

JSEALS

Journal of the Southeast Asian Linguistics Society

Managing Editor:
Paul Sidwell (Pacific Linguistics, Canberra)

Editorial Advisory Board:
Mark Alves (USA)
George Bedell (Thailand)
Gerard Diffloth (Cambodia)
Marlys Macken (USA)
Brian Migliazza (USA)
Keralapura Nagaraja (India)
Amara Prasithrathsint (Thailand)
Martha Ratliff (USA)
Sophana Srichampa (Thailand)
Uri Tadmor (Indonesia)
Justin Watkins (UK)

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. It is an electronic journal, distributed freely by Pacific Linguistics (www.pacling.com) and the JSEALS website (sealang.org). Hard copies may also be ordered (availability subject to reasonable demand).

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. Although we expect in practice that most JSEALS articles will have been presented and discussed at the SEALS conference, submission is open to all regardless of their participation in SEALS meetings. Papers are expected to be written in English.

Each paper is reviewed by at least two scholars, usually a member of the Advisory Board and one or more independent readers. Reviewers are volunteers, and we are grateful for their assistance in ensuring the quality of the publication. As an additional service we also admit data papers, reports and notes, subject to internal review.

JSEALS is published annually. Papers should be submitted to the Managing Editor, electronically (paul.sidwell@anu.edu.au or paulsidwell@yahoo.com) by December 31st for inclusion in the follow year's issue.

Contents

| Editorial | iii |
|--|-----|
| Papers | |
| Contact-Induced Change? Register In Three Cham Dialects Marc Brunelle | 1 |
| A Minimalist Approach To Amis Voice Markers Yi-Ting Chen | 23 |
| Identifying Prepositions In Thai Kitima Indrambarya | 37 |
| Deixis And Information Structure In Mon: The Multifunctional Particle kòh <i>Mathias Jenny</i> | 53 |
| Tone Neutralization Due To Consonants In Mulao Seunghun J. Lee | 73 |
| Eventivity And Stativity In Thai Verbs Natsuki Matsui | 85 |
| Vietnamese Passive Sentences From A Typological Perspective Nguyễn Hồng Cổn | 107 |
| Proto-Southwestern-Tai Revised: A New Reconstruction Pittayawat Pittayaporn | 121 |
| Complementizers And Verb Classification In Thai Amara Prasithrathsint | 145 |
| Village Names In Chiang Saen: An Ethnolinguistic Study Of Place Names In Thai Sorabud Rungrojsuwan | 161 |
| Arrival Expressions In Thai Kiyoko Takahashi | 175 |
| Data Papers / Reports / Notes | |
| A Preliminary Description Of Ende Phonology Bradley McDonnell | 195 |

| Aspect Marking And Modality In Child Vietnamese Jennie Tran | 227 |
|--|-----|
| A Note On Interference Of Thai Reduplication On Vietnamese Spoken In | |
| Udon Thani Province Of Thailand | |
| Songgot Paanchiangwong | 241 |

Editorial

Welcome to JSEALS Volume 2, the second issue of the *Journal of the Southeast Asian Linguistics Society*. From the inception of the Society in 1991, until 2006, papers presented at the annual SEALS meetings were published as proceedings volumes, first by the Arizona State University, and later by Pacific Linguistics and at the Australian National University. From 2008 onwards JSEALS has become the principal organ of the Southeast Asian Linguistics Society.

The first issue of JSEALS was released in March 2009. It had been our intention to have it out in 2008, but the transition from straightforward proceedings to a peer reveiwed journal proved to be quite a challenge, and it took nearly 2 years to make it happen. Happily we are now on track, and Volume 2 has come out in September 2009, a little later than a year after the 18th SEALS meeting, which was held at the Universiti Kebangsaan Malaysia (Bangi, 21-22 May 2008). Many of the papers appearing in this volume were first presented at that meeting.

I would also like to thank Mark Alves (Associate Professor, ESL/EAP and Chinese, Montgomery College), who so ably contributed to the first issue as Main Editor. Mark has now stepped back from that role, and I look forward to other SEALS supporters coming forward to support this important work. Thanks are also due to the Editorial Board members and others who have been reviewing papers and otherwise working to ensure the quality of the journal. The most important lesson that has emerged from the JSEALS experience so far is an appreciation of the extent to which it is a collective effort. To manage, assess, revise and publish a dozen papers requires academic input from as many as thirty or more scholars, all contributing their own time to the project. Thank you.

JSEALS will also continue to improve. With the 2010 issue we expect to include dates received and dates accepted for each paper. The fact that this has not been done so far reflects confusion in record keeping which was part of the steep learning curve associated with the birth of the journal. And we look forward to receiving suggestions and constructive critisism to further improve the journal. In particular, I would like to invite submission of book reviews, conference reports, and review articles. The electronic distribution of JSEALS makes it a widely accessible vehicle for scholarship that reaches way beyond printed journals. As this fact, combined with the high standard of papers, becomes more widely appreciated, I am confident that JSEALS will only grow from strength to strength.

Paul Sidwell (Managing Editor) September 2009

CONTACT-INDUCED CHANGE? REGISTER IN THREE CHAM DIALECTS¹

Marc Brunelle

University of Ottawa <marc.brunelle@uottawa.ca>

0 Abstract

The Chamic language family is often presented as a test case for contact-induced language change (Thurgood 1996; 1999). The register systems of Cham dialects have been claimed to evolve in the direction of neighbouring languages, becoming increasingly tonal when in contact with Vietnamese (Eastern Cham) or developing (or retaining) a more typical register system when in contact with non-tonal languages (i.e. Western Cham with Khmer). However, since instrumental evidence on the phonetic realization of register in Western Cham is limited (only Edmondson and Gregerson 1993), these proposals are partly speculative.

This paper aims at filling this void. A laryngographic and acoustic study of register in three Cham dialects was undertaken to uncover the similarities and differences between them. This study concludes that f0 (pitch) is the dominant property of register in all three dialects, shedding doubt on mechanistic models of contact and change. We argue for a low-level phonetic influence of Vietnamese on Eastern Cham register and for a more complex scenario of contact-induced sound change.

1 Cham register

Cham is an Austronesian language spoken in Mainland Southeast Asia. Because Cham and its sister languages have been in close contact with their Mon-Khmer neighbours for centuries, they have undergone a significant amount of linguistic restructuring (Lee 1974; Alieva 1984; Thurgood 1996; 1999). In this paper, we explore the phonetic realization of phonemic register (defined below) in three Cham dialects spoken in Vietnam and Cambodia. The goal of this study is to determine if contact with neighbouring languages can explain some of the differences found between dialects.

1.1 The development of Cham register

Register is a common type of phonological contrast in Mainland Southeast Asian languages (Henderson 1952; Gregerson 1973; Matisoff 1973; Huffman 1976; Pittman 1978; Ferlus 1979; Diffloth 1982; Jones 1986). It arises from the neutralization of voicing in onset stops and from the phonologization of phonetic properties originally associated

Marc Brunelle. 2009. Contact-Induced Change? Register in Three Cham Dialects. *Journal of the Southeast Asian Linguistics Society* 2:1-22. Copyright vested in the author.

I would like to thank Phú Văn Hẳn and Abubakar (Lebke) for their help in putting together the wordlists and in recruiting participants and Nguyễn Văn Huệ, for granting me the authorization to work at the University of Social Sciences and Humanities of Hồ Chí Minh City. Special thanks to Alexis Michaud for allowing me to use his MATLAB scripts, to Ricardo Tabone for his programming skills and to Sophia Stevenson and Marie-Claude Tremblay for helping me process the data. All errors are mine. This work was partly funded by a grant from the Social Science Research Council of Canada.

with voicing. Phonologically, register can be a property of the onset, the rime or the entire syllable, depending on the language. Phonetically, however, these same phonetic properties are realized on the rime (except durational cues, which are also present in onsets). They are listed in (1), where the low register derives from former voiced stops and the high register derives from former voiceless stops.

(1) Low register
Lower pitch
Laxer

High register
Higher pitch
Tenser

Higher vowels
Longer vowels
Shorter vowels

Aspirated onset stops (or longer VOT) Plain onset stops (or shorter VOT)

It is important to note that not all of these phonetic properties are found in all register languages. Languages typically retain a subset of the characteristic listed in (1). For example, despite evidence that all properties are present to some extent in Eastern Cham at a fine-grained phonetic level (Brunelle 2005b; 2006), this dialect mostly contrasts registers through pitch and voice quality.

Although it is difficult to date the development of register, the fact that most lowland Chamic languages have register systems suggests that it took place before the break-up of Coastal Chamic, a few centuries ago (Lee 1966; Thurgood 1993; 1996; 1999). A discussion of the formation of register systems in Chamic in general is beyond the scope of this paper, but it must be pointed out that in these systems the register contrast was extended to sonorants. In a nutshell, register spread from left to right through sonorant-initial syllables, as illustrated with an Eastern Cham example in (2). In Eastern Cham, sonorants initially received a default high register. Note that the low register is marked with a subscript circle, under the onset.

A variant of this pattern in found in Western Cham, where sonorants were originally assigned a low register, but underwent the same register spreading process (Friberg and Hor 1977; Thurgood 1996; 1999).

In section 1.2, we will review previous descriptions of the phonetic properties of register in Cham dialects. Contact with neighboring languages and its influence on the phonetic development of register will also be discussed.

1.2 Previous descriptions of Cham dialects

Cham is usually divided into two dialects: Eastern Cham and Western Cham. Eastern Cham (or Phan Rang Cham) is spoken by about 100,000 people² in the south-central Vietnamese provinces of Ninh Thuận and Bình Thuận, which were parts of Champa, the original Cham homeland (General Statistics Office of Vietnam 1999). Western Cham, on

² The census figure of 132,873 people also includes Western Cham and Haroi people, which are treated as Cham in the official Vietnamese ethnic classification.

the other hand, is spoken by about 300,000³ speakers in Cambodia and the Vietnamese Mekong delta (Gordon 2005). They are the descendants of Cham refugees who fled to Cambodia as Champa was conquered by the Vietnamese from the 15th to the 19th century. Although the term "Western Cham" suggests a certain homogeneity, Western Cham communities are not located in a continuous geographical area, as we can see in Figure 1. There are pockets of speakers all over Cambodia and in south-western Vietnam, close to the Cambodian border. Dialectal variation is not explicitly mentioned by linguists (with the exception of Baumgartner 1998), but Western Cham speakers themselves recognize that there are significant differences between regions.

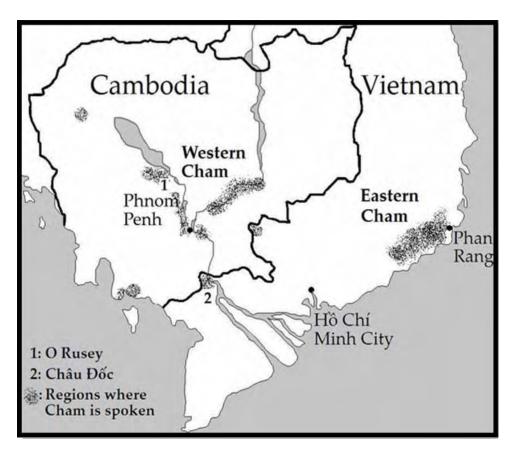


Figure 1: The geographic distribution of Cham dialects

1.2.1 Eastern Cham

Acoustic analyses of Eastern Cham have shown that its register contrast is largely based on pitch and voice quality (Phú *et al.* 1992; Brunelle 2005b; 2006). Minor differences in vowel quality and duration of VOT and rhymes are also present, but they do not seem to play a role in perception (Brunelle 2005b; 2006). The central role of pitch in the Eastern Cham register system and the fact that registers have coda-conditioned pitch allophones in

A figure of 220,000 speakers in 1992 is attributed to an unspecified "government figure" in *Ethnologue* (http://www.ethnologue.com). This estimation, which seems reasonable, is repeated in most sources without verification.

the speech of some speakers, as summarized in (3), have led a number of authors to claim that this language is either tonal or incipiently tonal.

(3) An example of a description of the pitch allophones of the two registers in Eastern Cham (Phú *et al.* 1992)

| pa | high pitch | 'to cross' |
|-----|---------------------------|--------------------|
| pa? | high pitch | 'four' |
| рa | low pitch, breathy | 'to carry' |
| pa? | low rising pitch, breathy | 'to wander (rare)' |

However, what different authors mean by "tone" is highly variable. The first modern descriptions often designated the register allophones as 'tones' or 'pitches' without claiming that they were phonemic (Blood 1967; Moussay 1971; Hoàng 1987). These authors recognized (sometimes implicitly) that there were only two distinctive registers contrasting in pitch, but that each of them had predictable variants conditioned by codas. Another group of authors has more recently adopted the view that the codas are either lost, in the process of being dropped, or reinterpreted as tonal elements, which comes down to claiming that the register allophones have become contrastive tones (Phú et al. 1992; Thurgood 1993; 1996; 1999). We have since gathered evidence that codas are still present and that although they are partly debuccalized, they preserve their patterns of contrast and show no sign of deletion (Brunelle 2005b; 2006). Moreover, the pitch curves of codaconditioned register allophones vary across speakers, making them unlikely contrastive elements (Brunelle 2005b; 2006). Finally, evidence from reduplication and a word game strongly suggests that register can be phonologically analysed as a segmental property of onset consonants (Brunelle 2005a; b). Until we have additional phonological evidence, and despite the central role of pitch in its register contrast and the fact that all the basic ingredients needed for tonogenesis are present, we must conclude that there is at the moment no evidence that the dialects of Eastern Cham that have been studied so far have phonemic tone or are incipiently tonal.

1.2.2 Western Cham (Châu Đốc)

There is only one description of register in a dialect of Western Cham spoken in Vietnam (Friberg and Hor 1977)⁴. Impressionistically, this dialect, spoken in Châu Đốc, seems similar to the Western Cham dialects spoken elsewhere in Vietnam (in Tây Ninh and in a small settlement near Xuân Lộc, in Đồng Nai province).

In this description, Friberg and Hor followed Gregerson (1973) in claiming that tongue root position is responsible for the register contrast in the dialect that they studied. They hypothesized that all low register consonants are produced with an advanced tongue root [+TRA], while other consonants are produced with a neutral tongue root position. Since the claim that the tongue-root is the main articulator involved in register production has since been abandoned, we will not discuss the technicalities of Friberg and Hor's featural analysis. However, their acoustic description of the effect of register on the vowels

⁴ Friberg and Hor are not explicitly saying that they investigated the Châu Đốc dialect. This information in mentioned in Headley (1991).

of Châu Đốc Cham, which is reproduced in (4), is still valuable. Note that short vowels seem to have been omitted from their description.

(4) Vowel system of Châu Đốc Cham (Friberg and Hor 1977)⁵

| Register: | High Low | | High Low | | High Low | | |
|-----------|----------|---------|-----------------------------|----------------|---------------------|--------------------|--|
| | į | i | ã | i | \mathbf{u}_{τ} | u | |
| | ę | e^{i} | $oldsymbol{\check{\Delta}}$ | 5 - | õ | $oldsymbol{o}^{u}$ | |
| | æ | e_{-} | a | a | Э | O_ | |

To sum up the content of (4), the Châu Đốc variety of Western Cham contrasts register in three ways. First, vowels always have a laxer ('breathier' is used on p.3) vowel quality in the low than in the high register, but the exact realization of the phonation contrast is relative and depends on vowel height. In high vowels, the high register is tense and the low register is modal. In mid-vowels, the high register is tense and the low register is lax. Lastly, in low vowels, the high register is modal and the low register is lax. The second phonetic element used to contrast registers is vowel quality. We see in (4) that all vowels except /i/ and /a/ have slight differences in height or in offgliding. Low register vowels are systematically higher than high register vowels. Finally, the low "register also displays a lower pitch in analogous vowels" (Friberg and Hor 1977, p.31).

1.2.3 Western Cham (Cambodia)

There are two descriptions of the register systems of Western Cham dialects spoken in Cambodia. The first one analyses the speech of a young male speaker from Kompong Thom (Headley 1991). The other one is based on recordings of a male speaker by Neil Baumgartner (Edmondson and Gregerson 1993). According to Kenneth Gregerson, the recordings are from Cambodia, but we have not been able to obtain more information on the speaker's dialect.

In his paper, Headley did not systematically explore the phonetic differences between registers, but he described the low register vowels (his *high* register) as "higher (in terms of tongue height), rather 'breathy' in voice quality, and associated with a low pitch". High register vowels on the other hand (his *low* register) are "lower in tongue height, often with lower on-glides, rather 'clear' in voice quality, and associated with higher pitch." (Headley 1991, p.106). This is rather similar to Friberg and Hor (1977)'s description of Châu Đốc Cham, although Headley's examples suggest that the exact vowel qualities are slightly different. An important observation here is that the dialect studied by Headley seems to have a vowel quality register contrast even for the vowel /i/.

Edmondson and Gregerson (1993) obtained similar results instrumentally. Although they could not find a clear pattern for voice quality (measured through intensity), perhaps because of limited data, they found that the high register has a slightly higher fo than the low register. However, the most central cue for register contrast in the dialect they studied seems to be vowel quality: low register vowels always have a lower F1 (higher

The choice of the diacritics for breathy phonation () and creaky phonation () was made for convenience only. Friberg and Hor (1977) describe the two types of phonation as lax and tense, respectively.

quality) than their high register counterparts and they tend to glide more (although results on gliding are not reported). No difference in F2 emerges from the data.

To sum up previous descriptions, all Cham dialects seem to base their register contrast on the same set of acoustic properties (pitch, voice quality and vowel quality). However, according to previous work, Eastern Cham tends to emphasize the use of pitch, while Western Cham dialects seem to focus on vowel quality. The fact that the different dialects have taken different phonetic directions has been attributed to contact with neighbouring languages (most explicitly by Thurgood 1996; 1999). We will see below that our experimental results diverge from previous phonetic descriptions of Western Cham (which could be due to the fact that different dialects/sociolects were studied). As for the claim about the role of contact in register development, it will be reviewed in detail in the next section.

1.3 Contact

Despite a considerable body of literature on contact-induced change, there is little detailed work on phonological or phonetic convergence, possibly because such work requires experimental data. For example, none of the papers in a recent book on contact linguistics devotes more than a page to phonological contact (Aikhenvald and Dixon 2006). The two types of contact-induced changes that are usually referred to are the borrowing of foreign phonemes along with lexical items and the phonemicization of allophones under the influence of a language in which they are contrastive (Thomason and Kaufman 1988). We will see in section 4 that Cham data argues for an additional type of contact-induced sound change.

Mon-Khmer has affected the phonological development of Cham register in several ways (Alieva 1984; Thurgood 1996; 1999), but we will focus on two changes that directly affected registers. The first one is registrogenesis itself, i.e. the initial development of register, that is often claimed to have been favoured or caused by contact (Friberg and Hor 1977; Thurgood 1996; 1999). This assumes that Cham registrogenesis was triggered by widespread bilingualism or a massive language shift from Mon-Khmer to Cham (Thurgood 2002). Although the exact mechanisms are unclear, the fact that registrogenesis itself is contact-induced is not very controversial (Matisoff 2001).

The second way in which Mon-Khmer might have influenced the evolution of Cham register is an effect on the selection of its phonetic correlates (Thurgood 1996; 1999). The implicit assumption of this model is that bilingual speakers are likely to transfer some properties of their second language into their first language, or to emphasize preexisting properties of their first language that are important in their second language. For example, Vietnamese is a tonal language that uses fine-grained pitch distinctions. Since most Eastern Cham speakers are very fluent in this language and therefore have a high degree of control over pitch production and perception, they are likely to increase the functional role of pitch in the register contrast of their native language. To quote Thurgood, "The Phan Rang [Eastern] Cham transition from a Western Cham style register system towards a Vietnamese style tone system represents more of an evolution that a revolution – the actual changes look to be little more than slight shifts in emphasis" (Thurgood 1999, p.196). We have argued, contra Thurgood, that Eastern Cham has a phonologically segmental, though phonetically pitch-based, register system rather than a full-fledged "Vietnamese style" tone system (Brunelle 2005a; b; 2006), but his basic incremental scenario remains sound.

