you, sg (peer)’		‘you, pl (peer)’

Proper names can be suffixed with the morpheme *hɤ˧* as well, in a construction known as the **associative plural** (Daniel and Moravcsik 2011). The expressions in (13) denote groups of people associated with the person whose name the suffix *hɤ˧* is attached to:

(13)

a ɑ˧lia˨-hɤ˧ ts^{h}ɿ˩ se˩
Alian come PERF
‘Alian and the others have come.’

b. ŋɤ˩ ɑ˧ka˧-hɤ˧ ndø˩
1SG Agang saw
‘I saw Agang and others.’

But the **proper name** in the associative plural construction may not undergo the tone change rule in (4):

(14)

a.	ɑ˧fa˧	→	*ɑ˧fa˨
	‘Afan’		‘Afan and others’
b.	ɑ˧ka˧	→	*ɑ˧ka˨
	‘Agang’		‘Agang and others’

Sequences comprising a (count) noun, a numeral and a classifier can be understood as singular or plural, depending on the meaning of the numeral. With the numeral *ndɯ˧* ‘one’, the whole noun+numeral+classifier expression is singular, and with other numerals greater than one, e.g., *sɿ˩* ‘three’, the whole expression is plural (CL=classifier):

(15)

a. sɿ˩dzɿ˧ ndɯ˧ *(kv˥)
teacher one CL
‘A teacher’

b. ndzɚ˩ ndɯ˧ *(ndzɚ˩)
tree one CL
‘A tree’

c. cɚ˧ ndɯ˧ *(ly˧)
cup one CL
‘A cup’

As can be seen in (16), plurality need no explicit expression in the sequence noun+numeral+classifier:

(16)

a. sɿ˩dzɿ˧ sɿ˥ kv˧
teacher three CL
‘three teachers’

b. ndzəˬ ŋgv˧ ndzəˬ
tree nine CL
'Nine trees'

c. cəˬ˧ wɑ˥ lyˬ˧
cup five CL
'Five cups.'

In fact, it is not possible to indicate plurality by *hɤˬ˧* -suffixation or tone change on the noun when it co-occurs with a numeral greater than one and a classifier:

(17) a. sɻˬdzɻˬ˧ sɻ˥ kvˬ˧
teachers three CL
'three teachers'

b. *sɻˬdzɻˬ˧-hɤˬ˧ sɻ˥ kvˬ˧
teachers three CL
'three teachers'

c. *sɻˬdzɻˊ sɻ˥ kvˬ˧
teachers three CL
'three teachers'

(18) a. çøˊseˬ˧ wɑ˥ kvˬ˧
student five CL
'five students'

b. *çøˊseˬ˧-hɤˬ˧ wɑ˥ kvˬ˧
students five CL
'five students'

c. *çøˊseˬˬ wɑ˥ kvˬ˧
students five CL
'five students'

The non-occurrence of the classifiers and explicit coding of plurality has led to a revived belief in the literature of recent years that the obligatory use of classifiers with numerals in nouns correlates with the lack of plural morphology (Chierchia 1996, 1998). We will return to this issue below (see section 4).

The base noun may appear before a partitive phrase, and the whole expression is semantically plural:

(19) a. sɻˬdzɻˬ˧ ndɯˬ˧ hwɑ˥
teacher one group
'A group of teachers'

b. çøˊseˬ˧ ndɯˬ˧ paˬ˧
student one class
'A class of students'

The noun before the partitive phrase may not be suffixed with the morpheme *hɤˬ˧* or undergo tone change for plurality:

(20) a. *sɻˬdzɻˬ˧-hɤˬ˧ ndɯˬ˧ hwɑ˥
teacher one group
'A group of teachers'

b. *sɻˬdzɻˊ ndɯˬ˧ hwɑ˥
teachers one group
'A group of teachers'

(21) a. *çøꓵse˧-hɤ˧ ndɯ˧ pa˧
student one class
'A class of students'
b. *çøꓵseꓶ ndɯ˧ pa˧
students one class
'A class of students'

The base noun, count or mass, can be followed by a quantified expression, e.g., *(ze˧ndɤ˩) ndɯ˧pe˥* 'many' and *ndɯ˧hɤ˧be˧* 'all' (ADV=adverb):

(22) a. sʅ˩dzʅ˧ (ze˧ ndɤ˩) ndɯ˧ pe˥
teacher very big one CL
'Many teachers'
b. sʅ˩dzʅ˧ ndɯ˧ hɤ˧ be˧
teacher one CL ADV
'All teachers'

(23) a. ɟi˩ (ze˧ ndɤ˩) ndɯ˧ pe˥
water very big one CL
'A lot of water.'
b. ɟi˩ dɯ˧ hɤ˧ be˧
water one CL ADV
'All water'

It is noteworthy that the classifiers that occur with particular nouns elsewhere cannot replace of the classifier *pe˥* used in conjunction with the quantifiers *(ze˧ndɤ˩) ndɯ˧* 'many'. For instance, the noun *sʅ˩dzʅ˧* 'teacher' co-occurs with the classifier *kv˥* elsewhere, but not together with in the quantifier *(ze˧ndɤ˩) ndɯ˧* 'many':

(24) a. sʅ˩dzʅ˧ ndɯ˧ kv˥
teacher one CL
'A teacher'
b. *sʅ˩dzʅ˧ ze˧ ndɤ˩ ndɯ˧ kv˥
teacher very big one CL
'Many teachers'

(25) a. ʂʮ˧-ka˩ ndɯ˧ kv˥
butcher one CL
'A butcher'
b. *ʂʮ˧-ka˩ ze˧ ndɤ˩ ndɯ˧ kv˥
butcher very big one CL
'Many butchers'

The quantifiers *(ze˧ndɤ˩)ndɯ˧pe˥* 'many' with the classifier *pe˥* being part of it and *ndɯ˧hɤ˧be˧* 'all' are fixed expressions.

The base noun in a quantified phrase may not be pluralized by suffixation or by tone change:

(26) a. sʅ˩dzʅ˧ ze˧ ndɤ˩ ndɯ˧ pe˥
teacher very big one CL
'many teachers'
b. *sʅ˩dzʅ˧-hɤ˧ ze˧ ndɤ˩ ndɯ˧ pe˥
teachers very big one CL
'many teachers'

c. *sɻ˨dzɻ˩ ze˧ ndɤ˨ ndɯ˧ pe˥
teachers very big one CL
'many teachers'

(27) a. sɻ˨dzɻ˧ ndɯ˧-hɤ˧-be˧
teachers one-ADV
'All teachers'
b. *sɻ˨dzɻ˧-hɤ˧ ndɯ˧hɤ˧-be˧
teacher- one-ADV
'All teachers'
c. *sɻ˨dzɻ˩ ndɯ˧hɤ˧-be˧
teachers all(one+CL)-ADV
'All teachers'

It is clear that numeral classifiers are in complementary distribution with explicit expression of plurality.

3. Plurality and definiteness

Noun phrases with explicit coding of plurality in Naxi are definite. Evidence for this comes from two sources. First, in contexts where a definite noun phrases are excluded, e.g., in the existential construction, the nouns cannot be suffixed with the morpheme *hɤ˧* or carry low-rising tone for plurality:

(28) a. çi˧ ndʑy˨
person be/exist
'There is/are someone/people. '
b. çi˧ ndɯ˧ kv˥ ndʑy˨
person one CL be/exist
'There is someone. '

(29) a. *çi˧-hɤ˧ ndʑy˧
people be/exist
'There are the people. '
b. *çi˩ ndʑy˧
person be/exist
'There are the people. '

The contrast between (28) and (29) shows that the nouns with the suffix *hɤ˧* observe the definiteness effect.

Second, definite noun phrases, e.g., those with a demonstrative, cannot appear as predicate nominals:

(30) a. ɑ˧ka˧ çi˧ tʂʰɻ˧ kv˥ ndø˨
Agang person this CL see
'Agang saw that person.'
b. *ɑ˧ ka˧ çi˧ tʂʰɻ˧ kv˥ wɑ˨
Agang person this CL be
'Agang is that man.'

Nouns explicitly coded for plurality with the suffix *hɤ˧* or tone change cannot occur in predicate position either (DEM=demonstrative, PL=plural, POSS=possessor):

(31) a. tʰɯ˧-ŋgɯ˨ [ŋɤ˧ gɤ˧ sɻ˨dzɻ˧ (*-hɤ˧)] wɑ˨
3PL 1SG POSS teacher be
'They are my teachers. '

b. tʰɯ˧-ŋgɯ˩ [sɿ˩˧dzɿ˧ (*-hɤ˧) sɿ˧ kv˥] wɑ˩
3PL teacher three CL be
'They are three teachers. '

c. tʰɯ˧-ŋgɯ˩ sɿ˥ *(kv˧) [sɿ˩˧dzɿ˧ (*-hɤ˧)] wɑ˩
3PL three CL teacher be
'They three are teachers. '

(32) a. *tʰɯ˧-ŋgɯ˩ ŋɤ˧ gɤ˧ sɿ˩˧dzɿ˩˧ wɑ˩
3PL 1SG POSS teachers be
'They are my teachers. '

b. *tʰɯ˧-ŋgɯ˩ tʰɯ˧ sɿ˩˧dzɿ˩˧ wɑ˩
3PL DEM teachers be
'They are those teachers. '

It is thus clear that *hɤ˧*-suffixation or tone change on base nouns by rule (4) in Naxi not only explicitly expresses plurality but also definiteness.

4. Some typological considerations

From a typological perspective, the expression of nominal plurality in Naxi is of special interest. The coding of plurality together with definiteness embodied in the same morpheme is apparently quite rare. To our knowledge, three other cases that are similar to Naxi with respect to the coding of plurality and definiteness is Chinese, Khmer (Ehrman 1970:43) and Maori (Bauer 1993:110) (the latter two are according to Dryer 2011). Noun phrases with the suffix *-men* in modern Chinese (cf. Li 1999) or the prefix *zhū* in Classical Chinese (Meisterernst 2012) are interpreted as both plural and definite. Just as in Naxi nouns suffixed with or undergone tone change for plurality, nouns with the suffix *-men* in modern Chinese cannot occur in the existential construction, as shown in (33a):

(33) a. zài kēshí lǐ yǒu xuéshēng(*-mén).
Be classroom in have student-PL
'There are students in the classroom.'

b. gū yù líng zhū ér gè jù yī zhōu yě.
I wish order PL son each occupy one province FIN
'I want each of the sons to occupy one of the provinces.'

In the example in Classical Chinese in (33b) (Meisterernst 2011:153), the noun phrase *zhū ér* 'sons' has to be understood to refer to a definite group of sons, i.e., those that have already been mentioned in the discourse. Our study of Naxi plurality adds one more language to this typogically rare class of languages that code plurality and definiteness in the same morpheme.

In many languages, definiteness is expressed independently of plurality. Thus, a definite noun phrase, typically marked in the form of a determiner or a demonstrative, can be either singular or plural. As well, a definite plural noun phrase contains two different morphemes, one for definiteness and one for plurality.

In Naxi, a noun phrase with a demonstrative cannot be pluralized by suffixation or tone change:[4]

(34) a. sɿ˩˧dzɿ˧ tʂʰɻ˧-hu˧
teacher this-PL
'These teachers.'

[4] Besides the proximal demonstrative *tʂʰɻ˧* 'this', the medial demonstrative *tʰɯ˧* 'that' and the distal demonstrative *ʔɤ˥tʰɯ˧* 'that over there, yonder' may also take the morpheme *-hu˧* for plurality. The combinations *ndɯ˧hu˧* 'some' ('one') and *ʔa˧tsɿ˧hu˧* 'what for things' (*ʔa˧tsɿ˧* 'what') are indefinite plurals. We thank Peter Jenks for drawing our attention to this point.

b. *sɿ˨˦dzɿ˧-hɤ˧ tʂʰʅ˧-hu˧
teacher this-PL
'These teachers.'

c. *sɿ˨˦dzɿ˨˦ tʂʰʅ˧-hu˥
teachers this-PL
'These students.'

The ungrammaticality of the examples in (34b, c) is perhaps due to definiteness being doubly expressed, by both the demonstrative and the suffix *hɤ˧* or tone change.

Definiteness and plurality are two independent semantic categories; the former having to do with discourse participants and the latter with number. To the extent that different semantic categories correspond to different syntactic categories, definiteness and plurality should be expressed by different syntactic categories. Our study of plurality in Naxi shows that languages may in fact deviate from this expectation in expressing these different semantic categories in one morphosyntactic unit.[5]

In a comparative study of several languages including English, Chinese, the Sino-Tibetan languages Qiang and Tsanglo, the indigenous Taiwanese languages Paiwan and Squlid Atayal and Amis in the Austronesian language family, Tang (2004) argued that the analysis of plurality should be related to the distinction between languages with a rich inventory of classifiers and those with a modicum of classifiers. She suggests that plurality in the former languages is part of an abstract feature system, while that in the latter is part of a morphological system. More specifically, plurality in classifier-rich languages is expressed morphologically as part of a feature complex. Along these lines, then, the Naxi suffix *hɤ˧* is not a marker exclusively for plurality, but has a feature complex consisting of the features [+definite] and [+human] in addition to the feature [+plural]. By contrast, the morphological expression of plurality in classifier-poor languages like many Indo-European languages is virtually exclusively for the number feature.

The complementarity of explicit expression of plurality and the presence of a numeral classifier in noun phrases (see section 2) has a direct bearing on Greenberg's (1974:25) claim regarding the correlation between numeral classifiers and lack of obligatory plural marking:

(35) Numeral classifier languages generally do not have compulsory expression of nominal plurality, but at most facultative expression.

Depending on how it is to be understood, the statement in (35) may or may not bear out the facts in Naxi.

If it is meant to assert that languages with numeral classifiers as a whole do not have obligatory plural marking, then Naxi is not a counterexample to the claim in (35). This situation is just the same as that in Northern Kam and Weining Ahmao (Bisang 2012). However, if it is read this way, then the claim in (35) is rather weak, as Bisang (2012:37) observed. It should nonetheless be pointed out that empirically weak statements in and of themselves are not particularly problematic, insofar as they are not contradicted by the empirical facts.

If the statement in (35) is understood to mean that when a numeral classifier appears, then no explicit expression of plurality is possible, then it may be problematic. It is indeed the case that numeral classifiers are in complementary distribution with pluralization by suffixation or tone change. But as Gerner (2006) showed, Northern Kam obligatorily employs plural forms of the classifiers for plural noun phrases. If the plural forms of the classifiers are taken, quite reasonably, to be the explicit expression for plurality, then the claim in (35) is empirically incorrect. To the extent that these plural forms, apparently, contain no segmental or suprasegmental morpheme that exclusively indicates plurality, the use of plural classifiers in plural noun phrases in Northern Kam are not necessarily counterexamples to Greenberg's claim in (35), if we take

[5] Corbett (2000:278) attributed to Dick Hayward (personal communication) the description that in the Omotic language Gamo any noun phrase marked as plural must be definite. As no example is given for the description, it remains unclear whether the two notions plurality and definiteness are expressed in the same morpheme in this language.

explicit expression of plurality to be by a specific morpheme that must occur whenever plurality is expressed.

If we take expression of nominal plurality to mean that it is expressed by a specific morpheme, then Naxi conforms to Greenberg's claim regarding the relation between numeral classifiers and compulsory expression of nominal plurality. It is indeed the case in Naxi that whenever a numeral classifier occurs, explicit expression of plurality is impossible.

5. Conclusion

In the foregoing sections, we provided a detailed description of the morphosyntax of nominal plurality in Naxi. The language does not apply a rule of pluralization, either by suffixation with the morpheme *hɤ˧* or by tone change, to all nouns, but only selectively to human nouns. We showed that plurality in Naxi is closely connected with the semantic property of definiteness.

Typologically, Naxi is of special interest, for belongs to a very small class of languages that code nominal plurality by tone. Moreover, the fact that plurality and definiteness may be embodied in the same morpheme in Naxi shows that particular languages do not necessarily realize different semantic notions in different morphosyntactic categories. Facts in Naxi are consistent with numeral classifiers not requiring explicit expression of nominal plurality. In fact, it is even stronger; the two cannot co-occur. The study of Naxi in comparison with other languages thus helps refine the understanding of Greenberg's universal in (35).

References

Bauer, Winifried (with William Parker and Evans), Te Kareongawai. 1993. *Maori*. London: Routledge.

Bisang, Walter. 2012. Numeral classifiers with plural marking. A challenge to Greenberg. In Dan Xu (eds), *Plurality and classifiers across languages in China*, 23-42. Berlin: Mouton de Gruyter.

Bradley, David. 1975 Nahsi and Proto-Burmese-Lolo. *Linguistics of the Tibeto-Burman Area* 2.1, 93-150.

Chierchia, Gennaro. 1996. Plurality of mass nouns and the notion of 'semantic parameter'. Ms. University of Milan.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Nature Language Semantics*, 6.4, 339-405.

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.

Daniel, Michael & Edith Moravcsik. 2011. The associative plural. In Matthew S. Dryer. & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. Available online at http://wals.info/chapter/36. Accessed on 05/10/2013.

Dryer, Matthew S. 2011. Coding of nominal plurality. In Matthew S. Dryer. & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. Available online at http://wals.info/chapter/33. Accessed on 05/10/2013.

Ehrman, Madeline Elizabeth. 1970. *Contemporary Cambodian: A grammatical sketch*. Washington D.C: Foreign Service Institute.

Gerner, Matthias. 2006. Noun classes in Kam and Chinese Kam-Tai languages: their morpho-syntax, semantics and history. *Journal of Chinese Linguistics* 34, 237-305.

Greenberg, Joseph H. 1974. Numeral classifiers and substantive numbers: Problems in the genesis of a linguistic type. In *Proceedings of the 11th International Congress of Linguistics, Bologna - Florence, Aug-Sept 1972*. 17-37. Bologna.

Haspelmath, Martin et al. 2005. *The world atlas of language structures*. New York: Oxford University Press.

He, Ji-Ren and Jiang Chu-Yi 和即仁、姜竹仪. 1985.《纳西语简志》[A Presentation of the Naxi Language], 民族出版社。

He, Zhi-Wu 和志武. 1987.《纳西语基础语法》[A Basic Grammar of Naxi], 云南民族出版社。

Li, Y.-H. Audrey. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8: 75-99.

Meisterernst, Barbara. 2012. Number in Chinese: a diachronic study of zhu from Han to Wei Jin Nanbeichao Chinese. In Dan Xu (eds), *Plurality and classifiers across languages in China*, 143-182. Berlin: Mouton de Gruyter.

Pinson, Thomas (ed.) 2012. Naqxi-Habaq-Yiyu Ceeqdiai. 纳西汉英词典. A Naxi-Chinese-English dictionary. Yunnan: Minzu Chuban She.

Tang, Chih-Chen Jane. 2004. Two types of classifier languages: A typological study of classification markers in Paiwan Noun Phrases. *Language and Linguistics* 5.2: 377-407.

Thurgood, Graham and Randy J. LaPolla. 2003. *The Sino-Tibetan languages*. London: Routledge.

NORQUEST, Peter, & Sean DOWNEY. 2013. *Expanding the PAN consonant inventory*.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:99-145
Received 25/9/2012, revised text accepted 10/9/2013, published October 2013
URL: http://hdl.handle.net/1885/10572
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

SOUTHEAST ASIAN LINGUISTICS SOCIETY SEALS
www.jseals.org / Volume 6 / 2013 / Asia-Pacific Linguistics
Australian National University

EXPANDING THE *PAN* CONSONANT INVENTORY[1]

Peter Norquest
University of Arizona
<norquesp@email.arizona.edu>

Sean Downey
University of Maryland
<sdowney2@umd.edu>

Abstract
This paper provides evidence for three Proto Austronesian (PAn) phonemes that are preserved in several distinct languages and subgroups. These include distinctions between *p and *f, *l and *l̩, and *k and *g. In addition, we assert that there is expanded evidence in Malayo-Polynesian for two currently recognized phonemes: *t̪ (PAn *C) and *c; evidence for the former has been restricted until now to the Formosan languages, and for the latter to a small group of languages in western Indonesia. These contrasts can be found in Nias (one of the Barrier Island languages off the northwest coast of Sumatra), Dohoi (a Northwest Barito subgroup of Borneo), the Western Central Malayo-Polynesian languages of Bimanese, Hawu, Dhao, Western Oceanic, and more sparsely in languages of North Sarawak, the Philippines, and Sumba. The findings presented in this paper highlight the importance of the above languages and subgroups for PAn reconstruction, and the new phonemes presented here are placed within the context of a wider PAn inventory which includes a total of seven places of articulation.
Keywords: Proto Austronesian, reconstruction, classification
ISO 639 language codes: nia, otd, bhp, hvn, nfa, kzi, kyi, ind, wew, lur, mvd, kod, akg, xbr, tgl, bik, ceb, mrw, uun, ssf, xsy, tay, sxr, xnb, dru, bnn, ami, ckv, pyu, pwn, sly, mlg, zlm, jav, bbc, mwt, kvh, mqy, ski, ksx

0.0 Introduction[2]

The Out-of-Taiwan (OoT) hypothesis was first proposed by Blust (1977) who used lexical evidence from Dyen (1963) to show that three conservative phonological distinctions were preserved in the Formosan languages, but in no other Austronesian languages. Under the OoT hypothesis nine first-order subgroups of Austronesian (An) are accepted for Taiwan (Blust 1999)[3]; conversely, all Austronesian languages spoken outside of Taiwan belong to a single Malayo-Polynesian branch which is defined by certain changes (*C > t,

[1] We would like to thank Steve Lansing for the use of his data collected during field trips to Nusa Tenggara in 2005 and 2007. Additional thanks to Leif Asplund, Brian Hallmark, Andrew Hsiu, Murray Cox, and Paul Sidwell for various comments on earlier drafts of this paper. Thanks also to the participants at the APLL5 and SEALS22 conferences for comments and suggestions on presentations which have been incorporated into this paper. Thanks to Robert Blust for making the ACD available online, without which much of this research would have been much more difficult. Finally, thanks to Linda Konnerth for offering valuable comments and suggestions on the final draft of this paper. This project was supported by the National Science Foundation (Awards #07254470 and #1030031).
[2] Abbreviations used in this paper include: (P)WCMP = (Proto) Western Central Malayo-Polynesian, PMS = Proto Macro-Sumba, PS = Proto Sumba, PHD = Proto Hawu-Dhao, PWOc = Proto Western Oceanic, PNS = Proto Northern Sarawak, PD = Proto Dayic, PSab = Proto Sabahan, B = Bimanese, H = Hawu, D = Dhao
[3] See Ross (2009) for a more nuanced view of higher-order Austronesian subgrouping.

*N > n, and *S > h). These features were purportedly innovated after the original speech community that left Taiwan immigrated into the Philippines.

The OoT model predicts that Malayo-Polynesian is defined in part by a set of phonological mergers that occurred at least by the time of the settlement of the Philippines. However, according to the principles of historical reconstruction, if any distinction exists at any node of the tree which cannot be explained as the result of internal factors (environmental conditioning) or external ones (such as contact), such a distinction must be reconstructed for the proto-language.

In this paper we present evidence for five as yet unrecognized phonological distinctions in Proto Malayo-Polynesian (PMP); these include contrasts between *p and *f, *l and *l̢, *k and *g (traditional *g reanalyzed below as a the voiced uvular stop *ɢ), as well as evidence for a additional distinctions between *t and *t̢ (previously only recognized in the Formosan languages) and *c and *s, expanding the domain of traditional *c. We argue that there are crucial correspondences between the Western Central Malayo-Polynesian (WCMP) languages, Dohoi (a branch of northern West Barito), Nias (a Barrier Islands language) and Western Oceanic – as well as other MP subgroups -- which are not innovations, but retentions. They are not the results of parallel sound change, but rather the retention of distinctions that have been lost in many other (but not all) parts of the Austronesian-speaking world. The phonemes discussed in this paper are bolded below in Table 1 (traditional reconstructions are placed in parentheses; see below for further discussion); those suggested for the first time are also italicized:

Table 1: Proposed additions to and expansions of the PMP consonant inventory

p	t	**t̢** (C)	**c**	k	q	ʔ
b	d	**d̢** (j)	**ɟ** (z)	***g***	ɢ (g)	
f	s					h
m	n		ɲ	ŋ		
	l	***l̢***				
w	r		j (y)		ʀ (R)	

The paper is organized into four major sections: Section 1 gives an introduction to the individual languages and subgroups which will form the core of the evidence provided in this paper. Section 2 discusses the evidence for the labiodental fricative *f, section 3 the retroflex series (*t̢, *d̢, and *l̢), section 4 the palatal series (*c, *ç, and *ʎ), section 5 the evidence for the voiced velar stop *g, and section 6 the uvular series (*ɢ and *ʀ). Section 7 concludes.

Description of key languages and subgroups

The languages and subgroups that comprise the primary evidence in this paper are Nias, Dohoi, the WCMP languages Bimanese, Hawu and Dhao, and Proto Western Oceanic. According to the most widely accepted version of the Austronesian phylogenetic tree (Adelaar & Himmelmann 2005), Nias and Dohoi are both Western Malayo-Polynesian (WMP) languages; Bimanese, Hawu and Dhao are members of the western half of the Central Malayo-Polynesian (CMP) subgroup; and PWOc, a subset of Oceanic, is part of the Eastern Malayo-Polynesian (EMP) subgroup. With the exception of the South-Halmahera-West-New-Guinea (SHWNG) group, then, witnesses are therefore found in all major Austronesian subgroups outside of Taiwan:

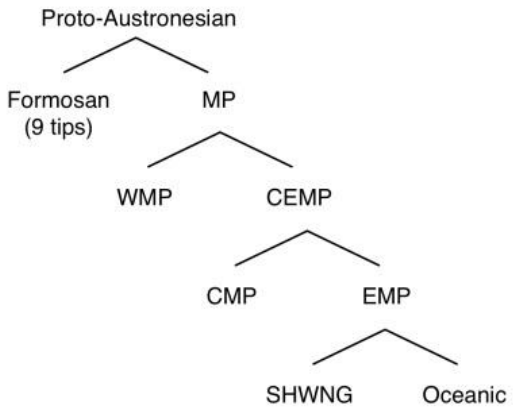


Figure 1: The Austronesian phylogenetic tree

The correspondences between them are presented in the following table, with our proposed reconstruction shown in a Revised Malayo-Polynesian (RMP) column:

Table 2: Consonant correspondences for five phonemic distinctions in PMP

PMP	RMP	Nias	Dohoi	Bima	Hawu	Dhao	PWOc
*p	*p	f-	-p-	p	p	p	*p
*p	*f	β-	-hp-	f	0	0	*β
*t	*t	t-	-t-	t	t	t	*t
*t	*ʈ	d-	-ht-	d	d	ɖ	*t
*s	*s	z-	-s-	s	h	s	*z
*s	*c	s-	-s-	c	h	c	*s
*k	*k	k-	-k-	k	k	k	*k
*k	*g	g-, -ʔ-	-hk-	h	0, -ʔ-	0, -ʔ-	*ɣ
*l	*l	l	-ɾ-	l	l	l	*l
*l	*ɭ	l	-ɾ-	r	r	r	*l

Note that distinct reflexes in Nias are restricted to initial position, with the exception of the split between *k and *g; conversely, all distinctions in Dohoi are restricted to medial position. Phonemic splits occur in the WCMP languages and PWOc in all positions.

Western Central Malayo-Polynesian

Blust (2008:98) argues for a relationship between Hawu-Dhao and the Sumbanese languages, but states the following in his conclusion: "In any case, it is clear that the most strikingly distinctive innovations shared by Bimanese with Kambera, Hawu-Dhao, or Manggarai cannot be taken as evidence for an exclusive "Bima-Sumba" group, and with this conclusion comes the end to a myth that has lasted longer than one might have imagined possible."

Blust is referring here primarily to the distinctions listed above in Table 2, and states that these are inexplicable parallel innovations in Bimanese and Hawu-Dhao that are not inherited from a common ancestor. The critical assumption he appears to make is that these cannot represent retentions from a shared ancestor since similar cases are not found higher up in the phylogenetic tree in Figure 1.

We agree with Blust that there is both evidence for a relationship between Hawu-Dhao and the Sumbanese languages on the one hand (a larger group which we call 'Macro-Sumba'), and current lack of evidence for a close relationship with Bimanese on the other[4]. This being the case, we assert that the striking

[4] There are at least a few instances of apparent shared lexical innovations, such as Bimanese *mami*, PMS *mami 'ripe'; Bimanese *weki* 'body' PS *βeki 'self'; Bimanese *haɓu*, Dhao *abo* 'pound'; Bimanese *male*, Dhao *ka-male* 'withered'. Further investigation is required to tell whether these are restricted to Bimanese and Macro-Sumba or shared by other CMP subgroups.

agreement between the reflexes in Bimanese and Hawu-Dhao indicates a case of shared retention as opposed to a collection of parallel, idiosyncratic splits; if it is true that Bimanese does not subgroup with Macro-Sumba, then this means that these distinctions must have been inherited from a more remote ancestor and been preserved independently.

1.1.1 Bimanese

Examples of the consonantal splits shown above in Bimanese are given below:

(1) Examples of phonemic distinctions in Bimanese

Split	Gloss	PMP	Bimanese	Gloss	PMP	Bimanese
*p/f	seven	*pitu	**p**idu	stingray	*paRih	**f**ai
*t/ʈ	three	*telu	**t**olu	sugarcane	*tebuh	**d**oɓu
*s/c	breast	*susu	**s**usu	spoon	*sidu	**c**iru
*k/g	scratch	*kaRaw	**k**ao	louse	*kutu	**h**udu
*l/ɭ	five	*lima	**l**ima	run, flee	*laRiw	**r**ai

These splits are unconditioned and do not seem to be the result of borrowing from Indonesian or any other language. When borrowing occurs, it nearly always reflects the original consonants of the donor. Note also that unlike the examples of native vocabulary above ('three' and 'sugarcane'), where PMP *ə is inherited as Bimanese [o], the Indonesian vowel [ə] is reflected as [a] in Bimanese:

(2) Examples of Indonesian loanwords in Bimanese

Gloss	Indonesian	Bimanese	Gloss	Indonesian	Bimanese
coach	pəlati	palati	slander	fitnah	fitina
trumpet	tərompet	tarompe	dice	dadu	dadu
nature	sifat	sifa	opium	candu	candu
lute	kəcapi	kacapi	result	hasil	hasi
drawer	laci	laci	bread	roti	roti

It should be noted that the distinctions between [p] and [f], [t] and [d], [s] and [c], and [h] and [g] are neutralized in one particular environment. In addition to a set of derivational prefixes (most notably *ka-* and *sa-*), Bimanese retains what appear to be the remnants of an older nasal derivational prefix (or prefixes). Although apparently synchronically unproductive, their original function can often be inferred:

(3) Examples of post-nasal neutralization in Bimanese

Gloss	Base	Gloss	Prefixed	Gloss	Base	Gloss	Prefixed
to close	pula	closed	mpula	to break	foka	broken	mpoka
straighten	tiri	straight	ntiri	egg	dolu	to spawn	ntolu
insert	coŋge	stuck	ncoŋge	nest	soɓu	to nest	ncoɓu
squint	giri	dazzled	taŋgiri	hook	hawi	to fish	ŋgawi

1.1.2 Hawu-Dhao

Proto Hawu-Dhao (PHD) inherited the same set of distinctions as Bimanese, with similar but non-identical reflexes. Examples are given below:

(4) Examples of phonemic distinctions in Hawu-Dhao

Gloss	PMP[5]	RMP	Bima	PHD	Hawu	Dhao
seven	*pitu	*p-	pidu	*piɖu	pidu	piɖu
rice	*pajay	*f-	fare	*are	are	are

[5] PMP forms are cited in accordance with the ACD. Note that PMP *e represents [ə].

three	*telu	*t-	tolu	*təlu	təlu	təlu
rope	*talih	*ʈ-	dari	*ɖari	dari	ɖari
breast, milk	*susu	*s-	susu	*susu	huhu	susu
nine	*siwa	*c-	ciwi	*ceo	heo	ceo
scratch	*kaRaw	*k-	kao	*kao	kao	kao
tree	*kahiw	*g-	hadʒu	*aɟu	aɟu	aɟu
five	*lima	*l-	lima	*ləmi	ləmi	ləmi
flee, run	*laRiw	*ɭ-	rai	*rai	rai	rai

Having accepted Blust's (2008) classification of Hawu-Dhao with the Sumbanese languages, it is natural to ask if the latter have preserved any of the same distinctions. As it turns out, while most of these had merged in Proto Sumba, two can still be reconstructed. Proto Sumba retains a distinction between *s and *c (where *c has lenited to *ç), which is preserved in the languages of northwest Sumba; a distinction between *k and *ɣ is preserved in all Sumbanese languages. The following correspondences and reconstructions can therefore be posited for Proto Macro-Sumba (PMS), the name we have given to the parent of Proto Sumba and Proto Hawu-Dhao:

Table 3: Phonemic splits in Proto Macro-Sumba

PMS	Proto Sumba	PHD	Hawu	Dhao
*p	*p	*p	p	p
*f	*p	*0	0	0
*t	*t	*t	t	t
*ʈ	*t	*ɖ	d	ɖ
*s	*s	*s	h	s
*c	*ç	*c	h	c
*k	*k	*k	k	k
*ɣ	*ɣ	*0, -ʔ-	0, -ʔ-	0, -ʔ-
*l	*l	*l	l	l
*ɭ	*l	*r	r	r

In Proto Sumba, *f, *ʈ and *ɭ merged with *p, *t and *l in parallel with most other Austronesian languages. It is likely that *f passed through a stage of affrication before merging completely with *p (*f > *pf > *p). *ʈ and *ɭ, on the other hand, apparently partook in a general constraint placed upon the entire retroflex series, where all members of this category merged with their alveolar counterparts. On the other hand, the same three phonemes underwent lenition in PHD; *f likely became a bilabial fricative before finally disappearing altogether (*f > *ɸ > *0), while *ʈ voiced to *ɖ and *ɭ delateralized, merging with *r. Examples of all five contrasts are given below:

(5) Examples of phonemic distinctions in Macro-Sumba etyma

Gloss	PMP	PMS	PSumba	PHD	Hawu	Dhao
warm, hot	*panas	*panas	*panas	*pana	pana	pana
rice	*pajay	*fare	*pare	*are	are	are
hood	*teduŋ	*təɖuŋ	*təɖuŋ	*təɖu	təɖu	tədu
egg	*qateluR	*ʈəlu	*təlu	*ɖəlu	dəlu	ɖəlu
breast, milk	*susu	*susu	*susu	*susu	huhu	susu
nine	*siwa	*ciwa	*çiwa	*ceo	heo	ceo
scratch	*kaRaw	*kaʔo	*kaʔu	*kao	kao	kao
wood	*kahiw	*ɣaju	*ɣaju	*aɟu	aɟu	aɟu
five	*lima	*lima	*lima	*ləmi	ləmi	ləmi
flee, run	*laRiw	*ɭai	*lai	*rai	rai	rai

The contrast between Proto Sumba *s and *ç has been maintained in only the languages of northwestern Sumba (Wejewa, Laura and Lauli). As stated above, all of the Sumbanese languages distinguish between *k and *ɣ:

Table 4: Reflexes of Proto Sumba *s/ç and *k/ɣ

PSumba	Wejewa	Laura	Lauli	Kodi	Mamboro	Anakalang	Lewa	Mangili	Kambera
*s	z	s	s	h	s	s	h	h	h
*ç	0	0	0	h	s	s	h	h	h
*k	k	k	k	k	k	k	k	k	k
*ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ	ɣ

Examples of Proto Sumba *s and *ç are given below:

(6) Examples of Proto Sumba *s and *ç

Gloss	PMP	PSumba	Wejewa	Laura	Lauli	Mamboro	Kambera
breast	*susu	*susu	zuzu	sussu	susu	susu	huhu
squeeze	---	*pasu	pazu	passu	pasu	pasu	pohu
one	*isa	*içа	ia	ia	ija	sisa	ha-u
heart	*pusuq	*puçu	pu:	pu:	pu:	pusu	puhu

There is some degree of variation in the reflexes of *ɣ; it is often deleted, and can also be substituted by the glides [j] or [w] depending on the vocalic context:

(7) Examples of Proto Sumba *k and *ɣ

Gloss	PMP	PSumba	Wejewa	Kodi	Mamboro	Anakalang	Lewa	Kambera
scratch	*kaRaw	*kaʔu	kaʔu	ka-jo	---	kau	kau	kau
tail	*ikuR	*ka-iku	kiku	kiku	kiku	kaiku	kiku	kiku
tree	*kahiw	*ɣaju	wazu	ɣai	jai	ai	e	ai
fish	*hikan	*iɣaŋ	ija	iɣja	ija	ijaŋ-u	ijaŋ-u	ijaŋ-u

1.2 Dohoi

Although the distinctions above occur in the WCMP languages in all positions, those in Dohoi occur only in intervocalic position in the final foot of the word (Inagaki 2005: 35). Hudson (1967: 53-4) comments on this in the following way:

"Dohoi is distinctive among all the Barito isolects for its medial preaspirated voiceless stops and affricate[6]. Whether this is a Dohoi innovation or the vestige of a proto-phoneme that has been lost in that environment in all other Barito isolects is indeterminate at present time. If it represents an innovation, it would appear that *T^{vl}(P-DM1) (i.e., voiceless stop and affricate proto-phonemes) became preaspirated in medial position (i.e., /V_V/) in Proto-Dohoi. This holds true for the vast majority of Dohoi forms that can be shown to represent P-B forms."

The situation which Inagaki (2005: 35-6) describes shows that in general, forms with underlying /-hC-/ are in free variation between [-hC-] and [-C-] at the surface level:

(8) Dohoi [-hC-] forms in free variation at the surface level

/ɲihpo/	[ɲipo] ~ [ɲihpo]	'tooth'
/ŋamuhto/	[ŋamuto] ~ [ŋamuhto]	'to shut the eyes'
/nohcot/	[nocot] ~ [nohcot]	'to shiver'
/borahkaŋ/	[borakaŋ] ~ [borahkaŋ]	'to sell'

[6] Dohoi contrasts one more pair of phonemes not shown above: -hc- and -c-. These correspond with PAn *z and *y, respectively; for example: Dohoi ruhca < *luzaq 'spit', uhcan < *quzaN 'rain', mahcu < *zauq 'far', but kacu < *kayu 'wood'.

However, forms with underlying [-hC-] are not in free variation if they contrast with minimal pairs having underlying [-C-], either native (9) or borrowed from Indonesian (10):

(9) Dohoi [-C-] forms in contrast with native [-hC-] forms

puti	'banana'	puhti	'white'
noto	'(personal name)'	nohto	'to see'
bakai	'unfinished'	bahkai	'monkey

(10) Dohoi [-hC-] forms in contrast with borrowed Indonesian [-C-] forms

From Indonesian		Native Dohoi	
kapan	'when?'	kahpan	'thick'
kita	'we (incl)'	kihta	'sap, resin'
toko	'shop'	tohko	'there are'

While it is apparent from other Northwest Barito languages that a distinction must be reconstructed between intervocalic plain and prenasalized voiceless stops, these categories have merged in Dohoi, which distinguishes instead between plain and preaspirated medial voiceless stops (data from Hudson 1967)[7]:

(11) Proto Northwest Barito plain intervocalic stops

	'to wash'	'centipede'	'to laugh'	'calf'	'sibling's child'
Proto-NWB	*mu**p**uk	*ɟɔli**p**an	*Ka-**t**au	*bɔ**t**ih	*a**k**un
Dohoi	mʊpʊʔ	ɟɔʎipan	ŋa-tau	bɔtih	akʊn
Murung 1	mopʊk	---	ka-tau	---	akɔ-m
Murung 2	mupʊk	ɟalipan	kɔ-tauʔ	bɔtih	---
Siang	mupʊk	ɲɔʎipan	kɔ-tau	botɪh	akʊn

(12) Proto Northwest Barito prenasalized intervocalic stops

	'spouse's parent'	'to swell'	'banana'	'elder sibling'	'forehead'
Proto-NWB	*u**mp**u	*lɔ**nt**iŋ	*pu**nt**i	*ɔ**ŋk**a	*li**ŋk**ɔu
Dohoi	upuʔ	ʎɔtɪŋ	putiʔ	ɔkaʔ	likou
Murung 1	ɔmpʊʔ	---	---	ɔŋka	---
Murung 2	ʊmpʊʔ	lɔntɪŋ	pʊntiʔ	ɔŋkaʔ	liŋkou
Siang	---	ʎɔntɪŋ	punti	ɔŋka	liŋkou

(13) Proto Northwest Barito preaspirated intervocalic stops

	'fire'	'thick'	'eye'	'louse'	'1sg'
Proto-NWB	*a**hp**ui	*ka**hp**an	*ma**ht**a	*ku**ht**u	*a**hk**u
Dohoi	ahpui	kahpan	mahtaʔ	kuhtuʔ	ahkuʔ
Murung 1	---	---	mata-m	---	---
Murung 2	apui	kapan	mataʔ	kutuʔ	akuʔ
Siang	apui	kapan	mata	kutu	akʊh

There is thus a three-way contrast that must be reconstructed for Proto Northwest Barito. It should be noted that Dohoi generally boyrrows Indonesian words with intervocalic plain stops as preaspirated, and prenasalized stops as plain (data from Inagaki 2005); however, what appear to be more recent borrowings of plain stops are also borrowed as plain. Likely examples of borrowings in all three categories include the following:

[7] Hudson's [ʎ] and [ɔ] are equivalent to Inagaki's [ɾ] and [o].

(14) Indonesian -C- loans into Dohoi

Indonesian		Dohoi	
ləpau	‘lounging-place’	ɾohpou	‘house’
lapik	‘lining, mat, base’	ɾahpik	‘sleeping mat’
upah	‘salary, reward’	uhpah	‘reward’
macam	‘kind, sort, type’	mahcam	‘(be) like’
akan	‘future’	ahkan	‘future’
pikir	‘think’	pihkir	‘think’
paku	‘(a) nail’	pahku	‘(a) nail’
pakat	‘discussion, agreement’	pahkat	‘friend’
pəŋakit	‘disease’	paŋahkit	‘disease’

(15) Indonesian -NC- loans into Dohoi

Indonesian		Dohoi	
pumpun	‘meeting place, center’	pupuŋ	‘peak, meeting’
tampar	‘slap’	tapar	‘slap’
kentut	‘fart’	kotut	‘fart’
lantiŋ	‘house built on a raft’	ɾatiŋ	‘raft’
rantang	‘hamper’	rataŋ	‘basket’
cinta	‘love’	sita	‘love’
aŋka	‘number’	aka	‘number’
leŋket	‘to cling’	ɾeket	‘to cling’
naŋka	‘jackfruit’	naka	‘jackfruit’

(16) (recent) Indonesian -C- loans into Dohoi

Indonesian		Dohoi	
capat	‘go fast’	capat	‘fast’
ketel	‘kettle’	keter	‘cooking pot’
utaŋ	‘debt’	utaŋ	‘debt’
garakan	‘mandated by’	garakan	‘omen’
tikus	‘rat, mouse’	tikus	‘rat, squirrel’

1.3 Nias

With one exception (*g), the distinctions between *p and *f, *t and *ţ, *l and *ļ, *c and *s, and *k and *g only occur in Nias in word-initial position. These have gone unrecognized in the past because they depend on the environment. Lase (2011: xxiv-xxv) describes these as "initial mutations", where the initial of a word undergoes a change when the word is in the middle or at the end of the sentence (i.e. in interphrasal position). It is not true for all lexical items, however, and it is our assumption that this environment preserves original alternations lost in other environments. The initial mutations that can occur in Nias are the following:

(17) Nias initial mutations

[f] > [β]	[b] > [mb]
[t] > [d]	[d] > [ndr]
[s] > [z]	[ʔ] > [g]
[k] > [g]	[ʔ] > [n]

Examples of non-mutating and mutating forms are given below:

(18) Non-mutating 'fortis' and mutating initials in Nias

Non-mutating			Mutating		
fao	[fao]	'with'	fiso	[βiso]	'ear'
taboi	[taboi]	'expel'	taroʔo	[daroʔo]	'hips'
sara	[sara]	'one'	simbi	[zimbi]	'jaw'
kiə	[kiə]	'urine'	kara	[gara]	'stone'
baxa	[baxa]	'inside'	baho	[mbaho]	'ravine'
dəhə	[dəhə]	'recover'	dela	[ndrela]	'bridge'
ʔato	[ʔato]	'crowded'	ʔeha	[geha]	'cough'
ʔirə	[ʔirə]	'tube'	ʔidanə	[nidanə]	'water'

The initials [b, d, g, mb, ndr, β, z, n] also occur independently in word-initial position, and contrast with the 'mutating' initials above:

(19) Non-mutating 'lenis' initials in Nias

basi	[basi]	'reap'
dudu	[dudu]	'release'
gaβu	[gaβu]	'sand'
mbadu	[mbadu]	'breath'
ndrohu	[ndrohu]	'sober'
βəʔi	[βəʔi]	'yes'
zizi	[zizi]	'blunt'
niha	[niha]	'person'

Our best understanding of the diachronic sources of these three categories of intials are shown below in (20):

(20) Fortis, mutating fortis, and lenis initials in Nias

Fortis	Mutating	Lenis
*p > f	*f > [β]	? > β
*t > t	*t̪ > [d]	---
*c > s	*s > [z]	*ɟ > z
*k > k	? > [g]	*ɢ > g
*ɓ > b	*b > [mb]	*mb > mb
*ɗ > d	*d > [ndr]	*nd > ndr
*ʔ > ʔ	*g > [g]	---
*ʔ > ʔ	? > [n]	*n > n

There are a couple of cases in which the source of the synchronic initial is unclear. For example, we have been unable to locate Austronesian etymologies for any forms with non-mutating [β]. One plausible candidate would be *w, but it is apparent from forms such as Nias *waʔa* 'root' (< PMP *wakaR) and *walu* 'eight' (< PMP *walu) that Nias has inherited PMP *w as [w].

We have also been unsuccessful in uncovering etymologies for any forms with mutating [g] from surface [k], although many (about 50 percent) of these forms are borrowings from Indonesian words with initial [k], i.e. Nias *[g]afalo* 'head' < Indonesian *kəpala* and Nias *[g]atafele* 'slingshot' < Indonesian *katapəl*. Forms with non-mutating [g] also include a fair number of Indonesian loans with either initial [k] or [g], i.e. Nias *galasi* 'glass' < Ind *gəlas*, Nias *gombəni* 'company' < Ind *kompəni*, suggesting that the absorption of Indonesian loanwords has added an extra layer of complexity within the overall phonological system, at least in the case of velar stops.

The source of mutating initial [n] is also presently opaque, although it's frequent occurrence in what might be considered inalienably possessed objects such as kinship terms, body parts, and domestic items

(*[n]ama* 'father' < PMP *ama, *[n]ixu* 'nose' < PMP *hiduŋ, *[n]omo* 'home' < PMP *Rumaq) suggests that at least one source of this alternation may be morphological.