The original impetus for this project is that during exploratory trips to the Vietnamese Mekong delta and Cambodia, we noticed that the register systems of the Cham dialects spoken in these areas appear to be based on pitch, just like the register system of Eastern Cham. If this is true, it would be problematic for the contact and change scenario just described. First, if Cham dialects that are not in contact with tone languages use pitch two distinguish registers, it becomes superfluous to attribute the central role of pitch in the Eastern Cham register system to contact with Vietnamese. Second, if bilingualism in Vietnamese favours an increasing role of pitch in the Eastern Cham register system, bilingualism in Khmer, a language with a rich array of vowel qualities, should lead to more emphasis on vowel quality in the register contrast of Cambodian Cham.

A phonetic study of the register systems of three Cham dialects that are in contact with different languages was therefore undertaken. It is described in section 2 and 3. Our results will lead us to reconsider some basic tenets of the language contact and change scenario described above in section 4.

2. Methods

In section 2.1, we will describe the dialects under investigation and the subjects who participated in the experiment. In section 2.2, we will briefly explain how we constituted the wordlist used for making recordings. Section 2.3 is a description of the experimental setup and section 2.4 summarizes the data analysis procedure.

2.1 Dialects investigated and participants

To have a good picture of the potentially very diverse realization of register in Cham, we looked at three Cham dialects. The first one, Eastern Cham, is spoken in south-central Vietnam and is in intensive contact with Vietnamese. The second one is a Western Cham dialect spoken around Châu Đốc, in the Vietnamese Mekong delta (number 2 in Figure 1). It is in contact with Vietnamese and has been in contact with Khmer until recently. The third one is a Western Cham dialect spoken in O Rusey village, in Kompong Chhnang province, Cambodia (number 1 in Figure 2). It is in contact with Khmer, but not with Vietnamese. Eastern Cham and Châu Đốc Cham speakers were recorded at the University of Social Sciences and Humanities in Hồ Chí Minh City, while Kompong Chhnang Cham speakers were recorded in their village.

Participants were recruited through contacts in the communities. Ten adult speakers (five men and five women) of each dialect, all younger than 55 years old, were recorded. Over all, our Eastern Cham and Châu Đốc Cham speakers are probably more educated and bilingual than average. Kompong Chhnang Cham speakers are representative of their village.

2.2 Wordlist

Subjects were asked to read a wordlist composed of all possible combinations of a set of onsets, vowels and codas susceptible to affect the phonetic realization of register. Onsets included labial, dental and velar plain stops and aspirated stops. Vowels in the three corners of the phonetic vowel space (/i/, /i//i/, /u/, /a/, /a/) were selected. Finally, we chose words with open syllables and with laryngeal codas, the type of codas which is supposed to have the most effect on pitch (/-i/2/, /-h/, /-2/). When no real open syllable word was available in the language, a word with a coda sonorant was included, based on the assumption that sonorants have a minimal effect on the acoustic properties of the vowels

preceding them. These three types of factors were combined to yield a list of monosyllables over which the two registers were superimposed. We then went through the entire list with speakers of each dialect and selected all the monosyllables that correspond to lexical words in the dialects under study. When monosyllables did not correspond to real words, we tried to find real sesquisyllabic words with a final stressed syllable matching the required monosyllable⁶. Gaps were not filled with nonce words. As a result, we obtained a wordlist composed of real words only, which allows us to avoid uncontrolled nonce word effects.

Besides this basic wordlist, a few minimal pairs of monosyllables contrasting in register and starting with the dental onset sonorants /n/ and /r/ were also included. Our goal was to determine if register is realized identically in these words and in monosyllabic words with initial onset stops (a few minor differences were found but will not be reported here). These words were easy to find for Eastern Cham, which has extended the register contrast to sonorant onsets (Brunelle 2005b; a), but only a few minimal pairs were found in Châu Đốc Cham and Kompong Chhanng Cham, which have a very restricted register contrast in words with onset sonorants.

In total, 70 words were included in the Eastern Cham and Châu Đốc Cham wordlists and 64 in the Kompong Chhnang Cham wordlist. Words differ a little across dialects, although there is a large amount of overlap because of common lexical items. All the words were embedded in a frame sentence meaning "I say ____ for him to hear". The frame sentence varies slightly between dialects:

(5) Eastern Cham:

/l̪āʔ dom kʰān ___ ka nu pāŋ/
I say word for he hear

Châu Đốc Cham:

/hulin mejaj ___ ka nu pāŋ/
I say for he hear

Kompong Chhnang Cham:

/lən mɨjej ___ prej kit pāŋ/
I say for he hear

2.3 Experimental set-up

Subjects had to read the randomized wordlist 10 times. The wordlist was presented to Eastern Cham and Châu Đốc Cham speakers in a Romanized script inspired form the Vietnamese alphabet. This decision was taken because only a handful of Cham speakers are able to read the Cham script fluently (Brunelle 2008). To ensure that the subjects did not confuse words while reading this ad hoc romanization, the wordlist was explained to them in detail before making the recordings and a Vietnamese translation of each target word was written in parenthesis before each line. Since Kompong Chhnang Cham subjects are not familiar with the latin alphabet and since there is no commonly accepted way of writing Cham in Khmer script, another strategy was used in Kompong Chhnang. The subjects were given the target word orally by a native assistant and had to embed it in the frame sentence before uttering it. Subjects were instructed not to pronounce words that they were not familiar with in order to avoid spurious renderings and nonce word effects.

Note that colloquial Eastern Cham has now become almost entirely monosyllabic (Brunelle 2005b; 2007; 2009). Formal sesquisyllabic words were thus excluded in this dialect to avoid unnatural pronounciations.

There is therefore a certain amount of variation in the number of tokens produced by each subject, with a tendency for educated speakers to skip fewer words.

The laryngeal activity of the subjects was recorded with an EG2-PC electroglottograph (EGG) while they were reading the wordlist. Their speech was also recorded acoustically via a microphone plugged into the EGG box. EGG (or laryngography) consists in placing electrodes on each side of the speaker's neck, at the level of the larynx box, and to drive a low level electrical current (less than one volt) between the two electrodes. Electrical conductivity will vary according to the degree of opening of the glottis (closed glottis: good conductivity; open glottis: resistance), which allows the researcher to measure the patterns of opening and closing of the vocal folds in real time and therefore to infer their rate of vibration (f0, the phonetic correlate of pitch) and the degree of closure of the glottis (open quotient, the articulatory correlate of voice quality). The EG2-PC laryngograph also allows the researcher to measure the vertical position of the larynx. However, these results are beyond the scope of this paper and will not be discussed here.

In order to obtain perfectly synchronized signals, the outputs of the EGG and of the microphone were channelled through a National Instrument data acquisition box. The signal was then acquired on a laptop computer at a sampling rate of 22,050 Hz with a MATLAB script.

2.4 Data processing and result extraction

The acoustic signal acquired with MATLAB was converted into a wav sound file that was used to determine the boundaries of the segments to be measured. The beginning and endpoints of the onsets, vowels and codas (if applicable) of each target word in the wordlist were labelled in Praat textgrids. Since the acoustic indicators investigated here (f0, open quotient, formants) can only be realized on voiced sonorants and since there are relatively few words with onset sonorants in the wordlist, only measurements made on vowels will be reported in section 3.

A MATLAB routine combining Alexis Michaud's MATLAB-based EGG analysis scripts and a Praat script was then used to extract the f0 (pitch), open quotient (% of the glottal cycle during which the glottis is open), the frequency of the first and second formants and the duration of the segments under investigation. Target segments were measured at five sampling points: their onsets, offsets and three intermediate points. For measurements made from individual glottal cycles (f0 and open quotient), the cycle that was measured was the one whose peak was the closest to the sampling point, while being within the same segment.

2.4.1 F0 (pitch)

The number of glottal cycles per second (f0), which is the articulatory correlate of pitch, was measured directly from the EGG signal (a second was actually divided by the duration of the target cycle). The major advantage of this type of measurement is that we can avoid inaccuracies that are common when pitch trackers are attempting to locate glottal cycles on waveforms that include the acoustic effect of the supraglottal cavity. A few suspicious measurements were removed from the dataset by excluding values that were more than two standard deviations away from the mean (calculated independently for each speaker). Since

f0 and open quotient were calculated from the same EGG signal, the open quotient values corresponding to excluded f0 values were also excluded.

2.4.2 Open quotient (voice quality)

Open quotient corresponds to the proportion of the glottal cycle during which the glottis is open. The open quotient is correlated with voice quality. A large open quotient corresponds to a breathier voice quality while a small open quotient corresponds to a creaky, tight or pressed voice quality (Ní Chasaide and Gobl 1997; Mazaudon and Michaud 2008 for good overviews).

The open quotient was extracted from the EGG results with a routine based on Alexis Michaud's MATLAB scripts (http://voiceresearch.free.fr/egg/). The routine smoothes the EGG signal, gets its derivative and calculates the open phase of the glottal cycle from the maximum peaks and valleys of the derivative. Once again, a few suspicious measurements were removed from the dataset by excluding values that were more than two standard deviations away from the mean (calculated independently for each speaker). Since f0 and open quotient were calculated from the same EGG signal, the f0 values corresponding to excluded open quotient values were also excluded.

2.4.3 Formant frequencies

The first and second formants, which are the acoustic correlates of vowel height and frontness-backness respectively, were calculated with a Praat script from the acoustic signal. However, since this signal is noisy and faint for most speakers, the Praat formant tracker returned numerous suspicious values. These values were excluded in the following way: Mean F1 and F2 values were calculated for each vowel quality in the speech of each subject. All values that were more than two standard deviations away for the mean F1 or F2 for a given vowel quality were then excluded.

2.4.4 Duration

As the speech rate of speakers tends to vary between utterances of the wordlist and even during the same wordlist, it is necessary to normalize duration values before comparing them. A common way to normalize duration is to select a fixed part of the frame sentence as a benchmark for controlling speech rate. Unfortunately, the frame sentence that we used for this experiment is not suitable for this type of normalization. Speakers know what the target word is and can lengthen it independently of the rest of the frame sentence. For this reason, and since duration has never been claimed to play a central role in the Cham register contrast, duration results will not be reported.

2.4.5 Normalization

In order to facilitate the comparison and visual representation of data from many speakers and dialects, all acoustic indicators were normalized using a Z-score method. This method consists in rescaling all the values for a given indicator so that they have an average of 0 and a standard deviation of 1. Z-scores for open quotient and f0 were computed over all of the measurements made in final syllable vowels. Z-scores for F1 and F2 were computed similarly, but independently for each vowel.

3. Results

In this section, results on the role of f0 (pitch), open quotient (voice quality) and F1 and F2 (vowel quality) in the register contrast will be reported for the three dialects under study.

3.1 f0 (pitch)

Figure 1 illustrates the mean normalized f0 of vowels in the high and the low registers. It is calculated by averaging the normalized f0 of all the final syllable vowels in the dataset, regardless of their onsets, codas or vowel qualities. Each line corresponds to a speaker. Full and dotted lines represent the high and low registers, respectively. Mean values were computed over a number of words ranging from 206 to 431 per register and per speaker⁷ (mean number of tokens: High register = 314,4, Low register = 253,6).

The results presented in Figure 1 illustrate commonalities and differences between dialects. First, we see that the mean f0s of the two registers seem to be relatively distinct in all dialects, although they tend to overlap at the end of Kompong Chhnang Cham vowels. However, we see that the two registers pattern much more distinctly in Eastern Cham than in the other two dialects and that Kompong Chhnang Cham has the least salient f0 difference between registers. The mean values plotted in Figure 1 obviously hide variation. Table 1 shows the standard deviations for each dialect and each register.

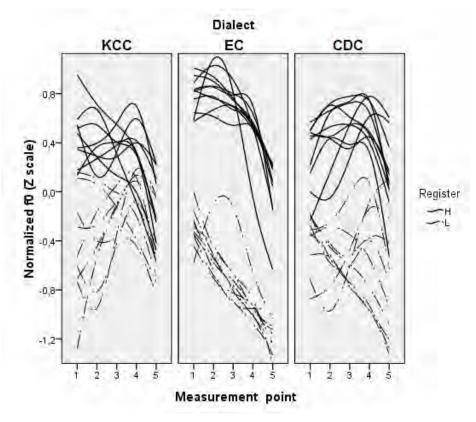


Figure 1: Mean normalized f0 in the two registers, for the three dialects (KCC: Kompong Chhnang Cham, EC: Eastern Cham, CDC: Châu Đốc Cham)

With the exception of one female Châu Đốc Cham speaker for which half of the recorded files are corrupted. The data that could be recovered includes 100 high register tokens and 81 low register tokens.

The standard deviations in Table 1 reveal that the f0 of the two registers tends to vary less in Eastern Cham than in the other dialects. This smaller standard deviation, combined with the relatively larger gap between registers, means that there is less overlap between registers in Eastern Cham than in Kompong Chhnang Cham and Châu Đốc Cham.

Table 1: Standard deviations in normalized f0, by dialect and register

| Register | KCC | EC | CDC |
|----------|------|------|------|
| High | 0,94 | 0,78 | 0,87 |
| Low | 0,98 | 0,73 | 0,92 |

A final observation must be made about the general shape of the f0 curves in the three dialects. In Eastern Cham, the two registers are distinct from the beginning of the vowel and their f0 curves fall in parallel, while remaining distinct. Most Châu Đốc Cham speakers follow a similar pattern, but a few seem to have rising rather than falling patterns (which could be due to the intonation with which they chose to read the wordlist). However, Kompong Chhnang Cham is qualitatively different in that most speakers start their vowels with a clear f0 contrast between registers and then let the two registers converge by, first by raising the f0 of the low register during the middle portion of the vowel and then by dropping the pitch of the two registers.

To sum up, pitch could be used to distinguish registers in all three dialects, but the pitch difference between registers is much more salient in Eastern Cham. Further, while all Eastern Cham speakers seem to conform to the same pattern of pitch contrast, there is a significant amount of variation in the other dialects.

3.2 Open quotient (voice quality)

Our results for open quotient are much less clear-cut than those for f0. They are reported in Figure 2, which illustrates the mean normalized open quotient of vowels in the high and the low registers, respectively. This mean normalized open quotient is calculated by averaging the open quotient of all the final syllable vowels in the dataset, regardless of their onsets, codas or vowel qualities. Mean open quotient values were computed over the same tokens as f0, as these values are paired (see 2.4.1 and 2.4.2).

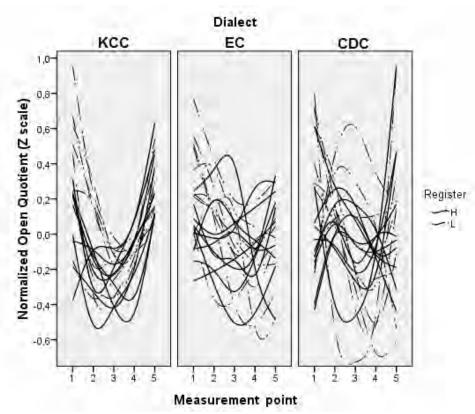


Figure 2: Mean normalized open quotient in the two registers, for the three dialects (KCC: Kompong Chhnang Cham, EC: Eastern Cham, CDC: Châu Đốc Cham)

At first sight, no clear pattern of contrast between registers emerges from Figure 2. However, a closer look reveals a few regularities. First, Kompong Chhnang Cham speakers tend to have a more open glottis at the beginning of low register vowels. There are even four speakers who show a relatively robust contrast at points 1 and 2. Eastern Cham speakers are far less regular, but once again, the low register tends to have a larger open quotient at the very onset of the vowel. Finally, the Châu Đốc Cham results, which seem hopeless at first glance, reveal an interesting tendency. Eight speakers out of 10 have a sharply rising open quotient at the onset of the high register and a dramatically falling open quotient at the onset of low register vowels.

Once again, the open quotient curves plotted in Figure 2 do not show all the actual variation found in the dataset. Standard deviations for registers and dialects gravitate around a value of 1 (between 0,96 and 1,05), which means that the amount of overlap between registers is important. This is not surprising since factors like onsets and codas, which are known to affect open quotient, have not been controlled for, but in any case, we can see that none of the dialects makes any special attempt to keep open quotient values overly distinct to mark the register contrast.

Surprisingly, high and low vowels have the same type of open quotient differences in all dialects, and the correlation between vowel height and phonation reported by Friberg and Hor (1977) is not found in the Western and Cambodian Cham dialects. High register vowels systematically have a smaller open quotient than low register vowels, but there is no evidence of a shift from a tense~modal contrast to a modal~breathy as we move from the high vowels to the low vowels.

In perceptual terms, i.e. to rephrase our results in terms of voice quality, it seems that the initial portion of low register vowels is very breathy in Kompong Chhnang Cham, whereas their high register counterparts are modal. Both registers then seem to become modal before ending on a breathier voice quality. Most Eastern Cham speakers follow the same general trend, but the amount of overlap between their registers is much greater. Finally, Châu Đốc Cham register seems to be "ballistic". The low register starts breathy and then become modal (or even tense), while the high register starts modal and then becomes breathy. It is difficult to isolate a clear pattern at the end of the vowel, but it suffices to say that the voice quality distinction seems lost by then. Overall, the salience of voice quality distinctions seems too low to be used alone for contrasting registers, but it could possibly reinforce a contrast in other phonetic properties.

3.3 F1 and F2 (vowel quality)

In this section, we will only report results for the long vowels /i, u, a/. Short vowels are excluded because there are relatively few of them in our wordlist. In any case, they show the same tendencies as long vowels and will be included in more complete statistical analysis that is beyond the scope of this short paper. The number of tokens included in the analysis for each vowel and dialect, after excluding outliers as described in section 2.4.3, is given in Table 2. The large number of /u/ tokens whose second formant was excluded is due to the large bandwidth and low energy of this formant, which caused the Praat formant tracker to skip it frequently.

| Table 2: Number of | f tokens used | for the formant | t analysis |
|---------------------------|---------------|-----------------|------------|
| | | | |

| | | /i/ | | /u/ | | /a/ | |
|-----------------------------|----------|------------|------|------------|-----|-----|-----------|
| Dialect | Register | F 1 | F2 | F 1 | F2 | F1 | F2 |
| Kompong Chhnang Cham | High | 714 | 714 | 577 | 367 | 839 | 900 |
| | Low | 761 | 813 | 385 | 260 | 963 | 1000 |
| Eastern Cham | High | 885 | 868 | 550 | 251 | 954 | 957 |
| | Low | 672 | 666 | 360 | 162 | 794 | 813 |
| Châu Đốc Cham | High | 947 | 1043 | 512 | 352 | 719 | 778 |
| | Low | 858 | 724 | 288 | 189 | 716 | 764 |

Figures 3-5 illustrate the distribution of normalized F1 and F2 values per dialect and register. Since the degree of overlap between registers is much greater for formants than for f0 and open quotient, plotting curves for individual speakers would have been misleading. We therefore decided to present the results as boxplots. We only show the midpoint of the vowels of word-final syllables because vowel onsets and offsets are too affected by consonant-to-vowel transitions to be meaningful when all the data is pooled together without controlling for onsets and codas.

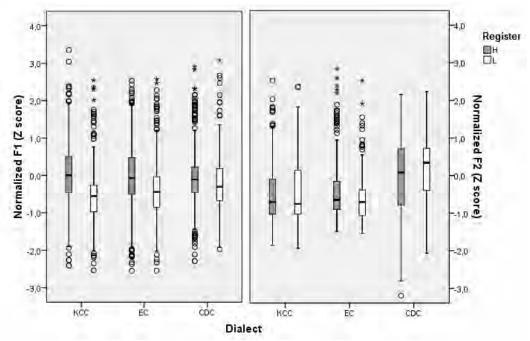


Figure 3: Mean normalized F1 (left) and F2 (right) of the long vowel /i/ in the two registers, for the three dialects (KCC: Kompong Chhnang Cham, EC: Eastern Cham, CDC: Châu Đốc Cham)

In Figure 3 and 4, we see that the F1 of the high vowels /i, u/ is higher in the high than in the low register in all three dialects. This means that these vowels are usually slightly more closed in the low than in the high register. No clear pattern emerges for F2, however, which means that high vowels do not differ in frontness/backness. The slight difference F1 is probably below perception because of the large amount of overlap between registers and because no difference in vowel quality has ever been reported for Eastern Cham, in which the F1 difference between register is comparable to other dialects. This means that the small formant differences would be a side-effect of another articulatory strategy rather than a phonetic property directly under speaker control. A likely candidate would be the lowering of the larynx, a common strategy to depress pitch (Ohala 1973; Hombert *et al.* 1979; Honda *et al.* 1999), that also has the effect of lowering F1 by lengthening the pharyngeal cavity.