1.4 Western Oceanic

Ross (1988) provides extensive discussion on a set of fortis/lenis contrasts in Western Oceanic for PAn *p, *s, and *k. Although limited examples are given of words with lenis reflexes (and fewer for fortis reflexes), the essential contrasts can be reconstructed as the following (we propose Proto Western Oceanic reconstructions based on the lower-level reconstructions in that book)[8]:

Table 5: Fortis and lenis consonants in PWOc

PWOc	PNNG	PPT	PMM	PSS
*p	*p	*p	*p	*p
*β	*β	*β	*β	*β
*t	*t	*t	*t	*t
*t	(*l)[9]	*t	*t	*t
*s	*s	*s	*s	*s
*z	*z	*z	*c	*z
*k	*k	*k	*k	*k
*ɣ	*ɣ	*q	*ɣ	*ɣ

Ross argues that the lenis reflexes of the Proto Oceanic series *p, *t, *s, and *k are secondary developments and cannot be reconstructed in Proto Oceanic itself[10]. Due to this, and the fact that the data in Ross (1988) is somewhat limited, the Western Oceanic data is not as strong as that presented from other languages and subgroups; we include it nevertheless as inspiration for further research in this branch of Austronesian.

Examples of our Proto Oceanic reconstructions based on Ross (1988) are as follows[11]:

(21) Split correspondences in Proto Western Oceanic

Gloss	POc	PWOc	PNNG	PPT	PMM
squeeze	*poRos	***p**oros	*poro	---	*poros
banana	*pudi	***β**udi	*βudi	*βudi	*βudi
outrigger float	*saman	***s**aman	*sama	*saman	*[s]ama
paddle	*pose	*βo**z**e	*[p/β]o[s/z]e	*βoze	*βoze
rat	*kusupeq	***k**u[s]uβe	*ku[s/z]uβe	---	*kusuβe
tree	*kaju	***ɣ**aju	*ɣai	*qaju	*ɣaju

2.0 The distinction between *p and *f

We now begin a more detailed discussion of each phonemic split, beginning with the one between *p and *f. As shown above, there are two sets of reflexes associated with PAn *p:

---

[8] PWOc = Proto Western Oceanic, PNNG = Proto North New Guinea, PPT = Proto Papuan Tip, PMM = Proto Meso-Melanesian, and PSS = Proto Southeast Solomonic. Reconstructions of PNNG are based on the following subgroups: PSCH (Proto Schouten), PHG (Proto Huon Gulf), PNg (Proto Ngero), PBEL (Proto Bel); those of PMM are based on PWZ (Proto Willaumez), PNI (Proto New Ireland), PNS (Proto North-West Solomonic), and PLN (Proto Lavongai/Nalik).

[9] This correspondence set is limited to the Bel languages and the closely related Nenaya and Roinji languages as well as languages of southeast Malekula and Epi. As Ross (1988) does not offer any clear examples between fortis and lenis *t, this distinction will not be pursued in this paper.

[10] For example, Ross (1988:50) argues that the variation between the PNNG forms for 'bow' (*pana(q)) and 'shoot' (*βana(q)) is due to morphological alternation, and this appears to be correct. However, the situation may be the reverse of what he suggests – rather than lenition having occurred in a prefixed form 'to shoot', the original initial may have been protected, instead undergoing *fortition* in the free form 'bow'.

[11] Ross (1988) does not provide any examples of contrasts between *t and *l in PNNG, so examples cannot be included here.

Table 6: Reflexes of *p and *f

PMP	RMP	Nias	Dohoi	Bima	Hawu	Dhao	PWOc
*p	*p	f-	-p-	p	p	p	*p
*p	*f	β-	-hp-	f	0	0	*β

We propose that the second set of reflexes associated with PAn *p be reconstructed as *f. These reflexes occur in all positions in Bimanese, Hawu-Dhao, and Western Oceanic; they are restricted to word-initial position in Nias and word-medial position in Dohoi.

Examples of initial and medial *p are given below[12]:

(22) Examples of initial *p

Gloss	PMP	Nias	Bima	Hawu	Dhao	PWOc	
causative	*pa-	fa-	pa-	pe-	pa-	---	
warm/hot	*panas	---	pana	pana	pana	---	(Ind *panas*)
fruit bat	*paniki	---	panihi	niʔi	---	---	
broken	*pataq	(a-fatə)	mpada[13]	pada	---	---	
squeeze	*peRes	---	pua	---	---	*poRos	
how many	*pija	---	pila	pəri	pəri	(*βiza)	
choose	*piliq	fili	---	pili	---	---	(Ind *pilih*)
seven	*pitu	fitu	pidu	pidu	piḍu	---	
dragnet	*puket	---	puka	pəku	pəku	---	(Ind *pukat*)
white	*putiq	(a-fusi)	---	pudi	puḍi	---	
sharp pain	*hapejes	([β]əxə)	pili	pəɗa	pəda	---	(Ind *pədas*)

(23) Examples of medial *p

Gloss	PMP	Dohoi[14]	Bima	Hawu	Dhao	PWOc	
fathom	*depa	---	ndupa	rəpa	---	---	
fold	*lepet	---	lipi	ləpa	ləpe	---	
pinch	*qapit[15]	---	api	ŋ-api	ŋ-api	*kapi	(Ind *apit*)
mat	*tepiR	---	dipi	dəpi	ḍəpi	---	

The Nias forms for ‘broken’ and ‘white’ have been placed in parentheses because, being affixed, their root initials are in an intervocalic environment where the distinction between [f] and [β] is neutralized. There is also a disagreement between the Nias form for ‘sharp pain’ and the PWOc form for ‘how many’ on the one hand, and the PWMP forms on the other.

Examples of initial and medial *f are given below:

(24) Examples of initial *f

Gloss	PMP	Nias	Bima	Hawu	Dhao	PWOc
paddy	*pajay	[β]axe	fare	are	are	---
to shoot	*panaq	---	fana	---	---	*βana
leucoderma	*panaw	---	fano	ano	---	---
bait	*paniŋ	---	(pani)	ani	ani	---
pandanus	*paŋdan	[β]andra	fanda	---	---	---
thigh	*paqa	[β]aha	---	---	---	*βaqa

[12] Indonesian parallels are given to the right when the phonotactics of some or all forms are consistent with a possible loan scenario.

[13] “exhausted”

[14] We have unfortunately not been able to locate any Dohoi forms with medial [-p-] which have cognates in the other languages used here.

[15] “tongs, anything used to hold things together by pinching”

chisel	*paqet	[β]ahə	(paa)	(pae)	(paa)	---	(Ind *pahat*)
hoarse	*paRaw	(a-fau)	fao	---	---	---	
stingray	*paRih	---	fai	---	---	*βari	
mango	*pahuq	---	foʔo	(pau)	(pao)	---	(Ind *pauh*)
turtle	*peŋu	[β]ənu	fonu	əŋu	əŋu	*βoŋu	
good	*pia	---	---	ie	ia	---	
ten	*sa-ŋa-puluq	[β]ulu	mpuru	ŋ-uru	ŋ-uru	---	
navel	*pusej	[β]usə	---	əhu	əsu	---	
heart	*pusuq	---	---	uhu	usu	---	

There are unfortunately not many instances of medial *p, and all are confined to WCMP with the exception of 'four' (below) in which there is variation: Dohoi and PWOc pointing to *əfat, but Bimanese and Hawu-Dhao pointing to *əpat. The PHD form *əpa might be explained as being inherited from a previous *əmpa (cf. Iban, Malay, Balinese and Sasak *əmpat*), since the regular reflex of PAn *mp is *p (i.e. PAn *əmpu 'grandchild', PHD *əpu). However, the Bimanese reflex of PAn *mp is normally [mp] (cf. Bimanese *ompu* 'grandchild'), so it may be necessary to reconstruct disjunct forms.

(25) Examples of medial *f

Gloss	PMP	Dohoi	Bima	Hawu	Dhao	PWOc	
thick	*ma-kapal	kahpan	---	me-aa	ma-aa	---	
dream	*h-in-ipi	nuhpi	nifi	nii	nii *m-niβi		
thin	*ma-nipis	mihpi	(nipi)	me-nii	ma-nii	---	(Ind *nipis*)
centipede	*qalu-hipan	(ɟoripan)	rifa	---	---	*qaliβan	
gall	*qapeju		folu	əɗu	ədu	---	
lime (for betel)	*qapuR	---	afu	ao	---	---	
sweep	*sapuh	---	cafi	hai	---	---	
fire	*hapuy	ahpui	afi	ai	ai	*(j)aβi	
four	*[h]epat	ohpat	(upa)	(əpa)	(əpa)	*βati	(Ind *əmpat*)
reward	*upaq	uhpa	ufa	---	---	---	

Some of the exceptions above are explainable. Nias *a-fau* 'hoarse' is intervocalic, so its underlying form has been neutralized. The Hawu and Dhao forms for 'mango' are likely borrowings from Indonesian *pauh*, and the same is true for Bimanese *nipi* 'thin' from Indonesian *nipis* (compare Bimanese *naŋi* and Indonesian *mənaŋis* 'weep', where the expected prenasalized form in Bimanese would be *ntaŋi*). The Nias forms for 'chisel' and 'sharp pain' have reflexes of *f where the WCMP forms have reflexes of *p; it is possible that the cases of 'chisel' in Bimanese and Hawu-Dhao are loans from Indonesian. This may also be the case in the Hawu-Dhao forms for 'sharp pain', although the vowels in the Bimanese form make this explanation less plausible

There is no apparent explanation for the discrepancy between Bimanese *pani* and PHD *ani 'bait'. In the case of Dohoi *ɟoripan* 'centipede', it should be noted that Hudson (1967) occasionally transcribed preaspirated medials as plain and some of these were corrected by Kazuya in his publications; in this case, Kazuya does not list a form for 'centipede' which could be used to cross-check Hudson.

In addition to the items above, the following words have been identified in individual languages which have either a *p or an *f reflex, but don't have cognates in other languages (many of these were identified in Blust (2008) or in his online Austronesian Comparative Dictionary). Some of them with Indonesian equivalents may actually be loans:

(26) Further examples of *p

	Gloss	PMP	Bimanese	
(a)	to rasp	*parud	paru	(Ind *parut*)
	knife	*pisaw	piso	(Ind *pisau*)
	other side	*hipaR	ipa	
	bitter	*paqit	paʔi	
	shoot, sprout	*qapucuk	pucu	(Ind *pucuk*)
	dew	*hapun	apu	

	Gloss	PMP	Hawu	
(b)	board	*papan	papa	(Ind *papan*)
	splash	*picik	pihi	

	Gloss	PMP	Dhao
(c)	fulfill	*pa-penuq	pa-pənu

(27) Further examples of *f

	Gloss	PMP	Nias
(a)	dove	*punay	[β]une

	Gloss	PMP	Dohoi
(b)	tooth	*ŋipən	ɲihpo

	Gloss	PMP	Bimanese
(c)	astringent	*apeled	firi
	temple (anat)	*pispis	mpifi
	blow	*upi	ufi
	fern	*paku	fahu
	base of tree	*puqun	fuʔu

	Gloss	PMP	Hawu
(d)	vulva	*puki	uʔi

Within the Flores-Lembata group of WCMP, there are two closely-related dialects of the same language spoken in the villages of Seranggorang and Nilanapo which seem to distinguish between *p and *f intervocalically: [-p-] being the reflex of the former, and [-hp-] of the latter in a way which bears a striking similarity to Dohoi. Although our data is limited, we can offer the following examples:

(28) Examples of *-p- in Seranggorang-Nilanapo

Gloss	PMP	Seranggorang	Nilanapo
what	*apa	aape	ape
pat, light slap	*pik	api[16]	api-ŋ
four	*epat	pa	pa

(29) Examples of *-f- in Seranggorang-Nilanapo

Gloss	PMP	Seranggorang	Nilanapo
fire	*hapuy	ahpe	ahpe
thin	*ma-nipis	mihpi	mihpi
tooth	*ipen	ihpe	ihpe

[16] *api* < Proto-Lembata *gəpik 'wing'

Regarding these two series (*p and *f) in the WCMP languages, Blust (2008: 93-4) states that "[…] much of this agreement is due to common conditioning. In particular, stops that otherwise undergo lenition in medial position tend to resist it when following a stressed schwa (PAn *ə), because this gave rise to phonetic gemination. Moreover, it is likely that the vocalic syncope in *ma- 'stative' before PMP bases that began with a labial stop either happened early in the history of the Central Malayo-Polynesian languages, or was recurrent, giving rise to prenasalized stops that also resisted lenition, but were subsequently reduced to simplex stops in some daughter languages."

However, it seems evident from the evidence above that neither of these conditions holds true absolutely. Not only are there exceptions to the rule that *p was protected after a schwa, there is little evidence for the hypothesis that *p remained [p] only after stative *ma-.

Blust goes on to say "[…] the conclusion that *p lenition was independent in Hawu and Bimanese also follows from the simple fact of subgrouping: because Sumba-Hawu is a fairly clearcut group and Kambera shows no evidence of *p lenition, the simplest hypothesis is that *p lenition in Hawu must have occurred independently of similar changes in Bima."

This makes the crucial assumption that reflexes of *f are the result of *p lenition. If, on the other hand, *f is assumed to be original, then it is languages in which *p and *f merged that were innovative.

Finally, "[t]he remaining phonological innovations that are shared exclusively by Hawu and Bima also fail to bear close scrutiny. Rather, parallel sound changes evidently have been unusually common in this part of Indonesia, and there are no clear grounds for using such evidence to propose a subgroup larger than that of Sumba-Hawu."

We assert that the evidence presented here is actually demonstrative of shared retentions, and it is rather that parallel mergers have actually occurred in many other Austronesian languages, particularly Western Malayo-Polynesian but also including the Formosan languages.

3.0 The retroflex series

Evidence for a series of retroflex initials can be found in the WCMP languages Bimanese, Hawu and Dhao. This includes a voiceless retroflex stop *ʈ and a retroflex lateral *ɭ. Additional evidence for *ʈ can also be found in initial position in Nias, as well as medial position in Dohoi. With this series established, traditional PAn *j can be interpreted phonetically as a voiced retroflex stop *ɖ. Finally, there is tentative evidence that a retroflex nasal *ɳ may have been preserved in Malay and the languages of northwest Borneo.

*3.1 Evidence for *ʈ*

The two series of reflexes of PMP *t are repeated below:

Table 7: Reflexes of PMP *t and *ʈ

PMP	RMP	Nias	Dohoi	Bima	Hawu	Dhao
*t	*t	t-	-t-	t	t	t
*t	*ʈ	d-	-ht-	d	d	ɖ

There is no apparent conditioning environment for the latter set of reflexes; given the correspondences above, it appears that a distinction needs to be reconstructed between *t and *ʈ. Examples of *t are given below:

(30) Examples of PMP *t in initial position

Gloss	PAn	PMP	Nias	Bima	Hawu	Dhao
natural cockspur	---	*tara	([ndr]ara)	tara	tara	(ɖara)
head covering	---	*teduŋ	---	todu	tədu	tədu
three	*telu	*telu	təlu	tolu	təlu	təlu
pierce	---	*tubak	---	tuɓa	təɓu	---

knock, peck	*tuktuk	*tuktuk	---	tutu	tutu	---	
pole for boat	*tuku	*tuku	---	tuku	tuku	ta-tuku	
old (people)	*tuqaS	*tuqah	---	tua	---	tua	(Ind *tua*)
knee	---	*tuqud	---	ta-tuʔu	tuu	tuu	

(31) Examples of PMP *t in medial position

Gloss	PAn	PMP	Dohoi	Bima	Hawu	Dhao
calf	---	*b[e/i]ties	botih	wisi	---	---
cut, chop	*Setek	*hetek	(n-ohtok)	---	əta	əta

Additional cases in Bimanese and Hawu-Dhao are given below:

(32) Examples of initial *t in Bimanese

Gloss	PAn	PMP	Bima	
not	---	*ati	ati	
flick, snap fingers	*betik	*betik	ɓeti	
explode	---	*betuʔ	wotu	
roof thatch	---	*qatep	ati	
low, of tide	---	*ma-qeti	moti ‘sea’	
brain	---	*qutek	iti	
alum	---	*tawas	tawa	(Ind *tawas*)
punting pole	---	*teken	tiki ‘staff, cane’	
pillar, mast, post	---	*tiqaŋ	tija	(Ind *tiaŋ*)
oyster	---	*tiRem	tire[17]	(Ind *tiram*)

(33) Examples of initial *t in Hawu-Dhao

Gloss	PAn	PMP	Hawu	Dhao	
split	*betak	*betak	wəta	---	
star	*bituqen	*bituqen	---	hətu	
earthworm	---	*kali-wati	kelati	---	
flea	*qatimela	*qatimela	teməla	---	
placenta	---	*tabuni	tewuni	---	
round shield	---	*tamiŋ	tami	---	(Ind *tameŋ*)
middle	---	*teŋaq	təŋa	---	(Ind *təŋah*)
knock, beat	---	*tetuk	dətu	ɖətu	
to blow	---	*tiup	tiu	tiu	
year	---	*taqun	təu	təu	(Ind *tahun*)
look upward	---	*tiŋadaq	teŋara	taŋara	
bend, curve	---	*-tuk	ŋətu[18]	---	

Upon comparison with PAn reconstructions, there appears to be a strong correlation between what we reconstruct here as *ţ and what is normally reconstructed as PAn *C. This is surprising, as evidence for *C has thus far been restricted to the Formosan languages:

[17] This is likely an Indonesian loanword since the normal reflex of *R in Bimanese is [0].
[18] ‘bow, curtsy’

(34) Examples of initial *ʈ

Gloss	PAn	PMP	Nias	Bima	Hawu	Dhao	
afraid	*[C/t]akut	*takut	(a-taʔu)	dahu	me-daʔu	ma-ɖaʔu	
rope	*CaliS	*talih	(tali)	dari[19]	dari	ɖari	(Ind *tali*)
bury	*Canem	*tanem	[d]anə	---	pe-dana	pa-ɖane	
earth	---	*taneq	[d]anə	dana	---	---	
feces	*Caqi	*taqi	[d]ai	(taʔi)	dei	ɖei	(Ind *tahi*)
indigo	---	*taRum	---	dau	dao	---	
sea	---	*tasik	---	dasi[20]	dahi	ɖasi	
person	*Cau	*tau	---	dou	dau	ɖau	
sugarcane	*CebuS	*tebuh	---	doɓu	dəɓu	ɖəɓu	
swallow	---	*telen	(tələ)	---	dəla	ɖəle	(Ind *təlan*)
mat	---	*tepiR	---	dipi	dəpi	ɖəpi	
elder	---	*tua	[d]ua[21]	dua[22]	---	---	
palm wine	---	*tuak	---	(tua)	due	ɖua	(Ind *tuak*)
index finger	*Cuzuq	*tuzuq	[d]uru	(turu)[23]	---	---	

(35) Examples of medial *ʈ

Gloss	PAn	PMP	Dohoi	Bima	Hawu	Dhao	
tree, log	---	*bataŋ	bahtaŋ	(ɓata)	---	---	(Ind *bataŋ*)
blind	*buCa	*buta	---	mbuda	ɓədu	bəɖu	
send	*pa-kaCu	*pa-katu	---	ŋgadu	---	pa-aɖu	
louse	*kuCu	*kutu	kuhtu	hudu	udu	uɖu	
eye	*maCa	*mata	mahta	mada	mada	maɖa	
die	*m-aCay	*m-atay	mahtoi	made	made	maɖe	
white	---	*putiq	puhti	---	pudi	puɖi	
ghost	*qaNiCu	*qanitu	(otu)[24]	---	nidu	niɖu	(Ind *hantu*)
liver	*qaCay	*qatay	ahtoi	ade	ade	---	
cucumber	---	*qatimun	---	dimu	dimu	---	
egg	*qiCeluR	*qateluR	(kotoɾuh)	dolu	dəlu	ɖəlu	
black	---	*ma-qitem	(mitom)	---	mədi	məɖi	
hundred	*RaCus	*Ratus	rahtus	(ratu)	---	---	(Ind *ratus*)
porcupine	---	*taRutuŋ	(tohotuŋ)	dudu	dudu[25]	ɖuɖu	

Of the four Dohoi exceptions above, two of them ('ghost' and 'black') descend from original prenasalized stops: Proto West Barito *həntu and *mintəm, respectively. As noted above, the phenomenon of Dohoi preaspiration is confined to the onset of the final syllable of the word, which explains the reflex in 'egg'. As for 'porcupine', the WCMP forms allow for the reconstruction of an initial *ʈ, but there is a discrepancy with Dohoi as to the place of articulation of the second stop. Since this example is drawn from Hudson's data and Kazuya does not provide an example of 'porcupine' to provide a crosscheck, it may be an example of Hudson failing to record the preaspiration. Bimanese 'tree, log' and 'hundred' are probably borrowed from Indonesian (the [r] in Bimanese *ratu* confirms this in the second case).

Additional examples of *ʈ are given below:

[19] 'strap'
[20] 'beach'
[21] 'grandfather'
[22] 'parent's sibling'
[23] 'point'
[24] 'corpse'
[25] The Hawu-Dhao forms both mean 'thorn'

(36) Examples of *ţ in Nias

Gloss	PAn	PMP	Nias
taro	---	*tales	[d]alə
ear	*Caliŋa	*taliŋa	[d]aliŋa
finger, toe	---	*taŋan	[d]aŋa
bone	*CuqelaN	*tuqelan	[d]əla

(37) Examples of *ţ in Bimanese

Gloss	PAn	PMP	Bima
blunt tip	---	*tampak	dampa
blunt	---	*tumpul	dumpu
callus, blister	*beCuʔ	*betuʔ	wodu[26]
testicles	*buCuq	*butuq	wudu[27]
kill	*ka-aCay	*ka-atay	hade
outsider	---	*qaRta	ada[28]

(38) Examples of *ţ in Hawu-Dhao

Gloss	PAn	PMP	Hawu	Dhao
beat on with both hands	---	*tambak	daba	---
come, arrive	---	*tekas	dəka	ḍəka
ear (wax)	---	*tilu	dilu	ḍilu
above	*i aCas	*di atas	dîda	deḍa
haft of knife	---	*utiŋ	udi	---

(39) Examples of *ţ in Dohoi

Gloss	PAn	PMP	Dohoi
sister	---	*betaw	bohtou
that	*i-Cu	*i-tu	ihtuh ‘this’, ahtuh ‘that’

There is a subset of forms in which *ţ corresponds to PAn *t:

(40) Correspondences between PAn initial *t and PMP *ţ

Gloss	PAn	PMP	Nias	Bima	Hawu	Dhao
fish poison	*tuba	*tuba	[d]uβa	duwa	---	---
freshwater eel	*tuNa	*tuna	---	duna	dəno	ḍəno
southeast wind	*timuR	*timuʀ	---	---	dimu	ḍimu

(41) Correspondences between PAn medial *t and PMP *ţ

Gloss	PAn	PMP	Dohoi	Bima	Hawu	Dhao
raw	*ma-qetaq	*ma-qetaq	---	mada	mada	---
stone	*batu	*batu	bahtu	wadu	wadu	waḍu
see	*(k)ita	*(k)ita	---	eda	-ədi	-əḍi
we (excl)	*kita	*kita	ihto	(ndai-ta)	di	əḍi
vomit	*um-utaq	*um-utaq	ŋ-uhta	---	mədu	məḍu
seven	*pitu	*pitu	pihtu	pidu	pidu	piḍu
penis	*qutiN	*qutin	---	---	udi	---

[26] ‘swell up slowly’
[27] ‘penis’
[28] ‘slave’

If we are accurate in interpreting PAn *C as *ʈ, then it must be assumed that in the examples above, *ʈ either shifted to *t in the Formosan languages, or *t shifted to *ʈ in PMP[29].
There are also five cases in which PAn *ʈ corresponds to PMP *t:

(42) Correspondences between PAn *C and PMP *t

Gloss	PAn	PMP	Nias	Bima	Hawu	Dhao	
foxtail millet	*beCeŋ	*beteŋ	---	witi	---	---	
weep	*Caŋis	*taŋis	---	---	taŋi	---	(Ind *taŋis*)
know	*Caqu	*taqu	---	---	tou	təu	(Ind *tahu*)
guts	*C-in-aqi	*tinaqi	---	---	tenei	tanei	
roast, burn	*CuNuh	*tunu	tunu	---	tunu	tunu	(Ind *tunu*)

Of the above, it is possible that the forms for ‘weep’, ‘know’, and ‘roast, burn’ are Indonesian loans.

3.1.1 Evidence from Seranggorang-Nilanapo

As in the case of *f (see (28) and (29) above), the retroflex stop *ʈ has been preserved in medial position (in the reflex [-ht-] as opposed to the [-t-] reflex of *t) in Seranggorang-Nilanapo. Forms such as the following show that medial [-t-] is contrastive with [-ht-] in Seranggorang-Nilanapo in a way similar to Dohoi:

(43) Seranggorang-Nilanapo reflexes of medial *t

Gloss	PAn	PMP	Seranggorang	Nilanapo
head	---	---	ette	ette
bad	---	---	dateŋ	datə
black	*ma-qitem	*ma-qitem	mite	mitaŋ

(44) Seranggorang-Nilanapo reflexes of medial *ʈ

Gloss	PAn	PMP	Seranggorang	Nilanapo
louse	*kuCu	*kutu	uhtu	---
die	*maCay	*matay	mahta	mahta
stone	*batu	*batu	wahto	wahto

Note that Seranggorang-Nilanapo agrees with the Dohoi and WCMP forms for ‘stone’ in presenting evidence for a retroflex stop, contrary to evidence from the Formosan languages for an alveolar stop. The reflexes for ‘black’, on the other hand, indicate PMP *t in contrast with the Hawu and Dhao forms in (35) which have reflexes of *ʈ.

3.2 PAn *ɖ

With evidence presented for the reconstruction of the voiceless retroflex stop above, attention can now be turned to its voiced counterpart. The PAn phoneme which seems most naturally to fill this gap is PAn *j. The distributional properties of *j support the idea that it was originally a retroflex. It is well-known that the acoustic properties of retroflexes are strongest in the vowel formants preceding the stops themselves (Ladefoged and Maddieson 1996: 28), and that word-initial retroflexes are therefore most prone to neutralization; this is exactly the pattern of PAn *j, which occurs only in medial and final position.

In Bimanese, the reflex of PAn *ɖ is *r* when following a low vowel, and *l* elsewhere, with the additional constraint that if the word begins with an *l*, the reflex becomes *r* via dissimilation (‘batten of loom’ in (45) and ‘day’ in (46)). In Macro-Sumba, *ɖ merged with *d and lenited to [r]:

[29] This may have happened, for example, in the case of *pitu ‘seven’, due to anticipatory assimilation of the retroflex feature from the retroflex lateral in ‘eight’ *waɭu (see section 3.3 below) in a counting sequence. It is presently unclear what the explanation might be for the discrepancy between Formosan *t and RMP *ʈ in the other forms above.

(45) Examples of PMP *ɖ (*j)

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao
to paddle	*aluja	[g]aluxa	---	---	---	---	---
batten of loom	*balija	---	---	lira	*lira	---	---
dry, evaporate	*maja	---	---	---	*mara	---	ɖasi mara[30]
foam	*bujaq	---	---	---	*βura	---	---
name	*ŋajan	---	aran	ŋara	*ŋara	ŋara	ŋara
rice	*pajay	[β]axe	paroi	fare	*pare	are	are
how much/many	*pija	---	pira	pila	*pira	pəri	pəri
when?	*p-ijan	---	mira	---	*piraŋ	pəri	pəri
nose	*qijuŋ	[n]ixu	uruŋ	ilu	*iru	---	---
ant	*sejem	---	---	sa-sili	---	---	---
to smell	*hajek	---	-arok	---	---	---	---
younger sibling	*huaji	[n]axi	---	ari	*ari	ari	ari

In words where *ɖ followed schwa, it was reinterpreted as an implosive in PMS, likely because of the common feature of retraction shared by retroflexes and implosives (see Haudricourt 1950, as well as Heine 1978 in which an identical shift of *ɖ > [ɗ] is posited between Proto-Sam and the Cushitic languae Boni). With the exception of 'day', these forms also have a reflex of [l] in Bimanese:

(46) PMP *ɖ > PMS *ɗ

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao
day	*qalejaw	---	ondou	liro	*ləɗo	loɗo	loɗo
gall	*qapeju	[g]aβəxu	---	folu	*pəɗu	əɗu	əɗu
sharp pain	*hapejes	[β]əxə	---	pili	---	pəɗa	pəda
stinging pain	*hapejiq	---	---	---	---	pəɗi[31]	pədi

Interestingly, the same change with the same conditioning environment occurred in the languages of northern Borneo. Examples from Proto Northern Sarawak (PNS) and Proto Sabahan (PSab) are given below (Borneo data are taken from Blust (1995b, 1997a, 1997b, 1998, 2000, 2001, 2002b, 2006b, 2007a, 2010)):

(47) Examples of PMP *ɖ > *d in languages of northern Borneo

Gloss	PMP	PNS	PSab
inform	*bajaq	*badaʔ	---
foam	*bujaq	*budaʔ	*budaʔ
chills	*dajem	*dadəm	---
younger sibling	*huaji	*(t)-(w)adi	(*aɗi)
when	*ijan	*idan (PK)	*ŋ-idan
how (much/many)	*kuja	*kuda	*kuda
name	*(ŋ)ajan	*(ŋ)adan	*ŋadan
riceplant	*pajay	*padaj	*padaj
how (much/many)	*pija	---	*pida
charcoal	*qajeŋ	*adəŋ	*adəŋ

(48) Examples of PMP *ɖ > *ɗ in languages of northern Borneo

Gloss	PMP	PNS	PSab
sting, smart	*hapejes	*pəɗəs	*pəɗəs
stinging pain	*hapejiq	*pəɗiʔ	*pəɗiʔ
day	*qalejaw	*əɗaw	*əɗaw

[30] 'low tide'
[31] PHD 'itch'

gall bladder	*qapeju	*pəɗu	*pəɖu
ant	*sejem	---	*səɖəm

Although PAn *ɖ was restricted to non-initial position as discussed above, there is reason to believe that the word 'two' had an original and exceptional retroflex initial in PWCMP. The evidence for this comes from PMS, in which 'two' can be reconstructed with an implosive (*ɗua), and from Proto Lembata, in which 'two' is reconstructed with a retroflex initial that is normally only reconstructed word-medially (the reflex of PAn *j):

(49) PWCMP evidence for initial *ɖ in 'two'

Gloss	PMP	PMS	PLembata
two	*duha	*ɗua	*ɖua

*3.3 PAn *ɭ*

Having postulated the existence of a retroflex stop series in PAn, we turn to the question of whether evidence exists for other retroflex phonemes. There is indeed evidence for a retroflex lateral.

*3.3.1 PPS evidence for *ɭ*

The WCMP languages exhibit another split series of reflexes for PAn *l; Paz (1981) reconstructs a distinction between front and back laterals for Proto Philippines:

Table 8: WCMP reflexes of *l and *ɭ

PMP	RMP	PPh	Bima	Dhao	Hawu
*l	*l	l	l	l	l
*l	*ɭ	ɭ	r	r	r

Since *r can be reconstructed independently for PAn, these series can be reconstructed as alveolar *l and retroflex *ɭ respectively. In comparison with the WCMP data above, the correspondence between *l and *ɭ is nearly exact. Examples of alveolar *l and retroflex *ɭ in Proto Philippines (PPh) are given below:
Examples of alveolar *l are given below:

(50) Examples of *l

Gloss	PMP	PPh	Bima	Hawu	Dhao	
return	*baliw	---	ɓali[32]	ke-ɓali[33]	---	(Ind *balik*)
buy	*beli	*bəlíʔ	weli	wəli	həli	
seaward	*lahud	---	ka-lau[34]	lou	ba-ləu[35]	(Ind *laut*)
sail	*layaR	---	lodʒa	lai	lai	(Ind *lajar*)
fold	*lepet	---	lipi	ləpa	ləpe	
five	*lima	*limáʔ	lima	ləmi	ləmi	(Ind *lima*)
day	*qalejaw	---	liro	loɗo	loɗo	
egg	*qateluR	*ʔiklúg	dolu	dəlu	ɖəlu	
three	*telu	(*ta-tɭúʔ)	tolu	təlu	təlu	
ear(wax), deaf	*tuli, *tilu[36]	*tulíʔ	---	dilu	ɖilu	

Additional examples of *l are given below:

[32] 'restore'
[33] 'again'
[34] 'unobstructed view'
[35] 'south'
[36] This comparison assumes vocalic metathesis in either Proto Philippines or PWCMP.

(51) Examples of *l

(a)	Gloss	PMP	PPh
	tongue	*dilaq	*dílaʔ
	back	*likud	*likúd
	neck	*liqeR	*liʔə́gʲ
	nit	*lisehaq	*lisáʔ
	ear	*taŋila	*talíŋaʔ

(b)	Gloss	PMP	Bima	
	broad, wide	*abelaj	wela	
	count	*bilaŋ	ɓila	(Ind *bilaŋ*)
	lamp, torch	*ilaw	ilo	
	lightning	*kilat	kila	(Ind *kilat*)
	exceed	*lalaw	lalo	
	walk, go	*lampaŋ	lampa[37]	
	sky	*laŋit	laŋi	(Ind *laŋit*)
	step, stride	*laŋkaŋ	laŋga	(Ind *laŋkah*)
	passageway	*lawaŋ	lawa	(Ind *lawaŋ*)
	calm (water)	*linaw	lino	
	eel sp.	*linduŋ	lindu	
	leech	*lintaq	linta	(Ind *lintah*)
	boil over	*luab	lua[38]	(Ind *luap*)
	outside	*luqar	lua	(Ind *luar*)

(c)	Gloss	PMP	Hawu	Dhao	
	saliva	*iluR	ilu	ilu	
	men's house	*kamali	kemali[39]	---	
	fold	*leku(q)	ləku	ka-ləko	
	cave	*liaŋ	lie	---	(Ind *liaŋ*)
	sour	*ɲilu	me-ɲilu	ma-ɲilu	(Ind *ɲilu*)
	dog flea	*qati-mela	teməla	---	
	wing	*qelad	əla	əla	
	to swallow	*telen	dəla	ɖəle	

Examples of retroflex *ɭ are given below:

(52) Examples of *ɭ in WCMP

Gloss	PMP	PPh	Bima	Hawu	Dhao	
reverse	*balik	---	wari	wari	hari[40]	
widow(er)	*balu	---	mbaru	(balu)	(ɓalu)	(Ind *balu*)
moon	*bulan	*búɭan	wura	wəru	həru	
ear of grain	*buliR	(*búlig[41])	wuri	wuri	huri	
body hair	*bulu	*buɭbúɭ	wuru	wuru[42]	---	
in(side)	*dalem	*ɖáɭəm[43]	---	ɗara	ɗara	
road	*zalan	*daɭan	---	ʄara	ʄara	

[37] 'road, run'
[38] 'pour'
[39] 'house(hold)'
[40] 'again'
[41] 'bunch, cluster'
[42] '(root) fibers'
[43] 'depth, bottom'

skin	*kulit	---	huri	k-uri	ka-ʔuri	
rat	*labaw	---	ka-rawo	---	ma-raho	
man, husband	*laki	*laɻáki	rahi	(laʔi)	(laʔi)	(Ind *lakí*)
ginger	*laqia	(*laʔújaʔ)	rea	---	---	
run away, flee	*laRiw	---	rai	rai	rai	
ten	*sa-ŋa-puluq	*púɻuʔ	mpuru	ŋuru	ŋuru	
pestle	*qahelu	*hákɻuŋ	aru	aru	---	
head(waters)	*qulu	---	uru	uru	uru	
error	*salaq	---	sara	(hala)	(sala)	(Ind *salah*)
torch	*suluq	*suɻúʔ	---	huru	suru	
forest	*halas	*háɻas[44]	ara	---	ɖua ara[45]	
snake	*hulaR	*ʔuɻág	---	---	əru-əru[46]	
rope	*talih	---	dari	dari	ɖari	
eight	*walu	*waɻúʔ	waru	aru	aru	

Additional examples of *ɻ are given below:

(53) Examples of *ɻ

(a)	Gloss	PMP	PPh
	house	*balay	*baɻəj
	rolled together	*balun	*báɻun[47]
	deaf	*bəŋəl	*biŋáɻ
	itch	*gatel	*gatáɻ
	mortar	*lusuŋ	*ɻasúŋ
	moss	*lumut	*ɻúmut
	leech	*qali-matək	*aɻimatək
	sour	*qaRsem	*kaɻsəm

(b)	Gloss	PMP	Bima
	astringent	*apeled	firi
	revolve, turn	*biliŋ	wiri
	trivet	*dalikan	riha ‘hearth’
	dig	*kali	ŋari (?)
	sesame	*ləŋa	riŋa
	Venus	*mantalaq	ntara ‘star’
	to plant	*mula	mura
	hand	*qalima	rima
	centipede	*qalu-hipan	rifa
	bamboo floor	*saleR	sari

(b)	Gloss	PMP	Hawu	Dhao
	pale, white	*balaR	wara	---
	housefly	*lalej	lara	lara
	withered	*ma-layu	me-raɟu	---

The correspondence between the Proto Philippines and the WCMP forms is generally quite good – there are only three discrepancies above (‘three’, ‘ear of grain’, and ‘ginger’). In the case of ‘three’, it may be

[44] ‘wild; snake’
[45] ‘palmwine’
[46] ‘worm’
[47] ‘provisions’

the case that the lateral in Pre-Proto Philippines *təlu assimilated to the initial retroflex initial in 'two' *ɖuha in counting.

*3.5 PAn *ŋ?*

Blust (2006) describes a small set of forms found in the Philippines, southern Sulawesi, northwest Borneo, and Malay which have liquid reflexes in the former two but sibilant reflexes in the latter two. Examples are given below:

(54) Examples of liquid-sibilant correspondences in WMP

Gloss	Tagalog	Bikol	Cebuano	Maranao	Kelabit	Malay
foam	bulá?	---	bulá?	bola?	---	busa
rotten	bulók	bulók	bulúk	---	---	busuk
to water	dilíg	---	---	---	---	diris
red	---	---	---	riga?	sia?	---
long hair	lúgaj	lúgaj	---	---	---	suraj
compel	pílit	pírit	pílit	---	---	pisit
disorder	guló	---	---	---	---	rusoh
sell	---	---	---	saliw	m-asiw	---
provoke	ulót	urót	---	---	---	us[o/u]t
flower		---	búlak	---	busak	---
blind	búlag	---	bulúg	---	busər	(bular)
tickle	---	gírok	gilúk	---	g-əl-isək	---

It is difficult to know how to interpret this data, and if it represents a legitimate proto-phoneme. Since the Philippine reflexes would normally reflect *j (*ɖ), some kind of retroflex is suggested. The gaps in the inventory include *ʂ, and *ŋ. *ʂ may seem possible at first glance, given the sibilant reflexes in Malay and the northwest Bornean languages, but the change from *ʂ to a liquid seems implausible. Although none of the reflexes include a nasal, *ŋ can be considered if the following changes are assumed:

Philippines/southern Sulawesi: ŋ > l > l, r

Malay/NW Borneo: ŋ > z > ʂ > s

Given the limited nature of the evidence, it is difficult to make a definitive statement, and this must remain a tentative hypothesis. One interesting corollary hypothesis, however, is that this would explain the unique change in the East Formosan languages of *j (*ɖ) to *n (see Blust (1999: 46)). Although there are no Formosan examples cited for the items above, *if* *ŋ was part of the PAn consonant inventory, then the following sequence of changes might be posited, in which East Formosan *ɖ merged with original *ŋ before shifting to *n:

*ɖ > *ŋ > n

3.6 Interim discussion on the retroflex series

One of the biggest reasons that the unity of the retroflex series has not been recognized is because of the different patterns of merger amongst its members. While *ɖ is reconstructed from a heterogeneous series of late mergers, *ʈ and *ɭ appear to have merged quite early in most PMP languages with *t and *l, respectively. In the case of the *l/*ɭ merger, it appears to have been motivated by the depalatalization of *ʎ to dental *l̪ (see below), which would have put a severe strain on perceptual contrast between the laterals. The same may have been true of the *t/*ʈ merger under pressure from *c. The western Indonesian evidence given in section 4.2.2 below suggests that the *c/*s merger occurred later than the *t/*ʈ merger, perhaps through a stage where it became a dental affricate. In this scenario, the following set of shifts and mergers would have occurred (in the non-Formosan languages):

c	>	t̪s̪	>	s̪
t, ʈ	>	t	>	t̪
ʎ	>	l̪	>	n
l, ɭ	>	ɭ	>	l

Overall, it appears that within Austronesian, evidence for the retroflex series has been preserved best at two geographic extremes – in the northwest in Taiwan and the Philippines, and in the southeast in WCMP. The cumulative reflexes for the retroflex series in the Formosan languages are given below:

Table 9: Reflexes of retroflex consonants in the Formosan languages

PAn	*ʈ	*ɖ	*l/ɭ
Pazeh	s	z, -t	r
Saisiat	s	z	ɾ
Thao	θ	ð	r
Atayal	ts	0	l
Saaroa	ts	ɬʲ	0
Kanakanavu	ts	l	0, l
Rukai	ts	g	ɭ
Bunun	t	0	h-, 0
Amis	t	n	ɭ
Kavalan	t	n	r, ɣ
Paiwan	ts	d	ɭ
Puyuma	ʈ	ɖ/d	ɭ

The most conservative Formosan language appears to be Puyuma, which has maintained retroflex reflexes in all cases. It is notable that even though the distinction between *l and *ɭ is not maintained in Taiwan, the Formosan languages still provide indirect evidence for the latter (as noted in Ross (1992)) since in the majority of cases, *l and *ɭ seem to have merged as *ɭ, with modern languages showing a combination of lateral and rhotic reflexes.

4.0 Palatals

A distinct palatal series has already been recognized in PAn, which includes the members *c, *z (*ɟ), *ɲ, and *N (*ʎ). These will be covered below in this order.

4.1 The PAn palatal stop

The reflexes of PAn *s and the *c proposed in this paper are given below:

Table 10: reflexes of *s and *c

PAn	RMP	Nias	Bima	Hawu	Dhao	PWOc
*s	*s	z-	s	h	s	*s
*s	*c	s-	c	h	c	*z

Although Dohoi distinguishes intervocalic [-c-] and [-hc-], one might expect it to pattern with the other key languages listed above in distinguishing *s from *c. However, comparative evidence shows that Dohoi [-c-] is a reflex of PMP *y while [-hc-] is a reflex of PMP *z:

(55) Dohoi [-c-] < *y and [-hc-] <*z

Gloss	PMP	Dohoi	Gloss	PMP	Dohoi
spit	*luzaq	ɾuhca	wood	*kayu	kacu
rain	*quzan	uhcan	crocodile	*buqaya	pacu (vocalic metathesis)
far	*zauq	ma-hcu			

There is more variation in this PWOc lenis series than the other two examined by Ross (*β < *f and *ɣ < *g), and as a result reconstructions are less secure.

(56) Examples of initial *s

Gloss	PMP	Nias	Bima	PSumba	Dhao	PWOc	
error	*salaq	(sala)	sara	*sala	(sala)	---	(Ind salah)
bamboo floor	*saleR	[z]alo	sari	*ka-[s]al[e]	---	---	
python	*sawa	[z]awa	sawa[48]	---	---	---	
douse fire	*sebu	---	suwu	---	səbu[49]	---	
ant	*sejem	[z]ixə	sa-sili	---	---	---	
elbow	*siku	[z]iʔu	(cihu)[50]	*siɣu	siʔu	---	
torch	*suluq	[z]ulu	---	*[s]ulu	suru	---	(Ind suluh)
breast, milk	*susu	---	susu	*susu	---	*zuzu[51]	(Ind susu)

(57) Examples of medial *s

Gloss	PMP	Bima	PSumba	Dhao	PWOc	
iron	*besi	ɓesi	*ɓəçi	bəsi	---	(Ind bəsi)
paddle	*beRsay	wese	*ɓose	---	*βoze	
contents	*isi	isi	*içi	---	---	(Ind isi)
cook	*nasu	---	*ma-na[s]u	nasu	---	
navel	*pusej	---	*pə[s]u	əsu	---	
breast, milk	*susu	susu	*susu	---	*zuzu	(Ind susu)
ripe	*ta-esak	ntasa	---	ma-ɖasa	---	
sea	*tasik	dasi	*tasik	ɖasi	---	

Proto Sumba forms with an *[s] do not have a Wejewa or other northwest Sumba cognate, crucial to the distinction between Proto Sumba *ç and *s; these are therefore unfortunately ambiguous and merely listed for completeness. The two Proto Sumba exceptions above, 'iron' and 'contents', both occur before a final [i] which indicates that *s palatalized to *ç in this environment.

(58) Examples of initial *c

Gloss	PMP	Nias	Bima	PSumba	Dhao	PWOc
climb	*sakay	---	---	*[s]aɣi	caʔe	*saɣe
bifurcation	*saŋa	---	caŋa[52]	*[s]aŋa	ka-caŋa[53]	*saŋa
comb	*saRu	---	cau	*[s]ai	---	---
nine	*siwa	siwa	ciwi	*çiwa	ceo	---
spoon, ladle	*suduk	---	ciru	*[s]uru	curu	---

(59) Examples of medial *c

Gloss	PMP	Bima	PSumba	Dhao	PWOc	
wet	*baseq	ɓeca	*mba[s]a	(basa)	---	(Ind *basah*)
one	*isa[54]	ica	*içа	əci	---	

[48] 'snake'
[49] 'smoke, vapor'
[50] Since all other examples of 'elbow' indicate *s instead of *c, this may have been an innovation in Bimanese; however, see the Proto Dayic form below.
[51] While *susu appears to be the dominant form, *suzu and *zuzu are also reconstructible based on various reflexes in daughter languages.
[52] 'split'
[53] 'branch of a river'
[54] See also Nias *sara* 'one'.

who	*i-sai	---	---	cee	*sai	
sated	*ma-besuR	mbocu	*ka-mbəçu bəcu	---		
salty	*ma-qasin	maci[55]	*maçi	(masi)	---	(Ind *(m)asin*)
mortar	*esuŋ	nocu	*ŋoçu	---	---	
heart	*pusuq	---	*puçu	(usu)	---	
smoke	*qasu	---	*maçu	---	*qasu	

Note that the following forms in Nias corroborate the hypothesis that in at least some instances, Nias [s] derives from an earlier [c] (the result of the palatalization of [t] before [i]):

(60) Nias si < *ci

Gloss	PMP	Nias		
white	*putiq	a-fusi	<	a-fuci
calf	*bities	[mb]isi	<	[mb]ici

These contrast with sequences of [ti] in Nias, which are largely, if not exclusively, the result of Indonesian loans, i.e. Nias *bufati* 'regent' < Ind *bupati*, Nias *gariti* 'curly' < Ind *kəritiŋ*, Nias *lati* 'train' < Ind *latih*, etc.

Additional examples of *s and *c are given below:

(61) Examples of PMP *s

(a)	Gloss	PMP	Bima	
	comb	*nusa	nisa	(Ind *nusa*)
	salt	*qasiRa	sia	
	pierce, stab	*rusuk	rusu	
	regret	*selsel	sinci	
	3sg	*si-ia	sia	

(62) Examples of PMP *c

(a)	Gloss	PMP	Nias
	comb	*sujud	suxu
	split	*silaq	sila
(b)	Gloss	PMP	Bima
	mercy, pity	*ma-qasiq	meci[56]
	sweep	*sapuh	cafi
	insert	*sipsip	cici
(c)	Gloss	PMP	PWOc
	nit	*lisehaq	*lisa

4.2 Proto Dayic

There is an additional correlation between PPS and Proto Dayic, a Northern Sarawak subgroup in northwest Borneo, where the following sound correspondences are observed:

Table 11: Correspondences between PPS and PD

PPS	PD
c	0/h
s	s

Where Proto Dayic forms are unavailable, Kelabit (K) forms are used instead. Examples of *s:

[55] 'sweet'
[56] 'love'

(63) Proto Dayic reflexes of *s

Gloss	PMP	PDayic
oar	*beRsay	bəsay (K)
heart	*pusuq	*pusuʔ
pierce, stab	*rusuk	n-usuk (K)
breast	*susu	tusu (K)
error	*salaq	salaʔ (K)

(64) Proto Dayic reflexes of *c

Gloss	PMP	PDayic
wet	*baseq	*bahaʔ
sated	*besuR	*baur
mortar	*esuŋ	iuŋ (K)
who	*i-sai	*iih
mercy, pity	*ma-asiq	*m-aheʔ
salty	*ma-qasin	m-ain (K)
navel	*pusej	*puəd
elbow	*siku	*ijuh
nine	*siwa	*iwaʔ
comb	*sujud	udud (K)

A crucial question can now be raised about the relationship between traditional PAn *c and the *c proposed in this paper. Since traditional *c is of such relatively low frequency[57] and retained in only a small set of western Indonesian languages, the extension of the domain of *c promises to solve a vexing problem in Austronesian comparative studies, summarized in the following table from Blust (2009b: 557):

Table 12: Arguments for and against assigning *c to PMP or PAn

Against	For
1. found only in western Indonesia	1. *c or massive unconditioned splitting
2. spread by borrowing from Malay	2. found in monosyllabic roots
3. not in basic vocabulary	3. patterns with other palatals (*z, *ɲ)

The particularly interesting point in the above table is that *c is "found in monosyllabic roots" in western Indonesia. We would add that it is not only found in monosyllabic roots, but in the prefixes attached to these roots as well. In order to illustrate this, we collected all of the words with reconstructed *c in the Austronesian Comparative Dictionary (Blust 1995e); the following are all of the instances in which *c appears outside of the final syllable. Forms in which there is an identifiable root are listed in (65), and those in which there is none in (66):

(65) PWMP words reconstructed with initial *c with identifiable roots

Gloss	PWMP	Root	Root meaning
take with the fingers	cu=bit	=bit	hook, clasp, grasp with fingers
shelter	ci(n)=duŋ	=duŋ	shelter, protect
quick, swift	ce=kas	=kas	swift, agile; energetic
choke, strangle	ce=kel	=kel	cough
seize, grasp	ce=kep	=kep	seize, grasp, embrace
catch with the hands	ci=kep	=kep	seize, grasp, embrace

[57] We have unfortunately not been able to uncover any cognate sets which include both forms with traditional PAn *c and forms with *c reconstructed from our data which would allow a correlation between the two to be (dis)proven.

sticky, stick to	ca(ŋ)=ket	=ket	adhesive, sticky
cover with a hollow container	cu(ŋ)=kub	=kub	cover
bent, curved	ci(ŋ)=kuk	=kuk	bent, crooked
close, cover	cu(ŋ)=kup	=kup	enclose, cover
to shine, of heavenly bodies	ci=lak	=lak	shine
sink, submerge	ce=leb	=leb	sink, disappear
soak, steep in water	ce=lep	=lep	sink, submerge
curved area; corner, angle	ce=luk	=luk	bend, curve
dive, immerse	ce(R)=neb	=neb	dive; sink, disappear underwater
immerse, submerge	ce=ŋeb	=ɲeb	dive, submerge
gape, open the mouth	ci=ŋaŋ	=ŋaŋ	amazed, gaping
seize in the mouth or beak	ce=ŋap	=ŋap	open, of the mouth
catch one's breath	ci=ŋap	=ŋap	open, of the mouth
to smack	ca(m)=pak	=pak	clap, slap; split, break
crack, split, break	ce=pak	=pak	clap, slap; split, break
pincers, tongs	ca(R)=pit	=pit	press, squeeze together; narrow
narrow	cu=pit	=pit	press, squeeze together; narrow
dull sound	ce=puk	=puk	thud, clap, break
part, separate, divide	ce=ray	=ray	separate
spotted, speckled	cu=rik	=rik	spot, freckle
mark, line, dash	cu=rit	=rit	scratch a line

(66) PWMP words reconstructed with initial *c without identifiable roots

Gloss	PWMP
rice variety (?)	candur
betel nut case	calapaʔ
small bird with shiny black plumage: *drongo*	cawi
stench	ceŋis
shrill, of the voice	cerik

Of the five words in (66) above, 'stench' and 'shrill' "look" like they could be prefix+root combinations, although no roots have yet been identified. Of the remaining three, 'rice variety' is a questionable reconstruction, 'betel nut case' is an item that could have been borrowed widely through trade, and 'drongo' is a faunal term (faunal terms are often ideophonic or known to be phonologically idiosyncratic in other ways (Smoll 2011)).

The next set of words is comprised of reduplicated roots:

(67) PWMP words reconstructed with *c with reduplicated roots

Gloss	PWMP
hissing, rustling sound	cik
skewer	cuk=cuk
insert	cu(l)=cul
sip, suck	cup=cup
suck noisily	cut=cut

Finally, the following are forms in which *c occurs in the penultimate syllable of the word. (68) lists all forms with identified roots, and (69) lists exceptions:

(68) PWMP words reconstructed with medial *c with identifiable roots

Gloss	PWMP	Root	Root meaning
muddy, waterlogged	ba=cak	=cak	muddy
muddy	bi=cak	=cak	muddy
step, tread on	i=cak	=cak	step, tread, trample
drive in, as a post, nail, or wedge	pa=cek	=cek	insert, stick into a soft surface
press in by force	pu=cek	=cek	insert, stick into a soft surface
spatter, fly out in all directions	be=cik	=cik	fly out, splash, splatter
splash, splatter	bi=cik	=cik	fly out, splash, splatter
fly off, of solid bits or droplets	le=cik	=cik	fly out, splash, splatter
splash, splatter	peR=cik	=cik	fly out, splash, splatter
splash, spray, sprinkle	pi=cik	=cik	fly out, splash, splatter
sound of splashing, etc	ri=cik	=cik	fly out, splash, splatter
speckled, stained	re=cik	=cik	fly out, splash, splatter
dent, dented	pi=cuk	=cuk	penetrate
hoe, chop up soil	ba=cuk	=cuk	penetrate
enter, penetrate	ha=cuk	=cuk	penetrate
squeeze, squirt out	be=cit	=cit	squirt out
squeeze, squirt out	pe(R)=cit	=cit	squirt out
squirt or gush out	pu(R)=cit	=cit	squirt out
squeeze out	pi=cit	=cit	squirt out
squeeze, squirt out	le=cit	=cit	squirt out
squeeze out, slip out	le=cut	=cut	squirt, squeeze, or slip out
squeeze, squirt out	lu=cut	=cut	squirt, squeeze, or slip out
spurt out	pu=cut	=cut	squirt, squeeze, or slip out

(69) PWMP words reconstructed with medial *c without identifiable roots

Gloss	PWMP		
hook used in lifting	gancu	sexually impotent	baciR
crush, press in	picak	budge, shift, move aside	icud
dove, pigeon	acaŋ	mountain peak	qapucuk
disturb	kacaw	mouth (of an animal)	mu(n)cuŋ
rotten, spoiled	bucek	slip or slide down	lu(ŋ)cuR
smooth, slippery	licin	take away from, dispossess	pucut
creep, crawl	icir		

There are more counterexamples in this group of words, but they are still outnumbered by those with identifiable roots. Several of the above forms also "look" like they could be prefix+root combinations; this list also includes 'dove, pigeon', another faunal term.

Given the evidence above, it can now be postulated that *c was originally more widespread than previously thought, not restricted to western Indonesia, and that it occurred in several basic vocabulary items. The question then becomes how to interpret the data from western Indonesia.

What seems to have happened is that evidence for *c has been preserved most conservatively in WCMP, Nias, Proto Dayic, and (erratically) in PWOc. Conversely, *c generally merged with *s in the languages of western Indonesia if they occurred in monomorphemic forms; it was preserved if the forms were bimorphemic, indicating a sensitivity to root structure (and the implication that it was still active when *c shifted to *s). It should be noted that this provides potentially valuable subgrouping evidence for a group that includes several Malayic languages, as well as languages spoken across the southern Indonesian chain of islands stretching from Sumatra to Lombok.

The fact that what is currently reconstructed as PMP *s derives from an earlier *c and *s may also serve to explain why the reflex of PAn *s in some languages (such as Kanakanabu, Bunun, Amis, or Manggarai) is *c* ([ts] or [tʃ]) – the existence of original *c meant that there was a pre-existing category into

which *s could merge. Whereas the normal pattern of merger in the majority of languages was *s, *c > s, the reverse pattern, *s, *c > c occurred in a minority of languages.

4.3 The voiced palatal stop

Blust (2009: 547) interprets PAn *z as a voiced palatal affricate, while Wolff (2010: 34) suggests that it was "a very forward voiced stop." We consider it most parsimonious to assume that PAn *z was the voiced counterpart of *c, namely the voiced palatal stop [ɟ].

4.4 The palatal fricative

Blust interprets PAn *s phonetically as the palatal fricative [ç] and *S as [s] (Blust 2009b: 547). Wolff (2010) agrees about *S but interprets *s as [c]. We suggest that PAn *s was [s], that *S was [ç], and suggest that *s remained stable in many languages with *ç shifting to *h* in all but some of the Formosan languages.

One of the reasons why Blust interprets *s as [ç] is because "in languages that preserve *ɲ as a palatal nasal and that have active systems of nasal substitution *s* is replaced by *ɲ* in certain word-formation processes […]" (2009b: 579). However, this could also be because of the fact that Proto-Austronesian *s was a dental sibilant with a domed (as opposed to grooved) tongue articulation, which would be a phonetic driver for a palatal nasal even if *s was not phonologically a member of the palatal series. Where a grooved tongue configuration would pull the tip of the tongue toward the alveolar ridge, a domed configuration would push it into a dental position, promoting interdentalization (Rukai) as well as affrication (Kanakanavu, Bunun and Amis) which would be an intermediate stage on the way to becoming a plain stop (Thao, Paiwan). This slot would then be available to be filled by the palatal fricative [ç] as it shifted forward.

(70) Formosan reflexes of PAn *s and *S (*ç)

PAn	Pazeh	Thao	Sai	Ata	Saa	Knknvu	Rukai	Bunun	Amis	Kav	Puy	Pai
*s	z	t	h	h, x	s, 0	c	θ, s, 0	c	c	s	s	t
*S	s	ʃ	ʃ	s	0	s	s	s	s	s	0	s

4.5 The palatal lateral

Evidence for PAn *N has been preserved, especially in word-medial position, most robustly in the Formosan languages – many of which have also retained a phonetic realization that supports its reconstruction as a palatal lateral *ʎ; this is supported by the fact that *ɲ has merged with *ʎ in all Formosan subgroups except for Tsouic (Blust 2009: 579).

The question of how exactly PAn *ʎ merged with *n in PMP is a thorny one. It would be simplest to argue that *ʎ first shifted to *ɲ which then merged with *n, but that leaves the question of why there are still etyma that can be reconstructed with *ɲ which did not undergo this shift. Since there is evidence that *t and *s were dental [t̪] and [s̪] (contrasting with alveolar *d, *n, and *l) (Blust 2009b: 552), it can be suggested that *ʎ passed through a stage where it depalatalized to a dental *l̪, remaining distinct from *ɲ (as well as alveolar *l) before its ultimate merger with *n.

There is reason to think that the shift to *n was not abrupt. This general trend toward de-palatalization is apparent in the variation which occurs in the handful of words that can be reconstructed with initial *ʎ and in which various non-Formosan languages preserve an initial [l]:

(71) Examples of PMP variation indicating original *ʎ

Gloss	PMP
swim	*[l/n]aŋuy
fish with poisonous dorsal spines	*[l/n]epuq
wound	*[l/n]uka
mosquito	*[l/ɲ]amuk

Examples of languages that retain [l] in these forms, selected from Wolff (2010), are given below:

(72) Examples of PMP forms which show reflexes of *ʎ

Gloss	Tagalog	Selayar	Malagasy	Malay	Javanese	T. Batak	Moken
swim	laŋoj	laŋe	lano	---	laŋuj	---	(naŋoj)
pois. fish	---	---	---	lepu	---	---	lepuuk
wound	---	loko	loka-na	luka	---	luha	lokaʔ
mosquito	lamok	---	---	(ɲamuk)	lamuk	---	(ɲamok)

There is another small group of forms which are more ambiguous, where Formosan languages provide evidence for *ʎ, but non-Formosan languages reconstruct to *l:

(73) Examples of variation between Formosan *ʎi- and PMP *li-

Gloss	PAn/PMP
sap	[N/l]iteq
water leech	qa[N/l]i-meCaq
den, pen	[N/l]ibu
bail out	[N/l]imas

Two interpretations of these data are possible. Either *ʎ was reinterpreted as *l before [i] in the PMP languages, or *l palatalized before [i] in the Formosan languages. A choice between these two options can't be made without more data, but this variation strengthens the interpretation of PAn *N as a palatal lateral more globally.

5.0 The distinction between *k and *g

Similar to the cases of *p and *t, there are two sets of reflexes associated with PAn *k:

Table 13: Reflexes of *k and *g

PMP	RMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao	PWOc
*k	*k	k	-k-	k	k	k	k	*k
*k	*g	g-, -ʔ-	-hk-	h	ɣ	0, -ʔ-	0, -ʔ-	*ɣ

We propose that the second set of reflexes associated with PAn *k be reconstructed as *g. These reflexes occur in all positions in Bimanese, Hawu-Dhao, and Western Oceanic; and in this case Nias shows this split in both initial and medial positions, although Dohoi continues to exhibit it in medial position only.

The development of the plain voiced stops in many WCMP languages is completely symmetrical, whereas it was asymmetrical in many other Austronesian languages. In this way, the original three-way contrast between voiceless, voiced, and implosive stops[58] was maintained. The majority of other Austronesian languages, conversely, experienced mergers of the bilabial and alveolar plain and implosive voiced stops on the one hand, and the plain voiced and voiceless velar stops on the other (the remaining palatal voiced stop then often merging with the voiced alveolar stop, although its development was more varied). We use the typical development of the WCMP languages as an example:

WCMP			Other languages		
ɓ	>	ɓ	ɓ	>	b
b	>	β	b	>	b
ɗ	>	ɗ	ɗ	>	d
d	>	r	d	>	d
k	>	k	k	>	k
g	>	ɣ	g	>	k

Examples of *k are given in the tables below:

[58] On the subject of Austronesian implosive stops, see Norquest & Downey (in preparation).

(74) Examples of *k in initial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
bite	*kaRat-i	---	---	*kati	kadi	kaɖi	
scratch	*kaRaw	---	kao	*kaʔu	kao	kao	
stand	*kideŋ	---	kidî	---	kədî	kədi	
belly	*kempuŋ	---	---	*kambu	kabu[59]	---	
pinch together	*kepit	---	kapi	*kəpit[60]	---	---	(Ind *kapit*)
lightning	*kilat	[mb]a-kila	kila	---	---	---	(Ind *kilat*)

(75) Examples of *k in medial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
split	*bekaq	---	---	*ɓəka	ɓəka	bəka	
open	*bukas	buko	---	*ɓukas	boka	bβoke	
joint, node	*buku	[mb]uku -	--	*ka-βuku	---	---	
bent	*ma-bukuq	---	mbuku	*ɓoko	buku	---	
tie, adhere to	*dekit	---	ɗiki	*ɗəkit	---	---	
dragnet	*puket	---	puka	*pukat	pəku	pəku	(Ind *pukat*)
come, arrive	*teka	---	---	*təka	dəka	ɖəka	

*g lenited to *ɣ in the WCMP languages (with reflexes which were often null or in variation with the glides [j] or [w] depending on vocalic environment) as well as in PWOc. Examples of *g in initial and medial position are given below:

(76) Examples of *g in initial position

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	PWOc
eat	*kaən	a[61]	ŋa-ha	*ŋa-ɣaŋ	ŋa-ʔa	ŋa-ʔa	*ɣani
elder sibling	*kaka	gaʔa	---	*ɣaɣa	aʔa	aʔa	---
tree	*kahiw	[g]eu	hadʒu	*ɣaju	aʄu	aʄu	*ɣaju
send	*(pa)-katu	---	ŋgadu	---	---	pa-ʔaɖu	---
see	*kita	---	eda	*ita	ŋ-ədi	ŋ-əɖi	---
1pl (incl)	*kita	---	ndai-ta	*ɣita	di	əɖi	*ɣita
cooking pot	*kuden	---	---	*urəŋ	əru	əru	---
skin	*kulit	[g]uli	huri	*ka-lit	k-uri	ka-uri	---
louse	*kutu	[g]utu	hudu	*utu	udu	uɖu	---

[59] 'chest'
[60] 'pinch between, insert'
[61] See also əʔa 'you eat', [g]iʔa 'he eats'

(77) Examples of *g in medial position

Gloss	PMP	Nias	Dohoi	Bima	PSumba	Hawu	Dhao	PWOc
monkey	*bakay	[mb]aʔe	bahkai	---	---	---	---	---
grime, dirt	*daki	[ndr]aʔi	---	---	*raɣi	raʔi	---	---
grass	*zukut	[ndr]uʔu	---	---	*rut	ʃuʔu	ʃuʔu	---
fish	*hikan	---	---	---	*iɣaŋ	---	iʔa	*iɣan
1sg	*i-aku	---	ahku	n-ahu	*jauwa	jaa	jaʔa	*[i]au
1pl (excl)	*i-kami	jaʔami[62]	ihkai	n-ami	*jamə	---	---	*ɣa[m]i
2pl	*i-kamu	---	ihkam	ndai-mu	*ɣəmi	mu	miu	*ɣa[m]u
2sg	*i-kahu	---	ihko	---	*ɣaʔu	au	au	*ɣo[e]
tail	*ikuR	[g]iʔo	ihku	---	(*ka-iku)	---	---	---
(finger)nail	*kuku	---	---	uhu	*[k/ɣ]uɣu	kuʔu	kuʔu	*ku[k/ɣ]u
man, husband	*laki	---	---	rahi	*laɣi	laʔi	laʔi	---
steal	*nakaw	---	nahkou	---	---	naʔo	naʔo	*pa[i]naɣo
wound	*nuka	---	---	---	*nuɣa	noʔe	nuʔa	---
fern	*paku	---	pahku	fahu	*paɣu	---	---	---
flying fox	*paniki	---	---	panihi	*paniɣi	niʔi	---	---
climb	*sakay	---	---	---	*[s]aɣe	haʔe	caʔe	*[s/z]aɣe
elbow	*siku	[z]iʔu	(siku)	cihu	*siɣu	hiʔu	siʔu	---
fear	*ma-takut	a-taʔu	---	dahu	*ma-ndaut	me-daʔu	ma-ɖaʔu	*mataɣut
root	*wakaR	waʔa	---	---	---	---	---	*wa[k/ɣ]ar

Additional examples of *k and *g are given below:

(78) Examples of *k

(a)

Gloss	PMP	Bima
little finger	*kiŋkiŋ	kiŋgi
mangrove	*bakaw	wako

(b)

Gloss	PMP	PSumba	
mangrove crab	*kali-maŋu	*kalimaŋu	
hook	*kawit	*kait	(Ind *kait*)
breadfruit	*kuluR	*kulu	
go	*lakaw	*lako	

(c)

Gloss	PMP	Hawu	Dhao	
men's house	*kamaliR	kemali	---	
grandfather	*baki	---	baki	
fold	*lekuq	ləku	ləko	
tie	*hiket	əki	əki	(Ind *ikat*)

(79) Examples of *g

(a)

Gloss	PMP	Nias
intestine	*bituka	[mb]etuʔa

(b)

Gloss	PMP	Dohoi
back	likud	ɾihkut

(c)

Gloss	PMP	Bima
hook	kawil	hawi

[62] 'you'

(d)	Gloss	PMP	PSumba
	octopus	kuRita	*βita

(e)	Gloss	PMP	Hawu	Dhao
	thick	*ma-kapal	me-aa	ma-aa

5.1 Flores-Lembata

The distinction between *k and *g is also preserved in Proto-Manggarai (in initial position) and some of the more remote languages of the Flores-Lembata group. Besides the already-mentioned Seranggorang-Nilanapo dialects, other languages include Komodo, Proto-Manggarai, Sika and Kedang. Examples are given below:

(80) Examples of *g in initial position

Gloss	PMP	Komodo	PManggarai	Sika	Ser-Nil	Kedang
eat	*kaən	ahaŋ	*ɣaŋ	a	a	(ka)
1pl (excl)	*kami	hami	*ɣami	ami	ome	(ke)
2sg	*kahu	---	*ɣau	au	---	o
thick	*kapal	---	---	apar	---	(kapal)
tree	*kahiw	---	*ɣazu	ai	---	ai
1pl (incl)	*kita	---	*ɣita	ita	ite	te
(finger)nail	*kuku	---	*ɣuku	---	---	---
skin	*kulit	---	---	ulit	uli	---
louse	*kutu	hutu	*ɣutu	utu	uhto	utu

(81) Examples of *g in medial position

Gloss	PMP	Komodo	Sika	Ser-Nil	Kedang
1sg	*aku	---	ʔau	---	---
fish	*hikan	---	iʔaŋ	iʔaŋ	iʔa
tail	*ikuR	---	ʔiur	iʔu	---
man, husband	*laki	lahi	laʔi-n	te-lae	---
steal	*nakaw	---	nao	---	---
climb	*sakay	---	haʔe	---	---
fear	*takut	---	(b-lau)	---	tauʔ

5.2 North Sarawak

Certain PNS languages have also preserved evidence for *g intervocalically via lenition: specifically Proto Dayic and Kiput. (Kelabit (K) forms are given in the absence of Proto Dayic reconstructions):

(82) Reflexes of PNS medial *g in Proto Dayic and Kiput

Gloss	PMP	PDayic	Kiput
large intestine	*bituka	*bətuəh	tufih
tail	*ikuR	*iur	cəuʔ
male	*laki	*də-laʔih	laaj
back	*likud	---	cut
steal	*nakaw	mə-no (K)	---
fern	*paku	*paʔuh	paaw
pain, sickness	*sakit	aʔit (K)	seet
elbow	*siku	*ijuh	ticəw
fear	*takut	*taʔut	---

Blust (2009b: 604) states that "Kelabit […] reflects *k as [ʔ] between unlike vowels provided that the first is not schwa, but as *k* elsewhere," implying a conditioned loss of *k intervocalically in Kelabit (and, by

implication, Proto Dayic). However, forms such as the following from the first part of Amster's 1995 Kelabit dictionary contradict this:

(83) Examples of Kelabit [k] between unlike vowels, the first of which is not schwa

Gloss	Kelabit
k.o. snail	akəp
agreement	akun
skin disorder	bakɛ
barren	bakəh
k.o. metal tool	bikuŋ
open	bukaʔ
stunted	bukəŋ

6.0 The uvular series

Although there has been a general longstanding acceptance of PAn *q as a voiceless uvular stop, the uvular series itself has always appeared generally impoverished, the only other member being *R which is sometimes interpreted variously as a uvular trill or voiced fricative. We propose that in addition to *q, the PAn inventory also included a voiced counterpart *ɢ as well as uvular liquid *ʀ (the latter of which nevertheless varied substantially in place of articulation in various Austronesian subgroups and daughter languages).

*6.1 Uvular *ɢ*

If the evidence presented here for the reconstruction of a novel PAn *g is accurate, then how should original PAn *g be reinterpreted? Traditional PAn *g is a phoneme which has been problematic because of the comparatively small number of forms in which it occurs as well as its general absence from any given core vocabulary list. Blust (2009b: 574-5) provides several examples in which *g is reconstructed, including examples that include the WCMP languages Manggarai and Sika, all of which have a [g] reflex: Manggarai *gəgər* 'shiver with chills, tremble' < PAn *gerger, *gərit* 'scratch, claw; scream' < PAn *geriC, *gagar* 'to like, have an appetite for (fighting, talk, sex)' < PMP *gagar 'bold', *pagal* 'hobble; heavy block hung from buffalo's neck to impede his movements' < PMP *pa(ŋ)gal 'neck shackle on domestic animal'; Sika gəmi 'pinch, shut, close (as in mouth)' < PMP *gemi 'hold on by biting'.

We propose that it be reinterpreted as an original voiced uvular *ɢ. This solution is attractive because it fills a gap in the reconstructed PAn phoneme inventory in which the voiceless uvular stop *q has always lacked a voiced counterpart; its relative infrequency may also be explained by the fact that uvular is a marked place of articulation.

Under this hypothesis, what is here reconstructed as PAn *g underwent two general developments: (1) devoicing and merger with *k (the case in the majority of Austronesian languages), or (2) lenition, sometimes to the point of deletion (as in the case of several WCMP languages and PWOc). *ɢ then filled this gap by shifting to the velar place of articulation:

*g > *k/ɣ
*ɢ > *g

In addition to the above, this hypothesis is also based on the following additional pieces of indirect evidence. First, as argued in the previous section, *g patterns with the other voiced stops *b, *d and *ɟ in undergoing lenition in the WCMP languages and PWOc. What we interpret here as *ɢ patterns with the implosive series (Norquest & Downey, in preparation) in resisting lenition, but is unlikely to be an implosive itself due to a general typological constraint on posterior implosives (Greenberg 1970, Maddieson 1984) as well as the fact that *ɢ may occur word-finally. Second, within the comparatively small part of the Austronesian lexicon in which *ɢ occurs, a disproportionate number of forms appear to be ideophonic reduplicants (*ɢapɢap 'stammer', *ɢuŋɢuŋ 'deep resounding sound', *ʈəɢʈəɢ 'pound, beat', etc), in which case a marked member of the phoneme inventory may be employed for effect. Finally, the

majority of disjunct/doublet forms listed in the ACD involve crossover with *k (*ɢisɢis ~ *kiskis 'scrape', *gərgər ~ *kərkər 'shiver, tremble', *ɢawaj ~ *kawaj 'tentacles', etc) which, without witnesses from the key languages listed in this paper, may also potentially be reconstructed with *g (i.e. *gisgis, *gərgər, *gawaj, etc.). This indicates a consonant that was close enough phonetically to *g to be in variation with it across subgroups; the most plausible gap in the PAn consonant inventory presented here is that of the voiced uvular stop.

*6.2 PAn *ʀ*

Wolff reconstructs *R as post-velar *ɣ. According to him (2010: 33), "[t]he change of *ɣ to [r] and [l] involves a change widespread in the world's languages, where a voiced velar spirant becomes a uvular trill and subsequently a tongue-tip trill, which may then merge with /l/." Blust, on the other hand, states (2009: 582-3) that "it is reasonable to suggest that […] *R [was] an alveolar trill, a hypothesis that is supported by known sound changes, since the shift of an alveolar trill to a uvular trill is well-attested in the history of French and other European languages, while a shift in the opposite direction is unknown."

We posit a third interpretation: that *R was originally a uvular trill [ʀ], of which retroflex [ɽ] was sometimes an allophone, and which contrasted with the alveolar trill *r. We consider it unlikely that *R was originally [ɽ], as [l̩] and [ɽ] would have been in close competition with each other phonetically; the existence of a retroflex series on the other hand may have provided an impetus for variation with [ɽ].

Nias, Bimanese, and PMS are consistent in their reflexes of PAn *r; however, there is a split in the reflexes of PAn *R in which they show unexpected reflexes of *r. The three sets of reflexes are the following:

Table 14: Rhotic reflexes

PMP	RMP	Nias[63]	Bima	PSumba	Hawu	Dhao
*R_1	*ʀ	0	0	0/ʔ	0	0
*R_2	?	r	r	r	r	r
*r	*r	r	r	r	r	r

Given the other split correspondence sets provided in this paper, it's tempting to wonder if the split in reflexes of *R indicate yet another reconstructible phoneme. However, unlike the cases of *p/f, *t/ʈ, *l/l̩, *s/c and *k/g, the two reflexes of *R don't show a strong correlation and instead appear to be somewhat random. The best explanation in this case is that items with the second reflex (*R_2) are borrowings from (or cross-contaminations with) Indonesian. Examples of *r and *R are given below:

(84) Examples of PMP *r

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
sound of grating	*kerit	---	---	*kərit	---	---	(Ind *kərit*)
insufficient	*kura	---	kura	---	---	---	
to rasp	*parud	---	---	---	---	paru	(Ind *parut*)
slice, cut fine	*qiris	---	---	*iris	---	---	(Ind *hiris*)
group	*quruŋ	---	uru	---	---	---	
ramie	*rami	---	---	*rame[64]	---	rame[65]	(Ind *rami*)
red	*rara	---	---	*rara(k)	rara	---	
slip/slide off	*rusrus	ruru	---	---	---	---	(Ind *rurut*)
stab	*rusuk	---	rusu	---	---	---	
write	*surat	[z]ura	sura	*surat	---	---	(Ind *surat*)
cockspur	*tara	[ndr]ara	tara	*tara[66]	tara	ɖara	

[63] *R in Nias additionally colored continguous non-high vowels to [o].
[64] 'k.o. high thicket'
[65] 'hemp'
[66] 'thorn, barb'

k.o. large bamboo	*teriŋ	---	---	*təriŋ	---	---	
directly through	*terus	---	---	*turus	---	---	(Ind *tərus*)

The majority of forms have null reflexes for *R[67]. However, a small residue of forms have reflexes of *r for PAn *R:

(85) Examples of PMP *ʀ

Gloss	PMP	Nias	Bima	PSumba	Hawu	Dhao	
new	*baqeRu	---	ɓou	*mbəʔu	wiu	hiu	
lung	*baRa	[mb]o	---	---	wa	haa	
swollen	*baReq	a-bao	---	*(m)baʔə	ɓai	bai	
brave	*baRani	(barani)	mbani	*ɓani	ɓani	bani	(Ind *bərani*)
rotten	*baRiw	---	mbai	*mbai	---	---	
hibiscus	*baRu	---	wau	*βau	wau	---	
rice	*beRas	([mb]əra)	(ɓura)	*βias	---	---	(Ind *bəras*)
give	*beRay	---	mbei	---	wie	hia	
paddle	*beRsay	---	wese	---	wohe	---	
wild taro	*biRaq	[mb]io	wia	*βia	---	---	
bear	*biRuaŋ	([mb]erua)	(ɓirua)	---	---	---	(Ind *bəruaŋ*)
blood	*daRaq	[ndr]o	raa	*raa	raa	raa	
earth	*daReq	---	rae	---	rai	rai	
bathe	*diRuq	---	ndeu	---	ƒiu	diu	
bone	*duRi	[ndr]oi[68]	ruwi	*rui	rui	rui	
west	*habaRat	(bara laowi)[69]	waa[70]	(*warat)	waa	haa	(Ind *barat*)
housepost	*hadiRi	---	rii	---	ge-rii	aga-rii	
needle	*zaRum	---	ndau	*roŋ	ƒau	ƒau	
bite	*kaRat-i	---	---	*kati	kadi	kaḍi	
scratch	*kaRaw	---	kao	*kaʔu	kao	---	
run	*laRiw	o-loi	rai	*lai	rai	rai	
heavy	*ma-beRat	a-bua	---	*mbuəto	---	bia	
red	*ma-iRaq	ojo	---	*mia	mea	mea	
come	*maRi	---	mai	*mai	mai	mai	
dry	*ma-Raŋaw	---	maŋo	---	maŋu	maŋo	
man	*ma-Ruqanay	---	mone	*mone	mone	mone	
k.o. tree	*naRa	---	(nara)	(*nara)	---	---	
hoarse	*paRaw	a-fau	ka-fao	---	---	---	
stingray	*paRih	---	fai	*pai	---	---	
slave	*qaRta	---	ada	*ata	---	---	
current	*qaRus	---	(aru)	(*aris) ?	---	---	(Ind *arus*)
salt	*qasiRa	[n]asio	sia	---	---	---	
bind, raft	*Rakit	---	(raŋki)	(*rakit)	---	---	(Ind *rakit*)
hundred	*Ratus	otu	(ratu)	(*rata)[71]	---	---	(Ind *ratus*)
root	*Ramut	---	amu	*amu	amo	amo	
thousand	*Ribu	---	(riwu)	(*riβu)[72]	(riwo)	(riho)	(Ind *ribu*)
house	*Rumaq	(ruma)	uma	*uma	əmu	əmu	(Ind *rumah*)

[67] Note that as pointed out in Blust (2008), the presence of a rhotic can often be inferred in PHD on the basis of *i in PHD.
[68] 'thorn'
[69] 'northwest'
[70] 'below'
[71] 'thousand'
[72] '10,000'

comb	*saRu	---	cau	*sai ʔ	---	---	
indigo	*taRum	---	dau	---	dao	---	
porcupine	*taRutuŋ	---	dudu	---	dudu	ɖuɖu	
tin, lead	*timeRaq	---	---	(*tambura)	(teməra)	(taməru)	
vein	*uRat	[n]uo	ka-ʔuʔa				
day	*waRi	(hari)	ai	---	---	---	(Ind *hari*)

Additional items in individual languages are given below:

(86) Regular reflexes of *ʀ

(a)	Gloss	PMP	Nias
	ember	*baRah	[mb]o

(b)	Gloss	PMP	Bima
	typhoon	*baRiuh	ɓadʒu
	squeeze	*peRas	pua
	evening	*Rabihi	awi

(c)	Gloss	PMP	PSumba
	octopus	*kuRita	*βita

(d)	Gloss	PMP	Hawu	Dhao
	rear	*buRit	wui	hui[73]
	fog	*Rabun	awu	ahu

(87) Irregular reflexes of *ʀ

(a)	Gloss	PMP	Bima	
	rotten	*buRuk	mburu	(Ind *buruk*)
	young girl	*daRa	ndara[74]	(Ind *dara*)
	wild duck	*ŋaRaq	ŋara	
	oyster	*tiRem	tire	(Ind *tiram*)

(b)	Gloss	PMP	PSumba	
	hibiscus	*baRu	*βaru	(Ind *baru*)
	pluck, pull out	*Rabut	*raβut	(Ind *rabut*)
	lines in palm	*uRat	*urat	(Ind *urat*)

(c)	Gloss	PMP	Hawu	Dhao	
	mangrove crab	*kaRakap	---	karaka	
	ribs	*Rusuk	ruhu	---	(Ind *rusuk*)

7.0 Conclusion

Evidence has been presented in this paper for three new phonemes (*f, *ɭ, and *g), as well as additional extra-Formosan evidence for *ʈ and an expanded domain for *c within PMP. The evidence comes from two subgroups on Borneo, as well as three of the four corners of the Austronesian-speaking world: Nias in the southwest, PWOc in the northeast, and WCMP in the southeast, with the conservative Formosan languages of Taiwan in the extreme northwest completing the picture.

According to the methodology of historical linguistics, whatever is reconstructed for PMP that is not the result of a conditioned split can be projected to the level of Proto Austronesian. The Out-of-Taiwan 'express train' hypothesis predicts that phonemic mergers should have occurred as the Austronesian

[73] 'stern'
[74] 'young'

expansion proceeded in time and space; the number of inherited phonemes for any node would be equal to or less than the number of those in the node above, and any secondary splits increasing the phoneme inventory which occurred in a lower node would be localized within that node with the conditioning factors likely remaining transparent.

As shown below in Table 15, however, this is not the pattern that appears. The Formosan languages are still unique in directly preserving evidence for the palatal phonemes *ç and *ʎ. Formosan evidence for *t̠, however, is now found in three other locations, and evidence for several other phonemes can be found in several other groups as well:

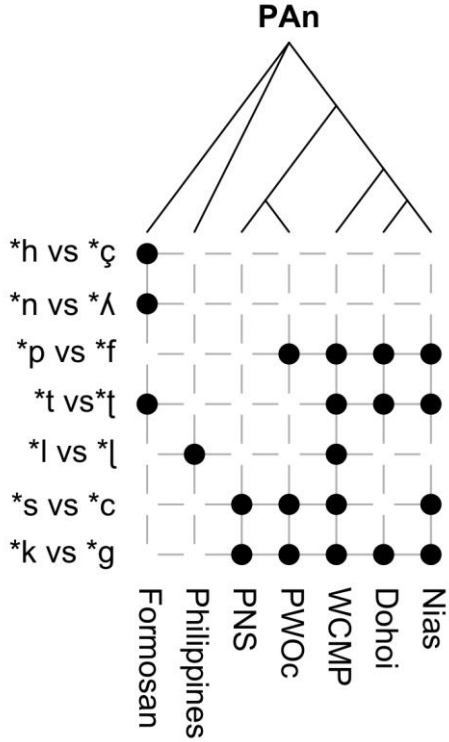


Figure 2: Archaic PAn phonemic distinctions

In terms of sheer number of distinctions preserved, WCMP is actually the most conservative group, followed by Nias. If one assumes CMP and EMP (SWHNG + Oceanic; See Fig. 1) to be the two lowest nodes of the Austronesian phylogenetic tree, then it is perplexing that they are more conservative than most WMP languages, the exceptions being the West Barito and North Sarawak groups on Borneo and the Barrier Islands group (to the extent that other languages of that region can be shown to subgroup with Nias[75]). How to interpret these data?

As mentioned above, the most conservative groups lie either on the periphery of the Austronesian-speaking world or on Borneo. Our present working hypothesis is that these languages represent an older layer of Austronesian languages that have been located in their present positions for some time. The WMP languages (excluding the Barrier Island languages), on the other hand were more recent expansions by various groups out of Borneo, possibly triggered by climate stress or other cataclysmic factors. The hypothesis that the Malayo-Chamic languages originated on Borneo (see for example Collins & Sariyan 2006) is well known; the South Sulawesi languages are related to the Tamanic group on Borneo, and the Philippine languages may subgroup with Sabahan (although this is still conjectural (Blust 1998)). If these examples are any indication, then it may be shown eventually that other WMP languages and subgroups originated on Borneo (cf. Blust 2010), and that immigration out of Borneo and into the surrounding islands

[75] We were quite fortunate to come across Lase (2011) in which the interphrasal distinctions in Nias word onsets were recorded for (what we believe to be) the first time. It is unclear whether similar distinctions may exist in other Barrier Islands languages, and we consider this to be one important aspect of future research.

has been occurring for quite some time, including quite possibly even the Philippine languages from Sabah in northern Borneo.

The phylogenetic tree in Figure 2 supports this conclusion. The tree was derived via a binary distance matrix based on phonological mergers and neighbor-joining. It suggests that the closest relationships between the easternmost Austonesian groups (WCMP and PWOc) are not to each other, but rather to discrete groups on Borneo, WCMP joining with Dohoi and PWOc joining with Proto North Sarawak. Although this phylogeny must remain tentative for now, we note that it is geographically consistent with two eastern migrations out of Borneo – one from southeast Borneo into the Nusa Tenggara region, and one from northern Borneo to the Bird's Head region of New Guinea which then spread eastward.

With this in mind, we propose an alternative to the traditional Austronesian expansion hypothesis: that the Formosan languages do form a subgroup, and that this group represents the first migration away from the original Austronesian homeland (and therefore the first split in the Austronesian phylogenetic tree, similar to the place of Anatolian within Indo-European). This Formosan-Malayo-Polynesian sister-group hypothesis would predict that retentions and innovations would be found in both subgroups, and not necessarily be constrained to WMP, as the OoT hypothesis would imply.

If it can be shown convincingly that the Formosan languages form a discrete innovation-defined subgroup and that the Formosan group is effectively a sister of PMP, then the question of Formosan origins becomes open – did the early Formosans migrate from mainland southeast Asia, as is commonly supposed, or might they have migrated from somewhere further south, perhaps ultimately from Borneo itself? Figure 2 above confirms that the Formosan languages still maintain a unique position in the Austronesian family tree, as they are the only languages to preserve concrete evidence for the phonemes we interpret as *ç and *ʎ. The same cannot be said for Proto Philippines, which retains evidence for *l̩ with WCMP (and evidence for *ŋ̩ -- if this hypothesis is valid -- with North Sarawak and Malayic). This question is ultimately outside the scope of this paper, but we hope to explore it in the future using both linguistic and non-linguistic evidence. The implications in regard to Southeast Asian prehistory and demographic processes are profound, and future work (especially interdisciplinary work) should bear this in mind as novel hypotheses are explored.

References

Adelaar, K. Alexander & Nikolaus P. Himmelmann. *The Austronesian Languages of Asia and Madagascar*. London/New York: Routledge.

Amster, Matthew H. 1995. Kelabit/English, English/Kelabit Glossary: A Concise Guide to the Kelabit Language. Kuching, Malaysia: Rurum Kelabit Sarawak.

Blust, Robert. 1977. The Proto-Austronesian Pronouns and Austronesian Subgrouping: A Preliminary Report. Working Papers in Linguistics 9 (2): 1-15.

_____. 1982. The Linguistic Value of the Wallace Line. *Bijdragen tot de tall-, land- en volkenkunde* 138: 231-50.

_____. 1984/85. The Austronesian Homeland: A Linguistic Perspective. *Asian Perspectives* 26: 45-67.

_____. 1988. Austronesian Root Theory: An Essay on the Limits of Morphology. Amsterdam: John Benjamins.

_____. 1991. The Greater Central Philippines Hypothesis. *Oceanic Linguistics* 30 (2): 73-129.

_____. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32 (2): 241-93.

_____. 1995a. The Prehistory of the Austronesian-Speaking Peoples: A View from Language. *Journal of World Prehistory* 9 (4): 453-510.

_____. 1995b. Notes on Berawan Consonant Gemination. *Oceanic Linguistics* 34 (1): 123-38.

_____. 1995c. The Position of the Formosan Languages: Method and Theory in Austronesian Comparative Linguistics. In *Austronesian Studeis Relation to Taiwan*, Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.). Taipei: Academia Sinica. 585-650.

_____. 1995d. Sibilant Assimilation in Formosan Languages and the Proto-Austronesian Word for 'Nine': A Discourse on Method. *Oceanic Linguistics* 34 (2): 443-53.

_____. 1995e. Austronesian Comparative Dictionary.

_____. 1997a. Ablaut in Northwest Borneo. *Diachronica* 14 (1): 1-30.

_____. 1997b. Nasals and Nasalization in Borneo. *Oceanic Linguistics* 36 (1): 149-79.

_____. 1998. The Position of the Languages of Sabah. In *PagtanaÏw: Essays on Language in Honor of Teodoro A. Llamzon*, Ma. Lourdes S. Bautista (ed.). 29-52. Manila: Linguistic Society of the Philippines.

_____. 1999. Subgrouping, Circularity and Extinction: Some Issues in Austronesian Comparative Linguistics. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, Elizabeth Zeitoun & Paul Jen-kuei Li (eds.). Taipei: Academia Sinica.

_____. 2000. Low-Vowel Fronting in Northern Sarawak. *Oceanic Linguistics* 39 (2): 285-319.

_____. 2001a. Language, Dialect, and Riotous Sound Change: The Case of Sa'ban. In *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society 1999*, Graham Thurgood (ed.). 249-359. Arizona State University Program for Southeast Asian Studies, Monograph Series Press, Tempe: Arizona State University.

_____. 2001b. Malayo-Polynesian: New Stones in the Wall. *Oceanic Linguistics* 40 (1): 151-55.

_____. 2002a. The History of Faunal Terms in Austronesian Languages. *Oceanic Linguistics* 41 (1): 89-139.

_____. 2002b. Kiput Historical Phonology. *Oceanic Linguistics* 41 (2): 384-438.

_____. 2005. The Linguistic Macrohistory of the Philippines: Some Speculations. In *Current Issues in Philippine Linguistics and Anthropology Parangal kay Lawrence A. Reid*, Hsiu-chuan Liao & Carl R. Galvez Rubino (eds.). 31-68. Manila: The Linguistic Society of the Philippines and The Summer Institute of Linguistics, Philippines.

_____. 2006a. Anomalous Liquid : Sibilant Correspondences in Western Austronesian. *Oceanic Linguistics* 45 (1): 210-16.

_____. 2006b. The Origin of the Kelabit Voiced Aspirates: A Historical Hypothesis Revisited. *Oceanic Linguistics* 45 (2): 311-38.

_____. 2007a. OÃma LoÏngh Historical Phonology. *Oceanic Linguistics* 46 (1): 1-53.

_____. 2007b. The Linguistic Position of Sama-Bajaw. *Studies in Philippine Languages and Cultures* 15: 73-114.

_____. 2008. Is there a Bima-Sumba Subgroup? *Oceanic Linguistics* 47 (1): 45-113.

_____. 2009a. The Position of the Languages of Eastern Indonesia: A Reply to Donohue and Grimes. *Oceanic Linguistics* 48 (1): 36-77.

_____. 2009b. *The Austronesian Languages*. Canberra : Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

_____. 2010. The Greater North Borneo Hypothesis. *Oceanic Linguistics* 49 (1): 44-118.

_____. 2012. Hawu Vowel Metathesis. *Oceanic Linguistics* 51 (1): 207-233.

Collins, James T. & Awang Sariyan (eds.). 2006. *Borneo and the Homeland of the Malays*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Dahl, Otto. 1976. *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London: Curzon Press.

Donohue, Mark & Charles Grimes. Yet More on the Position of the Languages of Eastern Indonesia and East Timor. *Oceanic Linguistics* 47 (1): 114-58.

Dyen, Isidore. 1963. The Position of the Malayopolynesian Languages of Formosa. *Asian Perspectives* 7 (1-2): 261-71.

_____. 1995. Borrowing and Inheritance in Austronesianistics. In *Austronesian Studies Relation to Taiwan*, Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng (eds.). Taipei: Academia Sinica. 455-519.

Greenberg, Joseph H. 1970. Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics* 36 (2): 123–145.

Greenhill, Simon, Robert Blust & Russell Gray. 2003-2012. Austronesian Basic Vocabulary Database.

Grimes, Charles. 2008. *Lii Hawu (Sabu) Online Dictionary*. Kupang: UBB-GMIT.

_____. 2010. Hawu and Dhao in Eastern Indonesia: Revisiting Their Relationship. In *East Nusantara: Typological and Areal Analyses*, Michael Ewing & Marian Klamer (eds.). Canberra: Pacific Linguistics.

Grimes, Charles, Tom Therik, Barbara Dix Grimes & Max Jacob. 1997. *A Guide to the People and Languages of Nusa Tenggara*. Artha Wacana Press. Kupang.

Haudricourt, André-Georges. 1950. Les consonnes préglottalisées en Indochine. *Bulletin de la Société de Linguistique* (46): 172-182.

Heine, Bernd. 1978. The Sam languages: a history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6.

Hudson, Alfred B. 1967. *The Barito Isolects of Borneo: A Classification Based on Comparative Reconstruction and Lexicostatistics*. Data paper: number 68, Southeast Asian Program, Department of Asian Studies, Cornell University, Ithica, New York.

Inagaki, Kazuya. 2005. Phonemic Sketch of Dohoi/Kadorih. *Kyoto University Linguistic Research* (24): 15-43.

_____. 2006. A sketch on the Morphosyntax of Kadorih (Dohoi: Austronesian). *Kyoto University Linguistic Research* (25): 41-65

_____. 2011. Word duplication in Kadorih. Asian and African Languages and Linguistics (6): 27-44.

Klamer, Marian. 1998. *A Grammar of Kambera*. Berlin/New York: Mouton de Gruyter.

_____. 2002. Typical Features of Austronesian Languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41 (2): 363-83.

_____. 2004. East Nusantara: Genetic, Areal, and Typological Approaches. *Oceanic Linguistics* 43 (1): 240-44.

_____. 2009. The Use of Language Data in Comparative Research: A Note on Blust (2008) and Onvlee (1984). *Oceanic Linguistics* 48 (1): 250-63.

Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lase, Apolonius. 2011. *Kamus Li Niha: Nias-Indonesia*. PT Kompas Media Nusantara. Jakarta.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Mansyur, Ismail, Azis Muhidin, Yakub Saleh, H. Taufik & Usman Kasim. 1985. *Kamus Bima-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.