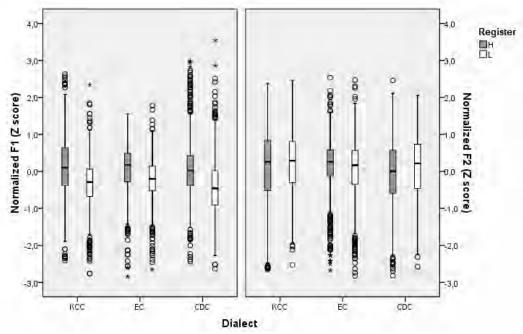


Figure 4: Mean normalized F1 (left) and F2 (right) of the long vowel /u/ in the two registers, for the three dialects (KCC: Kompong Chhnang Cham, EC: Eastern Cham, CDC: Châu Đốc Cham)

We see in Figure 5 that the vowel /a/ does not show clear patterns for either F1 or F2. Therefore, listeners could not rely on formant frequencies to distinguish low register and high register /a/'s.

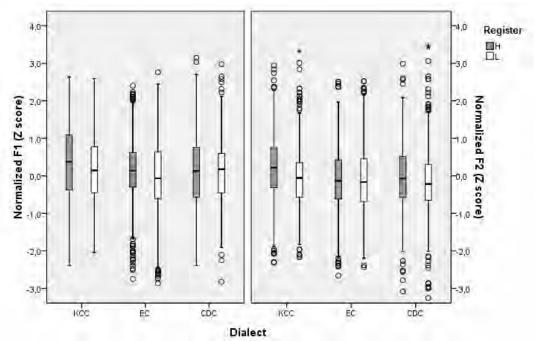


Figure 5: Mean normalized F1 (left) and F2 (right) of the long vowel /a/ in the two registers, for the three dialects (KCC: Kompong Chhnang Cham, EC: Eastern Cham, CDC: Châu Đốc Cham)

To sum up, vowel quality differences between registers are at best limited to the F1 of high vowels (at least in the set of three vowels tested here). Even then, these differences are so small and unsystematic that they are unlikely to be perceptible. It is therefore unlikely that vowel quality plays a crucial role in the phonetic realization of the register contrast.

4. Discussion

In this section, we will sum up the results and compare them to previous descriptions, discuss the relevance of our data for the hypothesis that Cham register systems are shaped by contact and put forward a fine-grained model of language contact and sound change.

4.1 The phonetics of Cham register

The first observation that can be made based on our results is that the three dialects under study are surprisingly similar with respect to the phonetic realization of their registers. First, f0 seems to be the most important property of register in all three dialects. However, the f0 difference between registers is greater in Eastern Cham than in Kompong Chhnang Cham and Châu Đốc Cham, which suggests that Eastern Cham might have phonologized it without making it contrastive, something that the other dialects have not done yet. In this sense, one could argue that Eastern Cham is transitional in that it has made a step further towards the development of tone.

The other phonetic properties measured here, open quotient (voice quality) and the frequencies of the first two formants (vowel quality), are much less conclusive. Open quotient differences between the registers could play a role in perception in all three dialects, but they seem less distinct and reliable than pitch differences. Therefore, open quotient seems to be an ancillary phonetic property, in the sense that it only reinforces a contrast that is primarily based on f0. Moreover, the exact implementation of open quotient differences between registers varies between dialects and speakers. Whereas most Eastern Cham and Kompong Chhnang Cham speakers maximize the open quotient difference at the beginning of the vowel and then neutralize it towards its end, most Châu Đốc Cham speakers exhibit a ballistic effect, where the high and low register open quotient curves cross paths and change positions at about the first third of the vowel. Finally, the role of vowel quality seems negligible. Only the first formant of the high vowels /i, u/ show register differences, and these differences are so small that it is unlikely that they can be used for register identification. At this point, a perceptual study conducted with the same stimuli in the three dialects is necessary to further understand the phonetic realization of Cham register.

Interestingly, our findings partially disagree with previous work. The Eastern Cham results are in tune with other phonetic studies (Phú *et al.* 1992; Brunelle 2005b; a), but Kompong Chhnang Cham and Châu Đốc Cham are different from other Western Cham dialects. Kompong Chhnang Cham, which is spoken in Cambodia, has a pattern of pitch and voice quality that matches previous descriptions of Cambodian Cham (Headley 1991; Edmondson and Gregerson 1993), but it does not seem to make much use of vowel quality contrasts, contrary to Kompong Thom Cham and the dialect investigated by Edmonson and Gregerson. More mysterious are our results for Châu Đốc Cham, which differ from those

obtained from the same dialect in the 1970s⁸ (Friberg and Hor 1977). Although we find a pitch difference between registers, like Friberg and Hor, the voice quality patterns that emerge from our data are much simpler (the low register is breathy and the high register modal for all three vowels) and the vowel quality differences that we uncovered do not seem robust enough to be comparable to the vowel system that they reported.

There are two possible ways of explaining these discrepancies. The first one is that there might be possible errors and misinterpretations in some of the analyses (including the present one). The other explanation is that there might be more variation than expected in Western Cham. The first type of variation is dialectal. As mentioned in Baumgartner (1998), Western Cham communities are scattered all over Cambodia and the Vietnamese Mekong delta and are not necessarily in close contact. This might have resulted in more variation than previously assumed. For this reason, it might be necessary to stop treating Western Cham dialects as a single entity (although our two Western Cham dialects, Kompong Chhnang and Châu Đốc Cham, are not very different). Another possible source of discrepancies that has been neglected so far is individual variation. Studies on Cham are typically based on a small number of consultants (often only one), and these consultants are often highly educated, linguistically aware and literate, contrary to the rest of their communities. Now that most areas of Mainland Southeast Asia are accessible and stable, it is high time we start paying attention to variation in small Southeast Asian languages, as we would in larger languages.

4.2 Are Cham register systems shaped by contact?

Our experimental results allow us to answer some of the questions raised in section 1.3. First, can we attribute the central role of pitch in Eastern Cham to Vietnamese influence? Not in a trivial sense. Pitch seems to be the dominant property of register not only in Eastern Cham, whose speakers are quasi-native in Vietnamese, but also in Kompong Chhnang Cham, a dialect of Cham that is not in contact with tone languages (let alone Vietnamese). Therefore, we cannot claim that the dominant role of pitch in Cham register systems is contact-induced.

However, Vietnamese influence seems to manifest itself in a more subtle way. As we have seen in section 3.1, the mean f0 curves of the two registers are much more distinct in Eastern Cham than in the two dialects that are less or not in contact with Vietnamese. Besides, standard deviations are smaller in this dialect, which supports the view of a more clearly defined centre of gravity for each register. This suggests that Eastern Cham has phonologized the pitch differences that are found in its register system, or has at the very least emphasized them, even if they still do not behave phonologically like lexical tone (Brunelle 2005a; b). This strengthening of the pitch distinction could be attributed to the familiarity of Eastern Cham speakers with pitch, which is crucial in a tone language like Vietnamese, a language that Eastern Cham speakers usually master natively or quasinatively. Nonetheless, even if the enhanced role of pitch in Eastern Cham register was originally caused by Vietnamese, it does necessarily follow that it is a change in progress. The enhancement of pitch could have taken place a few generations ago and have become a fossilized part of the native system. In fact, there is good evidence that it is Eastern Cham

⁸ Friberg and Hor (1977) do not explicitly mention the geographical origin of the dialect that they describe. This information is found in Headley (1991).

speakers who are least bilingual and bicultural who make the most use of pitch distinctions to contrast registers (Brunelle 2005b).

Our results also argue against a mechanistic approach to contact and change. First, the secondary role of voice quality in the register systems of Kompong Chhnang Cham and Châu Đốc Cham cannot be attributed to Khmer or Vietnamese, because Central Khmer, the dialect with which Kompong Chhnang Cham is in contact, has long lost voice quality distinctions and because Southern Vietnamese makes no use of breathiness. Further, if contact was automatic, we would expect Kompong Chhnang Cham, whose speakers are highly fluent in Khmer, to exploit vowel quality features for register contrast because Central Khmer has a very large vocalic inventory. We have seen that it is not the case.

In short, we must be careful when attributing sound changes to external factors. First, contact does not automatic entail sound change. Second, changes are not necessarily phonological and categorical, but can take place at a fine-grained phonetic level. Finally, even if a change was originally contact-induced, it is not necessarily in progress.

4.3 Implications for language contact and change

Our results allow us to refine the scenario according to which internal drift is influenced and shaped by external factors (Thurgood 1996; 1999). The picture that emerges is more nuanced than the usual categorical view. First, we argue that phonological similarities between neighbouring languages are not necessarily due to contact. For instance, despite the fact that Eastern Cham and Vietnamese both use pitch contrastively, the fact that Cham dialects that are not in contact with Vietnamese also make use of pitch suggests that it might be a basic property of Cham registers that can be reconstructed to an earlier stage of Cham, before its encounter with Vietnamese and the dialectal diversification that ensued. Sound changes are much more difficult to attribute to contact than lexical items because similarities are much more likely to arise by coincidence. That said, our results also argue for a low-level, fine-grained type of contact-induced phonetic change that might be much more common than usually assumed, but go unnoticed because of a lack of experimental data.

The main idea behind this claim is that highly fluent second language speakers become attuned to phonetic properties that are not found in their native language, or become more sensitive to phonetic properties that are more important for contrast in their second than their first language. The fact that they have an increased control over these phonetic cues could favour their enhancement in their first language, without necessarily leading to the development of new contrasts. For example, Eastern Cham speakers have a high level of pitch control and can perceive subtle nuances in pitch because of their knowledge of Vietnamese tones. They are therefore much more likely to start putting an emphasis on pitch as a phonetic property of the register contrast than speakers of other Cham dialects, who are less or not exposed to tone languages and are therefore less attuned to pitch. However, the fact that Eastern Cham speakers reinforce the phonetic role of pitch in their register system because of Vietnamese influence does not necessarily entail that they will go all the way and develop a complex tone system. Such a change is not impossible, but it would require the loss or the phonological reanalysis of laryngeal codas, a change that would be language-internal and would take place independently of Vietnamese, which has long lost its laryngeal codas and cannot be used as a model. While contact with Vietnamese would have contributed to set favourable conditions for tonogenesis by strengthening the role of pitch, the development of contrastive lexical tones

itself would not be contact-induced. In short, if we exclude new phonemes borrowed with loanwords, sound convergence must be phonetic in nature. Bilingualism can strengthen intrinsic phonetic tendencies and increase the likelihood of some types of phonological reorganization, but the presence of a phonological category in the "donor" language does not directly trigger the development of a similar category in the "receiver" language.

The question that begs for an answer at this point is why some phonetic factors are strengthened by contact while others are not. For instance, how can we explain the fact that Kompong Chhnang Cham has not developed a register system based on vowel quality contrasts, despite the fact that its speakers are highly fluent in Khmer and therefore have the ability to produce fine-grained vowel distinctions? The answer to this question is beyond the scope of this paper, but it might very well lie in language attitudes. A community can reject or favour innovations for a number of sociolinguistic reasons. Speakers might be aware that a change is triggered by contact with the dominant language and reject this change as a form of assimilation. For example, Eastern Cham aspirated stops tend to become voiceless fricatives in the speech of younger generations, which is attributed, accurately or not, to Vietnamese influence and frowned upon as contact-induced language decay. From yet another perspective, highly bilingual speakers, whose speech is more likely to be affected by contact, might be more socially successful thanks to their contacts outside the community and might therefore become linguistic trendsetters. In the end, large-scale sociophonetic studies of highly bilingual communities are needed if we want to move beyond speculation.

5. Conclusion

We have shown in this paper that three Cham dialects spoken in distant parts of Vietnam and Cambodia and having limited contact with one another have relatively similar register systems. As revealed by an experimental analysis combining laryngography with acoustics, the three register systems seem to rely primarily on pitch and to reinforce the contrast with voice quality cues, while vowel quality is unlikely to be a significant factor. The differences in the phonetic realization of the register contrast in the three dialects are quantitative in nature. The most significant of these differences is that the pitch curves of the two registers are much more distinct in Eastern Cham than in the other dialects.

While the central role of pitch in the register contrast of the three Cham dialects cannot be attributed to Vietnamese, the larger distinctiveness of pitch in Eastern Cham could be linked to Vietnamese influence. Eastern Cham speakers are highly bilingual in Vietnamese and therefore have a good command of the subtle pitch distinctions found in its tone system. Based on this observation, we develop a nuanced model of contact-induced sound change in which we distinguish the direct transmission of phonological categories, which we claim to be limited to loanwords, from the gradual strengthening of phonetic properties found in dominant languages, which is probably common but underdescribed. We propose that sociolinguistic factors such as language attitudes must play a central role in the selection or rejection of contact-induced sound changes.

References

- Aikhenvald, A and R Dixon 2006. *Grammars in Contact: A Cross-Linguistic Typology*. Oford, Oxford University Press.
- Alieva, NF 1984. "A Language-Union in Indo-China." *Asian and African Studies* XX: 11-22.
- Baumgartner, N 1998. A grammar sketch of Western (Cambodian) Cham. *Papers in Southeast Asian Linguistics no.15: Further Chamic Studies*. D. Thomas. Canberra, Pacific Linguistics. 89: 1-20.
- Blood, DL 1967. "Phonological Units in Cham." *Anthropological Linguistics* 9(8): 15-32. Brunelle, M 2005a. "Register and tone in Eastern Cham: Evidence from a word game."

Mon-Khmer Studies 35.

- Brunelle, M 2005b. *Register in Eastern Cham: Phonological, Phonetic and Sociolinguistic approaches.* Ph.D dissertation, Cornell Uni. Ithaca.
- Brunelle, M 2006. A phonetic study of Eastern Cham register. *Chamic and Beyond*. P. Sidwell and A. Grant. Sidney, Pacific Linguistics: 1-36.
- Brunelle, M 2008. "Monosyllabicization in Eastern Cham." SEALS XIV Volume 1: Papers from the 14th meeting of Southeast Asian Linguistics Society (2004), pp43-58.
- Brunelle, M 2008. "Diglossia, Bilingualism, and the Revitalization of Written Eastern Cham." *Language Documentation and Conservation* 2(1): 28-46.
- Brunelle, M 2009. Diglossia and Monosyllabization in Eastern Cham: a Sociolinguistic Study. *Variation in Indigenous Minority Languages*. J. Stanford and D. Preston. Amsterdam, John Benjamins: 47-75.
- Diffloth, G 1982. "Registers, Dévoisement, Timbres Vocaliques: leur Histoire en Katouique." *Mon-Khmer Studies*(XI): 47-82.
- Edmondson, J and K Gregerson 1993. Western Cham as a Register Language. *Tonality in Austonesian Languages*. J. Edmondson and K. Gregerson. Honolulu, U of Hawaii Press: 61-74.
- Ferlus, M 1979. "Formation des Registres et Mutations Consonantiques dans les Langues Mon-Khmer." *Mon Khmer Studies* VIII: 1-76.
- Friberg, T and K Hor 1977. Register in Western Cham phonology. *Papers in Southeast Asian Linguistics No.4*. D. D. Thomas, E. W. Lee and Đ. L. Nguyễn. Sydney, Australian National University. 48: 17-38.
- General Statistics Office of Vietnam. (1999). "Structure of population as of 1 April 1999 by ethnic group." from http://www.gso.gov.vn/default_en.aspx?tabid=476&idmid=4&ItemID=1841.
- Gordon, RG 2005. Ethnologue: Languages of the World. Dallas, SIL International.
- Gregerson, K 1973. Tongue-root and Register in Mon-Khmer. *Austroasiatic Studies*. P. N. Jenner, L. Thompson and S. Starosta. Honolulu, University Press of Hawaii. 1: 323-369.
- Headley, RK 1991. The Phonology of Kompong Thom Cham. *Austroasiatic languages essays in honour of H. L. Shorto.* J. Davidson: 105-121.
- Henderson, E 1952. "The main features of Cambodian prononciation." *Bulletin of the School of Oriental and African Studies* 14(1): 453-476.
- Hoàng, TC 1987. "Hệ thống thanh điệu tiếng Chàm và các kí hiệu." *Ngôn Ngữ*(1-2): 31-35. Hombert, J-M, JJ Ohala and WG Ewan 1979. "Phonetic Explanation for the Development of Tones." *Language* 55: 37-58.
- Honda, K, H Hirai, S Masaki and Y Shimada 1999. "Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control." *Language and Speech* 42(4): 401-411.
- Huffman, F 1976. "The register problem in fifteen Mon-Khmer languages." *Oceanic Linguistics special publication* Austroasiatic Studies, part 1(13): 575-589.
- Jones, RB 1986. Pitch Register Languages. *Contributions to Sino-Tibetan Studies*. J. McCoy and T. Light. Leiden, E.J. Brill: 135-143.
- Lee, EW 1966. *Proto-Chamic phonologic word and vocabulary*. Ph.D dissertation, Indiana Uni., Bloomington.

Lee, EW 1974. "Southeast Asian Areal Features in Austronesian Strata of the Chamic Languages." *Oceanic Linguistics* 13: 643-668.

- Matisoff, J 1973. Tonogenesis in Southeast Asia. *Consonant Types and Tone*. L. Hyman. Los Angeles, USC. 1: 71-96.
- Matisoff, J 2001. Genetic versus Contact Relationship: Prosodic Diffusability in South-East Asian Languages. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. A. Aikhenvald and R. Dixon. Oxford, Oxford University Press: 291-325.
- Mazaudon, M and A Michaud 2008. "Tonal Contrasts and Initial Consonants: A Case Study of Tamang, a 'Missing Link' in Tonogenesis." *Phonetica* 65.4:231-256.
- Moussay, G 1971. *Dictionnaire cam-vietnamien-français*. Phan Rang, Trung-tâm Văn hoá Chăm.
- Ní Chasaide, A and C Gobl 1997. Voice Source Variation. *The Handbook of Phonetic Sciences*. W. J. Hardcastle and J. Laver. Oxford, Blackwell: 427-461.
- Ohala, JJ 1973. How is pitch lowered? Manuscript accessed on August 22 2009 at http://linguistics.berkeley.edu/phonlab/users/ohala/papers/how-whole.pdf.
- Phú, VH, J Edmondson and K Gregerson 1992. "Eastern Cham as a Tone Language." *Mon Khmer Studies* 20: 31-43.
- Pittman, RS 1978. "The explanatory potential of voice-register phonology." *Mon-Khmer Studies* 7: 201-226.
- Thomason, S and T Kaufman 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley and Los Angeles, University of California Press.
- Thurgood, G 1993. Phan Rang Cham and Utsat: Tonogenetic Themes and Variants. *Tonality in Austronesian Languages*. J. Edmondson and K. Gregerson. Honolulu, U of Hawaii Press: 91-106.
- Thurgood, G 1996. "Language Contact and the Directionality of Internal Drift: The Development of Tones and Registers in Chamic." *Language* 72(1): 1-31.
- Thurgood, G 1999. From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change. Honolulu, University of Hawai'i Press.
- Thurgood, G 2002. Learnability and direction of convergence in Cham: the effects of long-term contact on linguistic structures. *Proceedings of the 2000 Western Conference on Linguistics*, 507-527.

A MINIMALIST APPROACH TO AMIS¹ VOICE MARKERS

Yi-Ting Chen

Arizona State University, Wenzao Ursuline College of Languages < <u>ychen77@asu.edu</u>>

0 Abstract

This paper relates underlying properties of Amis voice markers to each other in a Generative approach. This study suggests that the computation of Aktionsart in the head of the Inner Aspect phrase (InnAspP) is the key to correlate these underlying properties. This computation not only contributes to the default temporal, aspectual, and modal (TAM) reading of Amis voice markers but also differentiates undergoer voice markers from others. This study first assumes that the realization of the feature morpheme is determined by whether the argument it introduces is raised to the edge. Then agreement takes places, which explains why the voice marker and the nominative NP are co-indexed.