Norquest, Peter & Sean Downey. Forthcoming. Austronesian implosives.

Onvlee, Louis. 1984. Kamberaas (Oost-Soembaas)-Nederlands Woordenboek. Dordrecht: Foris.

Pawley, Andrew & Malcolm Ross. 1993. *Annual Review of Anthropology* 22:425-59.

Paz, Consuelo J. 1981. *A Reconstruction of Proto-Philippine Phonemes and Morphemes*. Dilliman, Quezon City: Cecilio Lopez Archives of the Philippine Languages and the Philippine Linguistic Circle, University of the Philippines.

Ross, Malcolm. 1988. Proto Oceanic and the Austronesian Languages of Western Melanesia. Canberra: Pacific Linguistics.

_____. 1992. The Sound of Proto-Austronesian: An Outsider's View of the Formosan Evidence. *Oceanic Linguistics* 31 (1): 23-64.

_____. 2009. Proto Austronesian verbal morphology: a reappraisal. In *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Alexander Adelaar & Andrew Pawley (eds.). Pacific Linguistics: Canberra. 295-326.

Smoll, Laetitia. 2011. Towards a Typology of Ideophones. http://www.academia.edu/2649993/Towards_a_Typology_of_Ideophones.

Wolff, John U. 2010. *Proto-Austronesian Phonology with Glossary* (2 volumes). New York: Southeast Asia Program Publications, Cornell University.

Yunus, H. Anwar. 1981. *Kamus Bahasa Bima-Indonesia, Indonesia-Bima (Nggahi Mbojo)*. Jakarta: Yayasan "Muhammad Salahuddin".

Appendix

The following list of PMP forms have been amended to include the evidence presented in this paper for the proposed phonemes *f, *ʈ, *ɭ, *c, and *g. Additional distinctions not discussed in this appendix have been made between plain and implosive *b/*ɓ, *d/*ɗ and *ɟ/*ʃ which will be the topic of another paper (Norquest & Downey, in preparation). Traditional PMP *e and *y have been rewritten as *ə and *j respectively, in conformance with the IPA. Forms with unexpected correspondences, or in which otherwise crucial distinctions are neutralized due to phonotactic environment, are placed in parentheses; cognates that are suspected to be Indonesian loans are placed in curly brackets. We have only cited forms that include at least two cognates from two branches. We have also reconstructed Proto Hawu-Dhao (PHD) on the basis of forms in both languages as well as single forms in each, as the two languages are phonologically quite close to each other.

PMP initial *p

Gloss	RMP	Nias	Bima	PHD	PWOc
causative	*pa-	fa-	pa-	*pa-	---
warm/hot	*panas	---	pana	*pana	---
fruit bat	*panigi	---	panihi	*niʔi	---
broken	*paʈaq	(a-fatə)	mpada	*paɖa	---
squeeze	*pəʀəs	---	pua	---	*poRos
how many	*[p/f]iɖa	---	pila	*pəri	(*βiza)
choose	*piliq	fili	---	*pili	---
seven	*piʈu	fitu	pidu	*piɖu	---
dragnet	*pukət	---	puka	*pəku	---
white	*puʈiq	(a-fusi)	---	*puɖi	---

PMP medial *p

Gloss	RMP	Dohoi	S-N	Bima	PHD	PWOc
fathom	*dəpa	---	---	ndupa	*rəpa	---
fold	*ləpət	---	---	lipi	*ləpa	---
pinch	*qapit	---	---	api	*ŋ-api	*kapi
four	*ə[p/f]ac	(ohpat)	pa	upa	*əpa	(*βati)
sharp pain	*ha[p/f]əɖəs	([β]əxə)	---	pili	*pəɗa	---
mat	*ʈəpiʀ	---	---	dipi	*ɖəpi	---

PMP initial *f

Gloss	RMP	Nias	Bima	PHD	PWOc
paddy	*faɖaj	[β]axe	fare	*are	---
to shoot	*fanaq	---	fana	---	*βana
leucoderma	*fanaw	---	fano	*ano	---
pandanus	*faŋdan	[β]andra	fanda	---	---
thigh	*faqa	[β]aha	---	---	*βaqa
chisel	*[p/f]aqət	[β]ahə	paa	*paa	---
hoarse	*faʀaw	(a-fau)	fao	---	---
stingray	*faʀih	---	fai	---	*βari
turtle	*fəŋu	[β]ənu	fonu	*əŋu	*βoŋu
ten	*ca-ŋa-fuɭuq	[β]ulu	mpuru	*ŋ-uru	---
navel	*fusəɖ	[β]usə	---	*əsu	---

PMP medial *f

Gloss	RMP	Dohoi	S-N	Bima	PHD	PWOc
thick	*ma-kafal	kahpan	---	---	*ma-aa	---
tooth	*nifən	ɲihpo	ihpe	---	---	---
dream	*h-in-ifi	nuhpi	---	nifi	*nii	*m-niβi
thin	*ma-nifis	mihpi	mihpi	{nipi}	*ma-nii	---
centipede	*qaḷu-hifan	(ɟoripan)	---	rifa	---	*qaliβan
gall	*qafəḍu	---	---	folu	*əḍu	---
lime (for betel)	*qafuʀ	---	---	afu	*ao	---
fire	*hafuy	ahpui	ahpe	afi	*ai	*(j)aβi
reward	*ufaq	uhpa	---	ufa	---	---

PMP initial *t

Gloss	RMP	Nias	Bima	PHD
natural cockspur	*[t]ara	([ndr]ara)	tara	*[t/ṭ]ara
head covering	*təḍuŋ	---	toḍu	*təḍu
three	*təlu	təlu	tolu	*təlu
pierce	*tuɓak	---	tuɓa	*təɓu
knock, peck	*tuktuk	---	tutu	*tutu
pole for boat	*tuku	---	tuku	*tuku
roast, burn	*tunu	tunu	---	*tunu
old (people)	*tuqah	tua	---	*tua
knee	*tuqud	---	ta-tuʔu	*tuu

PMP medial *t

Gloss	RMP	Dohoi	S-N	Bima	PHD
calf	*b[ə/i]tiəs	botih	---	wisi	---
cut, shop	*hə[t/ṭ]ək	(n-ohtok)	---	---	*əta

PMP initial *ṭ

Gloss	RMP	Nias	Bima	PHD
fear	*ṭagut	(a-taʔu)	dahu	*ma-ḍaʔu
rope	*ṭaḷih	{tali}	dari	*ḍari
bury	*ṭanəm	[d]anə	---	*pa-ḍana
earth	*ṭanəq	[d]anə	dana	---
feces	*ṭaqi	[d]ai	(taʔi)	*ḍei
indigo	*ṭaʀum	---	dau	*ḍao
sea	*ṭasik	---	dasi	*ḍasi
person	*ṭau	---	dou	*ḍau
sugarcane	*ṭəɓuh	---	doɓu	*ḍəɓu
mat	*ṭəpiʀ	---	dipi	*ḍəpi
elder person	*ṭua	[d]ua	dua	---
fish poison	*ṭuba	[d]uβa	duwa	---
freshwater eel	*ṭuna	---	duna	*ḍəno

PMP medial *ṭ

Gloss	RMP	Dohoi	S-N	Bima	PHD
stone	*baṭu	bahtu	wahto	wadu	*βaḍu
blind	*ɓuṭa	---	---	mbuda	*ɓəḍu
send	*(pa-)gaṭu	---	---	ŋgadu	*pa-ʔaḍu
louse	*guṭu	kuhtu	uhtu	hudu	*uḍu
see	*(g)iṭa	---	---	eda	*-əḍi
we (excl)	*giṭa	ihto	---	(ndai-ta)	*əḍi

raw	*ma-qəʈaq	---	---	mada	*maɖa
eye	*maʈa	mahta	---	mada	*maɖa
die	*m-aʈaj	mahtoi	mahta	made	*maɖe
seven	*piʈu	pihtu	---	pidu	*piɖu
white	*puʈiq	---	---	pudi	*puɖi
liver	*qaʈaj	ahtoi	---	ade	*aɖe
cucumber	*qaʈimun	---	---	dimu	*ɖimu
egg	*qaʈəluʀ	(kotoɾuh)	---	dolu	*ɖəlu
porcupine	*ʈaʀu[t/ʈ]uŋ	(tohotuŋ)	---	dudu	*ɖuɖu
vomit	*(um)-uʈaq	ŋ-uhta	---	---	*məɖu

PMP *l

Gloss	RMP	PPh	Bima	PHD
return	*ɓaliw	---	ɓali	*ka-ɓali
buy	*bəli	*bəlíʔ	weli	*βəli
seaward	*lahud	---	ka-lau	*ləu
sail	*lajaʀ	---	lodʒa	*lai
fold	*ləpət	---	lipi	*ləpa
five	*lima	*limáʔ	lima	*ləmi
day	*qaləɖaw	---	liro	*loɗo
egg	*qatəluʀ	*ʔiklúg	dolu	*ɖəlu
three	*təlu	(*ta-tɭúʔ)	tolu	*təlu
ear(wax), deaf	*ʈuli, *ʈilu	*tulíʔ	---	*ɖilu

PMP *ɭ

Gloss	RMP	PPh	Bima	PHD
reverse	*baɭik	---	wari	*βari
moon	*buɭan	*búɭan	wura	*βəru
ear of grain	*bu[l/ɭ]iʀ	(*búlig)	wuri	*βuri
body hair	*buɭu	*buɭbúɭ	wuru	*βuru
in(side)	*ɗaɭəm	*ɖáɭəm	---	*ɗara
road	*ɟaɭan	*daɭan	---	*ɟara
skin	*guɭit	---	huri	*ka-ʔuri
rat	*ɭabaw	---	ka-rawo	*ma-raβo
man, husband	*ɭagi	*la-ɭáki	rahi	(*laʔi)
ginger	*[l/ɭ]aqia	(*laʔújaʔ)	rea	---
run away, flee	*ɭaʀiw	---	rai	*rai
ten	*ca-ŋa-fúɭuq	*púɭuʔ	mpuru	*ŋ-uru
pestle	*qahəɭu	*hákɭuŋ	aru	*aru
head(waters)	*quɭu	---	uru	*uru
torch	*suɭuq	*suɭúʔ	---	*suru
forest	*haɭas	*háɭas	ara	*ara
snake	*huɭaʀ	*ʔuɭág	---	*əru
rope	*ʈaɭih	---	dari	*ɖari
eight	*waɭu	*waɭúʔ	waru	*aru

PMP initial*s

Gloss	RMP	Nias	Bima	PDayic	PSumba	PHD	PWOc
error	*saɻaq	(sala)	sara	salaʔ (K)	*sala	(*sala)	---
bamboo floor	*saɻəʀ	[z]alo	sari	---	*ka-[s]al[e]	---	---
python	*sawa	[z]awa	sawa	---	---	---	---
douse fire	*sə[ɓ/b]u	---	suwu	---	---	*səɓu	---
ant	*səɖəm	[z]ixə	sa-sili	---	---	---	---
elbow	*[c/s]igu	[z]iʔu	(cihu)	(*ijuh)	*siɣu	*siʔu	---
torch	*suluq	[z]ulu	---	---	*[s]ulu	*suru	---
breast, milk	*susu	---	susu	tusu (K)	*susu	---	*susu

PMP medial*s

Gloss	RMP	Bima	PDayic	PSumba	PHD	PWOc
iron	*ɓəsi	ɓesi	---	*ɓəçi	*ɓəsi	---
paddle	*bəʀsaj	wese	bəsay (K)	*ɓose	---	*βoze
contents	*isi	isi	---	*içi	---	---
heart	*pusuq	---	*pusuʔ	(*puçu)	*usu	---
pierce, stab	*rusuk	rusu	n-usuk (K)	---	---	---
breast, milk	*susu	susu	tusu (K)	*susu	---	*zuzu
ripe	*ʈa-əsak	ndasa	---	---	*ma-ɖasa	---
sea	*ʈasik	dasi	---	*tasik	*ɖasi	---

PMP initial *c

Gloss	RMP	Nias	Bima	PDayic	PSumba	PHD	PWOc
climb	*cagaj	---	---	---	*[s]aɣi	*caʔe	*saɣe
bifurcation	*caŋa	---	caŋa	---	*[s]aŋa	---	*saŋa
comb	*caʀu	---	cau	---	*[s]ai	---	---
nine	*ciwa	siwa	ciwi	*iwaʔ	*çiwa	*ceo	---
comb	*cuɖud	suxu	---	udud (K)	---	---	---
spoon, ladle	*cuduk	---	ciru	---	*[s]uru	*curu	---

PMP medial *c

Gloss	RMP	Bima	PDayic	PSumba	PHD	PWOc
wet	*ɓacəq	ɓeca	*bahaʔ	*mba[s]a	{*ɓasa}	---
one	*ica	ica	(*əɗəh)	*içа	*əci	---
who	*i-cai	---	*iih	---	*cee	*[s/z]ai
sated	*ma-bəcuʀ	mcocu	*baur	*ka-mbəçu	*ɓəcu	---
salty	*ma-qacin	maci	m-ain (K)	*maçi	{*masi}	---
mercy, pity	*ma-qaciq	meci	*m-aheʔ	---	---	---
mortar	*əcuŋ	nocu	iuŋ (K)	*ŋoçu	---	---
navel	*fu[c/s]əɖ	---	*puəd	*pə[s]u	(*əsu)	---
smoke	*(ma)-qacu	---	---	*maçu	---	*qa[s/z]u

PMP initial *k

Gloss	RMP	Nias	Bima	PSumba	PHD
bite	*kaʀaʈ-i	---	---	*kati	*kaɖi
scratch	*kaʀaw	---	kao	*kaʔu	*kao
stand	*kəɗəŋ	---	kidi	---	*kədi
belly	*kəmpuŋ	---	---	*kambu	*kambu
pinch	*kəpit	---	kapi	*kəpit	---
lightning	*kilat	[mb]a-kila	kila	---	---

PMP medial *k

Gloss	RMP	Nias	PDayic	Kiput	Dohoi	Bima	PSumba	PHD
split	*ɓəkaq	---	---	---	---	---	*ɓəka	*ɓəka
open	*ɓuka	buko	bukaʔ(K)	bukaːʔ ---	---	---	*ɓukas	*mboka
node	*buku	[mb]uku	*bukuh	bukaw	---	---	*ka-buku	---
bent	*ma-ɓukuq	---	---	---	---	mbuku	*ɓoko	*mbuku
adhere	*ɗəkit	---	*dəkət	dəkət	---	ɗiki	*ɗəkit	---
dragnet	*pukət	---	pukət(K)	---	---	puka	*pukat	*pəku
arrive	*ʈəka	---	---	---	---	---	*təka	*ɖəka

PMP initial *g

Gloss	RMP	Nias	Bima	PSumba	PHD	PWOc
eat	*gaŋ	a	ŋa-ha	*ŋa-ɣaŋ	*ŋa-ʔa	*ɣani
elder sibling	*gaga	gaʔa	---	*ɣaɣa	*aʔa	---
tree	*gaju	[g]eu	hadʒu	*ɣaju	*aʃu	*ɣaju
thick	*gafal	---	---	---	*ma-aa	---
send	*(pa)-gaʈu	---	ŋgadu	---	*pa-ʔaɖu	---
see	*(g)iʈa	---	eda	*ita	*-əɖi	---
1pl (incl)	*giʈa	---	ndai-ta	*ɣita	*əɖi	*ɣita
cooking pot	*gudən	---	---	*urəŋ	*əru	---
skin	*guʟit	[g]uli	huri	*ka-lit	*ka-ʔuri	---
louse	*guʈu	[g]utu	hudu	*utu	*uɖu	---

PMP
medial *g

Gloss	RMP	Nias	PDayic	Kiput	Dohoi	Bima	PSumba	PHD	PWOc
monkey	*bagaj	[mb]aʔe	---	---	bahkai	---	---	---	---
intestine	*bituga	[mb]etuʔa	*bətuəh	tufih	---	---	---	---	---
grime, dirt	*dagi	[ndr]aʔi	---	---	---	---	*raɣi	*raʔi	---
grass	*ɟugut	[ndr]uʔu	---	---	---	---	*rut	*ʃuʔu	---
fish	*higan	---	---	---	---	---	*iɣaŋ	*iʔa	*iɣan
1sg	*i-agu	---	---	(kaw)	ahku	n-ahu	*jauwa	*jaʔa	*[i]au
1pl (excl)	*i-gami	jaʔami	---	---	ihkai	n-ami	*jamə	---	*ɣa[m]i
2pl	*i-gamu	---	*gamih	(kamaj)	ihkam	ndai-mu	*ɣəmi	*miu	*ɣa[m]u
2sg	*i-gahu	---	---	---	ihko	---	*ɣaʔu	*au	*ɣo[e]
tail	*iguʀ	[g]iʔo	*iur	cəuʔ	ihku	---	(*ka-iku)	---	---
(finger)nail	*[k/g]ugu	---	---	---	---	uhu	*[k/ɣ]uɣu	*kuʔu	*ku[k/ɣ]u
man	*ʟagi	---	*də-laʔih	laːj	---	rahi	*laɣi	*laʔi	---
back	*ligud	---	---	cut	ɾihkut	---	---	---	---
steal	*nagaw	---	mə-no (K)	---	nahkou	---	---	*naʔo	*pa[i]naɣo
wound	*nuga	---	---	---	---	---	*nuɣa	*nuʔa	---
fern	*fagu	---	*paʔuh	paːw	pahku	fahu	*paɣu	---	---
flying fox	*panigi	---	---	---	---	pahihi	*paniɣi	*niʔi	---
climb	*cagaj	---	---	---	---	---	*[s]aɣe	*caʔe	*[s/z]aɣe
elbow	*sigu	[z]iʔu	*ijuh	ticəw	(siku)	cihu	*siɣu	*siʔu	---
fear	*ma-ʈagut	a-taʔu	*taʔut	---	---	dahu	*ma-ndaut	*ma-ɖaʔu	*ma-taɣut
root	*wagaʀ	waʔa	---	---	---	---	---	---	*wa[k/ɣ]ar

CARVETH, Vincent. 2013. Pathways of development for Qiándōng Hmongic aspirated fricatives.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:146-157
Received 12/3/2013, text accepted 3/5/2013, published November 2013
URL: http://hdl.handle.net/1885/10611
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics

PATHWAYS OF DEVELOPMENT FOR QIÁNDŌNG HMONGIC ASPIRATED FRICATIVES

Vincent Carveth
University of Calgary
<vcarveth@pncl.com>

Abstract
The Qiándōng dialects of Hmongic are characterized by the presence of multiple aspirated spirants (Carveth 2012, Jacques 2011, Wang 1979). This paper proposes three pathways of development for those fricatives, using Qiándōng data from Ma & Tai (1956) and Purnell (1970). The first, leading to alveolar and palatal aspirated fricatives in Qiándōng, is an extension of Wang's (1979) analysis of a chain shift in the Yǎnghāo dialect. Labiodental aspirated and unaspirated fricatives are reconstructed as having come from palatalized bilabial stops, akin to Pulleyblank's (1984) reconstruction of Middle Chinese. Finally, lateral aspirated fricatives developed from the spirantization of aspirated liquids.
Key words: Hmong-Mien, historical phonology, aspiration.
ISO 639-3 language codes: hmq, hea, hms.

1. Introduction

The focus of this paper is the Qiándōng Hmongic subgroup of the Hmong-Mien language family. Qiándōng speakers number roughly 1.4 million people in southeastern Guizhou province in southwest China (Niederer 1998:51). Qiándōng dialects are loosely associated with the Black Hmong/Miao ethnic subgroup in Guizhou (1998:77). However, the correlation is a loose one, as the Hmong/Miao ethnic group extends beyond Guizhou to much of southern China and northern Indochina, and other Black Hmong groups do not appear to speak Qiándōng varieties.

A major distinctive feature of Qiándōng Hmongic dialects is the presence of multiple aspirated fricatives, each contrasting with an unaspirated equivalent. This paper, drawing on Carveth (2012), aims to reconstruct the etiology of this fricative series. Such work is valuable in that it expands our understanding of both Hmong-Mien language history and the typology of sound change more broadly.

The paper is laid out as follows. Section 2 discusses the crosslinguistic presence of aspirated fricatives, the significance of their presence in Qiándōng, and their phonetic and phonological nature. Section 3 describes a diachronic chain shift creating aspiration distinctions in Qiándōng alveolar and palatal fricatives. Section 4 details the evolution of labiodentals with contrastive aspiration from palatalized labials. Section 5 posits that aspirated lateral fricatives in Qiándōng are the result of a spirantization of aspirated liquid laterals. Finally, section 6 concludes.

2. Typology of aspirated fricatives

Aspiration distinctions are very rare in fricatives. Only three languages in Maddieson's (1984a) study have such a contrast, and only in alveolars. Jacques' (2011:2) effort to catalogue other examples details roughly 20 more such languages with aspiration distinctions in spirants. Most languages with this quality are in Asia, with Tibeto-Burman possessing a relative surplus, as Burmese, Sgaw Karen and several Tibetan dialects have aspirated fricatives. Given the close geographic proximity of these and other languages involved, for instance, the Tai-Kadai language Shan and the Hmongic dialects described here, it is likely that aspirated

fricatives represents an areal development in mainland Southeast Asia. However, examples exist outside this region, such as parts of the Oto-Manguean family (Kirk 1966, Knapp 1996, Pike & Pike 1947, Rensch 1976, Silverman et al. 1995), the Siouan language Ofo (Rankin 1988, Rankin 2004) and Chumashan (Klar 1977:13-15). This indicates that while aspirated fricatives may have arisen areally in Southeast Asia, they also have arisen elsewhere in an unrelated manner, suggesting the circumstances that generated them in Asia were not unique.

Qiándōng is remarkable even among the small set of languages with aspiration distinctions in spirants in that it has fricative aspiration contrasts in labiodentals, alveolars, palatals and laterals. The Yǎnghāo dialect has five aspirated fricatives, the most of any language represented in Jacques (2011), and Qiándōng dialects on the whole have a similarly robust set of contrasts (Carveth 2012:12). As such, the Qiándōng group provides a unique opportunity to study the diachronic origin of a large array of aspirated spirants. This paper thus focuses on reconstructing sound changes involved in the creation of Qiándōng's aspirated fricatives, in order to add to the typological understanding of sound change as well as assist further historical work on Hmong-Mien languages.

It must be noted that the phonetic details of aspiration in fricatives are somewhat unclear. This is largely because phonetic studies of the phenomenon have not taken place. The main issue is that aspiration is traditionally associated with the [spread glottis] feature, but so are fricatives. Thus, difficulties arise when determining status of fricatives with regards to [spread glottis] in a language that has contrastive fricative aspiration (see Carveth 2012:46, Halle & Stevens 1971, Halle 1995, Vaux 1998). This is not merely a phonological problem; if fricatives are articulated with the glottis spread already, then the seemingly aspirated quality of the aspirated fricatives cannot be a simple matter of a spread glottis as it is in stops. Most likely, researchers transcribing aspiration on fricatives in these languages are hearing increased airflow as the primary cue, perhaps originating from a different degree of glottal compression. A phonetic study of aspirated fricatives would illuminate this topic greatly, but in the meantime this hypothesis is sufficient for reconstructive work.

3. Chain shift in alveolars and palatals

The question of how Qiándōng's aspiration distinction in alveolar and palatal spirants developed has been studied previously by Wang (1979), with regards to the Yǎnghāo dialect. Wang (1979) hypothesizes a chain shift between Proto-Hmong and modern Yǎnghāo. This chain shift combines the loss of prenasalization in consonants with a reductive process; presumably, prenasalization was lost as an articulatory simplification, which forced reduction in non-prenasalized equivalent phonemes to avoid ambiguity. Wang's Yǎnghāo chain shift is detailed below, the protoforms being his Proto-Hmong[1].

1) ^{n}ts- > s- > s^{h}-
^{n}tʂ- > ɕ- > ɕ^{h}-
^{n}ts^{h}-, -ts^{h} > s^{h}-
^{n}tʂ^{h}- > ɕ^{h}-

3.1 Second level heading

The chain shift ties together both of the major traits of Qiándōng dialects, the lack of the prenasalized consonants present in the rest of Hmongic and the presence of aspiration distinctions in fricatives, by suggesting that the loss of the prenasalized onsets in fact drove the development of aspiration distinctions in s- and ç-. As such, it has significant explanatory power. The possibility of extending Wang's (1979) analysis to the entirety of the Qiándōng group is then quite compelling. This approach is taken in this paper.

Listed on the following page are the relevant correspondence sets from Ma & Tai (1956) for Qiándōng alveolars and palatals, along with their reconstructions in Carveth (2012).

[1] The voiceless alveolo-palatal fricative noted by Wang (1979) here is transcribed in Ma & Tai (1956) as palatal and will be treated as such for the remainder of this paper, as are Wang's alveolo-palatal *^{n}*tɕ*$^{(h)}$ and *ɕ$^{(h)}$ forms. More fieldwork is necessary to determine the exact phonetic character of these items.

Table 1: Alveolar and palatal series in Ma & Tai.

Dialect	Ser. 1	Ser. 2	Ser. 3	Ser. 4	Ser. 5	Ser. 6	Ser. 7	Ser. 8
Reconstructed PQH Onset(s)	***ts**	***tç**	***tsʰ**	***tçʰ**	***ⁿts**	***ⁿtç**	***s/ⁿtsʰ**	***ç/ⁿtçʰ**
Cóngjiāng	ts	tç	tsʰ	tçʰ	s	ç	s	ç
Huánglǐ	ts	tç/tçʰ	tsʰ	tçʰ	s	ç	sʰ	çʰ
Jiābā	tç	tç	tçʰ	tçʰ	s	ç	sʰ	ç/kʰ
Jǐnpíng	ts	tç	tsʰ	tçʰ	ts	ç	s	çʰ
Jìngxiàn	ts	tç	tsʰ	tçʰ	s	ç	s	ç
Jīnzhōng	ts	tç	tsʰ	tçʰ	s	ç	sʰ	çʰ
Jiùzhōu	tç	tç	tçʰ	tçʰ	s	ç	sʰ	çʰ
Páitíng	ts	tç	tsʰ	tçʰ	s	ç	s	ç
Sāndū	ts	tç	tsʰ	tçʰ	s	ç	sʰ	çʰ
Sānsuì	ts	c	tsʰ	cʰ	ts	ç	s	ç
Táigǒngzhài	ts	tç	tsʰ	tçʰ	s	ç	sʰ	çʰ
Tàiyōng	ts	tç	tsʰ	tçʰ	s	ç	tsʰ	tçʰ
Wǔchà	s	tç	sʰ	tçʰ	s/sʰ	ç	sʰ	çʰ
Wūluò	ts	tç	tsʰ	tçʰ	s	ç	sʰ	çʰ
Xīnqiáo	tç	tç	tçʰ	tçʰ	s	ç	sʰ	çʰ
Xuānwēi	ts	tç	tsʰ	tçʰ	s	ç	sʰ	çʰ
Yǒng'ān	ts/tsʰ	tç/tçʰ	tsʰ	tçʰ	s	ç	sʰ	çʰ
Zhènyuán	s	tç	sʰ	tçʰ	s/sʰ	ç	sʰ	çʰ
Zhōuxī	l	tç/tçʰ	tsʰ	tçʰ	s/sʰ	ç	sʰ	çʰ

3.2 Extension of Wang's (1979) Chain Shift

Some alterations need to be made to expand Wang's analysis to this broader spread of dialects. This adds some complexity to the system, but does not disrupt the fundamentals of the analysis.

Instead of assuming that the derivation is from Proto-Hmong, this paper adopts Carveth's (2012:93) assessment that the prenasalized proto-forms are Proto-Qiándōng-Hmongic (henceforth PQH). However, given that all other fricative aspiration distinctions originate prior to the PQH stage, they may in fact also belong to a pre-Qiándōng stage of development. The changes for the family as a whole, drawing on Carveth (2012:94), are listed below.

2) PQH *ⁿts(ʰ)- > ts(ʰ)-, s(ʰ)- in differing dialects
PQH *ⁿtç(ʰ)- > tç(ʰ)-, ç(ʰ)- in differing dialects
PQH *ts(ʰ)- > tç(ʰ)-, s(ʰ)- in some differing dialects, no change in others
PQH *s- > sʰ-
PQH *ç- > çʰ-

In the figure on the following page, PQH consonants are shown on the left hand side, with arrows connecting them to their various reflexes in the modern dialects.

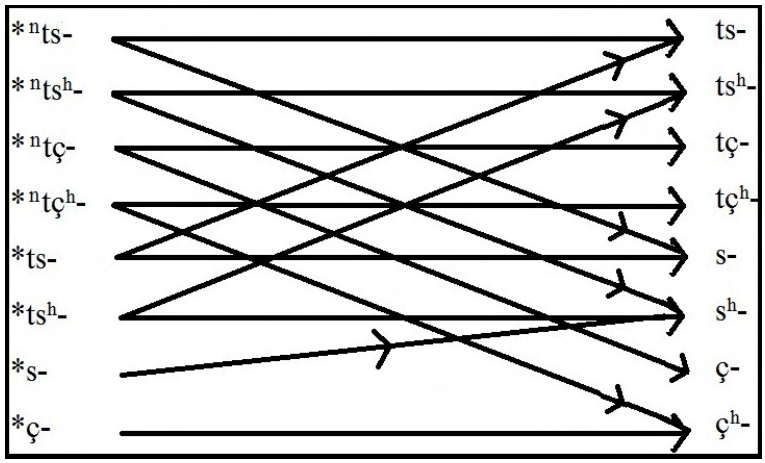


Figure 1: Diagram of Qiándōng palatal and alveolar reflexes.

Needless to say, this is more complicated. The non-prenasalized alveolar affricate noted by Wang (1979) not only does not always reduce to a fricative, but in many dialects moves backwards to become palatal. In other words, it changes independently from the reducing prenasalized affricates. Secondly, the prenasalized affricates in some dialects only reduce to affricates.

Still, the vast bulk of the system behaves according to the Yǎnghāo pattern. Only in three dialects, Jǐnpíng, Sānsuì and Tàiyōng, are there affricate reflexes rather than fricative ones for PQH's prenasalized fricatives (Carveth 2012:69). These are discussed further in section 6.1, and given the utility of Wang's (1979) analysis for the rest of the family, are ignored for the moment. In other dialects, the prenasalized affricates reduce to fricatives, as in Yǎnghāo.

3.3 Word sets demonstrating sound changes

Ma & Tai (1956) use very few actual examples to illustrate patterns, mostly relying on tables of sound correspondences, making specific etyma hard to come by at this time. This issue is compounded by the problem that in some cases, the series in Carveth (2012) have been produced by merger, such as the aspirate sʰ- series, which originates both from aspirated prenasalized affricates and unaspirated fricatives.

However, a few tables of examples can be shown to demonstrate the reconstructed onsets. Words behaving as part of series 7 and 8 in Table 1 are shown here in three representative dialects. Reflexes of the unaspirated prenasalized affricates, series 5 and 6, pattern similarly, but without the aspiration distinction.

Table 2: Series 7 and 8 reflexes.

Gloss	Cóngjiāng	Táigǒngzhài	Tàiyōng	Originating Series
'end'	[sa^{3}]	[sʰɛi^{3}]	[tsʰa^{3}]	7(*s/^{n}tsʰ)
'erase'	[ɕaŋ^{5}]	[çʰaŋ^{5}]	[tçʰaŋ^{5}]	8(*ç/^{n}tçʰ)
'song'	[ɕi^{7}]	[çʰa^{7}]	[tçʰa$^{6/7}$]	8(*ç/^{n}tçʰ)
'thousand'	[saŋ^{1}]	[sʰaŋ^{1}]	[tsʰaŋ^{1}]	7(*s/^{n}tsʰ)

(Ma & Tai 1956: 36)

The series reconstructed as non-prenasalized affricates are perhaps perplexing as they actually reduce to a fricative in fewer dialects than the prenasalized affricates do. However, evidence outside the family supports this reconstruction as well; 'five' and 'house' in the chart below have non-prenasalized postalveolar affricate cognates in the well known White Hmong variety of the Chuānqiándiān branch, tsib [tʃĩ^{1}] and tsev [tʃe^{3}] respectively. This begs the question of why the prenasalized series seems in so many dialects to have bypassed the purely oral affricate, rather than interact with it, and instead become a fricative. One possibility is that the loss of the prenasalization caused the loss of the affricate's closure in many cases, but the exact mechanism for this remains unclear.

Table 3: Alveolar affricate reflexes.

Gloss	Jiùzhōu	Táigǒngzhài	Wǔchà	PQH Onset
'fish basket'	[tɕʰa^5]	[tsʰa^5]	[sʰa^5]	*tsʰ
'five'	[tɕa^1]	[tsa^1]	[sa^1]	*ts
'house'	[tɕi^3]	[tsɛ^3]	[sa^3]	*ts
'stamp'	[tɕʰɔ^5]	[tsʰo^5]	[sʰɔ^5]	*tsʰ

(Carveth 2012:93, Ma & Tai 1956:37)

3.3 Nature of the Chain Shift

There is reason to believe that this is a push shift. First is the point that the loss of the prenasalization as a phonetic cue in stops would be a simplification of the sound, the removal of the nasalization gesture. The change at the level of the prenasalized affricate is thus quite phonetically natural. The change at the bottom of the system, the aspiration of unaspirated spirants, is not clearly so. Unaspirated fricatives are already produced with a spread glottis, so it is unclear what articulatory factors are involved in an aspirated fricative (Kingston 1990, Stevens 1991). Phonetic work on aspirated fricatives, unfortunately, is lacking (Carveth 2012:47). However, if such a change was phonetically motivated in and of itself, we would expect aspirated fricatives to be fairly widespread crosslinguistically, and as mentioned in section 2, they are not.

3.4 Burmese analogue to the Qiándōng Chain Shift

Two major analogues to this chain shift occur elsewhere. A chain shift occurred between Middle and Modern Burmese, prompted by the removal of [s] from the system, in which complex affricate clusters reduced in a similar way (Jacques 2011:8). Also, in southern Shan dialects, a change has occurred in which an unaspirated [s] aspirated and an affricate reduced to fill the gap (Edmondson 2008:197). Both resemble the Qiándōng shift in different ways.

Between Middle Burmese, that spoken around the fifteenth century, and the modern variety, a chain shift took place, the chronology of which is relatively well understood, thanks to the written record (Bradley 2011:4). The changes begin with development of Old/Middle Burmese *s-* into a dental affricate[2] sometime between the twilight of the 18th century and the middle of the 19th. Alveolar affricates appear to have reduced to fricatives, retaining their aspiration distinctions, between the 2nd Anglo-Burmese War in the 1850s and the 3rd in the 1880s. At some point in the latter half of the 19th century, velar stop-palatal glide clusters sibilantized to become affricates. These shifts are displayed below in chain form.

3) ts > s > tθ
   tsʰ > sʰ
   k(ʰ)y- > tɕ(ʰ)
   (Bradley 2011, Jacques 2011:7)

The most likely explanation is that the shift of the alveolar fricative to a dental affricate left Burmese with alveolar affricates, but no alveolar fricatives. As a result, the affricates' closure was no longer contrastive and thus was lost, the affricates becoming fricatives. However, the affricates originally had a contrast in aspiration, and this was maintained as they reduced, resulting in an aspiration contrast in alveolar fricatives in Modern Burmese. Subsequently, the gap in affricates allowed for the reduction and sibilantization of the velar stop clusters.

The reductive effects of the chain shift in Burmese are quite similar to those in Qiándōng; the clusters simplify to affricates and the simple affricates spirantize, with retention of aspiration distinctions. The major difference is that the timing suggests that the Burmese chain is motivated from the bottom, as the alveolar fricative removes itself from the system at the beginning of the shift, resulting in a drag chain.

In contrast, the Qiándōng shift is likely a push shift. The *s and *ç protoforms do not remove themselves from the system; instead, they aspirate. It is not clear that the gap this creates, a gap only in the

[2] The preaffrication is somewhat weak in Modern Burmese, and many sources inaccurately characterize the pronunciation as an interdental fricative (Bradley 2011:48).

unaspirated fricative, would motivate a reduction in both aspirated and unaspirated affricates. Moreover, the prenasalization contrast in Hmongic is crosslinguistically rare and loss of the distinction is thus far from strange; the Qiándōng shift could have started at the top with a simple lenition of the prenasalization, and all other elements would follow in a push chain shift.

3.5 Shan analogue to the Qiándōng Chain Shift

Shan also has a relevant and similar set of changes in its history. The following changes are known to have occurred in southern Shan (Edmondson 2008:197).

4) *s/z > sʰ
*tɕ/dʑ > s-

Northern varieties of Shan contrast the same word sets, but with the original affricate-fricative distinction rather than an aspiration distinction in fricatives. This is shown in the following table.

Table 4: Shan alveolar etyma.

Sibilant source	Gloss	Northern Shan	Southern Shan
*dʑ	'boy, male'	tsaai^{2}	saai^{2}
*dʑ	'elephant'	tsaaŋ^{4}	saaŋ^{4}
*dʑ	'artisan'	tsaaŋ^{6}	saaŋ^{6}
*tɕʰ	'prince, lord'	tsau^{3}	sau^{3}
*s	'three'	saam^{1}	sʰaam^{1}
*s	'four'	si^{5}	sʰi^{5}
*s	'unmarried girl'	sau^{1}	sʰau^{1}
*s	'tiger'	sɯ^{1}	sʰɯ^{1}
*s	'pillar'	sau^{1}	sʰau^{1}
*s	'tall, high'	suŋ^{1}	sʰuŋ^{1}
*z	'wash (hands)'	sak^{8}	sʰak^{8}
*z	'wash (surface)'	suk^{8}	sʰuk^{8}

(Carveth 2012:93, Ma & Tai 1956:37)

Unfortunately, written evidence from Shan is not as extensive as that from Burmese. The direction of this shift is largely known from previous Tai-Kadai reconstructive work, and philological evidence has little to say on the timing of the two changes. As deaffrication is phonetically motivated, being the weakening and loss of the stop closure, the deaffrication could have occurred for phonetic reasons and motivated the subsequent aspiration of the fricative. Alternately, the deaffrication could have resulted in a fricative distinct from the alveolar (perhaps something farther back, given the phonetic nature of the affricates), and distantiative effects may have caused the alveolar to aspirate, making aspiration and not place of articulation the distinctive cue.

Arguments for a pull chain in Shan are weaker. As discussed with Qiándōng, the phonetic naturalness of a spontaneous aspiration of s- is dubious. However, Shan is in close and intimate contact with Burmese, and the Burmese change may have influenced Shan. The geography supports this possibility as well, given that the southern dialects that were affected are those closest to the areas inhabited by ethnic Burmese. Still, this only removes the problem one step, in that the Burmese change's origin still needs to be pinned down.

Given the phonetic rationale, the Shan change likely began with deaffrication of the affricate onsets. In this matter, it is a better analogue for Qiándōng in that change also starts at the top of the system and moves through it via push chain effects. Shan, like Qiándōng, also includes the aspiration of unaspirated s-. However, Qiándōng's reduction of complicated clusters to affricates has no counterpart in Shan, unlike in Burmese.

3.6 Concluding remarks on the Qiándōng Chain Shift and its crosslinguistic counterparts

The Shan and Burmese changes suggest that the chain shifts like that posited for Qiándōng are viable pathways towards aspirated fricatives. They differ in their details, however, and likely in their actuation as well. In particular, Burmese appears to have developed its aspirated fricatives in a pull chain shift, whereas Qiándōng did so in a push chain. Further work on the cause of the Burmese chain shift would not only be informative in its own right, but might shed light on why Qiándōng behaved differently.

Wang's (1979) chain shift, given the alterations made in section 3.1, works well for the Qiándōng data at hand, in that it explains two of the Qiándōng branch of Hmongic's distinctive traits, its lack of prenasalized onsets and the presence of aspirated palatal and alveolar fricatives, with a single stroke. Furthermore, similar chain shifts exist in languages in other language families, adding to the plausibility of such a chain shift in Qiándōng. This paper accordingly proposes that Wang's (1979) Qiándōng Chain Shift, with this paper's modification, is indeed an accurate description of sound change in Qiándōng palatals and alveolars.

4. Labiodentals

Aspirated and unaspirated labiodental fricatives are both present in Qiándōng, with two series dominated by f- and f^h- reconstructed by Carveth (2012: 82) as *f and *f^h, respectively, in Proto-Qiándōng-Hmongic. The uniformity of these series suggests that the origin of this consonantal distinction lies in a stage prior to Proto-Qiándōng-Hmongic.

The chain shift analysis outlined in the previous section is insufficient to clearly elucidate the origins of the aspiration distinction in labiodental spirants. Any attempt to extend the Qiándōng chain shift to labiodentals would require positing a labiodental affricate of some kind, and there is no evidence anywhere in Hmongic for such a consonant, whereas the affricates involved in the chain shift are attested abundantly.

Furthermore, the chain shift produced aspirated affricate reflexes in some dialects, and no such remnant affricates are in evidence with labiodentals. This second problem could be circumvented by reconstructing only an unaspirated affricate and a Shan-type change as the origin of the distinction, but this raises the question as to why only the labiodental affricates would lack an aspiration contrast.

Finally, the aspiration distinction in labiodental fricatives might have been induced by its emergence in other fricatives in a solidarity chain shift, but no environment for a split is apparent. Without some sort of environment or condition to explain why some words aspirated and some did not, this option has little explanatory power and is not a compelling one.

An appealing alternative to extending the chain shift can be found in a pathway posited for Late Middle Chinese. According to Pulleyblank (1984), the Late Middle Chinese consonants *f and *f^h developed from Early Middle Chinese *p^j and *p^{jh}. Such a development is plausible in Qiándōng as well. From a phonetic standpoint, it is quite clear why such a change could occur; the movement of the tongue towards the palate during the palatalization gesture would tend to compromise the stop closure if the timing of the two gestures began to significantly overlap. Furthermore, there are similar changes elsewhere, most notably in the synchronic phenomenon of consonant lenition in Irish. Finally, palatalization is present elsewhere in Qiándōng, and palatalization and palatals are common elsewhere in Hmong-Mien (Carveth 2012:84).

Drawing on a proto-language as existing attestation of a pathway of change is potentially problematic, as proto-languages are subject to revision. However, this particular reconstruction is supported by Chinese rhyme tables indicating the existence of the aspirated spirants in Late Middle Chinese (Jacques 2011:4). Without these rhyme tables, the Late Middle Chinese consonants could potentially be reconstructed as affricates instead of fricatives. However, the phonetic naturalness of the change and the rhyme table evidence combine to create solid backing for this pathway in Middle Chinese, making its extension to Qiándōng tenable as well.

The spirantization of palatalized labial stops is thus phonetically natural and attested elsewhere. It also explains the data quite well, in that it would lead to an aspiration distinction in labiodental fricatives without any labiodental affricates in the system at any stage. A sample of etyma this account would explain is provided here.

Table 5: Labiodental etyma.

Original onset	Gloss	Táigǒngzhài	Gloss	Táigǒngzhài
*p^j	'melon'	[fa^1]	'bright'	[$faŋ^2$]
*p^{jh}	'be light'	[f^he^1]	'rub in hands'	[f^ha^1]

(Ma & Tai 1956: 35,39,47)

Accordingly, this paper proposes that such a retiming of gestures is indeed the origin of Qiándōng aspirated and unaspirated labiodental fricatives.

5. Laterals

Neither of the analyses developed in the previous two sections can account for the creation of an aspiration distinction in fricative laterals. There is no basis to posit a chain shift as in the alveolar and palatal fricatives, while the lenition of the palatalized stop is a direct result of its particular phonetic characteristics and cannot be applied to laterals. Instead, this paper argues that the presence of aspirated lateral fricatives is the result of a spirantization of aspirated liquid laterals. Furthermore, it is argued here that two aspirated lateral fricatives, palatalized and unpalatalized, exist due to a palatalization rule affecting laterals followed by a now-lost high vowel or onglide.

There are five series of laterals in modern Qiándōng, each one primarily dominated by a different one of the attested reflexes, and as such Carveth (2012:97) reconstructs *l, *l^j, *ɬ, *ɬ^h, and *ɬ^{jh} for Proto-Qiándōng-Hmongic. Following that analysis, this paper takes the origin of aspirated liquid spirants in Qiándōng to be prior to the PQH state. 5a and 5b below occurred in two distinct prior stages, henceforth called Early and Late Pre-Qiándōng.

5a) Early Pre-Qiándōng *l(^{h})[+high] > Late Pre-Qiándōng *l^j(^{h})

5a) Early Pre-Qiándōng *l(^{h})[+high] > Late Pre-Qiándōng *l^j(^{h})

5.1 Changes in laterals from a featural perspective

There are three contrasts among Qiándōng laterals: aspiration, palatalization and manner of articulation (fricative versus liquid). Some combinations are present in the data, while others are not. A look at which are available is revealing.

Table 6: Currently attested feature combinations in laterals.

	Unaspirated		Aspirated	
	Unpalatalized	Palatalized	Unpalatalized	Palatalized
Fricative	Attested	Unattested	Attested	Attested
Liquid	Attested	Attested	Unattested	Unattested

(Carveth 2012:105)

There are two gaps in the above paradigm. First, aspirated liquid laterals are not attested. More strangely, though, the unaspirated palatalized lateral fricative is also unattested. This is bizarre from an implicational standpoint, as the aspirated version is available. Given the rarity of aspirated fricatives, one would expect precisely the opposite.

One possible origin, the one chosen in this paper, would account for both anomalies. If the PQH aspirated fricatives originated from a spirantization of aspirated liquids, this implies a prior stage in which only liquids had palatalized variants. Such a stage is shown in Table 7 below.

Table 7: Available feature combinations in Late Pre-Qiándōng.

	Unaspirated		Aspirated	
	Unpalatalized	Palatalized	Unpalatalized	Palatalized
Fricative	Attested	Unattested	Unattested	Unattested
Liquid	Attested	Attested	Attested	Attested

At the Late Pre-Qiándōng stage, there is no implicational conflict with palatalized lateral fricatives simply because neither aspirated nor unaspirated ones exist. The spirantization created an aspirated palatalized lateral fricative after this stage, and the aspiration remained despite the lack of an unaspirated version due to the robustness of aspiration as a contrast in fricatives elsewhere in the language.

Moreover, this allows for a straightforward account of the application of palatalization. At the present stage shown in Table 6, it would be difficult to explain a palatalization taking place in aspirated fricatives and unaspirated liquids, but not in unaspirated fricatives. However, at the Late Pre-Qiándōng stage shown in Table 7, it becomes clear that palatalization only applied to liquids. Working backwards again produces the system shown in Table 8 for the Early Pre-Qiándōng stage.

Table 8: Available feature combinations in Early Pre-Qiándōng.

	Unaspirated	Aspirated
Fricative	Attested	Unattested
Liquid	Attested	Attested

The table below shows some etyma with their reconstructed Early Pre-Qiándōng onset, given the account of lateral development in this chapter.

Table 9: Early Pre-Qiándōng onsets for Táigǒngzhài laterals.

Gloss	Táigǒngzhài	Early Pre-Qiándōng Onset
‘a piece’	*ɬa^{8}	*ɬ
‘moon’	*ɬ^{h}a^{5}	*l^{h}
‘big’	*ɮ^{h}** ^{1}	*l^{h}
‘decay, rot’	*le^{2}	*l

** rime missing from text

(Ma & Tai 1956: 36,40,47)

5.2 Plausibility of aspirated liquid spirantization in crosslinguistic perspective

Given the utility of the spirantization account in explaining the patterns of liquid palatalization in Qiándōng and the lack of aspirated liquid laterals in the modern dialects, it is an attractive possibility for an origin of the fricative aspiration distinction. However, the question of plausibility of course arises with this account. Spirantization of an aspirated liquid to a lateral fricative with the aspiration distinction maintained intact has seemingly not been attested anywhere else (Jacques 2011).