1 Introduction

1.1 Purpose of the study

The purpose of this study is to use the Generative model to interpret the underlying properties of Amis voice markers, which have been investigated in many descriptive (e.g., T-L. Huang, 1988; Y-J. Huang, 1988; Wu, 2000) and Role and Reference Grammar (RRG) studies (e.g., Wu, 2005; 2006; 2007). Unlike previous research which has tried to reveal the entangled complexities of Amis voice markers, this study intends to correlate these underlying properties from the perspective of Generative Grammar and provide an interpretation of why they are connected.

1.2 What are voice markers?

Voice systems are commonly found in approximately 60% of western Austronesian languages (Himmelmann, 2005), a group to which Amis belongs. What are voice markers in terms of Generative linguistics? Many researchers (e.g., Guilfoyle, Hung, and Travis, 1992 (GHT hereafter); Rackowski, 2002; Chang, 2004; Pearson, 2005) have pursued an answer to this research question and provided various proposals. This study reviews several well-known proposals in the following section.

¹ Amis, a Formosan language, belongs to the western Austronesian language family. The speaker population is approximately 172,000, according to the census report by the Council of Indigenous People in Taiwan in November 2007.

Yi-Ting Chen. 2009. A Minimalist Approach to Amis Voice Markers. *Journal of the Southeast Asian Linguistics Society* 2:23-36. Copyright vested in the author.

24 Yi-Ting Chen

1.2.1 GHT (1992)

GHT (1992) proposes that there are two 'subject' positions, one in spec-VP and the other in spec-IP. The argument, which is not case assigned in VP, undergoes spec-head agreement with INFL for its case and moves to the spec of IP. In this proposal, voice markers are base-generated in different positions: undergoer in INFL and actor in V.

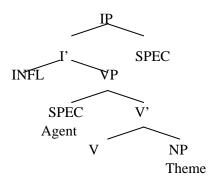


Figure 1: *GHT approach (GHT, 1992, p. 379)*

1.2.2 Rackowski (2002)

Rackowski (2002) suggests that voice markers are Spell-Outs of agreement. In her approach, the presence of an Extended Projection Principle (EPP) feature determines the appearance of the voice marker. She argues that if an EPP feature is present in the VoiceP, the patient must move up to the spec of VoiceP to check off the EPP feature and becomes the closest goal for agreement. Once agreement takes place, the patient feature is copied to the T and is further spelled out as the patient voice on T. On the other hand, if an EPP is not present in VoiceP, the patient remains in situ and the agent is the closest argument for agreement; thus T is spelled out as the actor voice.

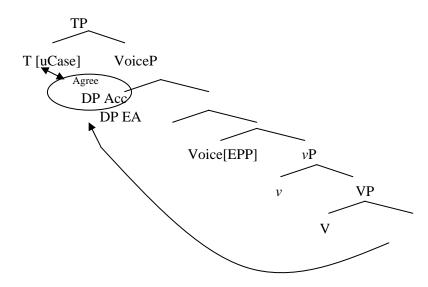


Figure 2: Rackowski approach (2002, p. 18)

Amis Voice Markers 25

1.2.3 Pearson (2005)

Pearson (2005) argues that the NP (Trigger in his terminology), which is co-indexed from the voice marker on the verb, is located in an A' position, parallel to TOPIC. This trigger is base-generated with a null operator in the spec of WhP raised from a 'case position'. The distribution of case positions is shown below. Since the trigger is not in an A' position, voice markers are not taken as the Spell-Out of agreement on the functional projection (such as T or INFL). Instead, Pearson (2005) suggests that these voice markers are indicators of the case position from which the null operator raises and their forms are dichotomized based on the weak or strong feature on the E(vent)P.

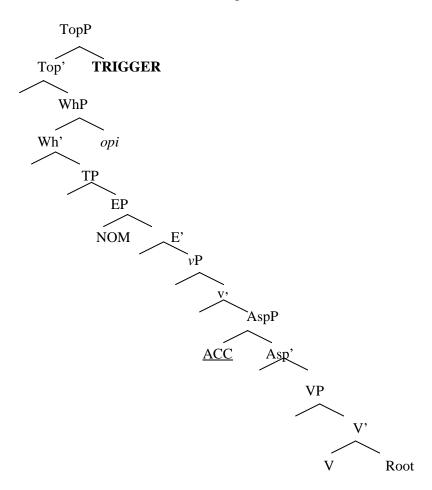


Figure 3: *Pearson approach* (2007, p. 385, 402)

1.2.4 Amis voice markers by Liu (2003)

As for Amis voice markers, Liu (2003), following Chang (1997), suggests that voice markers in Amis are in the head of VoiceP although they differ in their ability to assign case and theta-role. For instance, she argues "the AV marker *mi*- being functional does not assign any thematic role to spec-VoiceP nor does it assign case to the Actor" (Liu, 2003, p. 118); thus, in order to be case marked, the agent must move up to the spec of TP to check off the thematic feature in VoiceP. On the other hand, the undergoer voice marker, *ma*-, also located in VoiceP, can assign the genitive case to the agent in spec-VP and the theta role to patient in the spec of VoiceP.

26 Yi-Ting Chen

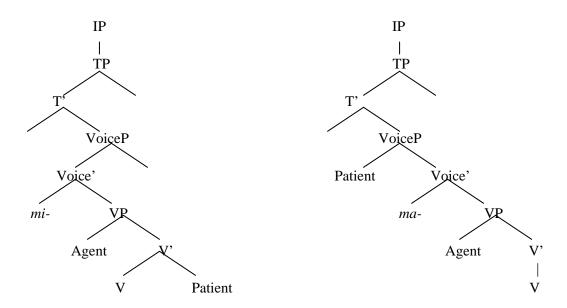


Figure 4. *mi*- (revised, Liu, 2003, p. 118) Figure 5. *ma*- (revised, Liu, 2003, p. 125)

The proposals reviewed so far all focus on the theta-agreement between voice markers and the nominative NP. In the remainder of this section, I review a proposal on morpheme deletion (Travis, 2005) by which this study proposes some adjustments for Amis.

1.2.5 Morpheme deletion (Travis, 2005)

Travis (2005) proposes that morpheme realization in Tagalog and Malagasy can be understood by morpheme deletion. For instance, by comparing Tagalog 'climb' *akyat* and 'X makes Y climb' *magpaakyat* or *papagakyatin*, it is found that if the causer is the subject, then the lower *pag*- (causative) is deleted whereas if the causee is the subject, the higher *pag*- is deleted.

magpaakyat m-pag-pa-pag-akyat Causer Subject papagakyat-in m-pag-pa-pag-akyat-in Causee Subject

Amis Voice Markers 27

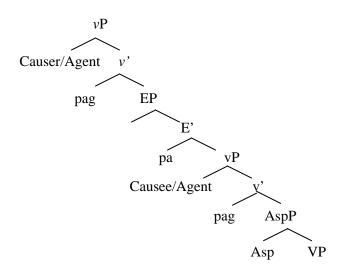


Figure 6: Morpheme deletion (revised, Travis, 2005, p. 178)

She also argues that in Malagasy, when the telicity marker a is present in an unaccusative sentence, an external argument can be introduced. Examples are shown below.

- (1) tafa.vor.in'ny mpampianatra ny ankizy
 TAFA.meet.Gen'the² teacher the children
 'The teachers were able to gather the children.' (Travis, 2005, p. 182)
- (2) *n.i.vor.in'ny mpampianatra ny ankizy
 PST.i.Meet'the teacher the children
 'The teachers were able to gather the children.' (Travis, 2005, p. 183)

To be specific, she argues that "telicity will take an Agent and turn it into a Cause (non-volitional agent) when attached to roots that have Agents in their theta-grids. When attached to a root with no external argument (such as an unaccusative or an adjective), it will add a Cause argument" (Travis, 2005, p. 183). Either a non-volitional agent or cause argument, according to Travis (2005), is in the spec of InnAspP. The key of arguing the non-volitional agent or cause being Spec of InnAspP is the occurrence of the [+telic] feature. However, whether Amis, under the same language family with Malagasy, might display similar theta-role changing process as Malagay does is beyond the scope of the current study and this study leaves this question for further research.

The present study adopts the generalized Doubly-Filled Comp Filter (Sportiche, 1992; Koopman, 1996) and extends Travis's research to further interpret Amis voice markers. Prior to proposing suggestions for Amis voice markers, I will introduce the major characteristics of Amis voice markers.

² Abbreviations: Gen: Genitive; Nom: Nominative; Dat: Dative; PST: Past tense; Asp: Aspect; AV: Actor voice; UV: Undergoer voice; Neut: Neutral voice; 1sg: First person singular; CN: Common noun marker; PPN: Personal noun marker; Prep: Preposition.

28 Yi-Ting Chen

1.3 Amis voice markers

Amis voice markers have been a popular research topic among scholars interested in Amis grammar. Until now, it has been found that they at least interact with theta-agreement, case assignment, lexical aspect, argument structure, and verbal classification. Below I briefly illustrate each of them individually.

The most prominent property of Amis voice markers is theta-agreement, where the theta-role of the nominative NP is reflected on the voice marker. For instance, if the verb is prefixed by an actor voice marker, the agent is assigned the nominative case.

Voice markers are also relevant to the case assignment principle in Amis. There are three cases in Amis, nominative, genitive, and dative, and their assignment is determined by voice markers. For instance, in a transitive sentence with an undergoer voice marker, the agent is assigned the genitive and the undergoer, the nominative; whereas in a transitive sentence with an actor voice marker, the agent is assigned the nominative and the undergoer, the dative.

Every voice marker is found to encode a default TAM reading. Wu (2006; 2007) argues that the default setting of TAM in voice markers is in fact traced back to the lexical aspect encoded in voice markers. For instance, ³-en and ma3- both carry the logical structure of BECOME so that their appearance infers an endpoint in the denoted event.

The meaning of a word can be altered by affixing different voice markers. For instance, *patay* 'die' in Amis becomes 'kill' (cause to die) when *patay* is prefixed by the actor voice marker *mi*-. On the other hand, the meaning becomes 'dead' when *patay* is prefixed by the undergoer voice marker *ma*-. In this example, the appearance of *mi*-introduces an argument, causer, or agent, but the prefixization of *ma*- doesn't have the same effect.

Many studies of Amis voice markers with focus on verb classification (e.g., Huang, 1988; Yan, 1992; Liu, 2003) show that each voice marker is likely to affix to some specific verbal classification. For instance, the actor voice marker, *mi*-, tends to mark dynamic/motional verbs, whereas *ma*- marks verbs ranging from psych-verbs to statives.

As reviewed, Amis voice markers are composed of features that are more than just voice or agreement markers. The proposal as in Liu (2003) seems to be insufficient to include the whole scope of Amis voice markers. A satisfactory explanation of what these voice markers are must take into consideration these possessed properties and have them connected. To be more specific, research questions would be: If there exists a connection among all these properties, why is there a correlation between the case assignment and the lexical aspect, varied by voice markers affixed?

2 Theoretical orientation

This study is presented within the theoretical structure of the Minimalist Program (MP), the latest development of Generative Grammar. MP can be regarded as a computational system, composed of finite sets of invariant principles exploitable to generate infinite sets of sentences from an abstract structural representation. Every articulated sentence is derived from the structure-building procedure, starting from selecting lexical items through merging and *remerging* (parallel to *internal merge* or *move* in the older terminology). Remerge dislocates positions of existing lexical items and is triggered by the need for feature-checking to ensure the full interpretation of all constituents. Physical production of

³ This study adopts the gloss from Wu (2006, 2007).

Amis Voice Markers 29

sounds and mental representation of meanings are connected by this computational system at two interface levels: Phonetic Form (PF) and Logical Form (LF). This study assumes that all projections formed by merge and remerge are imposed by the Linear Correspondence Axiom (Kayne, 1994), and are restricted to be binary and right branching only.

This study also assumes that case assignment is done by a Probe-Goal (P-G) relation (Chomsky, 1995; 2004; 2006). A finite T with an uninterpretable phi-feature looks down for a goal with a matching interpretable phi-feature and unvalued case. The search domain of P-G must be minimally conditioned by locality.

According to Chomsky (post 1995) and Larson (1989), the VP is no longer articulated as a single projection but split, composed of 'shells'. Observed from Amis verbal derivation, this study first assumes that verbs are derived from roots, projecting XP. Here this paper simply replaces VP with XP, because Amis verbs can be derived from various lexical categories of roots (e.g. adjectival, noun, etc.). XP can merge with the direct object, if one exists, then assigning a theme/patient theta-role to it. This paper also assumes that the external argument, actor or agent, is introduced in the Spec of upper VP, parallel to vP. Adopting Travis' argument (2005, in prep), this study also assumes that an additional functional projection exists, termed the inner aspect (InnAspP, hereby), which is sandwiched between the upper and lower VPs. This InnAspP is responsible for computing the situational aspect, which has been found actively in Amis voice markers. Again in accordance with Travis (1994), the upper VP is selected by the Event Phrase (EP) in which "[it] introduces or licenses the event argument of the verb, converting it into an eventdenoting constituent" (quote from, Pearson, 2005, p. 402). Finally, this paper assumes that EP is the edge, and that there is a dichotomized feature [+/- Actor] in the E head for which an argument with the matching feature must move up to check off this feature.

This study adopts the generalized Doubly-Filled Comp Filter (Sportiche, 1992; Koopman, 1996) in which a head can be spelled out if its specifier is empty. The present study also suggests that the [+ Actor] or [-Actor] promotes the corresponding argument to the Spec of EP position, and that the appearance of either feature determines which morpheme or voice marker is realized. Below is to illustrate how mi-, ma-, and -en, the three most commonly found voice markers, are realized individually.

3 Analysis

3.1 Actor voice mi-

The major voice markers in Amis are *mi*-, *ma*-, and *-en*, and they are what this study intends to cover. The actor voice marker *mi*- is associated with activity, either affixing to an activity verb as in (4) or deriving an event into activity as in (3) (Wu, 2006; 2007). Additionally, the affixation of *mi*- sometimes renders a purposive reading (Zeitoun et al, 1996; Wu, 2006; 2007).

(3) mi-nanum φ-ci aki t-u nanum AV-water Nom-PPN Aki Dat-CN water 'Aki is going to drink water.' or 'Aki is drinking water.' (Wu, 2007, p. 121)

30 Yi-Ting Chen

(4) mi-palu φ-ci sawmah ci mayaw-an AV-beat Nom-PPN sawmah PPN Mayaw-Dat 'Sawmah is going to beat Mayaw.' or 'Sawmah is beating Mayaw.' (Wu, 2007, p. 121)

Mi- can also convert a state predicate into a causative accomplishment verb (cf. Wu, 2007, p. 125), shown in (5).

(5) mi-tuniq k-u kuwaq t-u ti'ti
AV-soft Nom-CNpapaya Dat-CN meat
'The papaya will soften the meat.' (Wu, 2007, p. 125)

Here, I assume that mi- is the actor marker in the head of vP. If the external argument that mi- introduces moves to the spec of EP, assumed as the edge of computation, mi- is realized; otherwise, it is deleted. Once the external argument is raised to the spec of EP, it becomes the closest goal to agree with T and is therefore assigned the nominative case. I also assume the movement to the spec of EP takes place at LF.

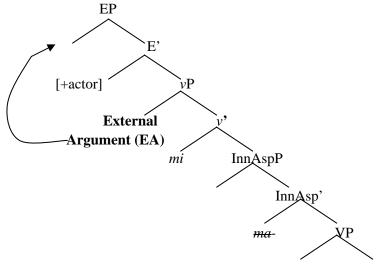


Figure 7: Proposed mi-

3.2 Undergoer voice -en

The undergoer voice marker, suffix -en, according to Tsukida (1993), can be used when the sentence meets three conditions: 1) there will be a positive occurrence of the event in the immediate future; 2) the agent's intention is to conduct the event; and 3) there is a clear image of the patient. Tsukida (1993) and Wu (2007) both note that, in -en sentences, the one who does the event must be animate, and the event cannot be progressive; that is, -en verbs imply an anticipatory endpoint within the event. For example, in (7), Tsukida (1993) explains that the image of the patient 'the shoes' has been on the agent's mind and the agent intends to buy that pair of shoes in the near future.

Amis Voice Markers 31

| (6) | a.*patay-en | aku | | | likes . | | | |
|-----|---------------------|------------|-----------|-----------|----------|----------|------------------|--|
| | Die-UV | 1sg.Gen | Nom-CN | I | mosquite |) | | |
| | b. ma-patay | aku | k-u | likes | | | | |
| | UV-die | 1sg.Gen | Nom-CN | I mosquit | .0 | | | |
| | 'I killed a mosqu | ito (unint | entionall | y).' | | | | |
| | 'A mosquito is de | ead becau | ise of me | ., | | (Tsukida | a, 1993, p. 131) | |
| (7) | a. cakay-en | | k-irana | | koco | | | |
| | Buy-UV | 1sg.Gen | Nom-tha | at | shoes | | | |
| | 'I will buy that pa | air of sho | es.' | | | (Tsukida | a, 1993, p. 137) | |
| | b. *mi-cakay | kaku | | tirana | koco | | | |
| | AV-buy | 1sg.Non | n | Dat-that | shoes | | | |
| | 'I will buy that p | air of sho | es.' | | | (Tsukida | a, 1993, p. 137) | |
| (8) | tuniq-en | aku/n-u | | kuwaq | k-u | ti'ti' | aca | |
| | Soft-UV | 1sg.Gen | /Gen-CN | papaya | Nom-CN | meat | a little | |
| | 'I/*papaya will s | often the | meat a li | ttle.' | | (Wu, 200 | 07, p. 127) | |

Based upon the properties of -en, Wu (2007) decomposes the logical structure of en into two essential properties: [+agentivity] and [+telic]. I would like to propose that, initially, there is an mi- in the vP due to agentivity and -en in the AspP due to the anticipatory [+telic]. I follow Doubly-Filled Comp Filter (Sportiche, 1992; Koopman, 1996), and also suggest that -en (1) promotes the patient to the spec of InnAspP, and (2) is realized when the patient is further raised to the edge. Thereafter, mi- is deleted since its specifier position is not empty.

After deletion, the external argument (EA) is left without any case and without the theta-role assigning head. The deletion of the agentivity morpheme mi- that introduces the EA makes the EA less agentive than an EA should be. This study assumes that this lesser degree of agentive EA must be bound with a genitive case that often marks the actor argument without any case. Thus, in -en sentences, the patient is assigned the nominative, and the agent the genitive.

32 Yi-Ting Chen

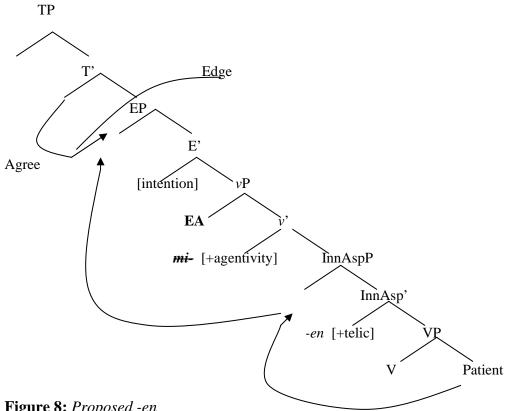


Figure 8: *Proposed -en*

3.3 Analysis of ma-

Amis ma- is not only an undergoer voice marker but also either an actor or a neutral one. Wu (2006; 2007) shows that ma-verbs can be classified into four categories. Irrespective of the logical structure, these ma-verbs all have one similarity, which is that the nominative case is assigned to the non-agentive argument. As shown in (9) to (12), the nominative NP is either the theme or the 'experiencer'. In some cases, the theme or the experiencer can be viewed as the non-volitional agent. For instance, in (10), the translation could be "the banana became ripe" and (12) could be read as "that pig became dead by Aki." In (9), the only argument is not a real actor either. Instead, it is a non-volitional actor. Below I will show each of them and their syntactic realization through the Minimalist Program and suggest that -a is a realization of a +/- telicity computation of the argument, which is raised up to the edge.

- (9) ma1-kerker φ-ci panay Neut-shiver Nom-PPN Panay (Wu, 2007, p. 130) 'Panay is shivering.'
- (10)ma2-ruhem (*ho) pawli (tu) k-u Neut-ripe (*Asp) (Asp) Nom-CN banana 'The banana is ripe.' (Wu, 2007, p. 131)
- ma3-laluk (11)(ho) (*tu) cingra Neut-diligent 3sg.Nom (Asp) (*Asp) 'He is (still) diligent.'

Amis Voice Markers 33

(12) ma4-patay n-i aki k-u-ra fafuy
UV-dead Gen-PPN Aki Nom-CN-that pig
'That pig was killed by Aki.' (Wu, 2007, p. 135)

Ma4- is the only ma- category glossed as an undergoer voice marker. In a ma4-sentence, the entity which undergoes the telic event moves to Spec of InnAspP first; then – ma- is realized, once the argument in the spec of InnAspP is further raised up to the edge covertly due to the requirement of [-Actor] in E. Finally, all the other morphemes which represent the other features are deleted, since their arguments remain in situ. This proposal is similar to the one for –en. Indeed, the case assignment principles for –en and ma4- are the same: nominative case to the patient and agent with genitive. Nevertheless, there are two differences between ma4- and –en. The first one is that –en denotes an anticipatory [+telic] feature, whereas ma4- simply denotes an endpoint. Second, -en is encoded with the agent's intention, whereas ma4- is not.

In the cases of *ma2*- and *ma3*-, two statives, all arguments are base-generated in VP and then raised to the spec of InnAspP, because of the computation of the Aktionsart of the event. *ma*- is realized when the argument is further promoted to the spec of EP to check off the [-Actor] feature in E. Example (10) for *ma2*- is tricky. In Amis, *tu* is often analysed as perfective marker (Zeiton et al, 1996) or sometimes inchoative aspect marker (Liu, 2003). However, in (10), there is no absolute ⁴endpoint but a result state.

The base-position of the actor in a mal- sentence is puzzling. However, the involuntary nature of the actor, which lacks full control of the action, leads me to postulate that this involuntary actor is not base-generated in the vP whose specifier position hosts a real agent/actor/causer. I suggest that this is why mi-, the actor/agent/causative marker, does not appear in sentences with involuntary reading. Here, I agree with Travis (2005, in prep) and hypothesize that the involuntary actor is merged into the structure in Spec of InnAsP. Later, when the [-Actor] in E triggers the involuntary actor to move further to the spec of EP, ma- is realized.

In short, from the research of *ma*-, it is suggested that in Generative framework, that existing argument is raised to the spec of InnAspP to measure out the endpoint or progressive status of the event. If the argument, which is either introduced or promoted to the spec of InnAspP, is further raised to the edge covertly, then *ma*- is realized. That is, *ma*- is interpreted to be spelled out of the head of the telicity computation whose specifier is promoted to the edge.

Wu (2007) indicates that verbs in example as (10) "all have an inherent ending point in the events they depict" (p. 132) and that explains why example (10) is compatible with *tu*, but not with *ho*, the incomplete aspectual marker.

34 Yi-Ting Chen

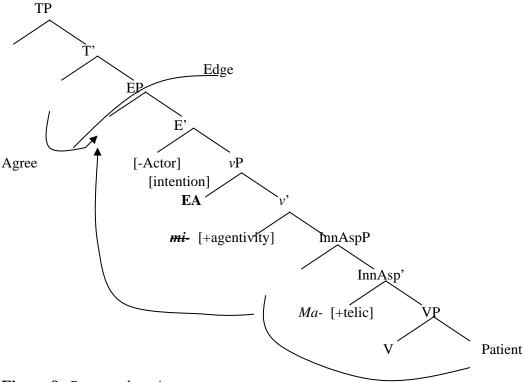


Figure 9: Proposed ma4-

4 Summary

I suggest that Amis voice markers are formed by a series of feature realization operations. A feature is realized only when the argument it introduces or promotes is raised to the edge. A morpheme is deleted if its argument remains in situ. The argument which is raised to the edge agrees with T. Features of that argument are copied to T, and T, on the other hand, assigns a nominative case to the argument in the edge. This is why the nominative NP is always co-indexed with the voice marker.

The [+telic] feature is particularly crucial in distinguishing undergoer voices from the others. Thus, of those with an agent theta role, only -en and ma4- verbs, which are encoded with the [+telic] feature, follow the ⁵case assignment of the genitive to the actor/agent/causer and the nominative to the theme/patient.

The [+telic] feature is also relevant to different degrees of agentivity shown by various voice markers. The [+agentive] or [+causative] feature introduces the external argument to the spec of νP , and makes mi- verbs most agentive. Then, once mi-, the agentivity feature, is deleted, the EA becomes less agentive; therefore voice markers, such as -en and ma4-, show a lesser degree of agentivity than mi-. Finally, the other ma- stative

One reviewer pointed out the problem of "case assignment" for *ma4*- and *-en* cases since the actor/agent is case-marked "genitive" when it is not assigned by any case. In Minimalism, the actor/agent is not assigned a case because T assigns nominative to the NP in EP and it cannot receive a case from the verb. Here I hypothesize that genitive case for actor/agent in *ma4*- and *-en* may be an inherent case. According to Chen (1987), "Amis operates with a mixed system of accusative and ergative constructions" (p. 268). Then, genitive case discussed here is somehow similar to ergative in Amis. Scholars, such as Massam (2002), Woolford (2006) and so on, indicate ergative is an inherent case, which is very often associated with certain theta position, specifically, the EA. Thus, genitive case for EA in *ma4*- and *-en* can possibly be explained as inherent case.

Amis Voice Markers 35

verbs, including *ma2*- and *ma3*-, whose non-agent arguments are base-generated in VP, of course, are the least agentive ones. Different degrees of agentivity coded in each voice marker can interpret why every voice marker tends to mark certain types of verbs. For instance, most of the *ma*- verbs are [-dynamic] (Liu, 2003) since most *ma*- verbs are either less agentive (*ma4*-), stative (*ma2*- and *ma3*-), or involuntary (*ma1*-).

To conclude, the [+/-Actor] feature in E determines which argument is moved at LF, and then agrees with T; the raise of argument also influences whether a head is deleted or realized under the assumption of the generalized Doubly-Filled Comp Filter (Sportiche, 1992; Koopman, 1996), and further affects the degree of agentivity. By these assumptions and proposals, why Amis voice markers are related to theta agreement, case assignment, and degree of agentivity is explained.

References

- Chang, Y-L. 1997. Voice, Case, and Agreement in Seediq and Kavalan. Hsinchu, Taiwan: National Tsing-Hua University Dissertation.
- Chang, Y-Y. 2004. Subjecthood in Tsou. Hsinchu, Taiwan: National Tsing-Hua University Dissertation.
- Chen, M. Verbal Constructions and Verbal Classification in Nataoran-Amis. Canberra, Australia: Australian National University Thesis.
- Chomsky, N. 1995. The Minimalist Program. Cambridge, MA: The MIT Press.
- _____. 2004. Beyond explanatory adequacy. *Structure and Beyond: The Cartography of Syntactic Structure Vol. 3*, ed. by A. Nelletti, 104-131. Oxford: OUP.
- _____. 2006. Approaching UG from below. Cambridge, MA: MIT, ms.
- Guilfoyle, E., Hung, H., and Travis, L. 1992. Spec of IP and spec of VP: Two subjects in Malayo-Polynesian languages. *Natural Language and Linguistic Theory* 10: 375-414.
- Himmelman, N. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. *The Austronesian Languages of Asia and Madagascar*, ed. by A. Adelaar and N. P. Himmelman, 110-181. New York: Routledge.
- Huang, T-L. 1988. Grammar of Amis in Taiwan. Fon-Yuan.
- Huang, Y-J. 1988. Amis Verb Classification. Taipei: Fu-Jen Catholic University Thesis.
- Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.
- Koopman, H. 1996. The spec head configuration. *Syntax at Sunset: UCLA Working Papers in Syntax and Semantics 1*, eds. By F. Lee and E. Garret, 36-65. Los Angeles: UCLA.
- Larson, R. 1988. On the double object construction. *Linguistic Inquiry* 19: 335-392.
- Liu, E-H. 2003. Conjunction and Modification in Amis. Hsinchu, Taiwan: National Tsing-Hua University Thesis.
- Massam, D. 2002. Fully internal cases: Surface Ergativity can be profound. AFLA 8. MITWPL 44: 185-196.
- Pearson, M. 2005. The Malagasy subject/topic as an A'-element. *Natural Language and Linguistic Theory* 23 (2): 381-457.

36 Yi-Ting Chen

Rackowski, A. 2002. The Structure of Tagalog: Specificity, Voice, and the Distribution of Arguments. Cambridge, MA: MIT Ph.D. Dissertation.

- Sportiche, D. 1992. Clitics, voice, and spec-head licensing. GLOW Newsletter 28: 46-47.
- Travis, L. 2005. Agents and causes in Malagasy and Tagalog. *The Syntax of Aspect: Deriving Thematic and Aspectual Interpretation*, ed. by Erteschuk-Shir, N. and Rapoport, T, 176-149. New York: Oxford University Press.
- _____. In Prep. *Inner Aspect*. Dordrecht, The Netherlands: Kluwer Academic Press.
- Tsukida, N. 1993. The use of the *-en* form in Fataan-Amis. *Asian and African Linguistics* 22: 123-140.
- Woolford, E. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37 (1): 111-130.
- Wu, L-L. 2000. Amis Reference Grammar. Taipei: Yuan-Liu Publishing Company.
- _____. 2005. *Ma* verbs in Amis: A role and reference grammar analysis. Paper presented at the 2005 International Course and Conference on Role and Reference Grammar, Taipei.
- _____. 2006. Verbal Classification, Case Marking, and Grammatical Relationships in Amis. Buffalo, NY: University of Buffalo Dissertation.
- _____. 2007. "Voice" markers in Amis: A role and reference grammar analysis. *Language and Linguistics* 8 (1): 95-142.
- Yan, Z-G. 1992. Syntactic Structure of Amis: A Study of Participants and Events. Taipei: National Cheng-Chi University Thesis.
- Zeitoun, E., Huang, M., Yeh, M., Chang, H., and Wu, J. 1996. The temporal, aspectual, and modal systems of some Formosan languages: A typological perspective. *Oceanic Linguistics* 35 (1): 21-56.

IDENTIFYING PREPOSITIONS IN THAI

Kitima Indrambarya

Kasetsart University < fhumkmi@ku.ac.th>

0 Abstract

Various linguists have examined words which are traditionally called prepositions in Thai. Warotamasikkhadit (1992) proposed that there are no prepositions in Thai. Other linguists maintain that there remain prepositions in Thai (cf. Savetamalya 1989; Indrambarya 1995; Prasithrathsint 2000). Words which appear in the following frame [NP V (NP) __ NP] have been widely discussed and classified into various categories: verbs (Indrambarya 1995), nouns (Savetamalya 1989; Indrambarya 1995, Prasithrathsint 2000), prepositions (Indrambarya 1995, Prasithrathsint 2000), adverbs (Indrambarya 1995) or even conjunctions (Warotamasikkhadit 1992).

With new findings in recent years, my aim in this paper is to take a closer look at each of the words which may appear in the above construction. Using precise and reliable set of syntactic criteria, I can synchronically identify prepositions in Thai. The data are drawn from present-day standard Thai corpus. The result shows that words which may appear in this position may have homophonous forms and they may belong to one of the four lexical categories, namely, prepositions, verbs, nouns or adverbs -- with the limited number of the last two classes.

1 Introduction

The topic of prepositions has been of interest among linguists. While Warotamasikkhadit (1992) proposes that there are no prepositions in Thai, other linguists (Kullavanijaya 1974; Savetamalya 1989; Indrambarya 1994, 1995; Starosta 1994; Prasithrathsint 2000; Takahashi 2005) maintain the existence of Thai prepositions. Words which are traditionally viewed as prepositions have been analyzed as nouns, verbs, prepositions, adverbs or even conjunctions. Some may have homophonous forms of different word categories (Singnoi 2000; Savetamalya 1989; Indrambarya 1994, 1995; Prasithrathsint 2000).

Some linguists have looked into these words diachronically and found that some of these words are in the process of grammaticalization from verbs to prepositions (Diller 2001; Takahashi 2005) or from nouns to prepositions, (Prasithrathsint 2000, Singnoi 2000:51) resulting in homophonous forms of different word classes.

In this paper, I reexamine words which appear in the following position and refine the criteria to synchronically identify their syntactic categories of the words traditionally referred to as prepositions in the following frame : [NP V (NP) _ NP].

2 Criteria to Identify the Syntactic Category of the Word in Question

Linguists have proposed a few criteria to distinguish prepositions from nouns. Here are the criteria which have been used to identify nouns.

Kitima Indrambarya. 2009. Identifying Prepositions in Thai. *Journal of the Southeast Asian Linguistics Society* 2:37-51.

Copyright vested in the author.

- a. Nouns but not prepositions can be subject of the sentence. (Indrambarya 1994, Prasithrathsint 2000)
- b. Nouns can occur with a determiner (Savetamalya 1989, Indrambarya 1995, Prasithrathsint 2000))
- c. Nouns can occur with the possessive marker /khŏɔŋ/. (Savetamalaya 1989, Prasithrathsint 2000)

I have found that using each of the criteria alone is not adequate to distinguish prepositions from nouns in Thai. Although the use of possessive marker in Prasithrathsint (2000) seems satisfactory to differentiate nouns from their homophonous prepositions. This criterion, like other criteria mentioned above, cannot be used singly to identify all types of nouns. For example, the words like *thaaŋ* 'way' and *dâan* 'side' may not easily be followed by the possessive marker $kh\check{\sigma}$. Yet, they can be modified by determiner day 'which' and can follow quantifiers $k\check{H}$ 'how many' and $l\check{a}ay$ 'several' like other nouns.

Noun: thaan 'way'

- (1) a. than pàtìbàt way practice *In practice*
 - b. * thaan khɔ̃ən pàtibàt way POSS practice Way of practice
 - c. thaaŋ day way which Which way?
 - d. kii thaan how many way How many ways?
 - e. lǎay thaaŋ several way Several ways

Noun: dâan 'side'

- (2) a. dâan sathăapàttayakam side architecture

 In architecture
 - b.* dâan khɔɔŋ sathäapàttayakam side POSS architecture Side of architecture
 - c. dâan day side which Which side?

- d. kɨɨ dâan how many side How many sides?
- e. lăay dâan several side Several sides

Therefore, it is necessary to use a combination of criteria to distinguish nouns from prepositions as well as to identify other syntactic categories. The following section discusses the criteria to identify the syntactic category of words. I also maintain some criteria used in Indrambarya (1995) and include some new ones.

2.1 Criteria to distinguish nouns and prepositions, from verbs and adverbs

Prepositions and nouns behave differently from verbs and adverbs. The stranding test reflects different characteristics of prepositions and nouns on the one hand, and verbs and adverbs on the other.

The Stranding Test (Indrambarya 1995)

Prepositions and nouns cannot be left behind when a following noun is topicalized while verbs and adverbs can.

Preposition: kàp 'with'

- (3) a. ?\(\xi\)? pay tal\(\text{aat}\) k a p j\(\text{jm}\)

 Ae go market with Joom.

 Ae went to the market with Joom.
 - b. * jǔm nà? ?é? pay talàat kàp Joom TOP Ae go market with As for Joom, Ae went to the market with her.

Noun: bâan 'house'

- (4) a. dεεη pay *bâan* phian Daeng go house friend Daeng went to his friend's house.
 - b. * phɨan nà? dɛɛŋ pay bâan friend TOP Daeng go house
 As for his friend, Daeng went to his house.

Verb: *s#i* 'to buy'

- (5) a. kháw *sɨi* khanŏm He buy sweets *He bought sweets*.
 - b. khanŏm nà? kháw sɨi sweets TOP he buy As for sweets, he bought them.

Adverb: yàanrûatrew 'quickly'

- (6) a. kháw dəən *yàaŋrûatrɛw* pay sathǎanii he walk quickly go station *He walked quickly to the station.*
 - b. pay sathǎanii nà? kháw dəən *yàaŋrûatrɛw* go station TOP he walk quickly *To the station, he walked quickly.*

As illustrated in (3) – (6), preposition $k \grave{a} p$ 'with' and noun $b \^{a} a n$ 'house' cannot be left at the end of the sentence while the verb $s \acute{n}$ 'to buy' and the adverb $y \grave{a} a \eta r \hat{u} a t r \epsilon w$ 'quickly' can.

Now that we can see how prepositions and nouns are different from verbs and adverbs, the following sections will discuss how to distinguish nouns from prepositions, and verbs from adverbs.

2.2 Criteria to identify nouns

I propose that one should use two tests: the determiner test and the quantifier test to distinguish nouns from prepositions.

2.2.1 The Determiner Test

According to Savetamalaya, (1989:103), determiners are, for example, $n\hat{i}i$ 'this', $n\acute{a}n$ 'that', day 'which' and $n\check{a}y$ 'where', ??in 'other'. I will use only the determiners day 'which' and ??in 'other' as a test for nouns since the determiners $n\^{i}i$ 'this', $n\acute{a}n$ 'that' and $n\check{a}y$ 'where' have homophonous pronouns which could be misleading if used as a test.

The Determiner Test:

Only nouns can be followed by determiners day 'which' and ?in 'other'.

Preposition: *kàp* 'with'

- (7) a. * dεεŋ khuy *kàp* day
 Dang talk with which *Which with did Dang talked?*
 - b. * deeŋ khuy kàp ?iin
 Dang talk with other
 Dang talked with others.

Noun: sìn 'thing'

- (8) a. thəə pràatthanăa sìŋ day you wish thing which What do you wish for?
 - b. the pràatthanăa sìŋ ?in you wish thing other You wish for other thing.

Verb: kin 'to eat'

(9) a. * kháw kin day He eat which

* Which did he eat?

b. * kháw *kin* ?ɨɨn
he eat other *He ate others*.

Adverb: yàaŋrûatrew 'quickly'

(10) a. * khaw ?àan *yàaŋrûatrɛw* day he read quickly which

* He read which quickly.

b. * khaw ?àan *yàaŋrûatrɛw* ?ìin he read quickly other

* He read other quickly.

As shown above, the noun sin 'thing' can be modified by the determiner day 'which' in (8a) and lin 'other' in (8b) while the preposition lin 'with' in (7), the verb lin 'to eat' in (9) and the adverb lin 'quickly' in (10) cannot be modified by any determiners.

2.2.2 The Quantifier Test

Quantifiers in Thai are words denoting amount and number which precedes nouns. They are, for example, *kii* 'how many' *baaŋ* 'some' and *lǎay* 'several'. Only nouns can be preceded by quantifiers. I then propose to use quantifiers such as *kìi* 'how many', and *lǎay* 'several' as a test to identify noun.

The Quantifier Test

Only Nouns can follow quantifiers such as kii 'how many' baaŋ 'some' and lăay 'several'.

Preposition: kàp 'with'

(11) a. * dεεη khuy lǎay kàp
Dang talk several with

* Dang talked several with.

b. * dεεη khuy kɨɨ kàpDang talk how many with

* How many with did Dang talk?

Noun: nánsži 'book'

- (12) a. *náŋs¥i* lǎay lêm kàw mâak book several CL old much *Several books are very old.*
 - b. thəə mii *náŋs¥i* k¥i lêm you have book how many CL *How many books do you have?*

Verb: *kin* 'to eat'

- (13) a. * kháw lǎay kin He several eat
 - * He several ate.
 - b. * kháw kìi kin he how many eat How many did he eat?

Adverb: yàaŋrûatrew 'quickly'

- (14) a. * khaw ?àan lǎay *yàaŋrûatrɛw* he read several quickly
 - * He read several quickly.
 - b. * khaw ?àan kìi *yàaŋrûatrɛw* he read how many quickly
 - * How many quickly did he read?