Spirantization of other aspirated sonorants with retention of aspiration distinction has been reported in two other cases (Jacques 2011:9). Cone Tibetan is one. Modern Cone Tibetan has [ʂ^{h}è], ‘coarse’, coming from Old Tibetan *hral*, probably pronounced [r^{h}] in Old Tibetan. This initial segment has no known common origin in Old Tibetan. However, unaspirated /ʂ/ in Cone Tibetan arose from the clusters *sr-* and *spr-* in Old Tibetan; for instance, ‘cloud’ is [ʂĩ] in Cone Tibetan and sprin in Old Tibetan. Thus, at least for ‘coarse’, the existing aspiration contrast in Cone Tibetan postalveolar fricatives is the result of a spirantization of an aspirated rhotic element with retention of the aspiration distinction.

The other case is in the development of Chiquihuitlán Mazatec from Proto-Mazatec. Kirk (1966:95) reconstructs **ʃ^{h}* and **s^{h}* as Proto-Mazatec phonemes, but Jacques (2011:10) believes this to be in error. Out of the twelve dialects in his study, only Chiquihuitlán Mazatec has /ʃ^{h}/ synchronically, the remainder having /h/. Moreover, a Chiquihuitlán Mazatec dictionary antedating Kirk (1966) transcribes <jy-> for Kirk’s (1966) /ʃ^{h}/. This suggests a pronunciation of [j^{h}] or perhaps [h^{j}] or [x^{j}] rather than [ʃ^{h}], suggesting transcription error

in the prior documents may have led Kirk's (1966) analysis astray. Jacques (2011:10) suggests that if the original phoneme was in fact **j^h*, then the Chiquihuitlán Mazatec reflex is explicable as a spirantization and the other reflexes as loss of the palatal gesture, in effect, the loss of all but the aspiration.

5.3 Concluding remarks on the evolution of Qiándōng laterals

In summary, the spirantization of aspirated liquid laterals is posited here as the origin of the aspiration distinction in lateral fricatives in Qiándōng. Doing so both accounts for the lack of aspirated liquid laterals in modern dialects and allows for a simple diachronic explanation of the palatalization distinction in laterals. Moreover, evidence from Tibetan and Mazatec exists that aspirated sonorants can spirantize and retain the aspiration distinction, so this analysis is in accordance with known typological facts about sound change.

6. Conclusion

Three separate pathways have been posited in this paper for the origin of aspiration distinctions in various fricatives in Qiándōng. Alveolar and palatal aspirated fricatives are the result of a chain shift. Labiodental aspirated fricatives are the product of the coarticulation of gestures in palatalized labial stops. Finally, aspirated lateral fricatives originate from a spirantization of aspirated liquid laterals. The first of these had already been suggested by Wang (1979), but the pathway reconstructed for labiodentals had heretofore only been found in a Middle Chinese reconstruction, and the spirantization of aspirated liquids as an origin for a spirant aspiration distinction is a new finding.

6.1 Dialects with aberrant reflexes

The three dialects with affricate reflexes for prenasalized affricate PQH consonants bear discussion at this point. In section 3's Table 1, Tàiyōng is listed as having affricate reflexes in series 7 and 8 instead of the expected fricatives. In Jǐnpíng and Sānsuì, the series of import predominantly is Table 1's series 5, composed largely of alveolar unaspirated fricative reflexes (Carveth 2012:69).

Tàiyōng also has no aspirated alveolar or palatal fricatives. One might thus be tempted to surmise that the prenasalized affricates only reduced to oral affricates, and given that we have interpreted the chain shift as a push chain, there was no impetus for the unaspirated fricatives to aspirate. However, the Tàiyōng reflexes of unaspirated prenasalized affricates in series 5 and 6 are unaspirated fricatives. It is as though only the unaspirated prenasalized affricates fully reduced.

Jǐnpíng and Sānsuì present their own difficulties. In these dialects, the reflex of *nts is ts-, and the reflex of merged *s/*nts^h is s-. As in Tàiyōng, the possibility again rears its head that the shift may not be completed in these dialects, at least in alveolars, with the prenasalized affricates reducing only to oral affricates and not inducing further changes. This would imply that the two dialects' fricative reflexes in Table 1's series 7 stem purely from PQH *s. Given the cursory nature of Ma & Tai's (1956) original study, this could indeed be the case; they may simply not have had adequate data to establish a contrast between Jǐnpíng and Sānsuì reflexes of PQH *s and *nts^h. However, unlike the alveolars, the palatals in Jǐnpíng and Sānsuì behave typically for Qiándōng and clearly take part in the chain shift normally.

No clear answers seem forthcoming on the behaviors of these three dialects, and more detailed data may be necessary to solve this puzzle. Nonetheless, the chain shift analysis accounts so well for all other dialects that it suffices to leave these dialects' particularities to future research.

6.2 A note on teleology

At first glance, an explanation in which multiple forces produce the same end result, aspirated fricatives, might seem all too convenient, even teleological. However, crosslinguistically it is not that strange. A variety of effects have been known to shape whole phonological systems in a similar way.

One such effect of note is that phonological systems tend towards symmetry (Hock 1986:152). This is likely an attempt at economization, in that symmetrical systems allow for the highest ratio of distinctions to features. In the case of Qiándōng, it may well be that once aspirated spirants became established in the language, converting other contrasts into aspiration distinctions in fricatives created a more efficient system. Thus, some of the changes discussed in this article may have been in part motivated by the neurological tendency to self-organize and simplify.

Also relevant is the notion, discussed in section 2, that the aspirated fricative is an areal feature for the Indochinese region. Areal tendencies may also have pushed the system to develop aspirated fricatives, with the symmetry motivation becoming a factor once some such developments were already in progress.

Hock (1986:165) discusses several 'conspiracy' changes, such as those in Pali and Slavic, where multiple effects seem to have acted irregularly to create a regular change to syllable structures from the sum of their effects. While these cases are clearly different than what has happened in Qiándōng, as the changes in this paper are themselves regular and do not involve syllable structure, they are notable in that they show system-wide motivations for change. He writes, "Cases like these seem to require the assumption that at a certain point the results of various 'tactical decisions' along the way may build up enough 'critical mass' to establish a clear goal for further changes," (1986:165). In Qiándōng, the areal tendencies and symmetry motivations alone, or in tandem with some of the changes already beginning, may have provided the 'critical mass' of which Hock speaks to initiate a shift towards aspiration in fricatives as a system-wide distinction. In any case, the question of teleology is surmountable.

6.3 Potential directions for future research

Future research could continue in several directions. As far as aspirated fricatives are concerned, a phonetic study to determine their exact articulatory nature would help in ascertaining what phonetic factors might play into their diachronic development. So far, the only language for which such a study has been conducted is Korean (Kim et al 2011). In Korean, tense unaspirated alveolar fricatives contrast with lax aspirated ones, and the study concluded that the contrast in Korean was in fact a fortis/lenis distinction in tenseness. The tense/lax contrast, widespread in Korean, is not present in other languages with aspirated fricatives, making Kim et al's (2011) conclusions' crosslinguistic applicability unclear. An articulatory study on a language possessing a more prototypical aspiration contrast in fricatives would be highly useful in understanding the sounds' production, which would inform future phonetic and phonological work.

With regard to Hmong-Mien historical phonology, further refinement of reconstructions of Proto-Qiándōng-Hmongic and Pre-Qiándōng protolanguages could help shed light on the findings of this paper. Reconstructions of other branches of Hmongic would be useful as a comparison for this paper's findings on Qiándōng. There is still some disagreement on the exact substructure of the Hmong-Mien family tree, and a clearer delineation of branch arrangements and boundaries may be a prerequisite for further subfamilial reconstructions (Carveth 2012, Ratliff 2010). Finally, reconstructive work on Qiándōng vowels and phonetic tone values would clarify Qiándōng's past sound changes.

Further reconstructive work on Qiándōng will likely require more and better fieldwork data on its dialects. In particular, the three dialects discussed in section 6.1 are prime targets for field linguistics. However, the sheer scarcity of Qiándōng vocabulary in Ma & Tai (1956) and Purnell (1970) makes any increase in the depth of data on these dialects invaluable.

Regardless of the approach chosen, there is clearly much work to be done. Nonetheless, the study of Qiándōng's aspirated fricatives has the potential to pay dividends in our understanding of phonetics, phonology, sound change, and Hmong-Mien language history.

References

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Bradley, David. Changes in Burmese Phonology and Orthography. Keynote presentation at 21st South East Asian Linguistics Society Conference, Kasetsart University, Thailand, May 2011.

Bybee, Joan. 2003. *Phonology and language use*. Cambridge: Cambridge University Press.

Carveth, Vincent. 2012. Explorations in the phonological reconstruction of Proto-Qiándōng-Hmongic onsets. MA thesis, University of Calgary.

Edmondson, Jerald A. 2008. "Shan and other northern tier southeast Tai languages." In: Diller, Anthony V.N., Edmonsdson, Jerald A., and Yongxian, Luo (eds.). *The Tai-Kadai Languages*. London: Routledge. Pp. 184-206.

Halle, Morris. 1995. "Feature geometry and feature spreading." In: *Linguistic Inquiry* 26:1. Pp. 1-46.

Halle, Morris, and Stevens, K.N. 1971. "A note on laryngeal features." In: *MIT Research Laboratory of Electronics Quarterly Progress Report* 101. Cambridge: Massachusetts Institute of Technology. Pp.

198-213. Reprinted in Halle, Morris (2002). *From memory to speech and back: papers on phonetics and phonology 1954–2002*. Berlin & New York : Mouton de Gruyter.

Hock, Hans Heinrich. 1986. *Principles of Linguistic Reconstruction*. Berlin: Mouton de Gruyter.

Iverson, Gregory K. And Salmons, Joseph C. 1995. "Aspiration and laryngeal representation." In: *Phonology* 12. Pp. 369-396.

Jacques, Guillaume. 2011. "A panchronic study of aspirated fricatives, with evidence from Pumi." In: *Lingua* 121:9.

Jamieson, Carole. 1996. *Diccionario Mazateco de Chiquihuitlán*. Oaxaca and Tucson: Instituto Lingüisco de Verano, A.C.

Kim, Chin-Wu. 1970. "A theory of aspiration." In: *Phonetica* 21. Pp. 107-116.

Kim, Hyunsoon, Maeda, Shinji and Honda, Kiyoshi. 2011. "The laryngeal characterization of Korean fricatives: Stroboscopic cine-MRI data." In: *Journal of Phonetics* 39. Pp. 626-641.

Kingston, John. 1990. "Articulatory binding." In: Kingston, John and Beckman, Mary E. (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press. Pp. 406-434.

Kirk, Paul. 1966. Proto-Mazatec phonology. PhD dissertation, University of Washington.

Mă, Xuéliáng and Tái Chānghòu. 1956. "A preliminary survey of the phonology of the Miao dialects in the southeast of Kweichou." In: *Yǔyán Yánjiū* 1. Pp. 265-282. Translated in Purnell, Herbert C. (ed.) (1972). *Miao and Yao Linguistic Studies: Selected Articles in Chinese*, *Translated by Chang Yü-Hung and Chu Kwo-ray*. Ithaca, NY: Cornell University. Pp. 27-55.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Niederer, Barbara. 1998. *Les langues Hmong-Mien (Miao-Yao): phonologie historique*. Munich: Lincom Europa.

Okell, John. 1969. *A reference grammar of colloquial Burmese.* London: Oxford University Press.

Pulleyblank, Edwin G. 1984. *Middle Chinese, a study in historical phonology.* Vancouver: University of British Columbia Press.

Purnell, Herbert C. 1970. Toward a reconstruction of Proto-Miao-Yao. PhD dissertation, Cornell University.

Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.

Stevens, Ken. 1991. Acoustic phonetics. Ms. Massachusetts Institute of Technology.

Vaux, Bert. 1998. "The laryngeal specifications of fricatives." In: *Linguistic Inquiry* 29:3. Pp. 497-512

Wáng Fǔshī. 1979. *Miáoyǔ fāngyán shēngyùnmǔ bǐjiào [A comparison of rhymes and initials in Miao dialects]*. Beijing: Zhōngguó shèhuì kēxuéyuàn..

Wheatley, Julian K. 2003. "Burmese." In: Thurgood, Graham and LaPolla, Randy (eds.). *The Sino-Tibetan Languages*. London: Routledge. Pp. 195-207.

PRANG Thiengburanathum. 2013. Thai *jùu* and *kamlaŋ*: where Tense and Aspect meet.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:158-188
Received 22/12/2012, accepted 24/4/2013, published October 2013
URL: http://hdl.handle.net/1885/10616
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics

THAI *jùu* AND *kamlaŋ*: WHERE TENSE AND ASPECT MEET

PRANG Thiengburanathum
Payap University
<prangt@gmail.com>

Abstract
This paper discusses the similarities and differences between *kamlaŋ* and *jùu* based on their own internal logic. It shows that *kamlaŋ* does not simply express the ongoing progression of an event, but also indicates a temporal relation between time of situation (T-SIT) and topic time (TT) (Klein 1994). Based on Klein's theory of tense and aspect, *kamlaŋ* serves as a temporal relator indicating that T-SIT coincides with TT. The fact that Thai is not a tensed language does not mean that the concept of reference point should be neglected. Reference time (or topic time) is the key to temporal interpretation even in a 'tenseless' language like Thai.
The so-called continuous marker *jùu*$_{3/4}$ is treated as a locator locating an event in various domains such as time, attribute, quantity, and possession. Continuity is the output of our experience of remaining in the same place through time. It is a secondary function, which can be backgrounded. Like *kamlaŋ*, *jùu*$_4$ serves as a temporal relator, where T-SIT is situated at TT.
Key words: tense, aspect, reference point, temporal location.
ISO 639-3 language codes: tha.

1. Introduction

Time deixis plays a crucial role in understanding temporal relations. Many languages of the world employ 'tense' in structuring and encoding time. Previous scholars claimed that Thai contains tense markers (Uppakitsinlapasarn 1964, Supanvanich 1973, among others). More recent scholars, however, argue that Thai is in fact tenseless—it lacks a grammatical means to express tenses. This tenseless language, it is said, encodes time by means of pragmatic context and temporal expressions.

Current studies of temporality in Thai have refused tense-based accounts (Boonyapatipark 1983, Muansuwan 2002, Srioutai 2006, among others). Thai scholars turn to aspect, which is another linguistic category pertaining to temporality. Aspect has become a dominant field of linguistic investigation of the study of temporality in Thai. Even though, there is no uniform and generally accepted theory of aspect, most Thai scholars share at least two perspectives on what aspect is—1) aspect is not relational; rather, it expresses the internal temporal contour of the event; 2) the most basic aspectual distinction is between perfective and imperfective (Comrie 1976). These western characteristics of aspect have become the foundation to the studies of aspect in Thai. The main explanation of these studies is to determine whether the word in question is, say, perfective or imperfective.

If Thai is tenseless (in the traditional sense), it still is equipped with some devices to deal with time, in addition to relying on context for determining the temporal setting of a state of affairs.

Like other languages, Thai employs temporal adverbial phrases[1] to assign temporal locations. Temporal expressions (e.g. *mɨawaan* 'yesterday', *pʰrûŋnii* 'tomorrow') establish a temporal relation with respect to the absolute locus, which is always the speech time (i.e., the here-and-now).

[1] These include both calendric expressions (e.g. *sìp mooŋ* '10 o'clock', *pii tʰîi lɛ́ɛw* 'last year', *kumpʰaapʰan* 'February') and non-calendric expressions.

(1) *mîawaan* *Dɛɛŋ* *paj* *talàat* | *tɔɔn* *fŏn* *kamlaŋ* *tòk*
yesterday Daeng go market | when rain PROG fall
'Yesterday, Daeng went to the market when it was raining.'

The word *mîawaan* 'yesterday' in (1) signals that both events ('Daeng's going to the market' and 'raining') precede the time of utterance (TU). And if the speaker continues talking about Daeng, the listener will infer that all the events occur one day before the time of speaking without repeating the word 'yesterday'.

It is, nevertheless, inadequate for a language to merely situate all events in time with respect to a fixed reference point (TU), due to complexity of time. Any language must be equipped with various tools to cope with this complexity.

One of the facets of time in language is the internal composition of an event. This internal facet is where aspect comes into play. The two clauses in (1) have different internal temporal contours. The main clause 'Daeng went to the market' implies that the event is a completed act. The other clause 'it was raining' expresses that the event is extended into a progressive event.

The question is, are these devices (i.e., temporal expressions and aspectual markers) sufficient for communication? What about temporal relations between events (i.e., 'Daeng's going to the market' and 'raining'), then? How is one event temporally related to another? One might say that conjunctions (e.g. *tɔɔn* 'when') could do the work. However, there can be something else, which is succinct enough to express such a ubiquitous experience as time without invoking another clause as *tɔɔn* does. I suggest that *jùu* and *kamlaŋ* do this job in Thai.

This paper aims to show that *jùu* and *kamlaŋ* are not 'pure' aspect markers. That is, they do not simply specify the internal contour of an event like *râəm* 'start', or *sèt* 'finish' do, but also signal how events are temporally related. That is, they serve as '**temporal relators**', i.e., signaling **the way the event in question is distributed in relation to another event**, which is the topic time (TT) in Klein's terminology (1994). TT is "the time span to which the speaker's claim on this occasion is confined" (1994: 4). TT span can be relatively long or short.

The main purpose of this paper is to offer a new account on the TAM markers *jùu* and *kamlaŋ*. Thai is tenseless in the sense that it does not have grammatical means to express a temporal relation between utterance time (TU) and topic time (TT). But it has grammatical devices (such as *jùu* and *kamlaŋ*) to express a relation between time of situation (T-SIT) and topic time (TT). This paper also presents some of the semantic and pragmatic subtleties of *jùu* and *kamlaŋ* and shows how these affect their grammatical behaviors.

Jùu will be discussed first in Section 3.1, and then *kamlaŋ* in Section 3.2. In Section 4, temporal relation the notion relevant to *jùu* and *kamlaŋ* will be discussed in more details. The analysis of Section 4 is based on Klein's model of tense and aspect, which will be reviewed in the beginning of the section. The co-occurrence *jùu* and *kamlaŋ* will be discussed in Section 5. The following section gives a brief overview of previous treatments of *jùu* and *kamlaŋ*.

2. Previous studies of *jùu* and *kamlaŋ*

In recent years, Thai scholars have agreed that *jùu* and *kamlaŋ* should not be treated as present tense markers. There is general consensus that *jùu* and *kamlaŋ* are aspect markers (Boonyapatipark 1983; Kullavanijaya and Bisang 2007; Tansiri 2005; Iwasaki and Ingkaphirom 2005; among others).

Following the framework of the viewpoint approach (Comrie 1976), Boonyapatipark (1983) proposes that the *kamlaŋ* marker is employed to indicate an on-going situation at a particular time; and that the *jùu* marker causes a situation to be viewed as accumulating through time.

She examines co-occurrence restrictions between the aspect markers and her proposed verb classes. It is suggested that *kamlaŋ* should be considered a progressive marker since it can combine with dynamic verbs. The progressive marker disfavours achievement verbs. It does not frequently occur with state verbs, especially permanent states.

As for *jùu*, it is treated as a continuative marker which expresses "the continuance of a situation at the reference time" (1983: 99). Like *kamlaŋ*, *jùu* does not appear with achievement verbs. It is compatible with temporary states, but it is usually incompatible with permanent states due to its property of temporariness.

Kullavanijaya and Bisang (2007) analyse *jùu* and *kamlaŋ* in the framework of Selection Theory.[2] They study all possible co-occurrences of the aspect markers with the five proposed states of affairs: totally stative, action, gradually terminative, totally terminative, and inceptive-stative.

They find that the progressive is incompatible with the totally terminative state[3]. The marker crucially relies on a potential time span on which it operates. As such, it does not prefer generic statements of totally stative.

They disagree with Boonyapatipark's treatment of *jùu*. In their view, accumulating through time is not necessarily part of *jùu*. The marker *jùu* describes that "a situation is continuous through time or along time without reference to boundaries" (2007: 74). For this reason, *jùu* does not appear with inceptive-stative and terminative states of affairs. Since the continuity of *jùu* does not imply permanence, it is incompatible with generic states (or permanent states [Boonyapatipark 1983]).

The no boundaries concept of *jùu* is supported by Tansiri (2005), who refers to *jùu* as a stative imperfective aspect marker. The *jùu* marker is compatible with both dynamic and static situations. When occurring with the static situation, it causes the situation to be construed as the state persisting at the reference time. When occurring with the dynamic situation, the progressive situation is referred to, being construed as static. He observes that the locative meaning still remains in the aspect marker.

As for *kamlaŋ*, its treatment agrees with the other scholars' analyses—*kamlaŋ*, "a dynamic imperfective aspect marker", highlights the dynamic phase of the situation and construes it as the on-going situation. As such, it is incompatible with static and punctual ones.

Like the other scholars, Tansiri puts an emphasis on the interactions between *jùu* and *kamlaŋ* and lexical aspect (transitory state, inherent state, activity, accomplishment, achievement and semelfactive). The analysis focuses on the lexical aspect of the situations denoted by alternating intransitive constructions.

Iwasaki and Ingkaphirom (2005) also analyse *jùu* as a continuous aspect, but treat *kamlaŋ* as a preverbal adverb. No detailed explanations are given. They simply point out that *kamlaŋ* and *jùu* can co-occur [*kamlaŋ* + VP + *jùu*] and emphasizes a continuous situation.

The previous studies have tended to impose linguistic labels such as 'progressive' and 'continuous' uncritically as a reflex of an Indo-European bias. This paper attempts to show that the category of aspect in Thai may not be maintained rigidly. It will argue that the concept of reference (or topic time) is required in understanding the nature of *kamlaŋ* and *jùu*.

3. Proposed treatment of *jùu* and *kamlaŋ*

In addition to *lɛɛw*, *jùu* and *kamlaŋ* are probably the most studied expressions in the Thai literature on aspect. They are considered as imperfective aspect markers. Both are often translated as '-ing' in English. This translation is problematic since in some contexts, the markers can be used interchangeably, but in some other contexts they have different meanings. They also have different grammatical behaviors. Consider the following sentences, where *jùu* can be used, but *kamlaŋ* cannot:

(2)

a. *Pìtì* *kʰít* ***jùu*** *samǎə*
Piti think **stay** always
'Piti always thinks (about it).'

b. **Pìtì* ***kamlaŋ*** *kʰít* *samǎə*
Piti **PROG** think always
'Piti is always thinking (about it).'

[2] It is the combination of the viewpoint approach and the time-schema approach.
[3] Initial and terminal boundary collapse into one/no situation.

(3)

a. *Pìtì jaŋ tʰamŋaan* ***jùu***
Piti still work **stay**
'Piti still worked.'

b. **Pìtì jaŋ* ***kamlaŋ*** *tʰamŋaan*
Piti still **PROG** work
'Piti is still working.'

The sentences in (2) show that *jùu* can occur with the frequency adverbial *samǎə* 'always' (2a), but *kamlaŋ* cannot (2b). In (3), *jùu* can occur with *jaŋ* 'still' (3a), but *kamlaŋ* cannot (3b).[4] Indeed, *jùu* and *kamlaŋ* behave differently syntactically. However, to arrive at a more insightful explanation of their grammatical behaviors, it is important to understand their semantic and pragmatic natures. The aim of this section is to present and explicate some of the semantic and pragmatic subtleties of *jùu* and *kamlaŋ* and to show how these affect their grammatical behaviors.

3.1 Proposed treatment of jùu

The word *jùu* can be considered to have (at least) four senses (lexical and grammaticalized senses), which are differentiated by subscript numbers in the following discussion.

(4) *Pìtì* ***jùu$_1$*** *bâan*
Piti **stay** home
'Piti stayed home.'

Lexical sense: *jùu$_1$*

Semantically, the main verb *jùu$_1$* 'be at, live, stay' takes two arguments: a located entity and a location. The relation between the predicate and its arguments is a 'locator relation', which can be formalized as LOCATOR (locatum, location). The verb *jùu$_1$* serves the function of '**locator'**, *having an effect of locating a locatum in a location* (i.e., '**locator effect'**). This relation is sketched in **Figure 1**.

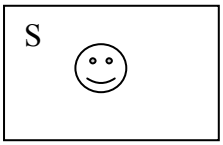


Figure 1: *Entity in Physical Space*

The box labelled S represents the space (i.e., location), while the face represents the locatum. The prototypical *locatum* of *jùu$_1$* is an entity, either animate or inanimate, and its prototypical location is a space. In (4), it denotes a relation between 'Piti' and 'house' such that 'Piti' is located at the house—LOCATOR (*participant, space*).

The Thai locative verb *jùu$_1$*, however, does not specifically convey how the entity is spatially related with the location. Frawley (1992: 254) describes that there are two kinds of spatial relations: topological and projective. Topological relations are constant under any change of the object—*coincidence* (on), *interiority* (in), and *exteriority* (out of). Projective relations are affected by viewpoint and thus variant—*inferiority*

[4] However, when *jùu* and *kamlaŋ* co-occur, the addition of *jaŋ* is possible, even though it is not frequently found: *jaŋ + kamlaŋ + VP + jùu*. Some speakers find this unacceptable.

(i) (hi5.com)
sǒŋsǎj jaŋ ***kamlaŋ*** *kin jùu*
suspect still **PROG** eat stay
'(He) probably is still eating.'

(below), *superiority* (above), *anteriority* (in front of), *posteriority* (behind), and *laterality* (between). **Table 1**: shows a list of common locative prepositions in Thai.

Table 1: *Locative markers*

bon	'on top of'
lâaŋ	'at the bottom of'
nâa	'in front of'
lăŋ	'behind'
naj	'inside'
nɔ̂ɔk	'outside'
tʰîi	'at'

The verb *jùu*$_1$ requires the occurrence of locative prepositions to complete spatial scenery, as exemplified in (5). Sentence (5b) illustrates that the deletion of the preposition *bon* 'on top of' results in an ill-formed sentence.

(5) (www.trekkingthai.com)

a. *nók* ***jùu*$_1$** *bon* *tônmáaj*
bird **stay** on tree
'Piti drew a picture/pictures at home.'

b. **nók* ***jùu*$_1$** *tônmáaj*
bird **stay** tree
'Birds stay the tree.'

There are some exceptions to this restriction. There are certain locations which *jùu*$_1$ can take without the need of these prepositions, for example, house, school, university, hospital, city names (e.g. Chiang Mai), country names (e.g. Thailand). This might be because the typical way a person is spatially in relation with these places is to be at the location. The preposition *tʰîi* 'at' thus can be omitted.

Note that there is a slight difference between, for example, *jùu*$_1$ *rooŋpʰayaabaan* 'stay hospital'[5] and *jùu*$_1$ *tʰîi rooŋpʰayaabaan* 'stay at hospital'. The former can be interpreted in two ways: 1) the participant is hospitalized and 2) the participant is physically located at the hospital. As for the latter, the preposition *tʰîi* 'at' places an emphasis on spatial relation—it does not imply the purpose of being there or the function of the hospital (although we can guess based on our encyclopedic knowledge). Table 2 shows the difference between *juu*$_1$ + LOC and *juu*$_1$ + *tʰii* + LOC.

Table 2: *The difference between jùu*$_1$ *+ LOC and jùu*$_1$ *+ tʰîi + LOC*

	***jùu*$_1$ + location**	**Meaning**	***jùu*$_1$ *+tʰîi* + location**	**Meaning**
Location	university	to study at the university level	university	to be located at the university
	Chiang Mai	to dwell in Chiang Mai	Chiang Mai	to be located at Chiang Mai
	house	to stay home	house	to be at the house

Grammaticalized sense: *juu*$_2$ (spatial locator)

As a grammaticalized verb, *jùu*$_2$ only occurs after a main verb or verb complex. The locatum can be semantically extended, from an entity to an event, as in (6), where the event is a 'drawing picture' kind of event, which is performed by Piti.

[5] This pattern of *jùu* (*jùu* + LOC) can only used with human (or human-like) subjects.

(6)

Pìtì	*wâat*	*rûup*	***jùu₂***	*tʰîi*	*bâan*
Piti	draw	picture	**stay**	at	home

'Piti drew a picture/pictures at home.'

What *jùu₂* does is to locate the event (i.e., drawing pictures) in a designated space (i.e., house), as shown in **Figure** 2. The circle labeled E represents the event.

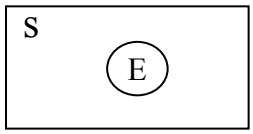


Figure 2: *Event in Physical Space*

The omission of *jùu₂* is possible[6], although it results in a different conceptualization—it appears to be 'generic'—less temporal and grounded.[7] It is also found in a different pragmatic and linguistic context (for example, an advertisement *ráp wâat rûup tʰîi bâan* 'teach drawing at home').

The preposition phrase (e.g. *tʰîi bâan* 'at home') designates a location, while the *jùu₂* -constituent (e.g. *jùu₂ tʰîi bâan* 'stay at home') designates a situation, specifically, a situation that obtains in a particular place. In (6), the noun expresses the spatial setting 'house' of 'Piti's drawing'. Here, *jùu₂* functions as a **spatial locator**—locating an event in space, LOCATOR (*event, space*).

Grammaticalized sense: *juu₃* (temporal locator)

Time[8] can be construed in terms of space. The spatial location word 'house' can be replaced by a temporal expression, for example, 'all day', as in (7). The *jùu₃*-constituent in (7) expresses the temporal setting of the event.

(7)

Pìtì	*wâat*	*rûup*	***jùu₃***	*tʰáŋ*	*wan*
Piti	draw	picture	**stay**	all	day

'Piti drew a picture/pictures for the whole day.'
(Piti's drawing holds all day.')

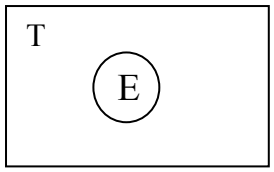


Figure 3: *Event in Temporal Space*

In Figure 3, the box labelled T represents a temporal space. The drawing event of (7) is located at a designated temporal location (i.e., *tʰáŋwan* 'all day'). This use of *jùu₃* functions as a **temporal locator**—the locator effect extends from space to time, LOCATOR (*event, time*).

---

6 The locative preposition is also predicative, as such it could occur without *jùu₂*.

7 The most equivalent English examples would be a) 'the picture above the sofa' vs. b) 'the picture was above the sofa'.

8 In addition to TIME, it is possible to have other target domains to locate the event, for example DANGER.

(ii)

kʰăw	*tòk*	***jùu***	*naj*	***ʔantaraaj***
3S	fall	**stay**	in	**danger**

'He is in danger.'

The expression in (7) specifies that the drawing event is anchored in time for the whole day. The word *jùu* indicates the all-day continuity of the event. Omitting *jùu*$_3$ is possible, but then (7) would simply mean Piti drew a picture/pictures all day. It does not profile on the relation between the located event and the temporal space. Additionally, it does not put much emphasis on the unchanging property which implies that Piti did not do anything else, but drew pictures all day. This semantic property will be discussed in more detail in Section 3.1.2.

Note that since space and time are logically parallel, it is not surprising to have a situation, as exemplified in (8a), where the same event is simultaneously located in time ('all day') and space ('house'). As such, it is possible to have [[*jùu*$_2$ + LOC] + [# + TEMP]][9] as a frame where the order of location and temporal constituents cannot be switched, as shown in (8b). The omission might be due to redundancy, since *jùu* can do double duty as a locative-temporal locator [*jùu*$_{2,3}$ + LOC + TEMP].

(8)

a. *Pìtì* *tʰamŋaan* ***jùu*$_{2,3}$** *bâan* *tʰáŋ* *wan*
Piti work **stay** house all day
'Piti worked at home for the whole day.'

b. **Pìtì* *tʰamŋaan* ***jùu*$_{2,3}$** *tʰáŋ* *wan* *bâan*
Piti work **stay** all day house
'Piti worked at home for the whole day.'

It is also possible to find contexts in which both *jùu*$_2$ and *jùu*$_3$ co-occur, although this co-occurrence is not frequently found. Sentence (9) demonstrates the structure of [[*jùu*$_2$ + LOC] + [*jùu* $_3$+ TEMP]]. TEMP of (9) refers to 'all the time'. The use of *jùu*$_3$ puts an emphasis on the whole period of time the speaker got to remain in the room.

(9) (my.dek-d.com)

cʰăn	*mâj*	*cʰâj*	*náktʰôot*	*náʔ*	*tʰĭŋ*	*càʔ*	*hâj*	*cʰăn*	*nâŋ*
1S	NEG	be	prisoner	Pt	CONJ	IRR	give	1S	sit

***jùu*$_2$**	*naj*	*hôŋ*	***jùu*$_3$**	*talɔ̀ɔt*	*weelaa*	*bɛ̀ɛp*	*níi*
stay	in	room	**stay**	all	time	like	this

'I am not a prisoner; (you could not tell) me to stay in the room all the time like this.'

More examples of *jùu*$_3$ are given in (10) and (11). Its occurrence is preferred for establishing the locational relation—locate an event in the temporal location.

(10) (www.santidham.com)

tʰân	*pen*	*săammáneen*	***jùu*$_3$**	*săam*	*pii*
3S	COP	novice	**stay**	three	year

'He was a novice for three years.'

(11) (pijitra.bloggang.com)

pʰŏm	*nɔɔn*	*cèp*	***jùu*$_3$**	*lăaj*	*cʰuâmooŋ*
1S.M	lie	hurt	**stay**	many	hour

'I was sick and lay down for many hours.'

Grammaticalized sense: *jùu*$_4$ (time-discourse locator)

As mentioned, the concept *jùu* inherently involves a location. Even in *jùu*$_4$, this facet of *jùu* is not lost. It is just extended to **temporal-discourse** use—the temporal location is contextually determined. The fourth

[9] # refers to *jùu*$_3$.

sense of *jùu* involves locating (a phase of) an event in reference time. To be more specific, it is LOCATOR (T-SIT, TT). That is to say, $jùu_3$ and $jùu_4$ (in the domain of time) indicate different kinds of time information. The temporal locator $jùu_3$ deals with *how long/how often an event lasts* (duration/frequency), while $jùu_4$ deals with *at what time (TT) an event is located.* As such, their locator effects are distinct. The locator $jùu_3$ locates an event 'in' a time frame, entailing that an event keeps going on or occurs in succession within the time frame. The locator $jùu_4$, on the other hand, locates a phase of an event 'at' a TT. Due to their difference, it is useful to make a terminological distinction. The term 'time frame' is employed to refer to the temporal location of $jùu_3$, while the topic time is for the temporal location of $jùu_4$.

Typically, TT is the moment of speaking encoded by temporal deixis. It can also be the moment another event is taking place as encoded by another clause. The temporal location of (12) is the time of speaking, which can be explicitly encoded by *tɔɔnníi* 'now'.

(12)

a. *Pìtì wâat rûup* ***jùu₄*** ***tɔɔnníi***
Piti draw picture **stay** **now**
'Piti is drawing a picture, now.'

b. ***tɔɔnníi*** *Pìtì wâat rûup* ***jùu₄***
now Piti draw picture **stay**
'Now, Piti is drawing a picture.'

The locator $jùu_4$ does not take any argument within a clause. It loses its verbiness[10] and functions as a grammatical marker establishing a relationship between the locatum and the location. The location or the temporal setting of the event can be represented in different ways. For example, it can be explicitly marked as an adverbial (e.g. 'now', 'when I arrived'), or it can be the time mentioned in the preceding context. More importantly, it does not have to immediately follow $jùu_4$. That is to say, *tɔɔnníi* 'can be fronted, as in (12b). This fronting operation is not allowed in the case of $jùu_3$, for instance, (7) and (10).

More examples of $jùu_4$ are given in (13) and (14). The temporal location of $jùu_4$ in (13) is the time the girl walked past Wisanu's room. In (14), $jùu_4$ locates the event when the hearer is told to end his/her romantic relationship.

(13) (Short Stories [CU Thai Concordance])

dèksăaw	*tɔ̀ɔp*	*léɛw*	*dəən*	*pʰàan*	*hôŋ*	*kʰɔ̌ɔŋ*	*Wítsanúʔ*
girl	answer	CONJ	walk	pass	room	POSS	Wisanu

hěn	*faj*	*jaŋ*	*pə̀ət*	***jùu₄***
see	light	still	open	**stay**

'The girl answered. Then, she walked past Wisanu's room. (She) saw the light still on.'

(14) (http://www.narak.com)

kʰít	*ʔaraj*	***jùu₄***	*tɔɔn tʰîi*	*tʰùuk*	*bɔ̀ɔk*	*lə̂ək*
think	what	**stay**	when	PASS	tell	cancel

What were (you) thinking, when (you) were told to break up?'

Note that we can insert a polite final particle (e.g. $kʰà$) in between $jùu_4$ and 'when (you) were told to break up?' in (14), or in between $jùu_4$ and 'now' in (12a). This is not allowed in the case of $jùu_2$ and $jùu_3$ (for example, in (7), we cannot say [draw picture + $jùu_3$ + $kʰà$ + all day]).

So far, we have seen examples of a straightforward relationship between locatum and location (i.e., locatum + *jùu*+ location). Example (15) shows that the temporal location (TT) can precede $jùu_4$.

[10] Important criteria for determining a verb class include negation and TAM markers.

(15) (Short Stories [CU Thai Concordance])

tɔɔn	*tʰîi*	*pʰǒm*	*tôɔnráp*	*lûukkʰáa*	***jùu₄***
time	that	1S.M	welcome	customer	**stay**

kʰun *Pʰɔɔncʰaj* *kôɔ* *jɨɨn* *jùu₂* *tʰěɛw* *níi* *duâj*
TA[11] Pornchaj CONJ stand stay area this also
'At the time I was welcoming the customer, Khun Pornchaj also stood around here.'

Here, the temporal location of *jùu₄* is not the time Khun Pornchai stood around. Notice that the temporal adverbial derives from [N_{temp} + *tʰîi* + NP + VP + *jùu₄*].[12] The temporal noun *tɔɔn* (lit. 'part, section') is the temporal location of *jùu₄*—at the time I was welcoming the customer. Here, *jùu₄* places an emphasis on that particular moment expressing that Khun Pornchai's standing around exists at the very time the speaker welcomed customers.

A summary comparison of the different senses of *jùu* is presented in Table 3.

Table 3: *Comparison of the different senses of jùu*

		jùu $_1$ [Ex. (4)]	*jùu* $_2$ [Ex. (6)]	*jùu* $_3$ [Ex. (7)]	*jùu* $_4$ [Ex. (12)]
grammatical function		matrix verb	subordinate verb[13]		grammatical marker
'locator' effect	locatum	entity	event	event	event
	location	space	space	time and beyond*	time-discourse and beyond*
syntactic construction		NP *jùu₁* LOC	Cl [*jùu₂* LOC]	Cl [*jùu₃* TIME]	Cl [*jùu₄* { Cl 'now' etc. }]

Note: the bracket { } indicates that temporal locations are not syntactic arguments of *jùu₄*.
* beyond the domain of time (this will become clearer in the next section)

The grammaticalized forms of *jùu* are far from semantically empty. They still have a clear relationship to the lexical source.

3.1.1 The locator effect of jùu

We have seen that the semantic content of *jùu* is molded into a grammaticalized *jùu* yielding the locator effect (for example, spatial locator, temporal locator). This section will show that the locator effect can be extended to more and more domains, especially in non-spatial-temporal domains.

Consider the following example.

(16) (http://bhudit.diaryis.com)

tʰɨ̌ŋmɛ́ɛwâa	*càʔ*	*mâj*	*dâaj*	*kwâaŋ*	*jàj*	*mâak*
even_though	IRR	NEG	get	large	big	much

[11] Term of address

[12] A similar pattern is also found with *jùu₂*: N_{place} + REL + NP + VP + *jùu₂*. For example:
(iii)
rooŋrian *tʰîi* *pʰǒm* *rian* *jùu₂...*
school that 1S.M study stay
'The school that I study at…'

[13] A subordinate verb modifies or adds to the meaning of main verbs. It takes a location as its argument forming a constituent.

tɛ̀ɛ	*kɔ̂ɔ*	*kwâaŋ*	***jùu₄***	*náʔ*
but	CONJ	wide	**stay**	Pt

'Even though it is not very big, it is big enough.'

This example is taken from an online diary. The writer keeps records of her family's (husband and son) activities. One day, the family went to an aquarium. The mother made a comment on the aquarium size which is sort of big. Here, the state of bigness is not located in time but on the scale of bigness itself—at the level of discourse expectation, as illustrated in Figure 4.

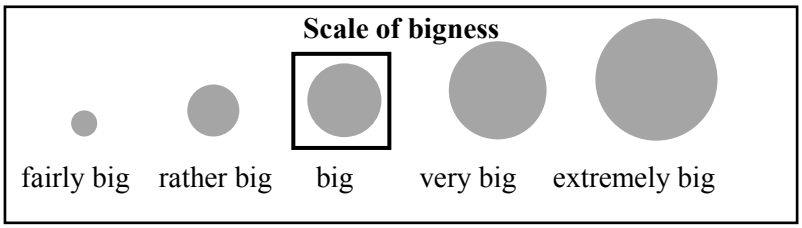


Figure 4: *The scale of bigness*

The square in bold denotes the speaker's subjective views about typical aquarium size. To assert the sentence is to say the aquarium meets discourse expectations for that property. The level of bigness can vary depending on the tone of the speaker.

(17) (www.meemodel.com)

pʰûujĭŋ	*léʔ*	*pʰûucʰaaj*	*kʰít*	*tàaŋ*	*kan*	*jáʔ*	***jùu₄***	*náʔ*
woman	and	man	think	differ	RECP	much	**stay**	Pt

'Women and men think quite differently from each other.'

This sentence expresses the difference in thinking processes between men and women. The difference in thinking is not construed as the state persisting at the reference time (i.e. at t_1 men and women think a lot differently, and at t_n they still think a lot differently), as suggested in previous studies. In my opinion, (17) has neither a continuous nor a stative imperfective reading. It involves the degree of difference in thinking—from a little to a lot. The locator effect of *jùu* causes the difference in thinking to be located on the scale of quantity. It expresses that the difference in thinking remains in the scope of 'a lotness'. It is neither a huge amount nor a little. It is somewhere in between. To put it another way, *jùu* does not profile the intermediate temporal phase of an event. Rather, it profiles the intermediate quantity scale. In (16) above, it profiles the intermediate attribute scale.

In order to further investigate the uses of *jùu₄*, we consider the following actual situation.

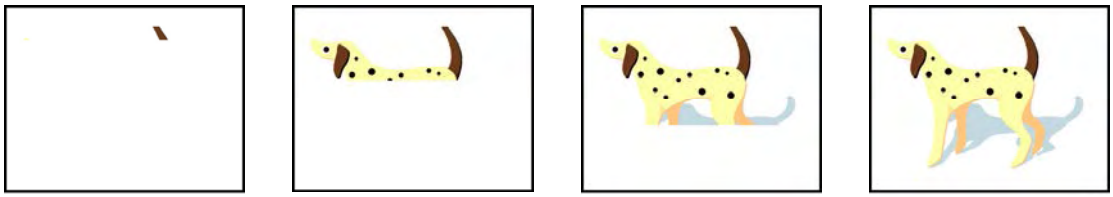
Situation: While auntie was taking a picture, my dog came and stood in front of everyone. Auntie said the dog ruined the picture because only its wagging tail could be captured. My uncle did not mind having the dog in the picture. So, he took turns to take pictures. He kneeled down so that he could capture both the people and the dog, although he was not sure if he could include the dog in the picture. Then, he instantly viewed the picture just taken. While doing that, he said:

(18) (Free conversation at grandma's house)

hěn	***jùu₄***
see	**stay**

'(I) saw (it_the dog).'

Sentence (18) is concerned with acceptable image quality.



(a) (b) (c) (d)

Figure 5: *Acceptable image quality*

The images in Figure 5a-d illustrate a range of perceptible images of the dog refered to in (18). There is the difference between visual perception and acceptable visual information. Our visual perception is the ability to interpret information from visible light reaching the eyes. However, not all visual information is considered acceptable or meaningful. What uncle actually 'wants to see' is a good photo of the dog. That is, the face, the whole body or the main part of the dog is captured—not just a tail. As such, only Figure 5c-d are acceptable[14]. The word *jùu* is employed to designate that the picture uncle just took is in the range of acceptable perception (i.e., the dog can be perceived).

Let us now turn to temporal use of *jùu*$_4$.

(19) (Free conversation at a restaurant)
Lek:

Námon,	*hěn*	*pʰrácan*	*jím*	*máj*
Namon,	see	moon	smile	Q

'Namon, Did you see the smiley moon.'

Namon:

hěn	***jùu*$_4$**
see	**stay**

'(I) saw (it).'

The smiley moon refers to a rare celestial trifecta of Venus, Jupiter, and the moon, which was witnessed in Thailand (and some other countries) on December 1, 2008. The conversation containing (19) took place on January 2, 2009. Lek had heard that Namon was out of town and might not have witnessed this spectacular event. She thus asked Namon if Namon had a chance to see this special phenomenon.

Unlike (18), (19) is temporally related. Tansiri (2005) suggests that *jùu* causes statives to be construed as persistent. Nevertheless, it would seem that what is focused here is not the persistence effect but the locator effect. Namon did not express that her seeing the smiley moon persisted at the reference time. Rather, (19) says that Namon's seeing the smiley moon existed at the reference time. She did witness the event when it happened.

Sentence (20) below illustrates the continuous use of *jùu*$_4$ which is given rise to by linguistic context.

(20) (www.songburi.com)

fáa	*mîit*	*lɛ́ɛw*[15]	*tɛ̀ɛ*	*kʰon*	*jaŋ*	*jə́ʔ*	***jùu*$_4$**
sky	dark	already	but	people	still	a lot	**stay**

'The sky is already dark, but there are still quite a lot of people.'

Unlike (17), which also contains the main verb *jə́ʔ* 'a lot', (20) conveys an aspectual meaning. It indicates that the number of people is unchanged. There were a lot of people before and at the reference time (i.e. at dusk). The cue word *jaŋ* activates the domain of time and the continuity value of *jùu*$_4$. Without *jaŋ*

[14] This is a matter of subjectivity. What is considered '*hěn jùu*$_4$' thus varies from one speaker to another. The point here is to show that *jùu*$_4$ does not simply function as a continuous marker, as previous studies claimed.

[15] *Lɛ́ɛw* is neither a perfect nor perfective marker, as previous studies suggested. It conveys an event transition (Thiengburanathum 2010). To avoid confusion from labelling, it is glossed as 'already'.

(and the context 'the sky is already dark'), the sentence is ambiguous (even incomplete). It could be interpreted as having a continuity reading (20) or a quantity reading as in (17).

If *jùu*$_4$ is a pure continuous marker (Boonyapatipark 1983; Iwasaki and Ingkaphirom 2005), it should be able to produce continuity interpretation regardless of inferential, pragmatic, or linguistic context. These examples show that Thai *jùu*$_4$ is not simply a grammatical aspect expressing temporal continuity.

Before moving to the next section, a brief discussion of *jùu*$_3$ is given. Like *jùu*$_4$, *jùu*$_3$ can cause an event to be located on a non-spatio-temporal scale.

(21) (www.komchadluek.net)

faj	*dàp*	***jùu*$_3$**	*sɔ̌ɔŋ*	*duaŋ*
light	extinguish	**stay**	two	CLF

'Two lights were out.'

(22) (Thai National Corpus)

pʰɔ̌m	*mii*	*lûukcʰaaj*	***jùu*$_3$**	*hòk*	*kʰon*
1S.M	have	son	**stay**	six	CLF

'I have six sons.'

In (21), *jùu*$_3$ is characterized against the domain of quantity (of concrete nouns). It focuses on the number of lights which went out in Soi Sukhumvit (Soi means 'a small lane'); the location is inferred from the previous discourse.

In (22), on the other hand, *jùu*$_3$ is conceptualized in the domain of possession. The occurrence of *jùu* is optional. It is used to place an emphasis on the number of sons existing in his possession.

One could argue that *jùu*$_3$ in (21) is actually understood against the domain of space (two lights went out at Soi Sukhumvit). A better example would be (23), which focuses on the number of dishes the speaker ate.

(23) (bubeexx.spaces.live.com/blog)

ʔaahăan	*tem*	*tóʔ*	*tɛ̀ɛ*	*kin*	***jùu*$_3$**	*caan*	*diaw*	*niâ*	*lɛ̀ʔ*
food	full	table	but	eat	**stay**	CLF	only	Pt	Pt

'There is a lot of food on the table, but (I) kept eating from one dish only.'

These examples show that the concept of location of *jùu*$_3$ is extended beyond time and space to quantity and possession. In the next section, the continuity effect of *jùu* will be discussed.

3.1.2 The continuity/unchanging effect of jùu

We have discussed the locator effect of *jùu*. What about its continuity value? How can the continuity property of *jùu* be accounted for? Let us recapitulate the semantic notion of *jùu*. The verb *jùu*$_1$ has the semantic effect of locating a participant in space. Moreover, it conveys that the participant remains in the same location without moving away throughout the period of time in focus.

The experience of remaining in the same place through time gives rise to the notion of **continuity**—the unbroken or consistent existence of an event over a period of time.

The notion of continuity has an '**unchanging**' value. To assert *Pìtì tʰamŋaan* ***jùu*$_4$** 'Piti is/was working' is to capture the current state of Piti, the fact that Piti was working rather than doing something else at the reference time. This continuity could be considered as a secondary function, which is not always active (even in the domain of time), as seen in the previous section (e.g. (19)). Together with the locator effect, the continuity effect has an influence on *jùu*'s grammatical behaviour, making it different from *kamlaŋ* (see the discussion of *kamlaŋ* in more detail in Section 3.2).

It should be mentioned that the 'unchanging' effect is not the same as 'stative' (contra to Tansiri 2005). *Jùu*$_4$ does not cause a dynamic verb to be construed as stative. A dynamic verb which co-occurs with *jùu* still involves action. This can be indicated by the following tests.

Table 4: *Criterion for dynamic verbs*

Criterion	VP + ***jùu***	Example
Occur with the progressive *kamlaŋ*	Yes	(24)
Occur with adverbials like *jàaŋkʰĕŋkʰăn* 'actively', *jàaŋkʰamàkkʰamên* 'diligently'	Yes	(25)
Occur with adverbials like *jàaŋruâtrew* 'quickly', *jàaŋcʰáacʰáa* 'slowly'	Yes	(26)

(24) (SEAlang Library Thai Corpus)
tɔɔnníi ***kamlaŋ*** *kin* *kʰâawpʰàtpʰrík* ***jùu₃***
now **PROG** eat fried_rice_with_chillies **stay**
'Now, (I) am eating fried rice with chillies.'

(25) (www.club4g.com/index.php?topic=174069.0;wap2)
ʔatʰíbaaj *jàaŋkʰamákkʰamên* ***jùu₃*** *kɨàp* *cʰuâmooŋ*
explain diligently **stay** almost hour
'(I) explained diligently for almost an hour.'

(26) (www.dharma-gateway.com)
rûup... *kàət* *kʰîn* *léʔ* *dàp* *paj* ***jàaŋruâtrew*** ***jùu₃*** *talɔ̀ɔt* *weelaa*
Rupa appear ascend and disappear go **quickly** **stay** all time
'Rupa... appears and disappears quickly all the time.'

According to Van Valin (2005: 33), dynamic events involve action, as indicated by the fact they can be modified by the progressive marker (test 1) and adverbs like *diligently* (test 2); *quickly* (test 3), as shown in Table 4. The fact that *jùu* can co-occur with these linguistic expressions suggest that *jùu* does not cause a dynamic verb to be construed as stative.

Because of this unchanging value, $jùu_3$ can take a manner adverbial such as *jàaŋníi* 'like this', *jàaŋdəəm* 'as previously' (while *kamlaŋ* cannot). To illustrate:

(27) (www.jamsai.com/Story/Part.aspx?PartID=125473)
tʰâa *tʰəə* *nâŋ* *rɔ́ɔŋhâj* ***jùu₃*** *jàaŋ* *níi*
if 2S sit cry **stay** like this

man *càʔ* *dâaj* *ʔaraj* *kʰîn* *maa*
3S IRR get what descend come
'If you continue to cry like this, what will you get?'

One might question why the co-occurrence between *jùu* and dynamic verbs is possible, since their nature involves change. For example, 'walking' involves lifting and setting down each foot in turn, as shown in **Figure 6**.

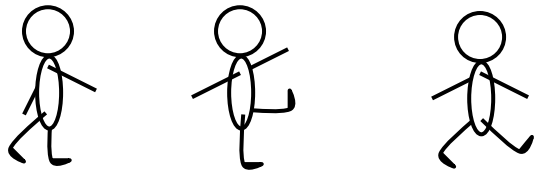


Figure 6: *Walking*

When *jùu* co-occurs with an activity verb, say *dəən* 'walk', it does not capture the change or dynamic property of the activity. That is, *jùu* does not track the changing state of walking through processing time. From a cognitive grammar perspective, such real-time observation is described as a sequential scanning (Langacker 2008). A dynamic experience, however, can be apprehended holistically. That is, the changing states are all captured in a single image. This summing capacity is called summary scanning (Langacker 2008). In this way, *jùu* can occur with dynamic verbs. The sentence *Pìtì dəən jùu* 'Piti is walking', for instance, expresses the fact that Piti is walking rather than doing something else at the moment.

3.1.3 Temporal location of jùu

Recall that *jùu*$_3$ and *jùu*$_4$ (in the domain of time) indicate different kinds of time information. The locator *jùu*$_3$ deals with how long/how often an event lasts, while *juù* $_4$ deals with at what time (TT) an event is located.

The time frame conceptualization is flexible depending on what type of temporal words occur with *jùu*$_3$. It should be noted that this time frame is not the same as the notion of temporal boundedness. This time frame is related to a particular period of time where an event exists. It is a set of consecutive time values. The idea of a beginning point and end-point is not necessarily entailed by the concept. As such, it can be either bounded or unbounded. The no boundaries concept assumed in the previous studies thus does not hold true (Tansiri 2005; Kullavanijaya and Bisang 2007).

Prototypically, the time frame of *jùu* is an interval construed as a whole or **bounded**, which can be linguistically further specified by, for instance, *tʰáŋ* (*kʰɨ̀ɨn)* 'all (night)', and *tâŋtèɛ…con* 'since…until'. This is illustrated by the following examples.

(28) (www.siamrath.co.th)

pʰǒm	*kɔ̂ɔ*	*rɔɔ*	***jùu*$_3$**	*tʰáŋ*	*kʰɨ̀ɨn*
1S.M	CONJ	wait	**stay**	all	night

praakòt	*wâa*	*nɔ́ɔŋ*	*kʰǎw*	*mâj*	*maa*
appear	COMP	3S	3S	NEG	come

'I waited all night. It turned out that she didn't come.'

(29) (www.pantown.com)

faj	*dàp*	***jùu*$_3$**	*tâŋtèɛ*	*sìp*	*mooŋ*	*cʰáaw*	*con*	*nɨ̀ŋ*	*tʰùm*
light	extinguish	**stay**	since	ten	o'clock	morning	until	one	o'clock (night)

'The light went out from 10 a.m until 7 p.m.'

We can construe time frame as a *series* of consecutive time values. This produces a habitual interpretation (**unbounded**). Examples of temporal words bringing out this reading include *pràcam* 'regularly', *samǎə* 'always', *bɔ̀j bɔ̀j* 'often', and *tʰúk (wan)* 'every (day)'. Example (30) illustrates a habitual reading.

(30) (www.t-pageant.com)

Təəj	*kʰít*	***jùu*$_3$**	*tʰúk*	*wan*	*wâa*	*jàak*	*pə̀ət*	*ráankʰǎajʔaahǎan*
Tei	think	**stay**	every	day	COMP	want	open	restaurant

'Tei (I) think every day that (I) want to open a restaurant.'

Note that this habitual reading is distinct from generic habituality (we can say *kʰít tʰúkwan* 'think everyday' without *jùu*$_3$). Sentence (30) means something like 'I keep thinking about opening a restaurant', where persistence over a time period is implied. It specifies that the thought rests on the mind every day. The word *jùu*$_3$ and *tʰúkwan* put a spotlight on the unvarying nature of the event.

The locator *jùu*$_4$, as mentioned, locates a phase of an event 'at' a TT. Tansiri (2005) suggests that *jùu* profiles only the intermediate phase of a situation without referring the boundaries. Consider the following examples.

(31) (http://diatv5.multiply.com/journal/item/27)

a.	*tɔɔnníi*	*pʰǒm*	*kɔ̂ɔ*	*râəm*	*kin*	***jùu*$_4$**
	now	1S.M	CONJ	start	eat	**stay**
	'I start eating it at the moment.'					

b.	* *tɔɔnníi*	*Pìti*	*kwàat*	*bâan*	*sèt*	***jùu*$_4$**
	now	Piti	sweep	house	finish	**stay**
	'Now, Piti is finishing sweeping.'					

Sentence (31a) refers to the beginning of taking antibiotic pills. The *jùu*$_4$ marker locates the beginning of taking medicines at the time of utterance—we are in the period of starting the treatment. The pattern [*râəm* + VP + *jùu*$_4$], although it does not frequently occur, is not impossible. In (31b), on the other hand, the occurrence of *jùu* is not acceptable. This, however, does not exclude my suggestion that *jùu* does not necessarily profile only the intermediate phase. The ungrammaticality of (31b) is partly due to the fact of difference in temporal points. The temporal location of *jùu* is the time of utterance, while the completion of sweeping means it has come to an end, i.e., it occurs before the time of utterance.

Tansiri (2005) further suggests that due to this intermediate profiling of *jùu*, it is incompatible with semelfactives (32), unless semelfactive is construed as iterative (33).

(32) (Tansiri 2005:122)

	faj	*nâa*	*rót*	*kapʰríp*	*nìŋ*	*kʰráŋ*	***jùu***
	light	front	car	flash	one	CLF	**stay**
	'The front light flashed one time.'						

(33) (Tansiri 2005:123)

	faj	*nâa*	*rót*	*kapʰríp*	***jùu***
	light	front	car	flash	**stay**
	'The front light flashed.'				

Nevertheless, it is found that *jùu* is in fact compatible with semelfactives if it occurs before a numeral phrase, as in (34). However, it is *jùu*$_3$ which is compatible with semlfactives, not *jùu*$_4$. Recall that it is *jùu*$_3$ which involves frequency/duration. Here, *jùu*$_3$ is characterized against the domain of frequency (and time) [*jùu*$_3$ + NUM CLF].

(34) (www.bnetshop.com)

faj	*sǐi*	*kʰiăw*	*kapʰríp*	***jùu*$_3$**	*nìŋ*	*kʰráŋ*
light	colour	green	flash	**stay**	one	CLF
'The green light (of a Canon printer) flashed one time.'						

Example (34) describes that at a particular period of time, the flashing occurred once. Note that *jùu* can be omitted here. Although its occurrence is not obligatory, it causes the sentence to be grounded in the timeframe (without *jùu*, it sounds like a factual statement irrelevant to the speech event).

It should be noted that this usage of *jùu* is not only constrained with semelfactives. It can occur with other kinds of states of affairs which can be repeated. The number of occurrences can be either specific (35)

or non-specific (36). Notice that the classifier *kʰráŋ* quantifies events in a similar way to the English 'times'. This differs from the noun classifier *duaŋ* in (21) which is used for counting lights (as well as stars, moons etc.).

(35) (www.bloggang.com/mainblog.php?id=g-unit&month=20)

càp	*cʰalàak*	*kan*	***jùu₃***	*săam*	*kʰráŋ*	*kwàa*	*càʔ*	*dâaj*	*pʰûu*	*cʰôokdii*
draw	label	together	**stay**	three	CLF	until	IRR	get	person	lucky

(We) drew the lotteries three times before we got a winner.'

(36) (www.komchadluek.net)

sòŋ	*siăŋ*	*kʰamraam*	*rɔ́ɔŋ*	*sàj*	*Cʰûaŋ Cʰûaŋ*	***jùu₃***	*lăaj*	*kʰráŋ*
send	sound	growl	cry	be_toward	Chuang Chuang	**stay**	many	time

(Linping) growled at Chuang Chuang (his Panda father) many times.'

These examples show that although *jùu* usually has an intermediate profiling, it is not the only possible phase of conceptualization of *jùu* (it is simply more entrenched) and the concept of intermediate profiling is irrelevant to the case of *jùu₃*.

3.1.4 Statives vs. topic time

It is not surprising if *jùu* cannot occur with all types of statives. Boonyapatipark (1983), together with Tansiri (2005), makes an insightful observation about *jùu*, that it is incompatible with permanent or inherent statives. This value is called "temporariness". According to Croft (to appear), an inherent state refers to the state which lasts for the entire history of the participant.

Inherent states can be further classified into original and acquired inherent state. Original inherent states are those that exist since the origination of the participant, for example 'be woman', and 'be stone'. Acquired inherent states refer to states which exist some time after the origination. Tall people, for example, were not born tall. But once they become tall, they remain tall for the rest of their life.

If *jùu* refers only to the intermediate phase of an event without referring to the boundaries, why does such a constraint exist? Why cannot *jùu* occur with all types of statives? This is the aim of the following discussion—to further explicate what Boonyapatipark (1983) and Tansiri (2005) have insightfully observed.

Thai *jùu* often occurs in transitory states (or temporary states as called by Boonyapatipark). It disfavours inherent states (Tansiri 2005: 125). To illustrate:

(37)

**Pìtì*	*pen*	*pʰûucʰaaj*	***jùu₄***[16]
Piti	COP	man	**stay**

'Piti remains a man.'

Recall that *jùu* inherits the value of location. This means a stative verb marked by *jùu* calls for a temporal location. In other words, *jùu* is employed to capture a state at **a topic time—**time under discussion (Klein 1994). This is thus contradictory with inherent states which is irrespective of time.

Complicating this constraint is the fact that an inherent state can be construed as a transitory state if we can establish a reference location where the inherent state can bear some temporal dependency on. Note that this is not possible for all inherent states.

(38) (www.thailife.de)

tɔɔnnán	*jaŋ*	*pen*	*pʰûucʰaaj*	***jùu₄***
then	still	COP	man	**stay**

'(She) was still a man then.'

[16] Note that this sentence is possible if it considered in terms of 'maleness'. However, the point here is to show temporal function of *jùu₄*, which is incompatible with inherent states.

This is part of an interview with a Thai transsexual posted on the Internet. The deictic time marker *tɔɔnnán* 'then' refers to the period of time she was a man, establishing a reference which causes the inherent state to be construed as a transitory state. This special circumstance gives rise to temporal location required by *jùu*.

This constraint (inherent vs. transitory state) is also applied to accomplishments and achievements which involve changes of state. Their resulting states can be either inherent or transitory (Tansiri 2005: 126-128). Only the interaction of *jùu* and achievements will be discussed here.

(39)

**kracòk*	*tɛ̀ɛk*	***jùu₄***
mirror	break	**stay**

'The mirror is still breaking.'

The verb of destruction in (39) is an example of an achievement with an inherent result state. This result state is irreversible and incompatible with *jùu*.

(40) (anne4seasons.multiply.com/journal/item/2)

faj	*dàp*	***jùu₄***	*náʔ*	*niâ*
light	extinguish	**stay**	Pt	Pt

'The light still went out.'

The achievement in (40), on the other hand, ends in a transitory result state, which is reversible and thus is compatible with *jùu*. Note that *jùu* in (40) is conceptualized against the domain of time; *jùu* in (21) against the domain of quantity.

Interestingly, a verb like *hàk* 'break' can be interpreted either way, depending on its argument ('bone' vs. 'tree branch').

(41) (www.pantown.com)

kradùuk	*hàk*	***jùu₄***	*tôŋ*	*kʰâw*	*fɨàk*	*tʰɨ̌ŋ*	*Pʰrɨ́tsàpʰaakʰom*
bone	break	**stay**	must	enter	plaster_cast	until	May

'The bone is still broken. (It) must be in a plaster cast until May.'

(42)

ʔkìŋmăaj	*hàk*	***jùu₄***
tree stick	break	**stay**

'The tree stick is still breaking.'

The fact that we can talk of (41) (as compared to the unnaturalness of talking about (42) rests on our knowledge of the participants. The knowledge of bone includes the fact that bone can regrow. The state of broken bone is thus not permanent but temporary. A broken tree stick, by contrast, is irreversible. Even so, one can imagine circumstances in which (42) can be viewed as a temporary state, e.g. a magical spell. All we need is a timeframe for (42) to situate providing it is pragmatically possible.

The main idea of this discussion is to point out the importance of topic time (i.e., the concept of location) in understanding the nature of *jùu*. The issue of topic time will be discussed in more detail in Section 4.

3.2 Proposed treatment of kamlaŋ

We have seen that *jùu* and *kamlaŋ* are two distinct forms. This section will investigate *kamlaŋ* in more detail in relation to its function and meaning.

Unlike *jùu*, no verbal use of *kamlaŋ* has been identified—i.e., it never serves as a main verb. What we have is the noun *kamlaŋ*, which means 'energy'—a Khmer loanword. This noun might be the lexical source from which the progressive marker *kamlaŋ* is derived, as exemplified in (43).

(43) (When my grandpa and grandma were young: volume 3: 68)
dèkdèk ***kamlaŋ*** *mâj* *mii* *pʰɔɔ*
children **energy** NEG have enough
'Children's energy is not enough.'

This nominal origin might be the reason why the progressive *kamlaŋ* is positioned before the main verb and where its dynamic property is derived from. The progressive use of *kamlaŋ* is shown in (44):

(44) (www.oknation.net/blog/print.php?id=254582)
pʰîi *Bɛɛm* ***kamlaŋ*** *wâat* *rûup* *kʰàʔ*
older_sibling Bam **PROG** draw picture Ppt
'Bam is drawing a picture.'

I suggest that the conceptualization of 'energy' is still found in *kamlaŋ.* Like *jùu*, some loss of meaning is involved (i.e., the physical and mental effort), but its dynamic sense still remains. By dynamic, it means that the process is characterized by constant change. This suggests that it should be considered a progressive marker. It is this very characteristic which motivates *kamlaŋ*'s grammatical behavior and distinguishes it from *jùu.*

Compare the following sentences:

(45) (Free conversation)
a. *cʰuâŋ* *níi* *duaŋ* ***kamlaŋ*** *tòk*
period this fortune **PROG** fall
'During this time, (my) fortune is falling.'

b. *cʰuâŋ* *níi* *duaŋ* *tòk* ***jùu₄***
period this fortune fall **stay**
'During this time, (my) fortune is down.'

Sentence (45a) and (45b) yield different interpretations. Sentence (45a) expresses that the speaker's fortune is moving downward at the reference time, while (45b) designates that his fortune remains at a lower level at the time of reference. They do not say when and how his fortune falls—gradually or instantly. The reference time or topic time is *cʰuâŋ níi* 'during this time' which is represented by TT.

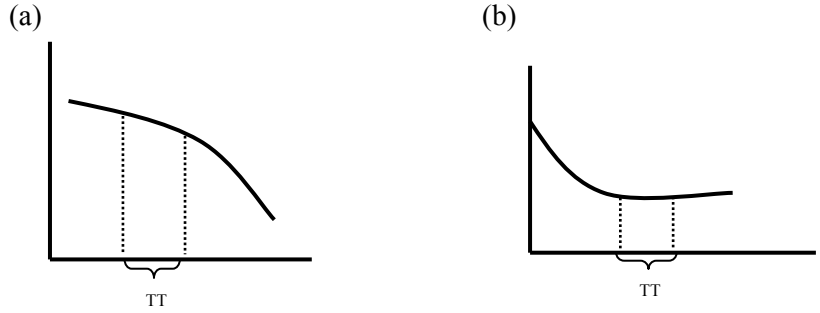


Figure 7: *Fortune*

The progressive *kamlaŋ* is preferentially connected to activities and iteratives which require energy for sustained physical and mental activity (i.e., dynamic processes—*run, walk, sweep, eat, cough, bounce*). It expresses the dynamic quality of actions that are in progress. To illustrate:

(46) (The Pear Story [Speaker 2])

kʰon	*kɛ̀ɛ*	***kamlaŋ***	*kèp*	*lûukpʰɛɛ*
person	old	**PROG**	pick	pear

'An old man was picking pears.'

Sentence (46) expresses the active movement of the old man's hands—taking hold of and removing pears from the tree.

(47) (www.thaiphone.com/forum)

tɔɔn níi	***kamlaŋ***	*ʔaj*	*mâj*	*jùt*
now	**PROG**	cough	NEG	stop

'Now, I am coughing non-stop.'

As for (47), *ʔaj* 'cough' is a typical example of a punctual process. Nevertheless, it is easy to interpret as an iterative process. As an iterative process, it denotes an extended, dynamic activity which composes of an unidentified number of iterations.

In order to obtain its compatibility with *kamlaŋ*, the number of instances of *ʔaj* 'cough' has to be left open. The end-point of *ʔaj* has to be unbounded. This explains why (48) is ungrammatical.

(48)

**cʰǎn*	***kamlaŋ***	*ʔaj*	*sɔ̌ɔŋ*	*kʰráŋ*
1S	**PROG**	cough	two	time

'I am coughing twice.'

This illustrates that *kamlaŋ* disfavours punctuality. It therefore cannot occur with achievement verbs such as *tɛ̀ɛk* 'break', *taaj* 'die', and *dàp* '(light) go out'.

It is interesting to note that the progressive *kamlaŋ* also occurs with state verbs.

(49) my.dek-d.com

cʰáaw	*wanníi*	*ʔaakàat*	***kamlaŋ***	*dii*	*náʔ*
morning	today	weather	**PROG**	good	Pt

bɛ̀ɛpwâa	*mâj*	*rɔ́ɔn*	*paj*	*mâj*	*nǎaw*	*paj*
somewhat	NEG	hot	go	NEG	cold	go

'This morning, the weather is just right. Not too cold, not too hot.

The verb 'good' inherently is a stative process, which involves little or no change—the process simply goes on. By this nature, it should not be able to occur with *kamlaŋ*, however, it does. If we take the notion of semantic flexibility into consideration, it will be easier to understand why this is possible. Typically, what the word 'good' encodes is the state of pleasantness. According to our encyclopaedic knowledge, however, we know that there are degrees of 'pleasantness'; as such the stative process can change over time, for example, from bad to good. What *kamlaŋ* does is bring out the potential range of a weather event which undergoes change over time, and it profiles or designates the pleasant state, as symbolized below (the profile indicated by the heavy line [Langacker 1987, 1991, 2008]).

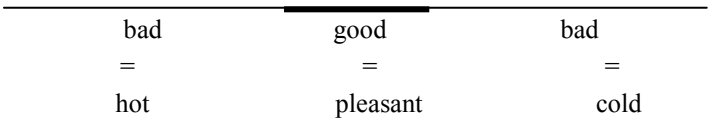


Figure 8: *Weather change*

In **Figure 8**, the line represents the possibility of weather change, while the state of goodness is indicated by the heavy line. In this way the progressive *kamlaŋ* can occur with a stative process.

Note that when a stative is progressivized, it does not express the same dynamic conception as a progressivized dynamic process. Consider the following sentence.

(50)

dɔ̀ɔkmáaj	***kamlaŋ***	*baan*
flower	**PROG**	blossom

'Flowers are/a flower is blossoming.'

Sentence (50) can express either the active opening of flowers/petals, or flowers' current state—the fact that flowers are in blossom. The former conveys a dynamic, unfolding movements through time (imagine the time-lapse camera movement). It is progressive since it requires change of flower production—bud, bloom, wither. The latter, on the other hand, illustrates a static-progressive. It is static because it focuses on the blossom state.

Note the pragmatic possibility of the following:

(51) (www.teana-club.com/webboard)

tèŋ	*bèɛp*	*níi*	***kamlaŋ***	*suǎj*
decorate	like	this	**PROG**	beautiful

'Decorating like this is beautiful.'

(52)

? Maalii	***kamlaŋ***	*suǎj*
Malee	**PROG**	beautiful

'Malee is beautiful.'

Without any context, it is acceptable to say Sentence (51), but less acceptable to say (52). The event in (51) is a car decoration situation, which can undergo change—a car can be decorated beautifully or terribly. For example, it is tacky if we decorate the car with too much or too little. But, if we do it just right, it looks attractive. The decoration scenario is construed as dynamic; hence the verb 'beautiful' can take the progressive *kamlaŋ*. As for (52), although a person's natural beauty can change over time, it is not as dynamic as (51)—it is construed as taking a longer time to change.

The progressive is also found to occur with other state verbs such as *rúu* 'know', *rák* 'love', *lŏŋ* 'lost', *cʰɨâmân* 'trust', and *cʰɨâ* 'believe'.

(53) (www.11news1.com)

wanníi	*kʰon*	*tʰaj*	***kamlaŋ***	*rúu*	*wâa*
today	person	Thai	**PROG**	know	COMP
cʰâat	*tʰaj*	*kàət*	*tʰɔ̀ɔrarâat*	*kʰɨ̂n*	*lɛ́ɛw*
nation	Thai	occur	tyrant	ascend	already

'Now, Thai people know that their nation has had a tyrant.'

(54) (http://webboard.mthai.com/5/2006-02-12/197819.html)

mɨâ	*raw*	***kamlaŋ***	*rák*	***kamlaŋ***	*lŏŋ*		***kamlaŋ***	*cʰɨâmân*
when	1P	**PROG**	love	**PROG**	crazy about		**PROG**	trust

kamlaŋ	*cʰɨâ*	*raw*	*mák càʔ*	*mɔɔŋ*	*pʰiaŋ*	*kʰêɛ*	*dâan*
PROG	believe	1P	often	look	only	just	side

dii	*kʰɔ̌ɔŋ*	*sìŋ*	*nán*
good	POSS	thing	that

'When we are loving, being crazy about, trusting, believing (something), we are likely to look only at the good side.'

The frequency of occurrence of the progressive with state verbs varies. State verbs which have a high potential to be changeable like *dii* 'good', and *ʔarɔ̀j* 'tasty' are found to occur frequently with *kamlaŋ*. State verbs which have less potential to be changeable like *rúu* 'know', and *cʰɨâ* 'believe' are less frequently found to occur with the progressive. As such, they are not well entrenched and might not be accepted by some speakers. Inherent states like *pen pʰûujĭŋ* 'be women', *pen kʰon tʰaj* 'be Thai' are normally incompatible with *kamlaŋ*.

It was mentioned in Section 3.1 that *jùu* can occur with a stative verb; however, its implication is different from that of *kamlaŋ* due to its different semantic value.

(55) (http://topicstock.pantip.com)

kʰon	*ʔaraj*	*suăj*	***jùu*$_3$**	*samăə*
person	what	beautiful	**stay**	always

'What a woman, she always stays beautiful.'

To assert (55) is to say that the participant, a famous Thai singer, was beautiful then, and is still beautiful now. Her beauty extends over a period of time, which began in the past, and which obtains at the present. This is the continuity or unchanging effect of *jùu*, which cannot be found in *kamlaŋ.*

Due to the value of dynamicity, *kamlaŋ* cannot occur with adverbials of duration. Examples of adverbials incompatible with the progressive *kamlaŋ* are *tʰáŋ (pii)* 'all (year)', *talɔ̀ɔt weelaa* 'all the time', *sàkkʰrûu* 'for a while'. The progressive is also incompatible with habitual adverbials such as *bɔ̀j bɔ̀j* 'often', and *tʰúkwan* 'everyday'. This dynamic value is opposite to the unchanging nature of *jùu* resulting in distinct syntactic patterns.

(56)

**dèk*	*dèk*	***kamlaŋ***	*kin*	*kʰâaw*	*tʰúk*	*wan*
child	REDUP	**PROG**	eat	rice	every	day

'Children are eating rice every day.'

Another difference between *jùu* and *kamlaŋ* lies in their scope of modification. The different scopes of *jùu* and *kamlaŋ* are explicitly shown in the following examples.

(57) (www.khaosod.co.th)

kamlaŋ	*dəən*	*paj*	*tʰamŋaan*
PROG	walk	go	work

'(He) was walking to work.'

(58) (www.bloggang.com)

tɔɔnníi	*jaŋ*	*dəən*	*paj*	*tʰamŋaan*	***jùu***	*ləəj*
now	still	walk	go	work	**stay**	Pt

'Even now, (I) still walk to work.'

The progressive *kamlaŋ* only takes scope over the first verb—'walk', while *jùu* modifies the whole (walk to work). That is to say, *kamlaŋ* tracks the changing state of walking through processing time (sequential scanning). As for *jùu*, it captures the change states in a single image (summary scanning). Apparently, due to their different scopings, *kamlaŋ* and *jùu* are compatible—i.e., they can co-occur. Examples of their co-occurrence will be discussed in Section 5.

4. Temporal relation

Based on Klein's model of tense and aspect (1994), the analysis of time involves three times, namely, time of situation (T-SIT), time of utterance (TU), and topic time (TT) (or 'reference time' according to the

Reichenbach (1947, reprinted in 2003) model). The TT is the time under discussion (Klein 1994, Klein et al. 2000). In conversation, it is typically the TU, but it is not compulsory. For instance, it is common to speak on the phone as (*kamlaŋ*) *tʰam ʔaraj jùu* 'what are you doing?' In this situation, the TT does not refer to the TU, but the time before the telephone conversation. The TT can be linguistically explicit, but it is usually implicit and inferred from the context. To illustrate:

TT is explicit
(i) At 4 p.m., my son was doing his homework. (TT = 4 p.m.)
(ii) What did you do when you saw him? (TT = the time of seeing)

TT is implicit
(i) *I forgot to turn off the oven!* (TT = the time before leaving the house)
(ii) (I smell smoke) *Were you smoking?* (TT = the time within the recent past)

According to Klein (1994), aspect indicates a temporal relation between the TT and the time of situation (T-SIT), while tense signals a temporal relation between the TT and the time of utterance (TU). The notion of temporal relation between TT and T-SIT is adopted in this study. It is argued that *jùu* and *kamlaŋ* serve as 'temporal relators', i.e., signaling the way an event in question (T-SIT) is distributed in relation to another event (TT).

4.1 Temporal relation of jùu$_4$

Recall the nature of *jùu*: LOCATOR (locatum, location). The concept *jùu* needs a spatial, attribute, or temporal location either explicitly or implicitly mentioned. Phrases like *tʰúkwan* 'every day', and *tʰáŋkʰɨɨn* 'all night' can be considered as examples of temporal locations of *jùu$_3$*.

What are the temporal locations of *jùu$_4$* then? They are contextually determined. The moment of speech is such an example. Even though *jùu$_4$* concerns contextual properties, it does so intrinsically. In spite of having discourse force, *jùu$_4$* is not external to semantics; it also has the semantic nature of 'location' in that it requires a place for an event to be located. This is taken as the frame of reference.

The semantic structure of *jùu$_4$* is thus a dependency between a locatum and a reference location. That is to say, *jùu$_4$* is a **temporal relator**, which requires a temporal relation between time spans—the time of the situation (T-SIT) (locatum) and the topic time (TT) (reference location). It indicates that **T-SIT is situated at TT**. Apparently, this function of *jùu$_4$* is inherited from its lexical source—'locator' nature. In the previous discussion (Section 3.1), this function is referred to as a time-discourse locator. That is, it relies on discourse context to determine the topic time.

(59) (Free conservation)

mɨâkíi	*hěn*	*mêɛ*	*tʰɨ̌ɨ*	*(krapăw)*	***jùu$_4$***
then	see	mother	carry	bag	**stay**

'(I) saw mother carry the purse just now.'

For example, the topic time of (59) is the time of witness which is the reference where the event mother's holding her purse is hooked on.

(60) (Short stories [CU Thai Concordance])

măa	*hâa*	*tua*	*tʰîi*	*nâŋ*	*kan*	***jùu$_4$***	*tɔɔn*	*níi*
dog	five	CLF	REL	sit	together	**stay**	time	this

mii	*cʰɨ̂ɨ*	*Túttùu*	*Deesîi*	*Ben*	*Bəənàat*	*lɛ́ʔ*	*ʔétdîi*
have	name	Tutoo	Daisy	Ben	Bernard	and	Eddie

'Five dogs which are sitting together are Tutoo, Daisy, Ben, Bernard, and Eddie

Here, the topic time is the time of speaking. Temporal deixis expressions like *tɔɔnníi* 'now', and *tɔɔnnán* 'then' are often found to occur with *jùu*$_4$.

4.2 Temporal relation of kamlaŋ

In addition to allowing an event to be construed as an event in progress, *kamlaŋ* also indicates the **coincidence** of the on-going event (T-SIT) and the contextual event performed at the time of the on-going event (TT)—**T-SIT coincides with TT**.

(61) (www.bloggang.com)

fŏn	***kamlaŋ***	*tòk*
rain	**PROG**	fall

'It is raining.'

The on-going event in (61) is a raining event. What is the contextual event of (61) then? Apparently, it is not linguistically expressed in this example.

In a given speech event, there would be at least two states of affairs: the speaker's utterance, and the utterance event. The utterance event is the on-going event, while the speaker's utterance is the contextual event, which is a precondition for the event in progress to emerge. In this particular example, it serves as the topic time (TT = TU) for purposes of establishing the relationship with the on-going event (T-SIT). This is a type of simultaneous relationship.

According to Grice's maxims of conversation[17] (1975), the speaker does not supply more information than is required (maxim of quantity). In a real time situation, as in (61), it is not necessary to assert that 'it is raining when the speaker is speaking'. When there is enough information, the contextual event is not linguistically encoded. It is grammatically and communicably sufficient to have only the on-going event in the independent clause, as in (61). Such an independent clause is often found in conversational discourse where there is enough contextual information.

When the contextual event is not the default time of speaking or cannot be inferred, the contextual event must be explicitly mentioned (following Grice's quantity maxim). Consider Sentence (62):

(62) (The Pear Story [Speaker 3])

mii	*dèk*	*pʰûucʰaaj*	*kʰon*	*nìŋ*
there is	child	male	CLF	one

kʰanà tʰîi	*kɛɛ*	***kamlaŋ***	*kèp*	*lûukpʰɛɛ*	*kʰèŋ*	*tʰîi*	*sǎam*
while	3S	**PROG**	pick	pear	basket	NuM	three

man	*kʰìi*	*rót càkkrajaan*	*maa*
3S	ride	bicycle	come

'There was a boy. While the old man was picking the third basket of pears, he rode a bike towards the old man.'

In (62), the old man was picking pears when a boy came by on a bicycle. This type of sentence is often found in a narrated story. The old man's picking pears is hooked up to the topic time—the time the boy biked. The temporal linker *kamlaŋ* makes a reference to accommodate another simultaneous event.

The omission of *kamlaŋ* will result in a vagueness of meaning, for example, [*fŏn tòk*] can be interpreted as 'it is raining now' or 'it rained'. Moreover, there are different ways in which the event [*fŏn tòk*] can be distributed in relation to another event: simultaneity, posteriority, and anteriority. To illustrate:

[17] There are four main maxims of conversation: quantity, quality, relation and manner.

(63)

cʰăn	*ʔàapnáam*	*tɔɔn/kɔ̀ɔn/lăŋ*	*fŏn*	*tòk*
1S	take_a_bath	when/before/after	rain	fall

'It took a bath, when/before/after it rained.'

However, there is only one way in which [*fŏn kamlaŋ tòk*] can be distributed to another event, that is, simultaneity. This simultaneity licenses the types of conjunctions *kamlaŋ* can occur with. In this example, only *tɔɔn* 'when' is allowed. This shows that the simultaneous relationship of *kamlaŋ* is not contextually derived.

Because of the coincidence nature of *kamlaŋ*, it tends to occur with temporal deictic expressions, (for example, *tɔɔnníi* 'at this time', *kʰanàʔníi* 'at this time', *tɔɔnnán* 'at that time', and *kʰanàʔnán* 'at that time',) and not with duration expressions (e.g. *tâŋnaan* 'for a long time, *talɔ̀ɔt weelaa* 'all the time', *sàkkʰrûu* 'for a while', *pépnɨŋ* 'for a second'), or two-time point expressions (e.g. *tâŋtèɛ* 'since', *con* 'until', *jaŋ* 'still' or 'up to and including the present or the time mentioned or an unspecified time'). To illustrate:

(64) (http://bbznet.com)

a.	*níik*	***jùu₃***	*tâŋ*	*naan*	*kʰraj*	*wàa*	*maa*	*tʰák*	*raw*
	think	**stay**	much	long	who	Pt	come	greet	1S

'(I) thought for a long time. Who came to greet me?'

b.	****kamlaŋ***	*níik*	*tâŋ*	*naan*	*kʰraj*	*wàa*	*maa*	*tʰák*	*raw*
	PROG	think	much	long	who	Pt	come	greet	1S

'(I was thinking for a long time. Who came to greet me?'

(65)

a.	**kʰáw*	***kamlaŋ***	*rɔɔ*	*tâŋtèɛ*	*cʰáaw*
	1S	**PROG**	wait	since	morning

'I am waiting since morning.'

(forums.popcornfor2.com)

b.	*kʰáw*	*rɔɔ*	***jùu₃***	*tâŋtèɛ*	*cʰáaw*
	1S	wait	**stay**	since	morning

'I stay waiting since morning.'

Situation: A man would like to get a betel nut which was chewed by a famous monk to worship. The monk answered to his request that:

(66) (board.palungjit.com)

a.	*rɔɔ*	*diăw*	*jaŋ*	*kʰiáw*	***jùu₄***
	wait	in_a_moment	still	chew	**stay**

'Just a moment. (I) still chew (betel nut).'

b.	*rɔɔ*	*diăw*	*jaŋ*	***kamlaŋ***	*kʰiáw*
	wait	in_a_moment	still	**PROG**	chew

'Just a moment. (I) am still chewing (betel nut).'

In the event coded by (66), *jaŋ* presupposes that the time frame of chewing a betel nut started some time in the past up to the moment of the request (i.e., TT). It is not simply a two-time point expression. It designates that the act of chewing remains unchanged at the TT. This is incompatible with *kamlaŋ*, which not only indicates the simultaneous connection between 'chew' (T-SIT) and 'request' (TT), but also dynamicity. It is acceptable for *jùu* which expresses continuity. The TT serves as a temporal location for the act of chewing to remain unchanged.

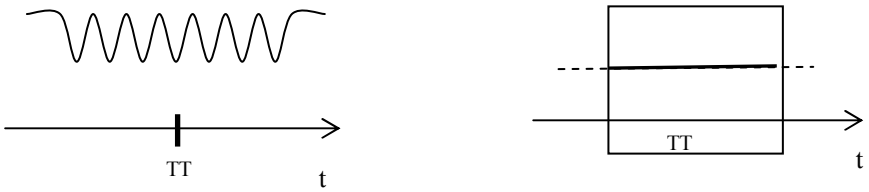
This property of *kamlaŋ* allows the speech participants to specify that the event in progress does not precede or follow the contextual event but at some point coincides with it. This is *kamlaŋ*'s grammatical requirement. Omitting the contextual event would result in an ungrammatical/incommunicable sentence, as in (67).

(67)

??miâwaannɨi	*pʰŏm*	***kamlaŋ***	*kʰàp*	*mɔ̀ɔtəəsaj*	*klàp*	*bâan*
yesterday	1S.M	**PROG**	drive	motorcycle	return	home

'Yesterday, I was riding a motorcycle back home.'

The difference between *kamlaŋ* and *jùu*$_4$, thus, is also found in the way they are distributed with respect to the topic time (TT), as in **Figure 9**. The dynamic property of *kamlaŋ* is represented by the wavy line, while the unchanging property of *jùu*$_4$ is symbolized by the straight line.



a) *kamlaŋ* b) *jùu*$_4$

Figure 9: *Distribution of kamlaŋ and jùu*$_4$ *with respect to their topic time*

The temporal linker *kamlaŋ* specifies that an event in progress coincides with TT. The temporal locator *jùu*$_4$, on the other hand, anchors a phase of an event (which typically but not necessary is the intermediate phase) in the temporal location (TT). The square represents the temporal location of *jùu*$_4$. The distinct distribution in relation to TT entails different temporal scopes, and thus requires different temporal modifiers, as seen above.

The foundation has now been laid to enable discussion of the extent of interchangeability between *kamlaŋ* and *jùu*$_4$. Although the temporal scopes of *kamlaŋ* and *jùu*$_4$ are distinct from each other, they both refer to the topic time. The temporal location is expandable from a point to a larger interval. When the time interval is precise (i.e., the event has started in close proximity to the reference point), the focus is on the locator effect. The continuity of *jùu*$_4$ is thus not active—the event is not presented as persisting over a significant amount of time. In this kind of context, *jùu*$_4$ is apparently similar to *kamlaŋ*, and thus *kamlaŋ* and *jùu*$_4$ can be used interchangeably.[18]

Situation: On the phone

Question:

halŏo	*tʰam*	*ʔaraj*	***jùu*$_4$**
Hello	do	what	**stay**?

Hello, What are you doing?

Answer:

(68)

a.

kamlaŋ	*tʰamŋaan*
PROG	work

'(I) am working.'

[18] The interchangeability is possible with activity verbs but not state verbs.

b. *tʰamŋaan* ***jùu₄***
work **stay**
'(I) work at the moment.'

The topic time of (68) is not the time of question, but the time prior to the question. Note that the speaker can teasingly adopt the time of question as the topic time. In doing that, he could say 'I am talking on the phone (with you)'.

5. Co-occurrence of *kamlaŋ* + VP + *jùu*

The co-occurrence, in the same clause, of *kamlaŋ* and *jùu* is possible. The question is how Thai utilizes this co-occurrence. The co-occurrence provides some special properties semantically or grammatically, which are different from the use of *kamlaŋ* and of *jùu* individually. Consider the following sentences.

Situation A: Conversation
Question: Speaker 1

ŋaan	*tʰîi*	*hâj*	*paj*	*tʰam*	*rɨ́*	*jaŋ*
work	that	give	go	do	or	yet

'Did you do the work I gave to you, or not?'
Answer: Speaker 2

(69)

a.	***kamlaŋ***	*tʰam*	***jùu₄***	*mâj*	*hěn*	*rǎə*
	PROG	do	**stay**	NEG	see	Q

'(I am) doing it [at this very moment]. Don't (you) see it?'

b.	*? tʰam*	***jùu₄***	*mâj*	*hěn*	*rǎə*
	do	**stay**	NEG	see	Q

'(I am) doing it [at this very moment]. Don't (you) see it?'

c.	*?? **kamlaŋ***	*tʰam*	*mâj*	*hěn*	*rǎə*
	PROG	do	NEG	see	Q

'(I am) doing it [at the very moment]. Don't (you) see it?'

All three answers are possible, although the co-occurrence (69a) is the most preferred and (69c) is the least likely. What Speaker 2 wants to communicate is not only that the event is in progress but also that S_2 is performing it at the very moment without doing anything else, i.e., *jùu₄* anchors the work in progress which is modified by *kamlaŋ* at the time of utterance, placing emphasis on the event. In (69b-c), although they are grammatical, they are not perceived as complete and firm, especially (69c)—it seems as if it were 'floating', as commented on by some native Thais.

In order to elucidate the special semantic/syntactic contribution of *kamlaŋ* and *jùu*, it is necessary to consider what type of *jùu* occurs in the *kamlaŋ…jùu* construction.

All types are possible, and each *jùu* requires a different type of location, as illustrated in **Table 5**:

Table 5: *Different types of location*

kamlaŋ + VP + $jùu_2$ + SPACE
kamlaŋ + VP + $jùu_3$ + TIME (or other abstract domains)
kamlaŋ + VP + $jùu_4$ + DISCOURSE EVENT/TIME (or other abstract domains)

The *kamlaŋ* + VP + $jùu_4$ construction will be discussed first, which is the focus of this section. The other types of combination will be discussed briefly.

5.1 kamlaŋ + VP + jùu$_4$ + DISCOURSE EVENT/TIME

Consider the following examples. Note that e_1 refers to the event modified by *kamlaŋ* and/or *jùu*$_4$ (T-SIT); e_2 refers to another event (TT).

(70) (Four Reigns [CU Thai Concordance])

a. *riâŋ* *ʔaraj* *kʰráp* *kʰunmêɛ* | *taa* *ʔân* *sǐŋ* *lúk* *càak* *tóʔ* |
story what Pt Mother | TA An who rise from table

léɛw *tʰăam* *kʰîn* | e_2
then ask ascend

tʰáŋ ***kamlaŋ*** *jɨɨn* ***jùu$_4$*** e_1
INCLUSIVE **PROG** stand **stay**
‘ “What is it about, Mother?” An who had stood up asked while he was standing.’

b. *? riâŋ* *ʔaraj* *kʰráp* *kʰunmêɛ* | *taa* *ʔân* *sǐŋ* *lúk* *càak* *tóʔ* |
story what Pt Mother | TA An who rise from table

léɛw *tʰăam* *kʰîn* | e_2
then ask ascend

tʰáŋ *jɨɨn* ***jùu$_4$*** e_1
INCLUSIVE stand **stay**
‘ “What is it about, Mother?” An who had stood up asked while he was standing.’

c. ** riâŋ* *ʔaraj* *kʰráp* *kʰunmêɛ* | *taa* *ʔân* *sǐŋ* *lúk* *càak* *tóʔ* |
story what Pt Mother | TA An who rise from table

léɛw *tʰăam* *kʰîn* | e_2
then ask ascend

tʰáŋ ***kamlaŋ*** *jɨɨn* e_1
INCLUSIVE **PROG** stand
‘ “What is it about, Mother?” An who had stood up asked while he was standing.’

Sentence (70a) is the most preferred form. The *kamlaŋ* + VP + *jùu$_4$* construction inherits the semantic values from both words. The semantic effect of *kamlaŋ* is to convert e_1 ‘stand up’ into a dynamic event (represented in **Figure 10** by a wavy line), and to indicate that it coincides with e_2 (represented by a line). The two events, however, simply occur simultaneously.

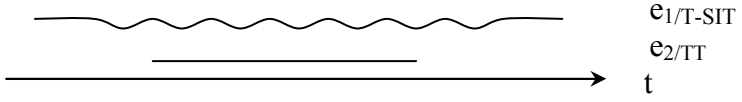


Figure 10: *The semantic effect of kamlaŋ*

The question is what does *jùu$_4$* contribute to the meaning? Is *kamlaŋ* not sufficient for indicating simultaneity? Since the two events simply co-occur, only *kamlaŋ* should suffice. However, the two events in

(70) do not simply co-occur. This is signalled by *tʰáŋ*[19] in the last clause of (70a). Even though, *kamlaŋ* and *jùu₄* are both temporal relators, it is *jùu₄* which inherits 'locator effect' from its lexical source. The function of *jùu₄* is to impose a scope on e_1, pinpointing that at the particular moment of e_1, e_2 occurs (indicated by the heavy line, and a box). It chains e_2 to e_1, i.e., the events are pooled to form a tighter relation (indicated by dashed lines) with the implication of emphasis. To put it in another way, *jùu₄* establishes the point in time TT_x (provided by e_2) where e_2 and a particular portion of e_1 occur.