The noun $n\acute{a}\eta s \check{r}$ 'book' can be preceded by the quantifiers $l\check{a}ay$ 'several' in (12a), and $k\check{r}$ 'how many' in (12b) while the preposition $k\grave{a}p$ 'with' in (11), the verb kin 'to eat' in (13) and the adverb $y\grave{a}a\eta r\^{u}atrew$ 'quickly' in (14) cannot follow any quantifiers.

2.3 Criteria to identify verbs

Since verbs and adverbs behave similarly, two ways to identify verbs in this position are 1) the co-occurrence with the negation word $m\hat{a}y$ test; and 2) the short answer to a yes-no question test.

2.3.1 The Co-occurrence with the Negation Word mây Test
Words which can co-occur with the negation word mây are verbs (Prasithrathsint 2000).

The Co-ocurrence with the negation word mây

Only verbs may be negated by mây.

Preposition: kàp 'with'

(15) * \(\xi\)? pay tal\(\hat{a}\)at m\(\hat{a}\)y \(\kat{k}\)ap j\(\hat{b}\)m

Ae go market NEG with Joom

Ae did not go to the market with Joom.

Noun: bâan 'house'

(16) * bâan nii mây *bâan* mêe house this NEG house mother *This house is not my mother's house.*

Verb: s# 'to buy'

(17) kháw mây *sɨi* khanŏm He NEG buy sweet *He did not buy sweets.*

Adverb: cháacháa 'slowly'

(18) * kháw dəən mây *cháacháa*He walk NEG slowly *He did not walk slowly*.

Only the verb $s\tilde{H}$ 'to buy' in (17) can be preceded by the negation word $m\hat{a}y$ while the preposition $k\hat{a}p$ 'with' in (15), the noun $b\hat{a}an$ 'house' in (16) and the adverb $ch\hat{a}ach\hat{a}a$ 'slowly' in (18) cannot.

2.3.2 The Short Answer to a Yes-No Question Test:

Words used in giving a short answer to a yes-no question could serve as a test for verbs (cf. Sindhvananda 1970).

The Short Answer to a Yes-No Question Test

Only a verb can be given as a short answer for a yes-no question.

Preposition: kàp 'with'

(19) A: dεεη khuy kàp thee rɨɨplàw Dang talk with you or not

Did Dang talk to you?

B: khuy / * kàp talk / with He did.

Noun: nám?àtlom 'soft drink'

(20) A: thee dim nám?atlom riiplaw you drink soft drink or not Do you drink any soft drink?

B: dɨm / * nám?àtlom drink soft drink I do.

Verb: hiw 'hungry'

(21) A: thəə *hǐw* rɨɨplàw you hungry or not *Are you hungry?*

B: hǐw hungry *I am.*

Adverb: cháacháa 'slowly'

(22) A: the khàp *cháacháa* riiplàw you drive slowly or not *Did you drive slowly?*

B: khàp / * *cháacháa* drive / slowly *I did*.

Only the verb *khuy* 'to talk' in (19), *dɨm* 'to drink' in (20), *hǐw* 'to be hungry' in (21), and *khàp* 'to drive' in (22), can be given as a short answer to a yes-no question. The preposition *kàp* 'with' in (19), the noun *nám?àtlom* 'soft drink' in (20), and the adverb *cháacháa* 'slowly' in (22), on the other hand, cannot.

The following table summarizes the application of each test to nouns, prepositions, verbs and adverbs.

| | Tests | Noun | Preposition | Verb | Adverb |
|-------|--------------------------|------|-------------|------|--------|
| | Stranding | - | - | + | + |
| Nouns | Determiner | + | - | - | - |
| | Quantifier | + | - | - | - |
| | Co-occurrence with the | - | - | + | - |
| Verb | negation word <i>mây</i> | | | | |
| | Short answer to a yes- | - | - | + | - |
| | no question | | | | |

Table 1: Application of each test to nouns, prepositions, verbs, and adverbs

3 Results

This section discusses the word category of the words appearing in the frame: [V (NP) $_$ NP] by applying the following tests: the stranding test in example (b); the determiner test in example (c); the quantifier test in example (d); the co-occurrence with the negation word $m\hat{a}y$ in example (e); and the short answer to a yes-no question test in example (f).

3.1 Prepositions

This study confirms the existence of 23 Thai prepositions in Indrambarya (1994, 1995), namely kìawkàp 'about', kàp 'with', kèe 'to' dèe 'to' tòo 'to', càak² 'from, wâa⁵ 'that', phâa 'for', sămràp 'for', yókwén 'except', chapó? 'in particular' dûay 'with', dooy 'by means of', thặn² 'about', phró? 'because' caròt 'until' con 'till' krathâŋ 'until, conkrathâŋ 'until' yan 'to' yaŋ 'to', sùu 'to', and ná? 'at'. With the criteria used in this paper, I found 12 more prepositions namely, phaaytây 'beneath', 'tây 'beneath', bon 'on', nặa 'above', rim 'edge', nâa 'in front of', lặŋ 'behind', nôɔk 'outside', nay 'in', tâŋtèe 'since', phróɔm 'together with' and khɔɔŋ 'of'.

Preposition nay 'in'

Examples (23a), (23b), (23c), (23d), (23e), and (23f) show the application of the stranding test, the determiner tests, the quantifier test, the co-occurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question to the word nay 'in'.

- (23) a. raw yùu *nay* bâan kháw We stay in house he *We live in his house*.
 - b. * bâan kháw nà? raw yùu *nay* house he TOP we stay in *As for his house, we live in.*
 - c. raw yùu *nay* * day / *?iin we stay in which / other
 - * Which in do we stay?
 - * We stayed in others.
 - d. * raw yùu kii / lǎay nay we stay how many / several in
 - * How many in do we stay?
 - * We stay several in.
 - e. * raw yùu mây *nay* bâan kháw we stay NEG in house he *We did not live in his house.*
 - f. A: kháw yùu *nay* bâan riiplàw he stay in house or not *Is he in the house?*

The word nay cannot be left stranding at the end of the sentence, just as prepositions and nouns in (23b). It cannot be modified by any determiners in (23c) or any quantifiers in (23d) and hence cannot be a noun. The word nay 'in' is not a verb because it cannot cooccur with the negation word may in (23e), nor can it be given as a short answer to a yesno question in (23f). The word nay is hence a preposition.

Preposition thin 'about'

The application of the stranding test in (24b), the determiner tests in (24c), the quantifier test in (24d), the co-occurrence with the negation word $m\hat{a}y$ test in (24e), and the short answer to a yes-no question test in (24f) are shown as follows.

- (24) a. kháw phûut *th*ặŋ nóoŋ chán he speak about younger sibling I *He spoke of my younger sibling*.
 - b. nóon chán nà? kháw phûut *th*ặŋ younger sibling I TOP he speak about *My younger sibling, he spoke of him/her.*
 - c. kháw phûut *th*ặŋ *day / *ʔɨɨn he speak about which / other
 - * Which of did he speak?

He spoke of others.

- d. * kháw phûut kìi / lǎay *th*ặŋ he speak how many / several about
 - * How many about did he speak?
 - * He spoke several about.
- e. * kháw phûut mây *th*ặŋ nóoŋ chán he speak NEG about younger sibling I *He did not speak of my younger sibling*.
- f. A: kháw phûut thặn nóon chán rɨiplàw he speak about younger sibling I or not Did he speak of my younger sibling?

 B: phôut /* thờn

B: phûut /* thặŋ speak about *He did*.

Example (23c) and (23d) show that $th + \eta_2$ 'about' does not reflect the characteristics of a noun because it cannot be modified by any determiners or any quantifiers. The word $th + \eta_2$ can be left stranding in (23b), a characteristic of verbs and adverbs. However, since it cannot be negated with $m\hat{a}y$ as in (23e) nor can it be given as a short answer to a yes-no question, $th + \eta_2$ is not a verb. Hence $th + \eta_2$ 'about' is regarded as a preposition. The fact that it can be stranded may show that $th + \eta_2$ is a preposition grammaticalized from the homophonous verb $th + \eta_1$ 'to reach' and that the process is not fully completed.

Preposition càak2 'from'

The application of the stranding test, the determiner test, the quantifier test, the cooccurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question are shown in (25b), (25c), (25d), (25e) and (25f), respectively.

- (25) a. kháw maa *càak* tʃiaŋmày he come from Chiangmai *He came from Chiangmai*.
 - b. * tʃiaŋmày nà? kháw maa *càak* Chiangmai TOP he come from *Chiangmai, he came from there*.
 - c. kháw maa *càak* *day / *?iin he come from which / other
 - * Which from did he come?
 - * He came from others.
 - d. * kháw maa kìi / lǎay *càak* he come how many / several from
 - * How many from did he come?

e. kháw maa mây *càak* tʃiaŋmày he come NEG from Chiangmai *He did not come from Chiangmai*.

f. A: kháw maa *càak* tʃiaŋmày rɨɨplàw he come from Chiangmai or not *Did he come from Chiangmai?*

B: maa /* càak come / from *He did*.

 $c\grave{a}ak_2$ cannot be stranded at the end of the sentence as in (25b). It can not be negated with $m\^ay$ as in (25e), nor can it be given as a short answer to a yes-no question in (25f). It is then not a verb. That it cannot co-occur with any determiners in (25c) or any quantifiers in (25d) show that $c\grave{a}ak_2$ 'from' is not a noun, but a preposition, having a homophonous verb $c\grave{a}ak_1$.

3.2 Nouns

Nouns in Thai appearing in this frame, are, for example, thîi 'place', tron 'straight, at', thaan 'way', dâan 'side', khâan 'side', mîa 'when' and yàan 'type'.

Noun thîi 'place'

The application of the stranding test, the determiner test, the quantifier test, the cooccurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question test are shown in (26b), (26c), (26d), (26e), and (26f), respectively.

- (26) a. raw phóp kan *thîi* mɨaŋnɔɔk
 We meet each other place abroad
 We met each other abroad.
 - b. * mɨaŋnôɔk nà? raw phóp kan thîi abroad TOP we meet each other place Abroad, we met each other at.
 - c. raw phóp kan *thîi* day / ?ɨɨn we meet each other place which / other *Which place did we meet?*We met other places.
 - d. raw phóp kan kii / lǎay thîi we meet each other how many / several place How many places did we meet?

 We met each other in a few places.
 - e. * raw phóp kan mây *thîi* mɨaŋnɔɔk we meet each other NEG place abroad *We did not meet each other abroad.*

f. A: raw phóp kan *thîi* mɨaŋnɔ̂ɔk rɨɨplàw We meet each other place abroad or not *Did we meet each other abroad?*

B: phóp / * thîi meet / place We did.

The word *thîi* 'place' does not show any characteristics of a verb since it cannot be negated by *mây* in (25e) and it cannot be given as a short answer to a yes-no question in (25f) either. The word *thîi* 'place' cannot be stranded as in (25b). It must be a preposition or a noun. The determiner test in (25c) and the quantifier test in (25d) reveal that it is a noun.

3.3 Verbs

Verbs which are traditionally referred to as prepositions are, for example, *klây* 'to be near', *klay* 'to be far', *th*#ŋ₁ 'to reach', *taam* 'to follow', *khâam* 'to cross', and *thûa* 'throughout'.

Verb klây 'to be near'

The application of the stranding test, the determiner test, the quantifier test, the cooccurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question test are shown in (27b), (27c), (27d), (27e), and (27f), respectively.

- (27) a. bâan kháw yùu *klây* bâan tʃán house he stay near house I *His house is near my house.*
 - b. bâan tfán nà? bâan kháw yùu *klây* house I TOP house he stay near *As for my house, his house is near.*
 - c. bâan kháw klây *day / *?ɨin house he near which / other
 - * Which near is his house?
 - * His house is other near.
 - d. bâan kháw kìi / lǎay *klây* bâan tʃán house he how many / several near house I
 - * How many near my house is his house?
 - * His house is several near my house.
 - bâan kháw yùu mây klây bâan t∫án e. house he stay **NEG** near house I His house is not near my house.
 - f. A: bâan kháw yùu *klây* bâan tʃán rɨɨplàw house he stay near house I or not *Is his house near my house?*

B: yùu / klây stay / near It is.

As illustrated above, $kl\hat{a}y$ 'near' cannot be modified by any determiners in (27c) or any quantifiers in (27d), and hence it is not a noun. $kl\hat{a}y$ 'to be near' can be left stranding at the end of the sentence in (27b) and can be negated with $m\hat{a}y$ (27d). It can also be given as a short answer to a yes-no question in (27f), a characteristic of a verb. The word $kl\hat{a}y$ 'near' then is a verb.

Verb thin, 'to reach'

The application of the stranding test, the determiner test, the quantifier test, the cooccurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question are applied as shown in (28b), (28c), (28d), (28e), and (28f), respectively.

- (28) a. kraproon yaaw *thɨŋ*₁ khàw skirt long reach knee *The skirt is at knee's length.*
 - b. khàw nà? kraproon yaaw *thťŋ*₁ he TOP skirt long reach As for knees, the length of the skirt reach it.
 - c. kraproon yaaw *th¥ŋ*₁ *day /*ʔ¥in skirt long reach which / other
 - * Which reach is the skirt long?
 - * The skirt's length reached others.
 - d. kraproon yaaw * kɨi / * lǎay thɨŋ₁ skirt long how many / several reach
 - * How many reach is the skirt's length?
 - * The skirt's length has several reach.
 - e. kraproog yaaw mây $th\check{\imath}\eta_1$ khàw skirt long NEG reach knee
 - * The skirt does not reach one's knees.
 - f. A: kraproon yaaw thěŋı khàw rɨɨplàw skirt long reach knee or not Does the skirt reach one's knees?

```
B: yaaw / thɨŋ<sub>1</sub> long / reach It does.
```

As illustrated, $th\check{t}\eta_1$ 'to reach' can neither be modified by any determiners in (28c) nor any quantifiers in (28d), and hence it is not a noun. The word $th\check{t}\eta_1$ 'to reach' can be left stranding at the end of the sentence in (28b) and can be negated with $m\hat{a}y$ in (28e). It can also be given as a short answer to a yes-no question in (28e), a characteristics of a verb. The word $th\check{t}\eta_1$ then is a verb.

3.4 Adverbs

This study finds that adverbs which may appear in this frame include $h\hat{a}y_6$ 'for' and $w\acute{a}y$ 'keeping'.

Adverb hây, 'for'

The application of the stranding test, the determiner test, the quantifier test, the cooccurrence with the negation word $m\hat{a}y$ test, and the short answer to a yes-no question are applied, as shown in (29b), (29c), (29d), (29e), and (29f), respectively.

- (29) a. kháw thii krapăw *hây*₆ nóɔŋ he carry purse for younger sibling *He carried a purse for his younger sibling.*
 - b. nóoŋ nà? kháw thii krapǎw *hây*₆ younger sibling TOP he carry purse for *As for his younger sibling, he carried a purse for her.*
 - c. kháw thii krapǎw *hây*₆ *day / *ʔɨin he carry purse for which / other
 - * Which for did the he carry a purse?
 - * He carried a purse for others.
 - d. kháw thii krapăw * kii / * lǎay *hây*₆ he carry purse how many / several for
 - * How many for did he carry the purse?
 - * He carried the purse several for.
 - e. * kháw th \tilde{H} krap \tilde{H} krap \tilde{H} mây \tilde{H} nóɔŋ he carry purse NEG for younger sibling
 - * He did not carry the purse for his younger sibling.
 - f. A: kháw thii krap<u>a</u>w *hây*₆ nóɔŋ ríiplàw he carry purse for younger sibling or not *Did he carry a purse for his younger sibling?*

```
B: th<del>ii</del> /* hây<sub>6</sub>
carry / for
He did.
```

As illustrated, $h\hat{a}y_6$ 'for' cannot be modified by any determiners in (29c) or any quantifiers in (29d), and hence it is not a noun. The word $h\hat{a}y_6$ 'for' can be left stranding at the end of the sentence in (29b), a characteristic of a verb and an adverb. However, it does not behave like a verb because it cannot be negated with $m\hat{a}y$ in (29e), nor can it be given as a short answer to a yes-no question in (29f). The word $h\hat{a}y_6$ 'for' has lost its original meaning 'to give'. It is regarded as an adverb grammaticalized from the homophonous ditransitive verb $h\hat{a}y_1$ 'to give'.

4 Conclusions

This paper illustrates that words which are traditionally referred to as prepositions are synchronically reanalyzed as various syntactic categories: nouns, prepositions, verbs and adverbs. It is found that there are 35 prepositions in Thai. namely, *kìawkàp* 'about', *kàp*

'with', kèe 'to' dèe 'to' tòo 'to', càak2 'from, wâa5 'that', phâa 'for', sămràp 'for', yókwén 'except', chapó? 'in particular' dûay 'with', dooy 'by means of', thɨŋ² 'about', phró? 'because' caròt 'until' con 'till' krathâŋ 'until, conkrathâŋ 'until' yan 'to' yaŋ 'to', sùu 'to', and ná?, phaaytây 'beneath', 'tây 'beneath', bon 'on', nɨa 'above', rim 'edge', nâa 'in front of', lǎŋ 'behind', nôɔk 'outside', nay 'in', tâŋtèe 'since', phróɔm 'together with' and khɔɔŋ 'of'.

References

- Diller, Anthony. 2001. Grammaticalization and Tai Syntactic Change. In Tingsabadh and Abramson (eds.) *Essays in Tai Linguistics*, 139-175. Bangkok: Chulalongkorn University Press.
- Indrambarya, Kitima. 1994. *Subcategorization of Verbs in Thai*. Ph.D. Dissertation. University of Hawai'i.
- _____. 1995. Are there Prepositions in Thai? *Papers from the Third Annual Meeting of the Southeast Asisan Linguistic Society*. Arizona State University Press. Ed. By Mark Alves. pp. 101 118.
- Kullavanijaya, Pranee. 1974. Transitive Verbs in Thai. Ph.D. Dissertation. University of Hawai'i.
- Prasithrathsint, Amara. 2000. From Nouns to Prepositions: A Grammaticalization Process in Thai. Paper Presented at the International Symposium on Parts of Speech in and across Languages.
- ______. To appear. Lak Samkhan Nay Karnjamnaek Chanit Khong Kham Nay Phasa Thai (Important Criteria in identifying Thai Word Class). In Prasithrathsint, Rungrojsuwan, and Sankaburanurak, (eds.). Proceedings of the Pan-dialectal Grammar of Thai Symposium on Diversity of Languages in Thai Society, Standard Thai, Thai Regional and Social Dialects: Issues and New Discoveries. Chiang Rai: Chopphim Publishing.
- Savetamalya, Saranya. 1989. *Thai nouns and noun phrases: a lexicase analysis*. Ph.D. Dissertation. University of Hawai'i.
- Sindhvananda, Kanchana. 1970. *The Verbs in Modern Thai*. Ph.D. dissertation. Georgetown University.
- Singnoi, Unchalee. 2000. *Nominal Constructions in Thai*. Ph.D. Dissertation. University of Oregon.
- Starosta, Stanley. 1994. Identification of Word Class in Thai. Ms. Revised version of a Paper presented at the Fifth International Conference on Thai Studies at the Center for Southeast Asian Studies. School of African and Oriental Studies in London. July 1993.
- Takahashi, Kiyoko. 2005. Allative Preposition in Thai. In Sidwell, Paul (ed.) *Papers from the 15th Annual Meeting of the Southeast Asian Linguistics Society*. 2005, 111-120.
- Warotamasikhadit, Udom. 1992. There are no prepositions in Thai. In Bamroongraks, et al. (Eds.) *The International Symposium on Language and Linguistics*. Thammasart University Press. Pp. 70 76.

DEIXIS AND INFORMATION STRUCTURE IN MON - THE MULTIFUNCTIONAL PARTICLE kôh

Mathias Jenny University of Zurich <jenny@spw.uzh.ch>

1. Abstract

The topic of this study is a multifunctional particle in Mon, viz. $k\partial h$. This morpheme is found in all stages of Mon since the 11^{th} century, where it appears as a deictic noun phrase marker indicating distal spatial and temporal deixis. The use and development of $k\partial h$ exhibits a number of semantic and pragmatic features that make it difficult to define its actual function(s). The aim of this paper is to present data from Old, Middle and Modern Mon and try to find explanations of the development paths and uses of this very frequent particle.