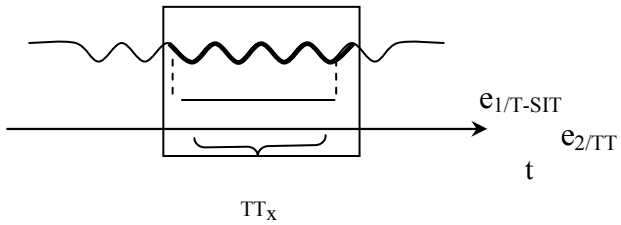


Figure 11: *Conceptual combination of kamlaŋ and jùu₄*

Because of this, Sentence (70b) does not sound natural since *kamlaŋ*, which marks simultaneity and progressive, is missing. As for Sentence (70c), it is the least acceptable due to the absence of *jùu₄*.

The requirement of this conceptual combination is motivated by several factors, for example, the pragmatic factor, as in (69) where sarcasm is indicated. The co-occurrence is also preferred when there are two events[20], and one event suddenly emerges. To illustrate:

(71) (Nick and Pim [2005: 85])

a.	*mŭu*	*pàa*	*tua*	*too*	*too*	***kamlaŋ***	*wîŋ*	*wîŋ*	***jùu₄***	e_1
	pig	wild	CLF	big	REDUP	**PROG**	run	REDUP	**stay**	
	kɔ̂ɔ	*lóm*	*taaj*	*kʰaa*	*tʰîi*					e_2
	CONJ	fall	die	stuck	place					
	'A big wild pig was running, and suddenly dropped dead.'									
b.	? *mŭu*	*pàa*	*tua*	*too*	*too*	*wîŋ*	*wîŋ*	***jùu₄***		e_1
	pig	wild	CLF	big	REDUP	run	REDUP	**stay**		
	kɔ̂ɔ	*lóm*	*taaj*	*kʰaa*	*tʰîi*					e_2
	CONJ	fall	die	stuck	place					
	'A big wild pig was running, and suddenly dropped dead.'									
c.	* *mŭu*	*pàa*	*tua*	*too*	*too*	***kamlaŋ***	*wîŋ*	*wîŋ*		e_1
	pig	wild	CLF	big	REDUP	**PROG**	run	REDUP		
	kɔ̂ɔ	*lóm*	*taaj*	*kʰaa*	*tʰîi*					e_2
	CONJ	fall	die	stuck	place					
	'A big wild pig was running, and suddenly dropped dead.'									

In (71), the two events are 'a wild pig was running' and 'it died' (actually there is another event, which is not mentioned here, that is the 'shooting' which is the reason causing the pig to die). The nature of the first event is an ongoing event, while the second is an interrupting event. What *jùu₄* does is to establish a position (a particular moment of e_1) for e_2 to take place.

[19] This word has different meanings, which can be glossed, for example, 'all' or 'together with'. Its crucial concept is **inclusiveness**, which requires a tight relation given by the co-occurrence of *kamlaŋ* and *jùu₄*.

[20] The reduplication appears to have an influence on the co-occurrence.

Another important factor is how the clauses of a sentence are combined. The *kamlaŋ* + VP + *jùu*$_4$ construction is often found to occur with a conjunction plus a demonstrative, for example *kʰanàʔ tʰîi* + VP + *nán*[21] 'while…that' (literally, it means 'at that time').

(72) (Thai National Corpus)

a.	***kʰanàʔ***	***tʰîi***	***kamlaŋ***	*rɔ́ɔŋhâaj*	*kʰrâmkʰruan*	***jùu*$_4$**	***nán***		e$_1$
	time	**at**	**PROG**	cry	lament	**stay**	**that**		
	siăŋ	*kʰɔ̌ɔŋ*	*nákrian*	*cʰaaj*	*săam*	*sìi*	*kʰon*		e$_2$
	sound	POSS	student	male	three	four	CLF		
	kɔ̂ɔ	*daŋ*	*kʰîn*	*maa*	*naj*	*sòotprasàat*			
	CONJ	loud	ascend	come	in	hearing			
b.	*?* ***kʰanàʔ***	***tʰîi***	*rɔ́ɔŋhâaj*	*kʰrâmkʰruan*	***jùu*$_4$**	***nán***			e$_1$
	time	**at**	cry	lament	**stay**	**that**			
c.	**** kʰanàʔ***	***tʰîi***	***kamlaŋ***	*rɔ́ɔŋhâaj*	*kʰrâmkʰruan*	***nán***			e$_1$
	time	**at**	**PROG**	cry	lament	**that**			

This conjunction structure prefers the co-occurrence of *kamlaŋ* + VP + *jùu*$_4$. The word *kʰanàʔ tʰîi* calls for an ongoing event which is given by *kamlaŋ*. It also serves as a temporal location. As for *nán*, it points to a specific moment of thinking, which in turn needs *jùu* to establish a path for it to refer to the thinking event.

These are not hard and fast rules. They are tendencies associated with the *kamlaŋ* + VP + *jùu*$_4$ construction.

5.2 kamlaŋ + VP + jùu$_3$ *+ TIME (or other abstract domains)*

In contrast to *jùu*$_4$, *jùu*$_3$ locates an event in non-topic time or other abstract domains. Sentence (73) illustrates an example of *kamlaŋ* + VP + *jùu*$_3$.

(73)

lôok	***kamlaŋ***	*tòk*	***jùu*$_3$**	*naj*	***júk***	***náamkʰɛ̌ŋ***
world	**PROG**	fall	**stay**	in	**era**	**ice**

'The world is being in the ice age.'

The temporal location is the ice age. Besides the time domain, examples of other abstract domains include *pʰawaŋ* 'trance', *monsakòt* 'spell', and *kʰwaamrák* 'love'.

5.3 kamlaŋ + VP + jùu$_2$ *+ SPACE*

Although the focus is on the temporal use of *jùu*, it is worthwhile to briefly discuss its spatial use. In contrast to other uses of *jùu*, *jùu*$_2$ locates an event in space. As a spatial locator, *jùu*$_2$ takes a spatial location, for example, 'in front of the cashier counter' as shown in (74).

[21] The word ***nán*** is a demonstrative designating an identifiable instance located away from the vicinity of the speaker. It occurs after the noun followed by the classifier: N + CLF + DEM, for example, *krapăw baj nán* 'bag CLF that' (*that bag*). It can also occur without head noun, with or without a preceding classifier. Prototypically, it is used to denote that the position of the located object is away from the speaker. This demonstrative use can be extended to function like the English definite article *the*. It designates an instance that the speaker has pointed out for attention (anaphoric and exophoric). In doing this, the speaker assumes that the hearer can identify the instance. That identification is possible may be due to various factors, one of which is the context of previous discourse. In order to state, 'that bag', it is likely that previous discourse between speaker and hearer has already established a unique referent for it (the bag). With respect to discourse structure, *nán* tends to refer backwards (anaphorically) to an event recently introduced by a narrator.

(74) (Four Reigns [CU Thai Concordance])

mɨâ	*mɔɔŋ*	*paj*	*tʰîi*	*ráan*	*kʰǎw*	*hěn*	*man*	***kamlaŋ***	*jɨɨn*	***jùu₂***
when	look	go	at	store	3S	see	3S	**PROG**	stand	**stay**

tʰîi	***nâa***	***kʰáwtôə***	***kʰítŋən***
at	**front**	**counter**	**cashier**

'When (he) looked at the store, he saw it was standing in front of the cashier counter.'

6. Conclusions

We have seen that *kamlaŋ* corresponds closely to the notion progressive expressing the dynamic quality of ongoing actions. It also has the potential to bring out a stative verb's dynamic range, if it is pragmatically possible and acceptable. However, rather than simply describing the internal temporal contour of an event, it indicates simultaneousity—T-SIT coincides with TT.

The grammaticalized *jùu₃/₄* has two semantic effects: locator effect and unchanging/continuity effect. The locator effect is a primary function. It locates an event in various domains such as time, attribute, quantity, and possession. Continuity is the output of our experience of remaining in the same place through time. It is a secondary function, which can be backgrounded. This 'unchanging' effect is not the same as 'stative' (contra to Tansiri 2005). This is evident by the fact that *jùu₄* does not cause a dynamic verb to be construed as stative. Like *kamlaŋ*, *jùu₄* serves as a temporal relator—T-SIT is situated at TT.

The notion of TT, together with other temporal concepts, is important to understand temporality, even in a 'tenseless' language like Thai. Unlike tense which conveys temporal information directly, TT is pragmatically inferred.

The properties of *jùu₃/₄* and *kamlaŋ* can be summarized as follows.

Table 6: *Summary of jùu₃/₄ and kamlaŋ*

	kamlaŋ	***jùu₃***	***jùu₄***
Can occur as main verb	No	Yes	
Position in syntax	Pre verb	Post verb (subordinate verb)	Post verb (grammatical marker)
Aspectual value	Changing/Dynamic	Continuity/unchanging	
Temporal relation	T-SIT coincides with TT	N/A	T-SIT is situated at TT
Compatibility with durative adverbials	No	Yes	N/A
Compatibility with temporal deictic expressions	Yes	N/A	Yes
Compatibility with two-time point expressions	No	Yes	Yes

References

Boonyapatipark, Tasanalai. 1983. *A study of aspect in Thai.* London: University of London dissertation.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Croft, William. to appear. *Verbs: Aspect and argument structure*. Oxford: Oxford University Press.

Frawley, William. 1992. *Linguistic semantics*. Hillsdale, New York: Lawrence Erlbaum Associates.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics* (3), 41-58. New York: Academic Press.

Iwasaki, Shoichi, & Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge: Cambridge University Press.

Klein, Wolfgang. 1994. *Time in language*. London: Routledge.

Kullavanijaya, Pranee, & Walter Bisang. 2007. Another look at aspect in Thai. *Manusya: Journal of Humanities* (13), 61-86.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald W. 1991. *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Supanvanich, Ing-Orn. 1973. *Tenses in Thai.* Bangkok: Chulalongkorn University thesis.

Tansiri, Kachen. 2005. *Interactions between grammatical aspect and lexical aspect: A case study of alternating intransitive constructions in Thai.* Bangkok: Chulalongkorn University MA thesis.

Thiengburanathum, Prang. 2010. Linguistic encoding of events in Thai. Melbourne: La Trobe University dissertation.

Uppakitsinlapasarn, Phaya. 1964. *Waiyakornthai [Thai grammar]*. Bangkok: Thai Wattana Phanit.

Van Valin, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

PATTAMA Patpong. 2013. Thematic progression of Thai Song Dam folktales.
Journal of the Southeast Asian Linguistics Society (JSEALS) 6:189-215
Received 4/8/2012, text accepted 6/10/2013, published October 2013
URL: http://hdl.handle.net/1885/10617
ISSN: 1836-6821 | Website: http://jseals.org
Editor-In-Chief Dr Paul Sidwell | Managing Editor Dr Peter Jenks

SOUTHEAST ASIAN LINGUISTICS SOCIETY SEALS
www.jseals.org | Volume 6 | 2013 | Asia-Pacific Linguistics
Australian National University

THEMATIC PROGRESSION OF THAI SONG DAM FOLKTALES[1]

PATTAMA Patpong
Research Institute for Languages and Cultures of Asia, Mahidol University
<pattama.pat@mahidol.ac.th>

Abstract
This paper reports on an investigation of Thai Song Dam folktales which is a linguistic part of *Textual data management of Thai Song Dam Ethnic Group*. It draws on Systemic Functional Linguistics to investigate the discourse structure of ten Thai Song Dam folktales. The paper explores thematic progression patterns in Thai Song Dam folktales spoken in Thailand. The thematic progression is associated with the method of text development which is based on the typical thematic selection realized by unmarked and marked topical Themes (cf. Fries, 1981/1983, 1995a, 1995b). In this current study, patterns of thematic progression are explored as a way of revealing the textual organization of the folktales and the local progression achieved by the thematic selection of successive clauses (cf. Daneš (1974)). The data for this study were drawn from ten Thai Song Dam folktales. They were taken from secondary sources (e.g., folklore researches' appendices, Thai Song Dan folktale collections). Based on Daneš's notion of thematic progression, patterns of Theme selection were explored. The study revealed that the most frequent pattern of thematic progression selected in the ten folktales was the continuous or constant Theme expressed by unmarked topical Themes. Both ellipsed and non-ellipsed unmarked topical Themes were selected and repeated as the point of departure of each clause. The second most frequent pattern is linear Theme pattern. The study also revealed that the primary methods of folktale development are those of temporal and spatial organizations.
Keywords: Folktales, Thematic progression, Thematic development
ISO 693-3 language codes:: soa, blt

1. Introduction

Folktales, as one variety of traditional narrative inherited in all societies, are texts that are created as a form of traditional story that tries to explain or understand the world. Imagination is vividly woven around talking animals, mythical creatures, supernatural beings, and magical objects (cf. Thompson, 1946). Folktales usually have no identified author and are orally passed down from generation to generation. In all societies, folktales are considered treasures of mankind. Telling folktales is a traditionally subtle method of teaching valuable lessons and mirror the values and culture of a society. In terms of medium, they can be either spoken or written. The nature of the narrative focused on folktales in particular which are a combination of

[1] This paper is a part of the sub-project "Textual data management of Thai Song Dam ethnic group" of the cluster research "Ethnicity: New paradigm in language and cultural transmission" led by Professor Dr. Somsonge Burusphat. This cluster research project is sponsored by the Research-Team Promotion Grant 2010-2013, Thailand Research Fund (TRF). This paper was presented at the 22nd Annual Conference of Southeast Asian Linguistics Society at Agay, France. May 31st-June 2nd, 2012. The author would like to thank the audiences for their comments and suggestions. The author would like to express her thanks to Mr. Richard Hiam for his help with the English version of this paper. My special thanks go to Mr. Sootawee Klinubon, a Thai Song Dam native speaker, for helping with the Thai translation of Reyrai et al's folktale collection. Many thanks go to Miss Sirinda Osiri for assisting with the IPA transcription.

character orientation and a series of events that unfold through chronological succession. The folktales usually involve first or third person. First person accounts tend to be more formal. As for third person accounts, there is an interesting matter of narrator viewpoint versus character viewpoint. Chronological linage is an important feature of folktales. In folktales, sequential events are told by using chronological resources of temporal conjunctions and temporal words. These resources are deployed to specify time when the folktales take place and develop chronologically. The folktales are told and retold through accomplished time encoding as past and present time (Longacre 1974; Martin 1992; Ochs 1997).

In discourse studies over the past two decades, there have been numerous studies on discourse, both spoken and in written texts, in a number of registers/genres: Halliday and Hasan (1985), Martin (1985), Ghadessy (1993, 1995b). The underlying assumptions in all these studies have been: (a) there are patterns of organization above the sentence — method of text development — and (b) these patterns influence the production and the comprehension of the texts concerned (cf. Ghadessy, 1995b). However, there is no research on ethnic languages spoken in Thailand in general or research on ethnic folktales in particular. If the notion of 'method of development' plays an important role in the organization of information in discourse, it would be of interest to explore how the thematic development is manifested in Thai Song Dam folktales.

2. Thai Song Dam and its affiliation

The Thai Song Dam are one of the numerous ethnic groups residing in Thailand. The original place of settlement of Thai Song Dam people in Thailand was Phetchaburi province. Later on, they moved to other provinces including Kanchanaburi, Ratchaburi, Nakornpathom, Samut Sakorn, Samut Songkhram, Suphanburi, Nakhonsawan, and Phitsanulok provinces. The Thai Song Dam have been known by various names such as Song, Lao Song, Tai Song, Thai Song Dam, Tai Dam or Black Tai[2] (see also Burusphat, 2012).

The Thai Song Dam language belongs to the Upper Southwestern Tai sub-branch of the Southwestern branch of the Tai-Kadai language family (cf. Li 1960, Hartmann, 1980). In the western regions of Thailand, Thai Song Dam is spoken widely in Phetchaburi, Nakornpathom, Ratchaburi, Kanchanaburi, Suphanburi, Samut Songkhram, and Samut Sakorn provinces (Somsonge, et al. 2010).

The goal of the present research is to explore thematic progression of Thai Song Dam folktales by applying Daneš's notion of thematic progression (cf. Daneš, 1974). In the later part of the paper, the method of text development of the Thai Song Dam folktales will be discussed.

3. Theoretical framework

3.1 Thematic selection: The system of Theme

Based on a metafunction perspective, language is interpreted as having meaning potential, diversified functionally into three simultaneous metafunctions/modes of meaning: ideational, interpersonal and textual. As this research deals with the textual organization of Thai Song Dam folktales, a textual analysis is used in this study.

Textual metafunction/meaning is a resource for presenting interpersonal and ideational meanings as the organized flow of information of a text in its local context. This textual resource creates the flow of information, involving two simultaneous orientations in text development — guiding, and enabling, both text producers (speakers/writers), and text receivers (listeners/readers) (Halliday, 1978; Matthiessen 1992, 1995a, 1995b). In the guiding orientation, the textual resources are utilized in a way that guides speakers (writers) in the process of creating texts, by providing them with the method for moving from one message to another. At the same time, in the enabling orientation, the textual resource is used to structurally direct the listeners (readers) in the process of interpreting the text as it unfolds. These two orientations of text development operate simultaneously.

[2] The term "Tai" is distinguished from "Thai". "Tai" refers to any speakers of the Tai language family residing outside the Kingdom of Thailand. "Thai" refers to speakers of Tai language family residing in the Kingdom of Thailand (Suriya Ratanakul, Khunying 1994).

One of the principle grammatical systems of textual metafunction is the Theme system. An analysis of the Theme system shows how clauses are organized as messages — as quanta of information in the flow of information created as a text unfolds. Theme serves as the point of departure of the clause as message, and it is typically realized by its initial position. What follows the Theme is known as the Rheme. Rheme is what is presented in the local context set by the Theme. The clause as a message is thus a configuration of the two thematic status, Theme + Rheme constituting 'thematic structure' or 'Theme-Rheme structure' (cf. Martin, Matthiessen, and Painter, 2010; Matthiessen, 1995a).

Halliday and Matthiessen provide characteristics of Theme and Rheme as follows:

> The **Theme** is the element which serves as the point of departure of the message; it is that which locates and orients the clause within its context. The remainder of the message, the part in which the Theme is developed, is called in the Prague school terminology the **Rheme**. As a message structure, therefore, a clause consists of a Theme accompanied by a Rheme; and the structure is expressed by the order — whatever is chosen as the Theme is put first.
> (Halliday and Matthiessen 2004: 64-65)

In order to explore the thematic structure, one needs to determine where the boundary between Theme and Rheme is located: i.e. Where does the Theme end and where does the Rheme begin? (i.e., identifying Theme-Rheme structure).

Matthiessen (1995a) notes that a clause can be contextualized in terms of all three of its metafunctional perspectives. The Theme of a clause can thus have textual, interpersonal, and ideational phases (the ideational stage is known as topical Theme). This research uses the term "topical Theme" to refer to one type of Theme.

As mentioned earlier, Theme serves as the point of departure of the clause. It is typically realized by its initial position. Matthiessen (1995a: 587-588) provides a typological discussion on the system of Theme. He states that many languages do have a category of Theme; for example, there is such a category related to the method of development in Chinese, French, German, Japanese, Tagalog and so on. Theme is often realized by initial position, as it is in e.g., Chinese, French and German (Li & Thompson (1976, 1981) refer to Chinese as topic prominent language.). In English, Theme is typically realized by its initial position; that is the Theme–Rheme structure is expressed by position in sequence. Theme is realized by initial position and Rheme is realized by non-initial position. However, this notion does not apply to Japanese. Theme in Japanese is achieved by means of a particle. For example, there is a special postposition *–wa*, which signifying that whatever immediately precedes it is thematic (further discussion see, Teruya, 2006). In Tagalog, topical Theme is indicated by the preposition *ang* and comes towards the end of the clause s unmarked Theme (further discussion see, Martin, 1996).

However, Thai, Thai Song Dam and Chinese systems of Theme differ from the English one in that the unmarked Theme of an indicative clause is not selected according to mood types or mood selection (see Patpong, 2006 and Halliday and Edward, 2004). That is, mood realizations do not occur in thematic position; rather, the end of the clause is interpersonal crucial and it is the final position of a clause that polar interrogative clauses are distinguished from declarative clauses by means of the interpersonal particle *maj*^{5} in Thai, *le*^{4} in Thai Song Dam, and *ma* in Chinese.

In this research, the Theme system at clause level will be explored in terms of thematic selection (Section 5). Textual resources expressed by the distribution of thematic selection will be briefly investigated in order to explore the system of Theme as it contributes to "the method of development of texts".

3.2 Thematic progression and method of thematic development

3.2.1 Thematic progression

Based on the Functional Sentence Perspective (FSP) work of Daneš (1974) and others, thematic progression means "the choice and ordering of utterance theme, their mutual concatenation and hierarchy, their relationship to the hyperthemes of the superior text units (such as paragraph, chapter, …), to the whole text, and to the situation. Thematic progression, thus, is viewed as the skeleton of the plot" (Daneš, 1974: 114). In other words, the thematic progression structurally reflects the internal organization of the given text. It is

concerned with the way in which the text develops. Thematic progression is concerned with where Themes come from — how they relate to other Themes and Rhemes in the text. Fries (1995c: 320) states that "Patterns of thematic progression are formed by a systemic relation between the Theme-Rheme selections and experiential selections in a text. Thematic progression may be investigated by exploring the cohesive ties which occur within the various Themes within a text and the locations of the items the locative ties presume". This process is confirmed by Daneš (1974: 114) who characterizes text connexity as being represented by thematic progression (TP).

In this current research, the method of text analysis is essentially based on Daneš's thematic progression. Daneš (1974) describes three typical types of thematic progression based on scientific and other professional texts in Czech, German and English. In the following section, each type of thematic progression is detailed. They are accompanied by graphical representations and text examples as illustrated by Daneš.

❶ **Simple linear thematic progression** (or thematic progression with linear thematization of rhemes, zigzag thematic pattern or Rheme → Theme progression)

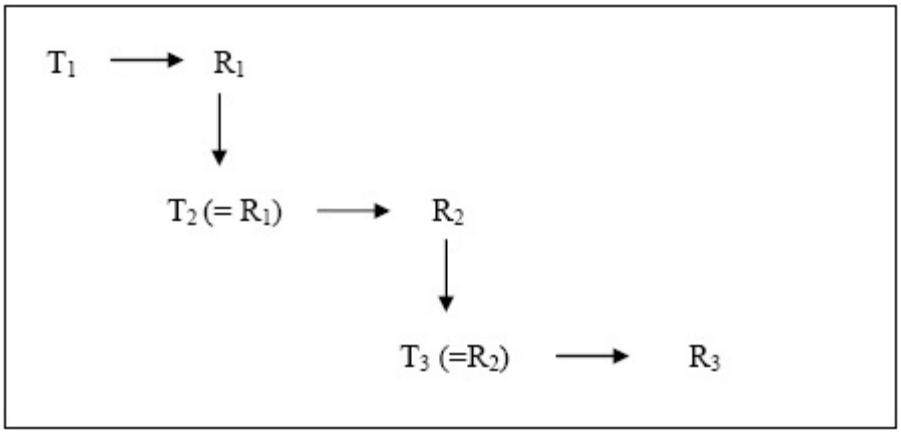


Figure 1: Simple linear Theme pattern

In a single linear thematic progression, the content of the Theme of the second clause (Theme 2) derives from the content of the previous Rheme (Rheme 1); the content of Theme 3 derives from Rheme 2 etc. In other words, the Rheme of one clause is taken up as the Theme of the subsequent clause. This thematic progression is termed "Rheme → Theme progression" (cf. Matthiessen, 1995a).

❷ Thematic progression with constant (continuous) theme (or theme iteration):

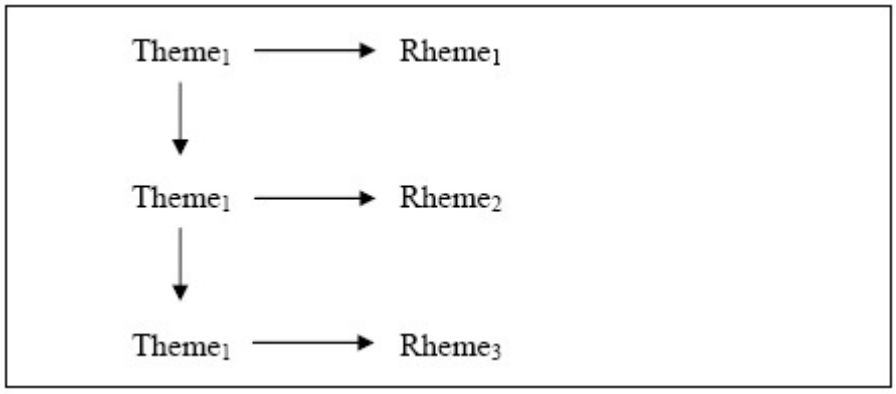


Figure 2: Constant (continuous) Theme pattern

In this type, the same Theme appears in a series of utterances, to which different Rhemes are linked up. This thematic pattern is known as "Theme → Theme progression" (cf. Matthiessen, 1995a). In this

pattern, the Themes in the text constitute a chain of (typically) co-referential items which extends through a sequence of clauses.

❸ Thematic progression with derived themes

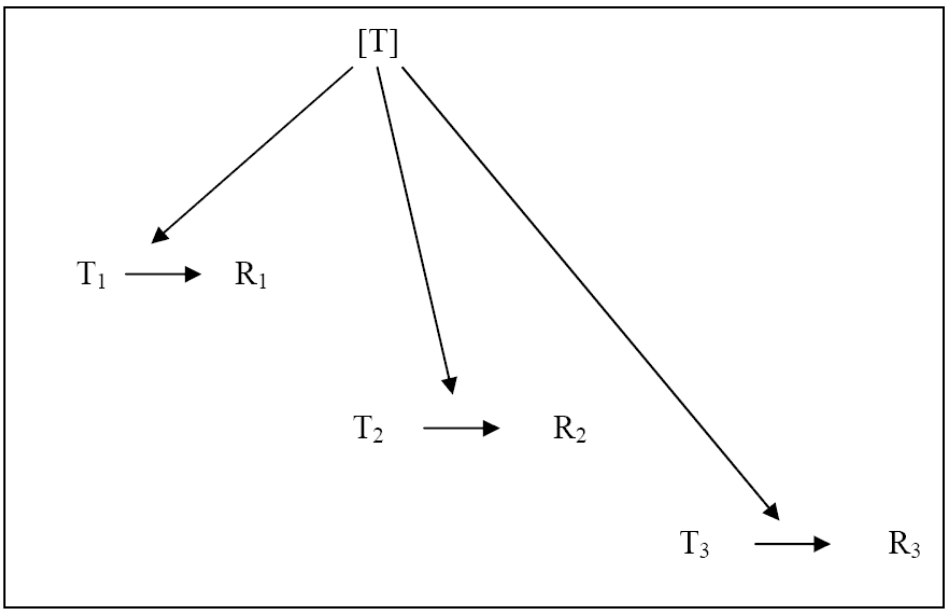


Figure 3: Derived Theme pattern

In this pattern, the passage as a whole concerns a single general notion (known as hypertheme), and the Themes of the various constituent clauses all derive from that general notion. This derived Theme pattern serves as a way of introducing New information which is taken up as the Theme of the subsequent clause (cf. Bloor and Bloor, 1995).

In the organization of the text, especially in long texts, it is possible that the three main patterns of thematic progression can be found combined within the same text because they can achieve several proposes. The combined thematic pattern is illustrated below.

❹ The split Rheme

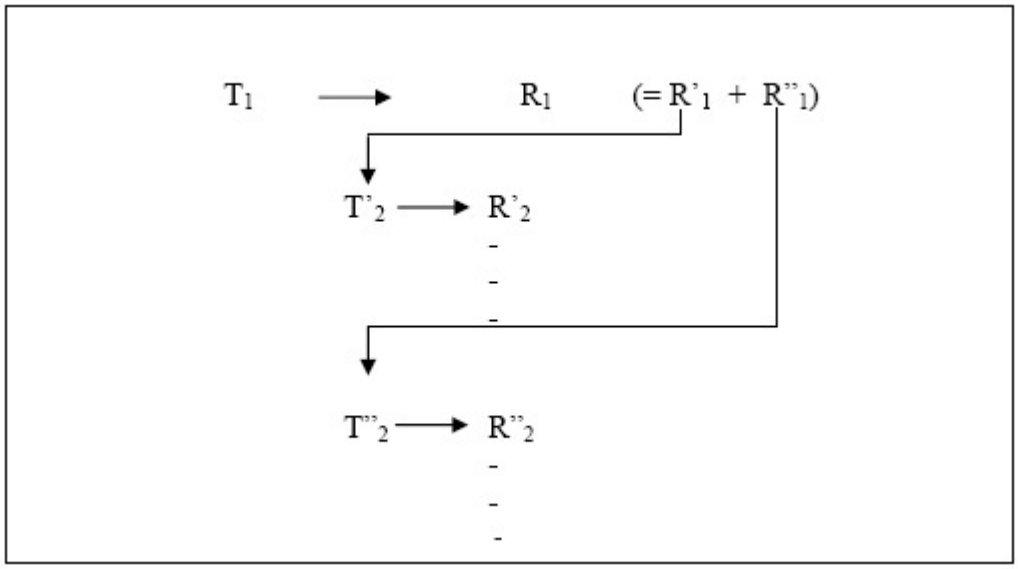


Figure 4: Split Rheme pattern

In cases of complex text structures, the thematic patterns can be combined and manifested in the organization of the text. The first and the second thematic progressions are combined and called a split Rheme. This pattern occurs when the Rheme of a clause has two or more components, each of which is taken in turn as the Theme of a subsequent clause. An example of split Rheme taken from Matthiessen (1995a: 582-583) is given below.

> The fuels of the body are **carbohydrates**, **fats** and **proteins**. These are taken in the diets. … **Carbohydrates** are the principle source of energy in most diets. … **Fats** make up the second largest source of energy in most diets. … **Proteins** are essential for the growth and rebuilding of tissue.

From the example above, the body fuels are composed of three sources of energy: carbohydrates, fats and proteins. Each of them is given thematic status one by one in subsequent clauses.

As pointed out by Daneš (1974), the organization of information in texts is determined by the progression in the ordering of utterance Themes and their Rhemes. The relationship between successive Themes and their Rhemes in the course of thematic selection provides an account of the 'method of development' of the text.

3.2.2 Methods of thematic development

The notion of "method of development" was introduced by Fries (1981/1983: 119). Method of thematic development, or thematic development, means that the choice of Theme for any individual clause will relate to the way in which information is developed over the course of the whole text (Martin, Matthiessen, Painter, 2010: 22).

Halliday and Matthiessen (2004: 98-99) state that, as a general principle of clause interpretation, the Theme of a clause extends up to the first element that has some representational function in the clause (i.e., the 'topical' Theme). The importance of the thematic pattern of a text reflects the importance of the Theme of the clause in the overall development of the text (cf. Fries, 1981/1983). Halliday and Matthiessen (2004: 99) emphasize that the choice of clause Themes plays a fundamental part in the way discourse is organized; it is this, in fact, which constitutes what has been called the 'method of development' of the text (see Ghadessy (1995b)'s contributions).

Method of development deals with the lexico-semantic content of Themes. Thematic content correlates with the method of development of a text and the nature of the text. In text analysis, the thematic context is illustrated by thematic selection. In Fries's further research on method of text development, he shows that thematic content correlates with different genres and with different generic structures within a text (cf. Fries, 1995c). There are a number of studies on thematic development in various English registers.

These previous researches include scientific texts (Hutchins, 1977); news reports, editorials, and letters of complaint (Francis, 1989, 1990); telephone conversation (Bäcklund, 1992); recipes and fables (Xiao, 1991); guide books (Long, 1991, Matthiessen, 1999b); advertisements (Fries, 1993); sports commentaries (Ghadessy, 1995a); and academic texts in relation to translation studies (Jalilifar, 2009). For further detailed discussion see Fries (1995c), Ghadessy (1995b).

According to the previous researches mentioned, to date, no research has been undertaken on methods of folktale development in ethnic languages spoken in Thailand in general, nor research on Thai Song Dam folktales in particular. Therefore, it is of interest to explore what strategies are deployed in Thai Song Dam folktales.

4. Data and text selection

In this paper, ten Thai Song Dam folktales were selected for analysis. Based on an index of folktale types (Thompson, 1946), folktales can be classified into two main types — simple, and complex. Simple tales are those with one motif, while complex tales are those with multiple motifs which involve multiple protagonists. For this study, all folktales selected were simple Thai Song Dam folktales taken from secondary sources (e.g., theses appendices, independent study appendices, field notes). The main sources for these folktales were Reyrai et al (1980), Sunant (1985), and Sootawee (2009). Table 1 lists the corpus of Thai Song Dam folktales used in this research, along with their folktale type, the name of the story teller and

where they live. Table 2 provides details of the number of complex and simplex clauses, found in each folktale.

Table 1: A corpus of ten Thai Song Dam folktales used in this research

No.	Folktale type	Folktales	Story teller (source)	Residence location
1.	Explanatory tales	The elephant and the ant	Mrs. Bunruan Bunnoy (Sootawee, 2009)	Yangyong Sub-district, Thayang District
2.	Explanatory tales	The dog and the cat	Mr. Sawing Loyceng (Sootawee, 2009)	Yangyong Sub-district, Thayang District
3.	Animal tales	Python	Miss Daeng Yaemmun (Reyrai et. al, 1980)	Nongprong Sub-district, Khaoyoy District
4.	Animal tales	Four friends	Miss Daeng Yaemmun (Reyrai et. al, 1980)	Nongprong Sub-district, Khaoyoy District
5.	Fables	Wishing crystal	Miss Saythip Chuenin (Sootawee, 2009)	Yangyong Sub-district, Thayang District
6.	Fables	Became rich because of us	Mrs. Sombun Thuanhit (Sootawee, 2009)	Yangyong Sub-district, Thayang District
7.	Jests & anecdotes	Father in law wanted a claver son in law	Mrs Or Caengcat (Sootawee, 2009)	Yangyong Sub-district, Thayang District
8.	Jests & anecdotes	The red duck	Mr. Sawing Loyceng (Sootawee, 2009)	Yangyong Sub-district, Thayang District
9.	Legends	The number 7 legend	Miss Saythip Chuenin (Sootawee, 2009)	Yangyong Sub-district, Thayang District
10.	Legends	The rice grain legend	Mrs. Sombun Thuanhit (Sootawee, 2009)	Yangyong Sub-district, Thayang District

Table 2: Number of complex clauses and simplex clauses of the selected folktales

No.	Folktale type	Folktales / Motifs	Complex clause	Simplex clause
1.	Explanatory tales	The elephant and the ant	22	65
2.	Explanatory tales	The dog and the cat	29	102
3.	Animal tales	Python	19	67
4.	Animal tales	Four friends	54	160
5.	Fables	Wishing crystal	56	236
6.	Fables	Became rich because of us	23	97
7.	Jests & anecdotes	Father in law wanted a claver son in law	12	38
8.	Jests & anecdotes	The red duck	47	141
9.	Legends	The number 7 legend	9	34
10.	Legends	The rice grain legend	12	38
		Total	283	992

In the following sections, thematic selection which manifests thematic organization of Thai Song Dam folktales is investigated. The selected Thai Song Dam folktales are annalysed clause by clause. Theme in each simplex clause is identified and profiled according to its type – textual, interpersonal and topical.

5. Thematic selection in the Thai Song Dam folktales

A mentioned above, the method of development is explored as a way of revealing the organization of the folktales and the local development achieved by the thematic organization of successive clauses which can

be explored through the thematic selection. In this section, the corpus of the ten Thai Song Dam folktales is explored in terms of lexcogrammatical analysis at the clause level. The Theme-Rheme structure of each clause is identified; the Theme of a clause may include elements from all three metafunctions: textual, interpersonal and ideational (topical) Themes. In the following sub-sections, three different types of Theme are discussed.

5.1 Textual Theme selection

Textual Theme includes three possible textual choices: continuative; relative element; and conjunction, used to express semantic progression (see Hutchins's discussion on thematic and semantic progressions in text organization (Hutchins, 1977)).

• *Continuatives*

Continuatives are discourse markers which signal that a new move is beginning: a response in dialogue, or a move to the next point if the same speaker is continuing (Halliday & Hasan, 1976; Halliday & Matthiessen, 2004: 81). The continuatives in Thai Song Dam folktales were found in the dialogic passages between two protagonists. An example is given below[3].

(1) Fable: Text 2 [4.1-4.2]

Theme					Rheme			
ʔan¹taː¹faːj³			*seːt³thiː¹*	*nan⁶*	*waː⁴ ciː⁴*		*lɔʔ⁴*	
as for			millionaire	that	say		Fp.	
waː⁴	*ʔəː³*	*haw²*	*naʔ⁴*		*caŋ²*	*laːj⁵*	*luaj²*	*ləj⁶*
that	well	we	part.		FUT.	get	be rich	Fp.

'That millionaire said that we would be rich.'

• *Relative elements*

Relative elements link the clause in which they occur to another clause forming an embedded clause. They are thematic in nature and, therefore, tend to occur in the initial position. Two specific relative items were found in the ten folktales under study: *thiː⁴* 'that, which', and *toː¹* 'who, that'. Among these relative elements, *thiː⁴* was the most commonly used in these Thai Song Dam folktales.

• *thiː⁴* 'that' as a marker of an embedded clause

(2) Explanatory tale: Text 2 [18.3-18.4]

Theme	Rheme				
(∅ = man²)	*hauu⁵*	*paj¹*	*ŋom²*	*ʔaw¹*	*kœːw⁵*
(3SG)	get	go	dive for	take	crystal
thiː⁴	*tok³*	*nam⁶*	*paj¹*		
that	fall	river	go		

'… (the cat) made (the cormorant) dive for the crystal that fell into the river…'

• *Conjunctions*

Conjunctions are words indicating a logico-semantic relationship between clauses in a clause nexus (structural conjunctions) or between a clause and some parts of the preceding discourse (cohesive conjunctions). Structural conjunctions can be divided into two sub-types: linkers (i.e., clauses linked

[3] The phonetic transcription of the illustrating examples is based on Orapin Maneewong (1987) and Somsonge Burusphat (2012).

paratactically) and binders (i.e., clauses linked hypotactically) (cf. Halliday & Matthiessen, 2004: Chapter 9, Section 9.3).

A linker relates two independent clauses in a paratactic clause nexus. It is used when the logico-semantic relation is one of expansion (e.g. *læʔ*^{4} 'and', *lɯː*^{1} 'or', *tæː*^{3} 'but').

A binder relates a dependent clause to the independent clause in a hypotactic clause nexus. Binders can be divided into three subtypes based on their logico-semantic relationship: elaboration (for example: *kɔː*^{5}*kɯː*^{2} 'that is'); extension (for example: *læʔ*^{4} 'and'); and enhancement (for examples: *læːw*^{6}*kɔː*^{5} 'and then', *poː*^{6} 'because').

Besides being used to link two clauses in the hypotactic status, binders are also used to link second clauses, projected by either verbal or mental lexical verbs, to the proceeding clause. From the selected corpus, one clause binder was found: *wa*^{4} 'that'.

• binder '*waː*^{4}' in verbal projection clause

(3) Explanatory: Text 2 [9.8-9.9]

Theme		Rheme
toː^{1}*nɯː*^{1}	*kɔː*^{5}	*bɔʔ*^{3}
rat	conj.	say
waː^{4}	*(∅ = man*^{2}*)*	*ɲɔːm*^{2} *pen*^{1}
that	(∅ = it)	be alive

'The rat then said that (it) would like to live.'

• binder '*waː*^{4}' in mental projection clause

(4) Animal: Text 1 [15.4-15.7]

Theme		Rheme			
ŋɯː^{2}*lɯam*^{1}	*nan*^{6}	*maː*^{2}	*nɔːn*^{2}	*khwaːŋ*^{1}	
python	that	come	lay down	obstruct	
(∅ = man^{2}*)*		*pəːn*^{1}	*kan*^{2}*kuː*^{2}	*haɯ*^{5}	*saw*^{2}
(∅ = it)		cover	ditch	give	them
(∅ = man^{2}*)*		*bɔʔ*^{3}	*hɯː*^{6}		
(∅ = it)		NEG.	know		
waː^{4}	*saw*^{2}	*laːj*^{5}	*nok*^{4}	*kan*^{1}	
that	they	chase	bird	together	

'That python laid there and blocked (the ditch), (it) covered the ditch for them. (It) did not know that he chased birds.'

5.2 Interpersonal Theme selection

An interpersonal Theme foregrounds the speaker's judgment or comments. It might include one or more of the following components: vocative; exclamatory element; and wh-interrogative. All of these interpersonal Themes were found in dialogic passages between protagonists.

• *Vocatives*

Vocative elements are used to directly address the listener in a dialogic passage. They can be expressed by personal names, kinship terms, occupational terms, or pronouns. Example 5 below illustrates a vocative element which is expressed by the kinship term 'grandma'.

(5) Jest and anecdote tale: Text 2: [20.1]

Theme	Rheme						
ʔem[1] thaw[5]	*ku:[1]*	*paj[1]*	*pɔ:[1]*	*to:[1]fa:n[1]*	*khaw[5]*	*to:[1]*	*nuŋ[3]*
grandma	I	go	find	barking deer	enter	CLF	one

'Grandma! I found a barking deer.'

• *Exclamatory elements*

Exclamatory elements occur at the beginning of clauses with thematic status. Functionally, they are used to express the interactant's emotion such as happiness, sadness, fear, surprise, or amusement. Exclamatory elements are expressed by exclamatory particles.

(6) Explanatory tale: Text 2: [5.2-5.3]

Theme				Rheme		
mot[4]dœ:ŋ[1]				*khuəŋ[2]*		
red ant				be angry		
wa:[4]	*hu:[6]*	*ca:ŋ[6]*	*man[2]*	*su:ŋ[1]*	*ni:[4]*	*wa:[3]*
that	aha	elephant	3SG	be tall	this	NEGOTIATOR

'"Aha" he said. "The elephant is tall."'

• *Wh-interrogatives*

Wh-interrogatives or Wh-items signal that an 'answer' is required from the addressee. Wh-interrogatives function as the unmarked Themes of Wh-interrogative clauses. These items have a twofold thematic value: they are at the same time both interpersonal and topical — interpersonal because they construe the mood, and topical because they represent the participant or circumstance (Halliday and Matthiessen, 2004:85). Examples of wh-interrogatives are ***sin[5]*** and ***ʔan[1]*** meaning 'why'.

(7) Jest and anecdote tale: Text 2: [9.1]

Theme	Rheme				
sin[5]	*maʔ[4]*	*hɔ:[3]*	*pha:[5]*	*ma:[2]*	*lə:j[6]*
Wh-element: why	NEG	wrap	cloth	come	NEGOTIATOR

'Why did not (you) wrap it with a piece of cloth?'

5.3 Topical (ideational) Theme selection

Topical (ideational) Themes are concerned with Theme selection — the selection between unmarked and marked Themes. The topical Theme is selected from one of the transitivity elements in the clause — a participant, a circumstance and / or a process.

5.3.1 Unmarked Themes

▪ Theme in declarative clauses

In a declarative clause, the unmarked Theme is the Subject. It is expressed by a nominal group, as in example 8, or a personal pronoun as in example 9.

(8) Explanatory: Text 1 [17.3]

Theme			Rheme			
***caːŋ*^{6}**	*kɔː*^{5}	*ləːj*^{2}	*muːn*^{2}	*taː*^{1}	*maʔ*^{4}	*khɯn*^{5}
elephant	conj.	conj.	open	eye	NEG.	ascend

'The elephant could not open its eyes.'

(9) Animal: Text 2 [48.5]

Theme		Rheme			
***saw*^{2}**	*kɔː*^{5}	*kaː*^{5}	*toː*^{1}*suaː*^{1}	*nan*^{6}	*taːj*^{1}
he	conj.	kill	tiger	that	die

'He then killed that tiger.'

▪ Theme in interrogative clauses

There are two main types of interrogative clauses: elemental interrogatives and polar, or yes/no interrogatives. Like Thai, Thai Song Dam interrogative clauses have no direct link between Theme selection and the nature of mood type since the realization of mood choice does not involve changes in the word order of the clause (cf. Patpong, 2006).

In this current research, only instances of Themes in elemental interrogative clauses were found. An example is shown below.

Wh/Subject/Theme

(10) Animal: Text 1 [7.5]

Theme	Rheme				
***phaw*^{4}**	*ʔiʔ*^{3}	*maː*^{2}	*waː*^{4}*kwaːm*^{2}	*hauː*^{5} *ʔaj*^{5}	*nok*^{4}*caw*^{5}
wh. interrogative: who	FUT.	come	judge a case	give title	heron

'Who will judge the heron's case?'

▪ Theme in imperative clauses

In an imperative clause, the unmarked Theme is the Predicator realized by a lexical verb of command. An example is provided below.

(11) Animal: Text 1 [5.1-5.2]

Theme			Rheme	
ʔan^{1}	*mœːw*^{2}	*kɔː*^{5}	*waː*^{4}	
as for	cat	conj.	say	
***jaː*^{3} *ʔet*^{3}**			*cen*^{4}	*nœʔ*^{4}
don't do			like that	Fp.

'As for the cat, it said not to do (like) that.'

Apart from unmarked Theme, existential clauses are a type of clause typically found at the beginning of folktale structure. Existential clauses are used to introduce the protagonists into the Placement Event of folktales (See Patpong, 2011 for lexicogrammatical realizations of the Placement Event or folktale setting). The unmarked Theme in an existential clause is realized by the lexical verb "*miː*^{2}" meaning exist/have.

(12) Fable: Text 1 [1.1]

Theme	Rheme			
mi:²	*phua¹*	*mia²*	*ku:⁴*	*nuŋ³*
exist/have	husband	wife	CL.	one

'There was a couple.'

5.3.2 Marked Themes

The marked Theme leads to the system of Theme Matter. The choice in the system of Theme Matter is concerned with the status of the Theme in the other metafunctional layers of the clause (Matthiessen, 1995a: 549). In a declarative clause, the marked Theme is concerned with thematic selection where elements other than the Subject are given thematic status. The marked Themes can be assigned either as transitivity role or as theme matter (also known as absolute theme). If the transitivity role is selected, the Theme is conflated with a participant, circumstance or process in the clause; if the theme matter or absolute theme is selected, the Theme is not structurally integrated through conflation with other elements of the clause (Matthiessen, 1995a: 549).

❶ Marked Theme: in transitivity role

Marked Themes in "indicative" or "bound" clauses (i.e. a non-imperative clause) are any element other than the Subject. This includes a (non-subject) participant (as Complement), and a circumstance (as Adjunct): the Theme is conflated with Complement, and/or Adjunct.

In Thai Song Dam, there is no process (or Predicator) functioning as a "marked" Theme as in English. That is, in indicative clauses (i.e. declarative and interrogative) with implicit Subject, although the Predicator/Process usually occurs in the clause-initial position, it is interpreted as a Rheme. In this research, both circumstance Themes and (non-subject) participant Themes were found. Examples are given below. Marked Themes are underlined.

(1) Marked Theme selection: non-subject (complement) participant Themes

A marked Theme can be a participant which does not serve as a Subject — instead it serves as a Complement. This participant is likely to be textually recoverable from the preceding discourse.