In the next section, examples from all recorded stages of Mon will be presented, followed by previous studies and definitions of the morpheme under discussion. Section 3 is concerned with demonstratives in general and the position of $k \partial h$ in the system of demonstratives in Mon. Section 4 considers the possibility of explaining $k \partial h$ as a marker of definiteness in Modern Mon. Section 5 takes a look at the pragmatic functions of $k \partial h$, i.e. as a marker of topic or identifiable information.

2. History and previous studies

2.1 The development of deixis in Mon

Old Mon seems to have had a two way distinction in its deictic system. Proximal deixis was expressed by wo?, distal by goh. The exophoric use of wo? is clearly present in expressions like (1), describing the scene in an accompanying picture.

(1) Old Mon (Ananda plaques)

wo? brumha ku ?in. this Brahma COM Indra

These are Brahma and Indra.

The originally probably distal deictic *goḥ* is used mainly anaphorically, either as noun phrase marker or in adverbial expressions such as *gun goḥ* 'therefore, by virtue of that' and *blaḥ goḥ* 'afterwards, having finished that'. The different uses of *wo?* and *goḥ* are seen clearly in the frequent expression *row wo?* 'thus, like this' to cataphorically introduce direct speech, while *row goḥ* anaphorically follows the quotation. This use is illustrated in sentence (2), where the first instance of *wo?* is cataphoric, while it is exophoric in the reported speech. This example of exophoric *wo?* also shows that it is not primarily speaker-centred, but rather spatial proximal.

(2) Old Mon (Shweizigon inscription)

| smāñ | row | wo? : | ma | tirla | pa | k∢ir>?im | wo? | ci |
|----------|-------------|--------------|-----------|--------|----------|------------------|-------|------------------|
| ask | manner | this | ATTR | master | do | <nml>smile</nml> | this | EMPH |
| mu | het | yo? | row | goḥ | tarley | ?ānan s | smāñ | da. ¹ |
| what | reason | Q | manner | that | master: | 1s Ananda a | ısk | FOC |
| He ask | ed like t | his: Who | at is the | reason | for this | smile you h | nad o | on your face? |
| Like tha | at our lore | d Anando | ı asked. | | | | | |

The deictic system of Old Mon is summarised in table 1.

Table 1: Demonstratives of Old Mon

| Form | Exophoric uses | Discourse functions |
|------|----------------|---------------------|
| wo? | proximal | cataphoric |
| goḥ | ? | anaphoric |

By Middle Mon, the deictic system was expanded by two new elements, the distal *gah*, i.e. the MM spelling of OM *goḥ*, now competing with *te?* and the proximal *wo?*, variously spelt *wo?*, *wwo?*, *wwa?*, *wa?*, *wwå?* in MM, being replaced by the adverbial *?ano?* 'here' in some contexts. The exact use of these new deictics is not known and their occurrence is not very frequent in Middle Mon texts.

It is not certain whether the forms *te?* and *?ano?* were in fact absent from Old Mon or whether their absence from the documents is due to the incompleteness of the data. They may have belonged to a less formal register of the language and were therefore not used in literary documents. It should be noted that in closely related Nyah Kur, widely believed to be a direct offshoot of Old Mon in Thailand, there are the forms *te?*, *te?* and *tee?*, marking three degrees of distal deixis (L.Thongkham 1984:173f). The proximal *?ano?* has good cognates in other Austroasiatic languages (cf. Shorto 2006:90f) and appears in Nyah Kur in the expression *potam noo?* 'tonight' (Diffloth 1984:147).

In MM wo? is still used exophorically as proximal deictic and cataphorically to introduce direct speech, and gah anaphorically closes quotations. In addition, gah is used increasingly to mark NPs in prominent position, e.g. fronted objects. As Burmese influence becomes stronger during the Middle Mon period around the 14th century, the frequent fronting of objects, originally probably for pragmatic reasons, can be seen as influence of Burmese OV structures. This is illustrated in (3).

In transliteration of Old and Middle Mon, I basically follow the common spelling conventions as used in Shorto 1971, with the following two exceptions: For the glottal stop \mathfrak{B} I use the IPA symbol in all positions and for the Mon-Burmese digraphic vowel symbol $\mathfrak{S}_{\mathbb{Q}}^{\circ}$, I use the transliteration $\langle iu \rangle$ as suggested by Yanson (2002), which corresponds to the ordering of the symbols in indigenous spelling in both Mon and Burmese (in spite of the tradition among Western (and consequently also Burmese) scholars to use $\langle ui \rangle$). A still better solution would be to use a separate symbol in transliteration altogether (e.g. $\langle \mathfrak{S} \rangle$, which is probably close to the intended sound in Old Burmese and Old Mon), similar to the use of $\langle \mathfrak{S} \rangle$, which is made up of $\langle \mathfrak{S} \rangle$ and $\langle \mathfrak{S} \rangle$ in Indic scripts.

Mon Particle kòh

Middle Mon (Shwedagon inscription) (3) ?idhi gah dhāt swok gah ket ket tuy. power that relic INSTR that take FINISH hair TAKE With that power he took the hair relic by himself.

That *gah* was not grammaticalised as a marker of objects is shown in sentence (4), where it is the subject that receives the marker.

(4) Middle Mon (Shwedagon inscription)
tapussa **gah** go? sotāpatiphiuw ra.
Tapussa that get conversion FOC
Tapussa achieved religious conversion.

Another frequent use of *gah* in Middle Mon is at the end of adverbial expressions, similar to Old Mon usage, functioning as a kind of phrase boundary marker, as in sentences (5) and (6), both formally relative clauses (both from the Shwedagon inscription).

- (5) het ñaḥ ma ha go? liṅwor pūjau **gah** kium nḍa. reason person ATTR NEG get pay.homage worship that ADD FOC *It was because they could not worship and pay homage [to the relics].*
- (6) brau sāmanī truh sāmanī brau khā gamī truh gamī female novice male novice female time monk male monk gah, khā ma nwom tuy gah that time that ATTR exist **FINISH** When there are male and female monks, male and female novices, at that time...

The deictics of Middle Mon are summarised in table 2.

Table 2: Demonstratives of Middle Mon

| Form | Exophoric use | Discourse function |
|-------|---------------|--------------------|
| Pano? | 'here' | ? |
| wo? | proximal | cataphoric |
| gah | ? | anaphoric |
| te? | distal | ? |

In modern spoken Mon, a three way distinction has emerged with $k\partial h$ functioning as medial deictic. The form wu'? from Middle Mon wo'? is used only in literary and formal style where it is apparently freely interchangeable with no?. It is replaced by no? in the spoken language. The exophoric uses of no?, $k\partial h$ and tv'? are rather straight forward, the choice being based on relative distance from the point of reference. The cataphoric-anaphoric distinction made in Old and Middle Mon is no longer followed in the spoken language, though there are traces of it in literary Mon. Sentences (7a-c) illustrate the exophoric deictic use, with the object placed at an increasing distance from the point of

reference. Sentence (8), uttered after the speaker finished telling a story, shows the anaphoric use of no?.

- (7) a. ket 6è? no?.² take REF PROX *Take this one.*
 - b. ket 6è? **kòh**. take REF MEDL Take that one.
 - c. ket 6\(\epsilon\)? **tx?**.
 take REF DIST

 Take that one over there.
- (8) ?əkhak **nɔ?** ra? lèə kጵ? ra?. manner PROX FOC tell GET FOC That's how I can tell [stories].

While $n\partial$? and tv? behave like typical demonstratives (see section 3 below), $k\partial h$ has a number of features and functions that differentiate it from its proximal and distal counterparts. In any given text in Mon, $k\partial h$ is easily the most frequent word. It occurs in a large number of sentences, often more than once in a single sentence. Among the functions of $k\partial h$ are the marking of noun phrases as in (9), including complex NPs as in (10) with an unmarked relative clause, and the marking of clauses as complements (11) or adverbial clauses (12).

- (9) mòŋ kàh pγ ηèə Gcb pəlen kòh. **STAY** frog MEDL bottle watch LOC **MEDL** They were watching the frog in the bottle.
- (10)kỳ? 6è? tèh kon həkao? klày kòh. ηèə REF child get HIT frog body search **MEDL** He found the young frog he was looking for.
- (11)nèh hù? məkèh hù? kỳ? çiə? kòh, ?uə kp if person NEG give NEG **GET** eat MEDL 1s ra?. tem mòŋ know **STAY FOC**

I know that I cannot eat anything if they don't give me [food].

Spoken Mon data are transcribed according to the phonological system developed in Jenny 2005 (pp. 23-42). Where quoting other sources, the original spelling is retained.

Mon Particle kòh

| (12) | kla before | deh 3 | hù? NEG | tèh HIT | ?op surrender | kðh , MEDL | deh 3 | priəŋ prepare | | |
|------|----------------|----------|------------|------------|-------------------|----------------------|----------|------------------|--|--|
| | là senat.həlàk | | kòh. | | | | | | | |
| | DEPOSIT | cannon | | MEDL | | | | | | |
| | Before i | they had | to surrei | nder, they | v prepared that o | cannon. | | | | |

In some cases it is not clear what the scope of $k \partial h$ is in an expression, as shown in (13).

(13) la? həmèə pən həcpt thp? rəə kwan poy kðh when Burmese shoot kill DISCARD group village 1p MEDL when the Burmese shot the people in our village

In this sentence, $k\partial h$ can have scope over the pronoun poy 'we', over the NPs $kwan\ poy$ 'our village' or $r\partial a$ $kwan\ poy$ 'the people in our village', or over the whole adverbial clause.

The deictic system of Spoken Mon is summarised in table 3.

Table 3: The demonstratives of Spoken Mon

| Form | Exophoric use | Discourse function |
|------|---------------|-----------------------|
| no? | proximal | anaphoric, cataphoric |
| kòh | medial | anaphoric |
| tx? | distal | anaphoric |

The frequency of use of $k\partial h$ suggests that it is more strongly grammaticalised than the proximal and distal demonstratives. The functions of grammaticalised $k\partial h$ will be discussed in sections 4 and 5. I will now first turn to an overview of earlier studies and explanations of this particle.

2.2 Previous studies and definitions of kh

The early western grammars and dictionaries of Mon describe *kh* variously as a (deictic) pronoun or adjective 'that' (Halliday 1955 [1922]:115), or as

that; also a sign of the accusative case before the verb; it is often emphatic, laying stress on the word or phrase which precedes it. [...] As an emphatic particle = the Burmese $\langle k\bar{a}h\rangle$ kà;³ very often it is simply indicative of the subject, and is equivalent to the nominative case. (Duroiselle1962 [1921]:171f)

Shorto (1971) gives the following definition for *goḥ* in Old Mon and *gah* in Middle Mon:

[&]quot;kà: 1 (Phr~) as for Phr; however, but; highlighting the subject or topic of a sentence, or contrasting it to a different topic." (Okell and Allott 2001:5)

goḥ, goḥ, goḥ /goḥ/ ns. & n $(ns.)^4$ deictic, that, and otherwise with nn. and n.phr. with mooted referent, esp. at head of sentence, the, then often serving to mark end of complex n.phr. [...] **gah**, rarely **gāh** ns., n phr. marker, rarely deictic except in such fixed phr. as $kh\bar{a}$ gah then, het gah ra therefore [...]. (pp.82f)

In his dictionary of Spoken Mon (1962), he explains $k \partial h$ as

Particle marking nn. and nominal phr., esp. in initial prominence position, and occasionally complete sents. [...] kòh is occasionally deictic, esp. in phr. following **doa** ['in, at'], **nù** ['from'], where it stands as n. (pp.78f).

The most extensive study of Mon is Bauer's Morphology and Syntax of Spoken Mon (1982). He dedicates three pages to $k\partial h$, which he calls a clitic or particle marking "determination". Other nominal categories listed by Bauer are possession, plurality, and deixis. The markers of all these categories follow the head noun in this order, with determination occupying the final slot. Bauer does not elaborate on his use of the term "determination", but from the discussion and examples it looks like he is referring to definiteness: "The most versatile clitic and, on some contexts, particle of the nominal phrase is $\underline{k\partial h}$; in most environments, it translates into English 'the'." (Bauer 1982:322f) Another function of $\underline{k\partial h}$ mentioned by Bauer is "marking a boundary of any nominal phrase" (p. 323). Unlike Halliday, Bauer does not see $\underline{k\partial h}$ as a marker of grammatical relations or case marker:

In no case can $\underline{k\grave{o}h}$ function as overt marker of grammatical subject or object. [...] in a sequence of subject verb object where subject and object may be simple or compound nouns with any number of noun-clitics either noun (subject or object) may be followed by $\underline{k\grave{o}h}$. (p. 324)

In other cases $k\partial h$ is explained as marking a noun phrase as emphatic or in "prominence position", while Bauer's translations ('as for ...') suggest that $k\partial h$ functions to mark a nominal expression as topic. The use as emphatic marker "is particularly obvious in cases where the noun position, preceding $\underline{k\partial h}$, is occupied by a personal name or a personal pronoun" (p. 325).

Later in his study, Bauer states that $k\partial h$ is the "only native and 'natural' device to nominalize verbs or incorporating verbal phrases into complex NPs" (p. 331). He concludes that, while $k\partial h$ can be used to mark subordinate (relative) clauses,

it is inappropriate to call these sentence types 'relative clauses' [...] but rather to interpret them as nominalized VPs and to retain <u>kòh</u> simply as a determining and nominalizing (or noun) particle. (p. 332)

We have seen that for earlier authors $k \partial h$ was part of the deictic system with some special functions, Bauer explains $k \partial h$ as belonging to a distinct category, i.e. "determination". The main functions of $k \partial h$ as given by Bauer are

_

⁴ n. = noun, n.phr. = noun phrase, ns. = noun suffix

Mon Particle kòh

- 1. Determination (i.e. definiteness?)
- 2. Emphasis
- 3. Marking of boundary of nominal phrase
- 4. Nominalisation of verbal phrases.

In the following sections more data from Modern Mon will be given and checked against the theoretical properties of the categories to which $k\partial h$ may belong or has been analysed as belonging to.

3. Demonstratives

3.1 Theoretical overview

The most comprehensive survey of demonstratives is probably Diessel (1999), which covers morphological, syntactic, semantic and pragmatic aspects of demonstratives in a crosslinguistic perspective and also explores grammaticalisation paths involving demonstratives. Diessel gives three criteria that are relevant for the notion of demonstratives:

First, demonstratives are deictic expressions serving specific syntactic functions. [...] The notion that I will use [...] subsumes not only demonstratives being used as pronouns or noun modifiers but also locational adverbs such as English *here* and *there*.⁵

Second, demonstratives generally serve specific pragmatic functions. They are primarily used to focus the hearer's attention on objects or locations in the speech situation [...], but they may also function to organize the information flow in the ongoing discourse. [...] Demonstratives are often used to keep track of prior discourse participants and to activate specific shared knowledge. [...]

Finally, demonstratives are characterized by specific semantic features. All languages have at least two demonstratives that are deictically contrastive: a proximal demonstrative [...] and a distal demonstrative. (Diessel 1999:2)

According to Diessel, demonstratives have two main functions, viz. exophoric and endophoric (Diessel 1999:6). The exophoric use is seen as more basic and historically predating the endophoric uses. Languages vary in the distinctions they make in exophoric demonstratives. Most common are languages with a two-way or three-way distinction, based either on the relative distance from the point of reference or on the closeness to speaker/hearer/other person. Typically exophoric demonstratives are accompanied by a pointing gesture towards the object of reference in the discourse situation.

The endophoric use is divided into anaphoric, discourse deictic and recognitional. Diessel (1999:93) gives the following definitions of these terms:

Anaphoric and discourse deictic demonstratives refer to elements in the ongoing discourse. [...] Anaphoric demonstratives are coreferential with a prior NP; they keep track of discourse participants. Discourse deictic demonstratives refer to propositions; they link the clause in which they are embedded to the proposition to which they refer.

⁵ It is noteworthy that Diessel does not include other adverbial demonstratives (e.g. temporal, manner) in his survey.

Recognitional demonstratives do not refer to elements in the surrounding discourse; rather, they are used to indicate that the hearer is able to identify the referent based on specific shared knowledge.

This classification of demonstratives is largely identical to one proposed by Himmelmann (1996), with some differences in terms of labels rather than content. Himmelmann uses "situational use" for what Diessel calls exophoric, and "tracking use" for Diessel's anaphoric. Diessel and Himmelmann do not mention the cataphoric use of demonstratives as seen in example (2) above. If we add this notion, we get the following possibilities for demonstratives:

- 1. Exophoric: Proximal, medial, distal (language specific scales)
- 2. Endophoric: Anaphoric, cataphoric, discourse deictic, recognitional

Demonstratives can occur either as pronouns or as attributes to nouns (adnominal demonstratives). In some languages there are different forms for adnominal demonstratives and pronominal demonstratives, while others use some kind of derivation to derive pronouns from adnominal demonstratives. The combination of an adnominal demonstrative with a relator noun or a demonstrative pronoun with an adposition can result in an adverbial expression, as in English 'this way, that way' and 'like this, like that'. For lack of adequate data, Diessel mentions this adverbial use of demonstratives but does not give any details or further discussion (Diessel 1999:74). It is especially in this use, though, that the cataphoric – anaphoric distinction is relevant in many languages, including older stages of Mon.

Demonstratives tend to grammaticalise along different but consistent paths across languages. The most common endpoints of these grammaticalisation paths include

- 1. Pronouns (third person and relative)
- 2. Complementisers
- 3. Sentence connectives
- 4. Possessives

from pronominal demonstratives and

- 1. Definite articles or noun class markers
- 2. Boundary markers of postnominal relative clauses or attributes
- 3. "Determinatives"

from adnominal demonstratives (Diessel 1999:119ff).

"Determinatives" according to Diessel (1999:135) are demonstratives that mark the nominal head of a relative clause or function as head of a relative clause. This use is clearly different from Bauer's label "determination" for the Mon particle $k\partial h$ seen above.

I will now turn to more data from Modern Mon to see how the particle $k\partial h$ fits in the category of demonstratives.

3.2 Demonstratives in Modern Mon

As seen above, Modern Mon has a deictic system with three members based on relative or absolute distance from the point of reference, usually i.e. the speaker in the spatial dimension and the present time in the temporal dimension. The basic forms in Mon, i.e.

Mon Particle **kòh** 61

 $n \ni ?$, $k \ni h$ and $t \not r ?$, are mainly used adnominally though they may occur as quasipronominals after prepositions like $d\theta\theta$ 'LOCATIVE' and $n\dot{u}$ 'ABLATIVE'. All three can be used to refer to objects in the real world located at different distances from the place of the speaker. They therefore count as demonstratives according to Himmelmann's defining characteristic, i.e. "the element must be in a paradigmatic relation to elements which when used exophorically – locate the entity referred to on a distance scale: as proximal, distal, etc." (Himmelmann 1996:210f), as will be seen below, the Mon demonstratives also satisfy Diessel's characterising criteria given above. The medial form can be used to indicate closeness to the hearer, but this does not necessarily have to be the case. If the speaker is located between the hearer and the object referred to, he still uses $k\partial h$ if the object is some distance away from him. In the temporal dimension all three deictics can refer to a point in the past or future, but only $k \partial h$ seems to be used with relative time reference. As relative time reference always implies anaphoric use, the use of $k\partial h$ as the most frequent anaphoric particle is not surprising in this context. In some contexts, exophoric tx? seems to be used merely to express a great distance in time or space without pointing to a specific referent, as seen in (14). This use is not possible with $k \partial h$ and $n \partial l$, which always point to a specific referent when used exophorically as in (15) and (16).