(13) Joke and Anecdote: Text 2 [7.1-7.2]

Theme	Rheme				
pet¹ dœ:ŋ¹	*ku:¹*	*cap³*	*la:j⁵*		
duck red	I	catch	ASP.:pfv.		
lœ:w⁶ ku:¹	*phuʔ³*	*phak³boŋ⁵*		*sau³*	
conj. I	tie	morning glory		put	
	ha:ŋ¹	*man²*	*ma:²*		
	tail	it	come		

'I caught a red duck, and I tied it with a piece of morning glory.'

(2) Marked Theme selection: circumstance Themes

In Thai Song dam folktales, marked circumstantial Themes are typically location, in time and space, since these relate to the method of narrative development. In this research, folktales are organized by a series of events that unfold through chronological succession. This chronological sequence is expressed by marked circumstantial Themes, especially temporal locations. They are used to foreground a temporal sequence in

the discourse, and they are exploited to indicate a rhetorical break in the discourse. They are essential for introducing new episodes relating to the development of the text.

• marked circumstantial Theme: temporal location

(14) Animal: Text 2 [15.2]

Theme		Rheme			
kam⁴	***ni:⁶***	*saʔ³maːj¹*	*kin¹*	*læ:ŋ²*	*haŋ² kua¹*
evening	this	they	eat	dinner	Wh: what

'This evening, what will we have for dinner?'

• marked circumstantial Theme: spatial location

(15) Fable: Text 1 [3.2]

Theme			Rheme		
pɔʔ³waː⁴	*haŋ⁵*	*ta⁶leː²*	*ʔaʔ³*	*miː²*	*paː¹*	*ʔeː¹*
because	in	sea	part.	exist/have	fish	many

'... because in the sea, there were many fish.'

❷ Marked Theme: in theme matter (or absolute themes)

Theme matters play a role in the textual structure of the clause only. They are not structurally integrated with other elements of the clause through conflation (Matthiessen, 1995a: 552). The theme matter is specified as a textual "subject matter" and in Thai Song Dam, the marking of a theme matter is typically made apparent by two resources: (i) a set of prepositions (e.g. *faːj³* and *ʔan¹*, 'as for'), occurring at the beginning of the clause; and/or (ii) a thematic particle (e.g. *ʔaʔ³, naʔ⁴*). An example of theme matter marked by a thematic particle is given below.

(16) Animal: Text 1 [8.1]

Theme				Rheme				
ʔaj⁵	***mæːw²***	*naʔ⁴*	*kɔː⁵*	*paj¹*	*haː¹*	*ʔaj⁵*	*fon¹*	*lom²*
title	cat	thematic part.	conj.	go	see	title	rain	wind

'The cat then went to see the Rain and the Wind.'

Table 3 presents the total frequency and percentage of topical Themes in the ten Thai Song Dam folktales.

Table 3: Topical Themes in the ten Thai Song Dam folktales

Topical types	**Participant role**	**instances**	**Percentage**
Unmarked (86.90 %)	Subject as Theme in a declarative clause (non-ellipsed – 434 instances; ellipsed – 386 instances)	820	82.66 %
	Process as Theme in a declarative clause	16	1.61 %
	'existential' verb in an existential clause	26	2.62 %
	Total	862	100
Marked (13.10 %)	Theme Matter or Absolute Theme	20	2.02 %
	Non-subject (complement) participant	6	0.61%
	Circumstantial marked Theme	104	10.48 %
	Total	130	100

Table 3 summarizes the Theme selection of the ten Thai Song Dam folktales. Unmarked Themes were selected in 86.90 % of instances, whereas marked Themes were selected in 13.10 %. Among the unmarked Themes selected, Subjects/ Participants serving as Themes in declarative clauses were found in 82.66 % of instances. Processes serving as Themes in imperative clauses were selected in 1.61 % of instances and 2.62 % of all selections were existential verbs serving as Themes in existential clauses. Selection of unmarked Themes is highly favoured in constructing and organizing folktales, as this method of folktale development is constructed and developed around the protagonists. Structurally, folktales are organized by systemically selecting the Theme which is realized by Subjects/Participants as unmarked Themes.

6. Thematic progression

As far as the method of development is concerned, the flow of textual information is achieved by typical thematic selection which in turn contextualizes the thematic progression of the given text (cf. Daneš, 1974, Fries 1981/1983). Table 4 presents the frequency of occurrence of various types of text progression and their percentages, as found in this research.

Table 4: Thematic progressions in the ten Thai Song Dam folktales

No	**Tale types**	**Text**	**Thematic progression**			**Frequency of occurrence of types of text progression**
			Constant Theme	**Simple Linear Theme**	**Split Rheme**	
1	Explanatory tales	Text 1	9	8	—	17
		Text 2	20	5	—	25
2	Animal tales	Text 1	10	6	1	17
		Text 2	33	11	—	44
3	Fables	Text 1	40	21	1	62
		Text 2	16	4	—	20
4	Jests & anecdotes	Text 1	5	4	—	9
		Text 2	13	20	—	33
5	Legends	Text 1	3	4	—	7
		Text 2	7	5	—	12
		Total	156 (63.41 %)	88 (35.78%)	2 (0.81%)	246 (100%)

Based on Table 4, three types of thematic progression are revealed: constant Theme, linear Theme and split Rheme. An analysis of thematic progression reveals that the constant Theme progression pattern was

dominant in this current study (63.41 percent), whereas the simple linear Theme progression pattern was less frequently found (35.78 percent). Split Rhemes occurred with the lowest frequency (0.81 percent). The high frequency selection of constant and simple linear Theme progression patterns was expected given that Daneš (1974) refers to them as the basic thematic patterns.

In the constant Theme pattern, the same Theme is picked up and repeated at the beginning of each subsequent clause. These repeated Themes are realized by zero references in Thai Song Dam. This signals that each of the clauses focuses on the same protagonist, until the focus of the tales is shifted to either other protagonists or other points of interest. The constant Theme progression is associated with a key motif of folktale text development — actor orientation (cf. Longacre, 1974; Martin 1992; Ochs 1997). Actor or protagonist orientation is the vital feature of folktales as folktales are organized around the protagonists and protagonists' activities.

The simple linear Theme pattern occurs when the subject matter in the Rheme of one clause is taken up in the Theme of a following clause. In this research, this thematic pattern is used to provide a wide variety of actions performed by the protagonists that have been identified in the Rheme. Later, it is structurally picked up in the subsequent Theme. In addition, an object located in the Rheme part is given a thematic status in the subsequent clause.

A split Rheme occurs when the Rheme of a clause has two or more components, each of which is selected in turn as the Theme of a subsequent clause rendering split Rhemes comparatively rare. This study found split Rhemes at the beginning of folktales and they were used to introduce two protagonists into the folktale setting.

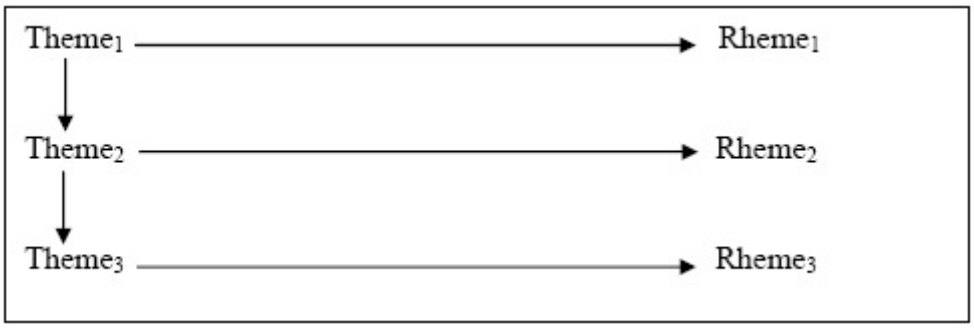
In this study, the spilt Rheme was found after the main protagonists were introduced into the folktale setting. During the Sequent Event, the folktales got complicated (see Patpong, 2011 for a detailed discussion on Generic Structural Potential of Thai Song Dam folktales). The complications that the protagonists mentioned encountered were further elaborated in the successive clauses. In this research, the Rheme of a clause consisted of two groups of protagonists (as in Fable 1, two couples) and the two protagonists mentioned (as in Animal tale 1, Fish and Heron). In the course of folktale development, the actions of each protagonist were elaborated in turn in the following Themes.

It is interesting to note that the number of folktale texts used in analyzing the thematic progression may account for the low frequency of the split Rheme (0.81 %). Moreover, this research is based on an analysis of the simple tale type, which has one motif and involves only a few key protagonists, so it is possible that only a few instances of the split Rheme are found.

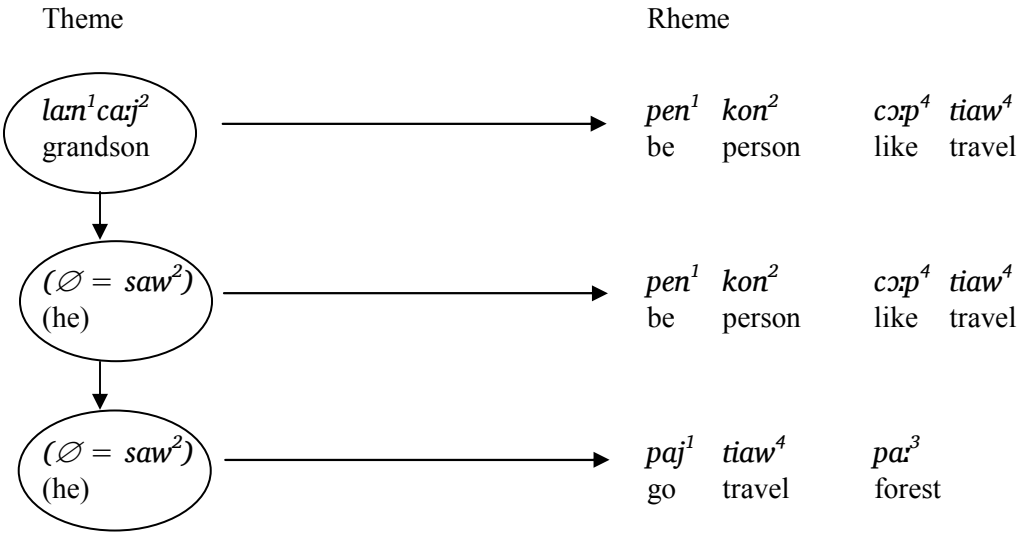
As shown by the analysis of the ten Thai Song Dam folktales, Theme selection reflects the flow of information in the unfolding text. There are three possible types of thematic progression in an unfolding text: constant Theme (i.e. Theme → Theme progression), simple linear Theme (i.e. Rheme → Theme progression), and split Rheme.

Following, are examples of thematic progressions based on the frequency of progression patterns: constant Theme, simple linear Theme and split Rheme respectively.

(i) Constant Theme progression pattern (Theme to Theme progression)



(17) Joke and Anecdote: Text 2 [2.1-2.3]



'Grandson was a person who liked to travel, (he) travelled in a forest.'

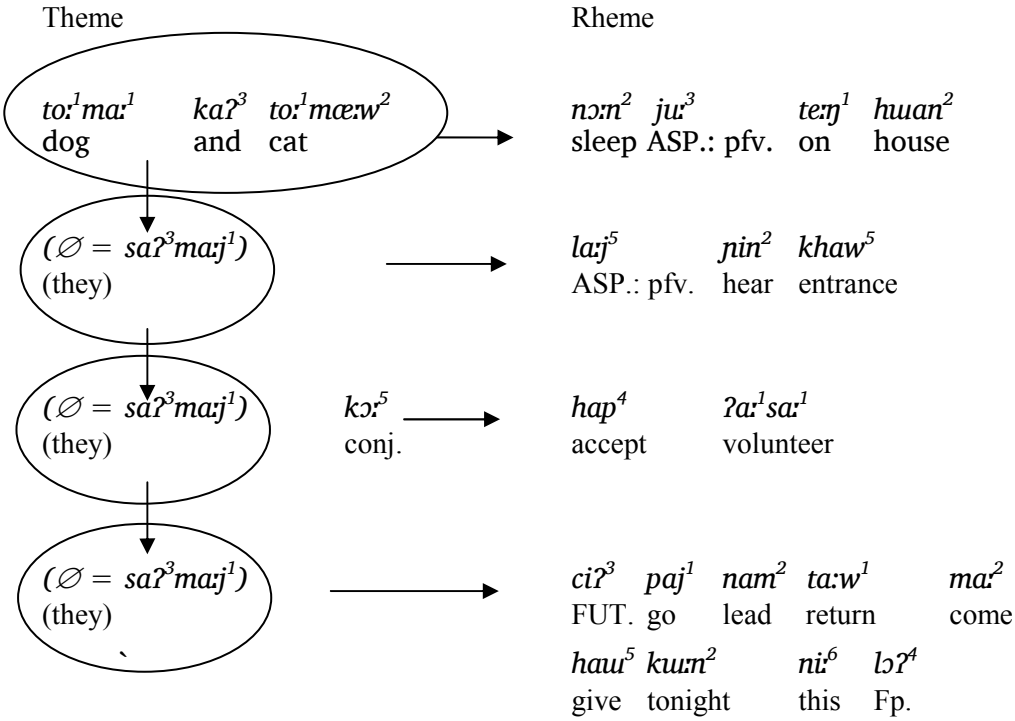
Example 17 illustrates that, after the main protoganist is introduced into the folktale setting, the folktale progresses by selecting the zero pronoun Themes or zero reference Themes 'saw^{2}' meaning 'he' serving as Themes (as in clauses 2.2-2.3). It is typical for folktales to be constantly developed by the selection of the main protogonists functioning as Themes in subsequent clauses.

In realizing a zero reference in folktales, and especially in clause complexes, hypotactic interdependency with the first clause can be constructed without mentioning the Subject of the clause. The same Subject from the previous discourse still exists, unless a new Subject is introduced into the discourse. Thus, if the Subject is elided, it will then be presumed in the following clause.

In folktales, one can track a referent from the context of the discourse. The referent is presumed by implicit anaphora and serves as Subject. What follows in a referencing chain will thus also be an implicit Subject. That is, if the Subject remains the same in successive clauses, it will be implicit; if a new Subject is introduced, the explicit new Subject will take over.

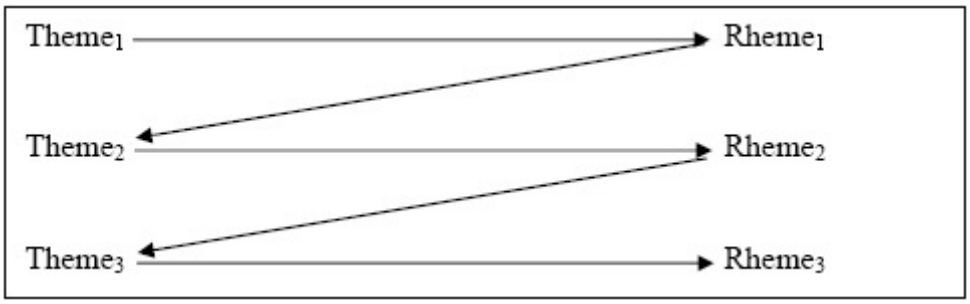
In terms of thematic progression, Explanatory Text 2 is developed mainly by the constant Theme progression pattern. This instance confirms that the constant Theme progression or iteration Theme progression is the key method of text development; the protagonists of the folktale are given thematic prominence. When they are introduced in a given folktale, they are further given the Theme positions in the subsequent clauses. Successive Themes are typically established by way of ellipsed personal pronouns (or zero pronouns) which are represented by '∅' symbol.

(18) Explanatory: Text 2 [7.1-7.4]



'The dog and the cat slept in the house, (they) heard (what the owner complained) and (they) volunteered to return the crystal glass tonight.'

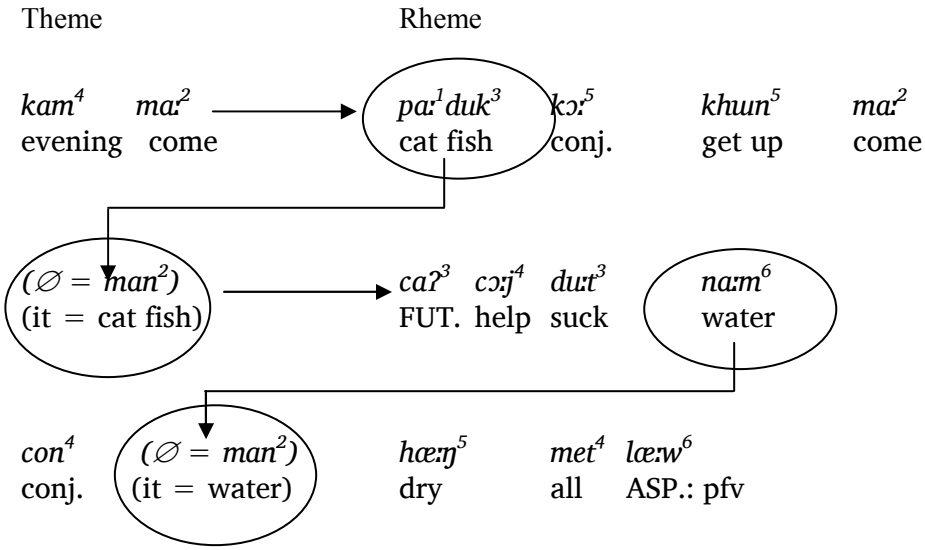
(ii) Simple linear Theme progression pattern (Rheme to Theme progression)



As mentioned above, in the second type of progression, the Theme of a new clause is picked up from the Rheme of the previous clause. Here, either a Subject or a Complement within the Rheme of one clause can be picked up by the Theme of the next clause.

• Subject within Rheme is picked up by the new Theme

(19) Explanatory: Text 1 [12.1-12.3]

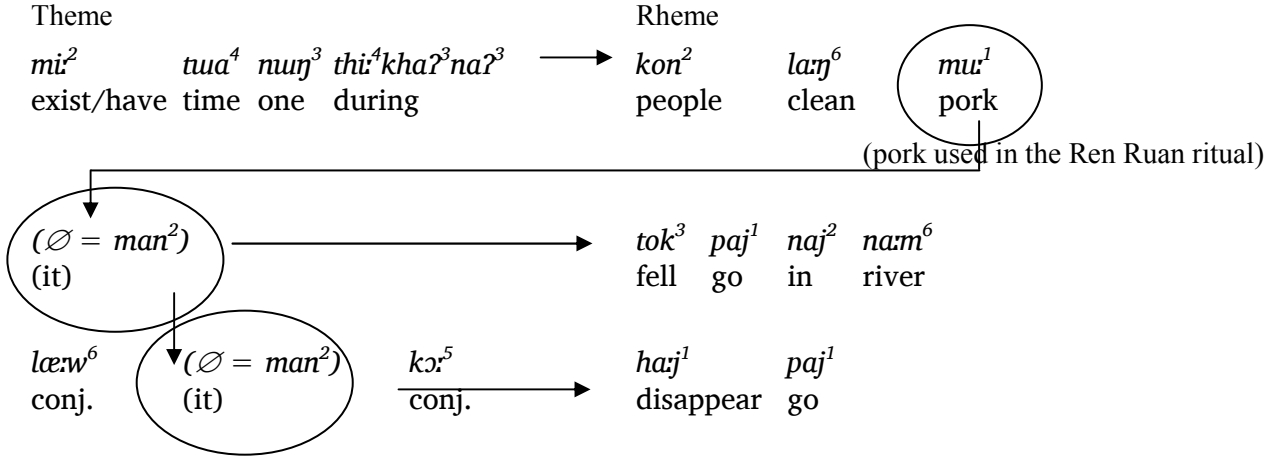


'In the evening the cat fish surfaced. (It) helped to suck the water (out of the pool), until the water was all gone.'

In example 19, the subject (a cat fish) in the Rheme part was selected as a Theme in the following clause (clause 12.2). At the same time, the object (water) of clause 12.2 was picked as a Theme in the successive clause (clause 12.3). Both Themes of clauses 12.2 and 12.3 are realized by zero pronouns.

• Complement within Rheme is picked up by the new Theme

(20) Legend: Text 1 [3.1-3.3]



'Once, while people were cleaning pork, (it) fell into the river and disappeared.'

Example 20 illustrates that the object (pork) in clause 3.1 was selected as a Theme in clause 3.2 which is realized by a zero pronoun.

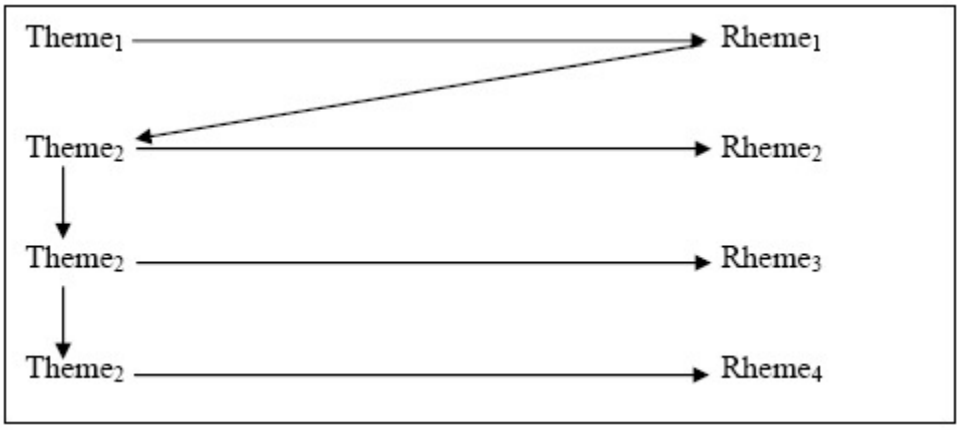
In addition, this type of thematic progression may also operate locally in the form of an embedded clause. A Rheme, which is embedded by an embedded clause, is picked up by the Theme of the successive clause.

(21) Explanatory: Text 2 [18.3-18.4]

Theme		Rheme				
toː¹ mæːw²	→	*hauu⁵*	*paj¹*	*ŋom²*	*ʔaw¹*	*kœːw⁵*
cat		give	go	dive for	take	wishing crystal
thiː⁴		*tok³*	*naːm⁶*	*paj¹*		
that		fell	river	go		

'The cat dived for the wishing crystal that fell into the river.'

From the corpus of folktales under study, it was common to observe an extension of the simple linear Theme progression pattern (19.92 percent of linear cases). That is, once the simple linear Theme pattern was instantiated in the first clause, the following clause was extended by the constant Theme progression pattern, as illustrated in example 22 below.



(22) Legend: Text 1 [4.1-4.4]

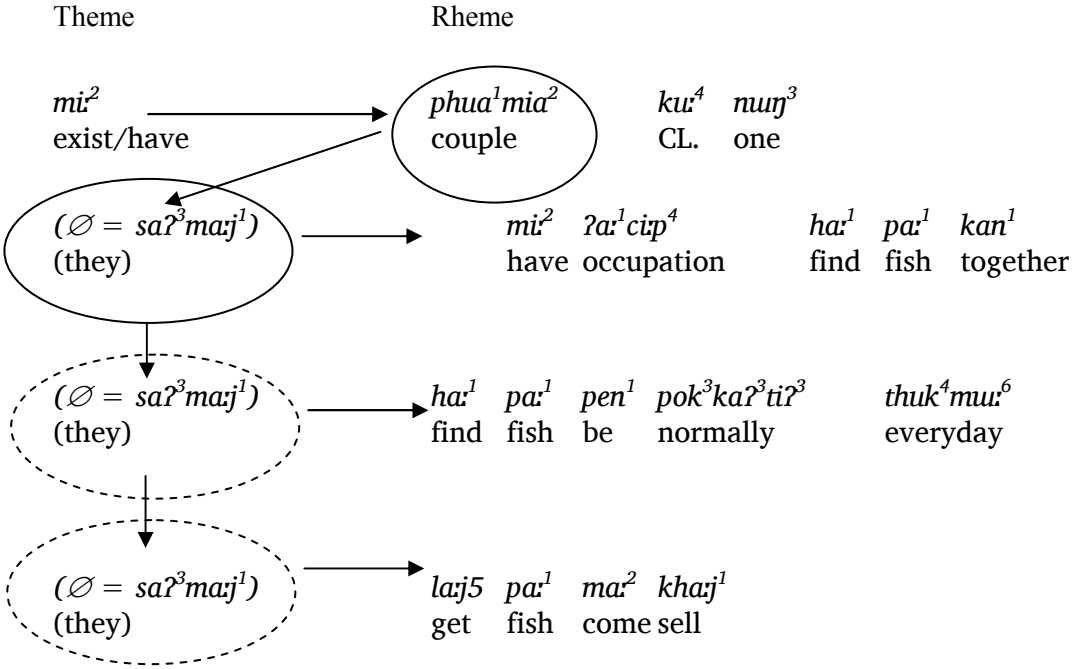
Theme			Rheme			
pɔː²diː¹	*miː²*	→	*kon²*	*cet³*	*kon²*	*ʔaʔ³*
just	exist/have		person	seven	CL.	Fp.
(∅ = saʔ³maːj¹)		→	*kam¹laŋ²*	*kwaːŋ⁵*	*hœː¹*	
(they)			ASP.: Ipfv	cast	net	
(∅ = saʔ³maːj¹)	*ləːj²*	→	*paj¹*	*kwaːŋ⁵*	*hœː¹*	
(they)	conj.		go	cast	net	
(∅ = saʔ³maːj¹)		→	*lɔː⁴*	*toː¹muː¹*		
(they)			see	pig		

'There were seven people. (They) cast a net. (They) saw a pig.'

In terms of the flow of information, most of the folktales began by relating a new Theme to the previous Rheme (Rheme → Theme progression). The selection of this thematic progression is highly motivated as it is thematically used to introduce a protagonist who is made apparent in an existential clause. Later, the folktales change and relate a new Theme to the previous one (Theme → Theme progression). Within the Theme-Rheme structure, this extended thematic pattern is expressed by either explicit or implicit unmarked topical Themes functioning as subject of the clauses. In addition, the constant Theme progression pattern, as the extended segment of the thematic pattern, is used to provide various routines and activities/performances done by the protagonists (as shown in example 23) or characteristics of the mentioned protagonist (as shown in example 24).

• The simple linear Theme progression pattern with an extended constant Theme pattern describing the protagonists'acts.

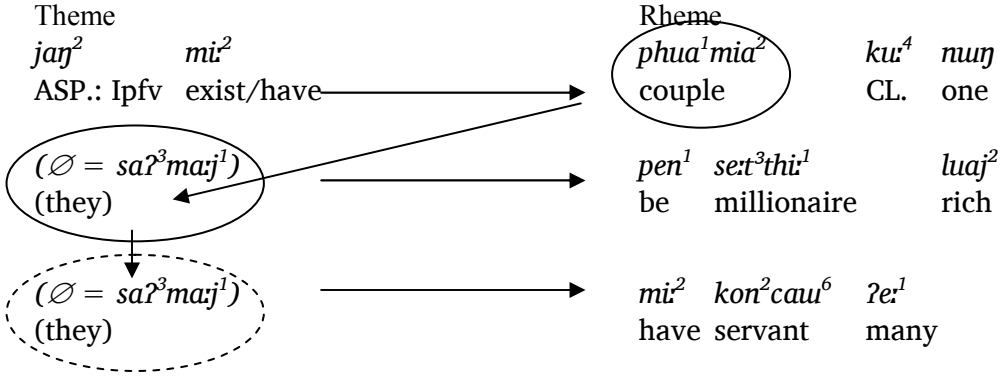
(23) Fable: Text 1 [1.1-1.4]



'There was a couple. (They) earned their living by fishing. (They) fished everyday and sold them.'

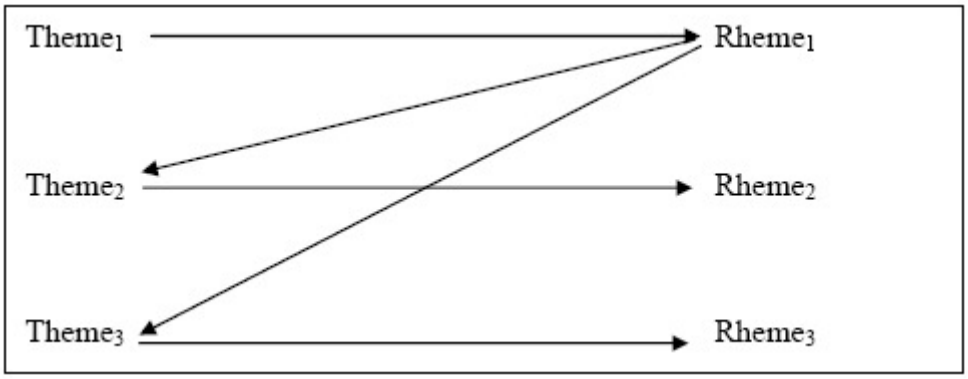
• The simple linear Theme progression pattern with an extended constant Theme pattern describing the protagonists' characteristics

(24) Fable: Text 2 [1.1-1.3]



'There was a couple. (They) were millionaires. (They) had many servants.'

(iii) Split Rheme progression pattern



In the split Rheme progression, the Rheme of a clause has two or more components, each of which is taken in turn as the Theme of a subsequent clause (See the illustrating example in Section 3.2.1).

(25) Fable: Text 1 [38.1-38.4]

Theme Rheme

theː²waʔ⁴daː¹	*kɔː⁵*	*maː²*	*nuk⁴*				
god	conj.	come	think				
(∅ = saw²)	*kɔː⁵*	*maː²*	*hen¹*	*phua¹mia²*	*sɔːŋ¹*	*kuː⁴*	
(he)	conj.	come	see	couple	two	CL.	
		ʔiʔ³	*lœːw⁶*				
		again	ASP.: pfv				
kuː⁴	*lœːk⁴*	*haw²*	*hauu⁵*	*kœːw⁵*	*nan⁶*	*paj¹*	
CL.	first	I	give	crystal	that	go	
kuː⁴	*niː⁴*	*kɔː⁵*	*maː²*	*ʔet³*	*ciʔ⁴*		
CL.	this	conj.	come	do	same		

‘The god then thought that (he) saw two couples, (and thought to himself) “I gave the wishing crystal to the first couple. I did the same with the second couple”.’

Example 25 illustrates a split Rheme progression pattern. In this fable, the god imagined and saw two couples (clauses 38.1- 38.2). Then, each couple is futher mentioned in the following clauses (clauses 38.3-38.4). The first couple is the couple to whom the god gave a wishing crystal (clause 38.3). The second couple comes to see the god and wishes the god to give them the wishing crystal since he gave it to the first couple (clause 38.4).

In addition, because the study sample size (10 simple folktales) was rather small, the occurrence of split Rheme progression was, accordingly, constrained. Therefore, a study of thematic progression of a larger corpus of folktales is required in order to confirm the results of this study.

An exploration of the distribution of the thematic progression patterns of the ten Thai Song Dam folktales revealed that constant and simple linear thematic progressions occurred frequently in all the selected folktales. This finding was expected given that Daneš (1974) refers to them as basic thematic patterns. Only two instances of the split Rheme progression were found and they were constrained by location in which they appeared in the text organization. The thematic progression patterns initiated in Thai Song Dam folktales clearly show the relationship between the method of folktale development and a key characteristic of narrative discourse — actor orientation.

Focusing on Generic Structure Potential or text structure, two thematic progression patterns (i.e., the constant and the simple linear thematic progressions) occurred with the greatest distribution. These two patterns were found in all the folktales under study (see Patpong 2011 for a discussion on Generic Structure Potential of Thai Song Dam folktales). However, there were some differences in terms of distribution in the course of text development. The constant thematic pattern dominated at all generic stages of the folktale organization in this study, whereas the simple linear pattern, including the extended linear thematic progression, occurred with greater frequency in the Sequent Event (or the Body stage of the folktales) than at other generic stages. The argument presented here is that this tendency is prompted and common in that one particular pattern is preferred over another. This may account for the conventionally schematic structure of the folktales analysed; folktales are oriented and developed by the protagonists (cf. Longacre, 1974).

7. Methods of folktale development

In the course of text analysis, Themes are selected, clause by clause, to indicate the development of the given text in question. This thematic development, manifested over the course of a text, is referred to as the text's method of development (cf. Fries 1981/1983). Therefore, the contribution of the method of development can be determined by an analysis of the thematic structure of independent clauses — Theme-Rheme structure.

Theme is selected in such a way that it relates the clause to the method by means of which the text is being developed. This well-constructed environment is referred to as the method of development of the text. Thus, in a particular discourse, there is an interrelationship between the thematic selection and the method of development. Theme is selected to reflect and establish the method of development (Fries, 1981/1983, 1993).

With respect to methods of text development, textual resources are concerned with enhancing resources as episode knowledge. Matthieesen (1995b: 39-40) characterizes narratives as being different from expository texts in that they depend on the process of ideational knowledge construction in instantial systems, and the textual resources of Theme play an important role in constructing narratives. The kind of knowledge that is constructed in a narrative is episodic. Episodic construction is achieved by two textual resources: temporal sequence and enhancing temporal motif. The temporal sequence is typically used to construct episodes. It is made up of sequences of events where the main characters take on different participant roles. These sequential events are made apparent by temporal marked Themes. The enhancing temporal motif is realized by conjunction resources, that is: 'and then, and then, and then'. Examples are provided below.

▪ Temporal sequence of episodic knowledge realized by temporal marked Themes (Temporal marked Themes are underlined.)

(26) Fable: Text 1 [14.1-14.3]

Theme		Rheme				
<u>muː^{6}niː^{6}</u>	*<u>ləːj^{6}</u>*	*naːm^{6}*	*jaŋ^{2}*	*maʔ^{4}*	*hæːŋ^{5}*	*lɔʔ^{4}*
today	conj.	river	ASP.: Ipfv.	NEG.	dry	Fp.
<u>puʔ^{4}niː^{6}</u>		*kɔːj^{4}*	*maː^{2}*	*maɯ^{3}*		
tomorrow		just	come	again		

puʔ⁴niː⁶	*paj¹*	*lɔʔ⁴*
tomorrow	come	Fp.

'Today the river is not dry. Tomorrow (we) will come back (here). Tomorrow (we) will go (there) again.'

(27) Fable: Text 1 [32.1-32.8]

Theme	Rheme				
ʔan¹dap³lœːk⁴	*soːŋ¹*	*kon²*	*phua¹*	*mia²*	*jaʔ³*
first (wish)	two	CL.	husband	wife	want
	laːj⁵	*kœːw⁵*	*wœːn¹*	*ŋəːn²*	*tɔːŋ²*
	get	crystal	ring	money	gold

(∅ = saʔ³maːj¹)	*kɔː⁵*	*jaʔ³*	*laːj⁵*	*ŋəːn²*	*tɔːŋ²*	*ʔaʔ³*	*naʔ³*
(∅ = they)	conj.	want	get	money	gold	Fp.	Fp.

jaːŋ³thiː⁴sɔːŋ¹	*kɔː⁵*	*jaʔ³*	*laːj⁵*	*huan²*
second (wish)	conj.	want	get	house

(∅ = saʔ³maːj¹)	*jaʔ³*	*laːj⁵*	*huan²*
(∅ = they)	want	get	house

(∅ = saʔ³maːj¹)	*jaʔ³*	*laːj⁵*	*thiː⁴juː³ʔaː¹saj¹*	*naʔ³*
(∅ = they)	want	get	residency	Fp.

lœʔ⁴kɔː⁵pɔː²	*jaːŋ³thiː⁴saːm¹*	*(∅ = saʔ³maːj¹)*	*kɔː⁵*	*nuk⁴*	
conj.	third (wish)	(∅ = they)	conj.	think of	
(∅ = saʔ³maːj¹)		*jaʔ³*	*laːj⁵*	*kon²cau⁶*	
(∅ = they)		want	get	servant	
(∅ = saʔ³maːj¹)		*jaʔ³*	*laːj⁵*	*kon²cau⁶*	*baːw³paj⁴*
(∅ = they)		want	get	servant	servant

'Firstly, the husband and the wife wanted to have money and gold. (They) wanted to have money and gold. Secondly, they wanted to have a house. (They) wanted to have a house. (They) wanted to have a house to live in. And thirdly, (they) thought, (they) wanted to have servants. (They) wanted to have servants.'

▪ Enhancing temporal motif realized by conjunctions (Conjunctions are underlined.)

(28) Joke and Anecdote: Text 2 [24.1-24.3]

Theme		Rheme					
saw²	*kɔː⁵*	*kwaː⁶*	*maːj⁶*	*lœːm¹*			
he	conj.	grab	stick	sharp			
(∅ = saw²)		*jɔːŋ⁴*	*paj¹*	*tœːŋ²*	*bɔː⁵*	*cek⁴*	*thaw⁵*
(∅ = he)		walk	go	stab	person	Chinese	old
lœːw⁶ (∅ = saw²)	*kɔː⁵*	*bœʔ³*	*ʔaw¹*	*maː²*	*huan²*		
conj. (∅ = he)	conj.	carry	bring	come	house		

'He then grabbed a sharp stick, (he) walked stealthily and stabbed that old Chinese man, and then (he) carried him back home.'

As shown in the above examples, the textual resources of the Theme selection are instantiated through temporal topical marked Themes construed in folktales together with temporal textual Themes. The marked topical Themes are all temporal and locative. The temporal topical Themes are marked and highly selected in Thai Song Dam folktales. They are commonly used to construct the folktales' episodes. Locative or spatial marked Themes are selected to set the scene for the folktales to be developed chronologically.

8. Conclusion

This paper has explored thematic selection and thematic development in a corpus of ten Thai Song Dam folktales. Thematic selection is textually analysed by lexicogrammatical realization of the Theme through the Theme system. Theme is viewed as a point of departure for the discourse and is seen as a means of systemic linkage to previous discourse. The method of thematic development of folktales is achieved by systemic patterns of thematic selection. The study also reveals the relationship between thematic selection and the method of text development.

This study also illustrates that thematic selection in discourse plays a significant role in organizing folktales. Themes are selected to guide the audience/addressee in interpretation of a clause as a message, by relating it to the method for developing the text. In this research, the method of development of Thai Song Dam folktales were thematically organized and developed through the semantic motif of time which is a means used to construct the folktale temporal sequence. It is given thematic status both as a textual Theme, expressed by conjunction resources, and as a marked topical Theme, expressed by circumstantial location of time. In folktales, the unmarked (topical) Themes are selected from among the participants or protagonists. These protagonists can be either main or associated characters, animate or inanimate. In terms of themaization, the Theme is often also the topic or subject of the course of text development. Each Theme indicates the point of the rhetorical development of the folktale text. In addition, different stages (Generic Structural Potential) within folktales may also vary in selecting favoured types of Theme, because they exploit different methods of text development.

Focusing on the distribution of thematic progressions, the study reveals that all the selected folktales show a frequent use of the constant Theme or iteration Theme progression (Theme → Theme progression) (63.41 percent) over the simple linear thematic progression (Rheme → Theme progression) (35.78 percent). There are some instances of extended linear thematic progression instantiated in the selected folktales. There are two instances of split Rheme progression found in this research (0.81 percent).

The Theme constant progression, the thematic progression with the highest frequency, was selected and organized around the protagonists doing and acting different activities, whereas the simple linear thematic progression was organized and based on new information located in the Rheme, which the content of the Theme of the second clause (Theme 2) derived from the content of the previous Rheme (Rheme 1).

The present study is based on a very modest sized folktale corpus. It represents only one subtype of Thai Song Dam folktale — simple tales. Nevertheless, the text analysis shows that even a small corpus embodies a number of interesting issues on thematic organization of folktale text type at the discourse level. For comparative purposes, this thematic progression approach is potentially applicable to other similar corpora in the Southeast Asian region. For further research, the current folktale corpus could be enlarged and investigated to include an exploration of methods of text development of complex Thai Song Dam folktales and/or a comparative study of methods of text development of oral and written Thai Song Dam folktales.

9. Acknowledgements

I would like to express my thankfulness to Professor Somsonge Burusphat for her inspiration on narrative discourse studies and her great leadership. I am greatly indebted to Professor Christian Matthiessen and Assistant Professor Kazuhiro Teruya for their continued support and encouragement.

10. References

Primary sources

Reyrai Suebsuk, Suksaman Yodkaew, and Rachadapan Settawat. (1980). *The verbal folk literature of Thai Song dam, Amphur Khaoyoy, Changwat Phetchaburi*. Phetchaburi: Phetchaburi Teacher Collage. (in Thai)

Sootawee Klinubon. (2009). *Folklore of Ban Thalo Muu 5 Yangyong Sub-district, Thayang District, Phetchaburi*. Independent Study Paper. Nakhonpathom: Faculty of Education. Silpakorn University. (in Thai)

Sunant Udonwech. (1985). An annotated collection of Thai Song Dam folktales Khaoyoy District, Phetchaburi Province. Phetchaburi: Phetchaburi Teacher Collage. (in Thai)

Secondary sources

Bäclund, Ingegerd. 1992. Theme in English telephone conversation. *Language Sciences*. Volume 14, 4: 545-565.

Bloor, Thomas & Meriel Bloor. 1995. *The functional analysis of English: a Hallidayan approach*. London: Edward Arnold.

Burusphat, Somsonge. 2012. Tone variation of Thai Song Dam. Journal of Southeast Asian Linguistics Society. *Journal of Southeast Asian Linguistics Society*. Volume 5: 32-48. http://www.jseals.org/seals21/burusphat11toneh.pdf

Burusphat, Somsonge Sumittra Suraratdecha, Sujaritlak Deepadung, Patama Patpong, Narong Ardsmiti, and Pichet Setapong. 2010. *Research Report: Ethnicity Language Culture and Ethnic Tourism Development*. Mahidol University (MS.) (in Thai)

Daneš, F. 1974. Functional sentence perspective and the organisation of the text. In F. Daneš (ed.) *Papers on Functional Sentence Perspective*. The Hague: Mouton. 106-128.

Francis, Gill. 1989. Thematic selection and distribution in written discourse. *Word*, 40, 1-2: 201-221.

Francis, Gill. 1990. Theme in the daily press. *Occasional Papers in Systemic Linguistics*. 3, 2: 199-244.

Fries, Peter. H. 1981/1983. On the status of theme in English: arguments from discourse. *Forum Linguisticum*. 6.1. [Reprinted in Janos S. Petöfi & Emel Sözer (eds.) 1983, Micro and Macro Connexity of Discourse. Hamburg: Buske (Papers in Text Linguistics 45)].

Fries, Peter H. 1993. Information flow in writing advertising. In James Alatis (ed.). *Language, Communication and Social Meaning*. Georgetown University Round Table on Language and Linguistics, 1992. Washington DC: Georgetown University Press.

Fries, Peter H. 1995a. A personal view of Theme. In M. Ghadessy (ed.), *Thematic development in English texts*. Amsterdam: John Benjamins.

Fries, Peter. H. 1995b. Patterns of information in initial position in English. In Fries, P.H. & M. Gregory (eds.) *Discourse in Society: functional perspectives*. Norwood, N.J.: Ablex (Meaning and Choice in Language: Studies for Michael Halliday).

Fries, Peter. H. 1995c. Themes, methods of development, and texts. In R. Hasan & P.H. Fries (eds.), *Current issues in linguistic theory*. Volume. 118. On Subject and Theme. A Discourse Functional Perspective. Amsterdam: John Benjamins. 317-359.

Ghadessy, Mohsen. (ed.) 1993. *Register analysis: theory and practice*. London: Pinter.

Ghadessy, Mohsen. 1995a. Thematic development and its relationship to registers and genres. In Mohsen Ghadessy (ed.), *Thematic Development in English Texts*. London: Pinter. 129-146.

Ghadessy, Mohsen (ed.) 1995b. *Thematic development in English texts*. London: Pinter.

Halliday, M.A.K. 1978. Language as social semiotic: the social interpretation of language and meaning. London: Edward Arnold.

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

Halliday, M. A. K. and Edward McDonald. 2004. Metafunctional profile of Chinese. In Caffarel, Alice, James R. Martin and Christian M.I.M. Matthiessen (eds.), *Language typology: a functional perspective*. Amsterdam: John Benjamins.

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

Halliday, M.A.K. and Ruquiya Hasan. 1985. *Language, context, and text: aspects of language in social-semiotic perspective*. Geelong, Victoria: Deakin University Press. (Language and Learning Series).

Hartmann, John F. 1980. A model for alignment of dialects in Southwestern Tai. *Journal of Siam Society*. 68: 72-86.

Hutchins, John. 1977. On the structure of scientific texts. *UEA Papers in Linguistics* 5, 18-39.

Jalilifar, Alireza. 2009. Thematic development in English and translated academic texts. *Journal of Language and Translations*. 10,1: 81-111.

Li, Fang Kuei. 1960. A tentative classifications of Tai dialects. *Culture in History Essay in Honor of Paul Radin*, edited by Stanley Diamond, 951-959. New York: Columbia University Press.

Li, Charles N., Thompson, Sandra A. 1976. Subject and Topic: A New Typology of Languages. In Li, Charles N. (ed.). *Subject and Topic*. New York/ San Francisco/London: Academic Press. 457-490.

Li, Charles N., Thompson, Sandra A. 1981. Madarin Chinese: A Functional Reference Grammar. California: University of California, Berkeley.

Long, Nie. 1991. An analysis of texts of different genres in terms of thematic selections and process types. Unpublished manuscript.

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

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

Martin, James R. 1985. Systemic functional linguistics and an understanding of written text. *Proceedings of the 1984 Working Conference on Language in Education*. Bartlett, Brendan and Carr, John., Brisbane: Brisbane College of Advanced Education.

Martin, James R. 1992. *English Text: System and Structure*. John Benjamins: Amsterdam.

Martin, James R. 1996. Transitivity in Tagalog: a functional interpretation of case.

In M. Berry, C. Bulter & R. Fawcett (eds.). Meaning and Choice in Language: Studies for Michael Halliday. Volume 2: Grammatical Structure a functional interpretation. Norwood, NJ: Ablex.

Martin, James R, Christian M.I.M. Matthiessen and Clare Painter. 2010. *Deploying Functional Grammar*. Beijing: The Commercial Press.

Matthiessen, Christian M.I.M. 1992. Interpreting the textual metafunction. In Martin Davies and Louise Ravelli (eds.), *Advances in systemic linguistics: recent theory and practice*. London: Pinter.

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

Matthiessen, Christian M.I.M. 1995b. THEME as an enabling resource in ideational 'knowledge' construction. In Mohsen Ghadessy (ed.), *Thematic Development in English Texts*. London: Pinter. 20-55.

Matthiessen, Christian M.I.M. 1999a. THEME and conjunction resources. MS. Department of Linguistics, Macquarie University.

Matthiessen, Christian M.I.M. 1999b. The textual metafunction: Theme and Conjunction. MS. Department of Linguistics, Macquarie University.

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

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

Matthiessen, Christian M.I.M, Kazuhiro Teruya, and Marvin Lam. 2010. *Key Terms in Systemic Functional Linguistics*. London: Continuum.

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

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

Patpong, Pattama. 2011. A Generic Structure Potential Analysis of Thai Song Dam Folktales. A paper presented at the 21st Annual Conference of Southeast Asian Linguistics Society at Kasetsart University, Thailand, May 11th-23rd 2011.

Suriya Ratanakul, Khunying. 1994. *Languages in Southeast Asian.. Part* I. Nakhonpathom: Institute of Language and Culture for Rural Development. (in Thai)

Teruya, Kazuhiro. 2006. A systemic functional grammar of Japanese. UK: Continuum.

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

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

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

Xiao, Qun. 1991. Toward thematic selection in different genre: fables and recipes. Unpublished manuscript.