- (14) kla **tx?** ... before DIST

 Long time ago ...
- (15) pətəm nù **nɔ?** start ABL PROX from now on
- (16) pətəm nù **kòh** start ABL MEDL from then on, from that time

All three demonstratives have derived nominal forms functioning as demonstratives pronouns or locative adverbs. In most cases the prefix 2i?- forms pronouns and 2a- forms locative adverbs, but there is some inconsistency in the use of the forms, i.e. the forms with 2i?- prefix are sometimes used as locative demonstratives. The general adverbial and temporal forms originate in collocations of a nominal head with the adnominal demonstratives, sac 'manner' for the former and la? (from Mon/Pali $k\bar{a}la$ 'time') for the latter. While the presence of a separate lexeme hamù h 'now' explains the gap in the proximal-temporal slot, no explanation can be given for the gaps in the distal-adverbial and distal-temporal slots. I will return to the gap in the medial-plural slot below. Table 4 summarises the demonstratives and their derivate forms in Modern Mon.

The prefix ?i?- has other grammatical functions besides nominalisation. It is realised as -y- infix after some initials, including the ones occurring in the demonstratives, causing palatalisation (cf. Jenny 2003).

The proximal-adverbial has another, less common form, viz. kənə?, obviously contracted from kiəŋ nə? 'this habit, this custom'.

| | Basic form | Nominalised | Locative | Adverbial | Temporal | Plural |
|------|------------|--------------|----------|-----------|-----------|-----------|
| PROX | no? | ?i?nɔ? / ɲɔ? | ?ənə? | hənə? | (həmùh) | tənə? |
| MEDL | kòh | ?i?kòh / còh | ?əkòh | həkòh | ləkòh | (tɔʔ kòh) |
| DIST | tr? | ?i?tr?/cr? | ?ətr? | (sac tx?) | (la? tx?) | tətx? |

Table 4: *Basic demonstratives and derivate forms in Mon*

As seen above (table 3), the use of all three demonstratives has been extended to anaphoric use, but only $k\partial h$ has also acquired discourse deictic and recognitional functions. In cataphoric use, only the proximal $n\partial l$ seems to be used. In anaphoric function, $k\partial h$ is by far the most frequent. The proximal and distal particles are chiefly used to explicitly indicate closeness or distance either in the discourse or in the real world, i.e. there is a close relationship between exophoric and anaphoric use.

Although no?, kòh and tr? share a number of common features, there are some characteristics that distinguish $k\partial h$ from the other two demonstratives. The first important difference is the singularising effect that is common to n3? and tx?, but not $k \ge h$. Mon has two nominal plural markers, viz. to? for definite or inclusive plural and həlàn for indefinite or open plural. If the number of referents of a nominal expression is either known from the context or irrelevant to the discourse, plurality is not overtly marked, except for personal pronouns. The pure nominal klp can therefore mean '(the/a) dog' or '(the/some) dogs'. If the noun is modified by a proximal or distal demonstrative, it becomes singular. Plural referents must in this case obligatorily be marked by either the definite or the indefinite plural marker. This is not the case with the medial demonstrative, which does not necessarily imply singularity of the referent, irrespective of the use as exophoric or anaphoric deictic. The expression klp no? can only be interpreted as 'this dog'; 'these dogs' is always klp tənə? or klp həlàn nə?. The same is true for klp ty? 'that dog' and its plural forms klp tətx? and klp həlàn tx?. Compare with these the use of kh in sentences (17) and (18), both from a recorded conversation. While the plural is overtly marked in (17), it is implied in (18) by the use of the quantifier hə?vt 'all'.

- (17) ?εŋkəlòc **tɔ? kòh** lε ?a ləkòh. English PL MEDL ADD go then *Then the Englishmen went away*.
- (18) la **kòh** tho? tho? hə?ot. donkey MEDL discard DISCARD all *They left all the donkeys there*.

The fact that the plural marker is obligatory with $n\mathfrak{d}$ and $t\mathfrak{d}$ but not with $k\mathfrak{d}h$ explains why only the former two occur in the contracted form with the definite plural marker as prefix $t\mathfrak{d}$ -, while $t\mathfrak{d}$ is never shortened before $k\mathfrak{d}h$. That the collocation of a noun with a demonstrative entails singularity of the referent is a common feature in many languages of

_

⁸ There is a special pronoun for the first person plural, viz. *poy*. Second and third person pronouns always receive the definite plural marker with plural referents.

Mon Particle **kòh** 63

Southeast Asia, including Burmese, which has two grammatical morphs marking plurality (optional in most contexts except for pronouns) and Thai, which lacks fully grammaticalised plural markers. This fact sets $k\partial h$ apart from the other demonstratives, not only in Mon, but also in an areal perspective.

Secondly, $k \partial h$ can occur after constituents other than nominal expressions, including verbs, adverbs and whole clauses, whereas $n\partial^2$ and tv^2 are restricted to nominals. Relevant examples were given above in (11) and (12). There are some exceptions to this rule with $n\partial^2$ and tv^2 occurring after prepositions, as seen in sentences (14) and (15) above. In these cases, the adnominal forms function like pronominal forms.

The third difference is that $k\partial h$ can occur after the other demonstratives, including $k\partial h$ itself. Bauer explains the function of $k\partial h$ in this position as merely emphatic. Relevant examples are given in sentences (19) - (21). In (19) the distal and proximal demonstratives must be interpreted as exophoric, reinforced by the deictic directional 2a 'go' in the second part. Neither Chiangmai not Three Pagoda Pass are mentioned in the preceding discourse, so anaphoric reading is excluded here. In (20) an anaphoric reading is more natural (the hill has been mentioned in the preceding sentence). In (21), the proximal is anaphoric, while the medial has another discourse function, most likely to indicate the element which the subsequent discourse is about.

| (19) | cèk march | ceh go.dowi | nù 1 ABL | kəpac part | cèŋmày Chiangmai | tx? DIST | | mùə one | ləpac, side |
|------|-----------------------|------------------------------|-------------|---------------|--------------------------|------------------------------|-----------|-------------|--------------------|
| | cèk march | ?a go | nù ABL | klòŋ way | kyac.ppə? Three.Pagoo | yac.ppə? hree.Pagoda.Pass | | no? PROX | kòh MEDL |
| | one mùə One par | part kəpac. rt [of the | army] m | arched d | own from Ch | iangm | ai, one j | part mar | ched |

(20) dɔə kp bè? tỳ hmop.plày kɔ̀h kɔ̀h LOC OBL REF hill prince MEDL MEDL

(21) dɔə ʔərèə nɔʔ **kɔh**LOC matter PROX MEDL

in this matter

at the Prince Hill

from here at the Three Pagoda Pass.

The function of $k\partial h$ in these sentences is clearly no longer demonstrative (or deictic) in these expressions, i.e. the distance from the point of reference is no longer relevant. An analysis as anaphoric (as opposed to exophoric for the first of the two in each expression) is excluded by the fact that the localities in (19) are both not mentioned before in the text.

A last feature distinguishing $k\partial h$ from the other demonstratives is stress. While $n\partial l$ and tyl are always stressed, $k\partial h$ can be either stressed or unstressed. There seems to be

In newer publications such as Newspapers, one can find other combinations of demonstratives like N $wu\lambda$ $n\partial$. This development seems to be very recent and may or may not be an extension of the secondary use of $k\partial h$ to other demonstratives.

some regularity in the distribution and function of stressed and unstressed $k\partial h$, but it is not clear at the present stage of research whether stress can be assigned to certain functions of the particle.

Obviously $k\partial h$ has some grammatical or pragmatic functions not shared by the other demonstratives. That $k\partial h$ is more grammaticalised, i.e. has extended its area of applicability to more contexts, is further supported by the fact that it is far more frequent that the other two. In a short narrative text of one page, $k\partial h$ occurs 86 times, while there are only two instances each of $n\partial l$ and tr l in the same text. As seen above, Bauer lists $k\partial h$ as a marker of "determination", presumably meaning definiteness. The next section takes a closer look at definiteness and tries to answer the question whether $k\partial h$ is a kind of definite marker.

4. Definiteness

4.1 Defining definiteness

Definiteness is a grammatical category present in some but not all languages. Definiteness is typically a nominal category, closely associated with the noun phrase. According to Lyons (1999:1), "in many languages a noun phrase may contain an element which seems to have as its sole or principal role to indicate the definiteness or indefiniteness of the noun phrase." Definite markers regularly arise from grammaticalised demonstratives, while the unstressed numeral 'one' becomes an indefinite marker in many languages. In Mon, there is a tendency to add unstressed mùə 'one' to noun phrases when they are first mentioned in a discourse, i.e. mùo can be taken as a kind of indefinite marker. Though the semantics of definite and indefinite markers seems to be straight forward, it is not easy to reach a crosslinguistically valid definition. Three features appear to be involved in definiteness, viz. uniqueness or inclusiveness, identifiability or familiarity and specificity or referentiality. In his study of definiteness, Lyons comes to the conclusion that the features 'identifiable' and 'unique/inclusive' are expressed in many languages in a single morpheme. grammaticalisation grammaticalised This of 'identifiability' 'uniqueness/inclusiveness' is definiteness. There seems to be no reason to see these two concepts as related, but "the evidence for identifiability and inclusiveness being distinct features is lacking" (Lyons 1999:158). If a language expresses only 'identifiability' with a grammatical morpheme, this is taken to be definiteness in that language (Lyons 1999:278). Some languages use definite expressions for specific referents only, while others can use definite noun phrases also in generic contexts. That means that specificity is not a prototypical feature of definiteness.

Lyons only briefly mentions the use of definiteness "beyond the noun phrase" (Lyons 1999:45f). He does not pursue this topic, but in a later chapter (pp. 60ff) he returns to nominalising and other functions of definite articles: "An important aspect of the behaviour of definite articles is their use other than with nouns." (p. 60). In many languages the definite article can be used to nominalise adjectives and verbs.

More strikingly, a definite article can sometimes serve to introduce an entire finite clause, thus functioning somewhat as a complementizer. [...] This applies particularly to subordinate clauses with an argument function. (ibid.).

Mon Particle **kòh** 65

According to Lyons definitions and extensions of use of definite articles, it seems plausible that $k\partial h$ in Mon is a grammaticalised definite article which originates in and coexists with medial demonstrative.

4.2 Is kòh a definite article?

In this section I will give more examples of the different uses of $k\partial h$, trying to determine whether they can be united in the single category of definiteness. In addition to the data already presented in sections 2 and 3, the data presented in this section will concentrate on the features typically associated with definiteness, i.e. identifiability/familiarity, uniqueness/inclusiveness and specificity/referentiality. It should be noted from the beginning that, unlike definite articles in many languages that have them, $k\partial h$ in Mon is never syntactically obligatory.

Definite noun phrases are used to refer to entities that are either known from the discourse context or that have been mentioned in the previous discourse. They are not used in presentational expressions ('there was a/*the NP'). In most cases, $k \partial h$ is attached to nominals that are identifiable, either linguistically or extralinguistically. Personal names, pronouns and place names are inherently semantically definite, so that in many languages they are not explicitly marked as such. In Mon, $k \partial h$ can be freely suffixed to names of people and places as well as to pronouns of all persons. This also suggests that the medial demonstrative value of $k \partial h$ has been lost in this function. Otherwise the collocation of $k \partial h$ with $\partial u \partial h$ 'I' and $\partial h \partial h$ we' would be contradictory. That the (extralinguistic) identifiability can be based on shared knowledge or on the notions associated with a given frame is illustrated in sentence (22). While the Karen are not mentioned previously they are associated with the place name Mesali, obviously a Karen village.

| (22) | tèh | nì?mòn | Gcb | mèsəlì | tx?, | cao | kp.ppn | 6pt | tek |
|------|-------------|-------------------|-----|------------|---------|-----------|------------------|----------|--------------------|
| | HIT | invite | LOC | Mesali | DIST | return | eat | about | beat |
| | pon four | | | | | | khənə, noodle | | khənə" noodles" |
| | çiə? eat | hənəm. noodles | | i over the | one and | wa want l | naak to oo | rt about | |

We were invited to Mesali over there, and we went back to eat about four o'clock, and the Karen called us "?3 khən, ?3 khən", [that means] 'eat noodles'.

Sentence (23) occurs at the beginning of a story, introducing the main participants in the text. The use of unstressed $m\dot{u}$ resembles an indefinite article here, but the function of $k\partial h$ is again not clear. Its cooccurrence with the quasi indefinite article $m\dot{u}$ clearly excludes an interpretation as a definite marker/article. The referent of the nominal expression is clearly not identifiable in this context, neither within the discourse nor extralinguistically.

_

One might try to explain expressions like $\partial u \partial k \partial h$ as indicating a kind of emotional distance, but this does not seem to be the case in Mon.

(23)kla tx? siən, nùm mòŋ kon.ŋàc mùə kòh ... child before DIST be.so exist STAY one MEDL Long time ago, right, there was a boy ...

As can be seen from this sentence, $k\partial h$ is not restricted to identifiable or known referents but can be used with a completely new actant if it is relevant in the subsequent discourse.

We have seen above that, unlike the other demonstratives, $k \partial h$ does not imply singularity of the referent. The same is true for inclusiveness. An expression like $kon.n \partial c$ $l \partial n \partial c h$ 'some of the children' is perfectly grammatical in Mon. Therefore uniqueness or inclusiveness is obviously not part of the semantics of $k \partial h$.

The following sentences show that $k\partial h$ is not restricted to specific (as opposed to generic) referents either. The examples are taken from Ketumati's translations of English proverbs into Mon (Ketumati 1965), they are therefore representative of Literary Mon rather than the spoken variety, but there are no obvious differences in use in this respect.

(24) nèh pyỳ **kòh** tòh nèh thiə.

person hungry MEDL be person angry *A hungry man is an angry man*. (p. 130)

In (24), the use of $k\partial h$ could be motivated by the complex NP including an attributive verb, but this does not explain why it is not present in the second part of the sentence. Clearly the expression 'a hungry man' is not specific, but generic. The second NP has predicative function and is non-referential, i.e. there is a difference in referentiality involved here. Possibly $k\partial h$ is only used with referential noun phrases.

In sentence (25), the first noun is generic and the second specific, but both have the marker $k \partial h$.

(25) mənìh **kòh** 6pn khyəp, này tèwè? **kòh** prao?.priəŋ pənəp. man MEDL though consider master god MEDL prepare command *Man proposes, God disposes.* (p. 168)

Lyons states that definite articles can develop into markers of complement clauses. That this is true for Mon $k\partial h$ is illustrated in the following sentence, again taken from Ketumati's translation of proverbs. Here $k\partial h$ is used to mark argument (i.e. non-predicative) function of a verbal expression.

kàh kàh (26)hpm lùə klon wàt plon. lòn, peak MEDL easy exceed do **MEDL** difficult again Easier said than done. (p. 144)

This use of $k\partial h$ does not make it a full fledged nominaliser, as the resulting expression does not have all the syntactic possibilities of a typical nominal expression. It can, for example, not be modified by an attributive verb or demonstrative: *hom k\partial h khoh 'good talk', *hom k\partial h no?' this talk'. Rather, k\partial h is used to mark the boundary of a complex expression, as already seen above with noun phrases modified by relative clauses.

Mon Particle kòh

It appears that $k\partial h$ does have some of the semantic properties of a definite article, but it can hardly be seen as a grammaticalised marker of definiteness, unless we are ready to accept a more vague definition of definiteness than the one given by Lyons. The fact that $k\partial h$ is used to mark a verbal expression as an argument rather than the predicate indicates that there is a possible connection with the topic – comment distinction, i.e. $k\partial h$ might be used to mark topics. This means that the function of $k\partial h$ is one of information structure rather than syntax or semantics. To this topic we will turn in the next section.

5. Information structure – Topic and focus

5.1 Topicality

Different authors have suggested different definitions of the term 'topic'. Haiman gives the following definition:

The topic represents an entity whose existence is agreed upon by the speaker and his audience. As such, it constitutes the framework which has been selected for the following discourse. (Haiman 1978:585)

Dik states that "a linguistic expression will [...] usually contain some *given* information and some *new* information" (1989:265ff). Given information is what the speaker assumes to be available to the addressee as pragmatic information, which consists of "*general*, *situational*, and *contextual* information" (Dik 1989:9, 265). He goes on that "partially corresponding to the "given"/"new" distinction, we may distinguish the dimension of *topicality* and *focality*" (p. 266). Topicality tends to coincide with given information, and focus with the most salient or important piece of new information that is given about the topic.

Dik distinguishes different kinds of topics: The discourse topic denotes the entity which the discourse is "about". A discourse may have different discourse topics with different degrees of centrality to the discourse. While a topic usually refers to an entity that is known or given, i.e. mentioned in the previous discourse ("GivTop", Dik 1989:267), new referents may be introduced to the discourse as "NewTop" (ibid.). A NewTop denotes an entity that is not mentioned before but that is relevant to the following discourse. A topic can be reactivated after a stretch of discourse. It is then called a "resumed topic" (ResTop; ibid.).

Other authors (e.g. Erteschik-Shir 2007:) take up the features of topic a "givenness" and "aboutness". Givón (2001:254) states that "topicality involves two aspects of referential coherence, one anaphoric, the other cataphoric". The anaphoric aspect is "referential accessibility" and the cataphoric aspect is "thematic importance".

Lambrecht (1994) says that in his use of the term, "the topic of a sentence is the thing which the proposition expressed by the sentence is ABOUT" (p.118). Also, the topic must be relevant to the present discourse and the predicate must add some new information about it. "The definition of topic in terms of aboutness and contextual relevance entails that there is an inherent relationship between topic and pragmatic PRESUPPOSITION" (p. 150). That means according to Lambrecht's definition, the topic is part of the presupposition of the utterance. This is related to the pragmatic (or information-structural) status of the NP referred to by the topic expression. Lambrecht (p. 109) distinguishes different degrees of identifiability:

68 Mathias Jenny

Unidentifiable: anchored, unanchored

Identifiable: inactive, accessible (textually, situationally, inferentially), active

According to Lambrecht, topics (as part of the presupposition) must be identifiable and may be active or activated.

While the definitions of topic (and focus) employed by different authors vary to some degree, they largely agree in that topics must be accessible to the hearer in some way and must be relevant to the ongoing discourse. Dik's NewTop seems to contradict the prerequisite of accessibility, but it can be seen as activated by its introduction to the discourse.

5.2 Topicality and focality in Mon

Mon has a focal particle, viz. ra?, originating in a weak form of the copula das in Old Mon (s. Jenny 2005, 2006). This focus marker contrasts in some contexts directly with $k \partial h$. While $k \partial h$ marks an expression as argument or non-predicative, ra? can be used to mark the same expression as a predicative clause, as seen in (27).

- (27) a. mənìh klrŋ çiə? prŋ **kòh** man come eat rice MEDL the people coming to eat
 - b. mənìh klrŋ çiə? prŋ ra?.
 man come eat rice FOC
 The people are coming to eat.

As $k\partial h$ is mostly used anaphorically, it marks known or accessible referents (nominal, verbal or clausal). Many sentences have the form $[X \ k\partial h] [Y \ ra?]$, with X being the presupposed part of the sentence, i.e. the theme or topic, and Y the predicate or comment, as in (28) and (29).

- (28) mənìh kòh klyn çiə? pyn ra?. man MEDL come eat rice FOC That man is coming to eat.
- (29)ppə?-klom-ppə? kàh ďεh hù? cao nεm pùh, three-hundred-three MEDL 3 NEG return yet NEG 6pt ďεh kàh (<?i?-ra?).kya? kya? ?a ya? iust 3 lose **MEDL** lose GO **NSIT** (< PFV-FOC)*In 1303, they had not retreated yet, but they had already lost [the war].*

In other cases, $k\partial h$ is used to mark a constituent as antitopic or afterthought, which is usually nominal, but may be verbal or adverbial. In this function, $k\partial h$ is always unstressed. This use is illustrated in example (30), where the negation particle $p\hat{u}h$ marks the end of the basic sentence.

Mon Particle kòh

(30)?uə hù? ket ra? pùh, lòc kòh. NEG take FOC 1 NEG text **MEDL** I don't want it anymore, that book.

This may easily lead to an analysis of $k\partial h$ as a topic marker. If we take the definitions given in section 5.1 for topic, there is indeed a large degree of overlap in the use of $k\partial h$ and topicality. As seen in the examples in earlier sections, $k\partial h$ marks a constituent which is either mentioned in the previous discourse or is pointed at in the discourse situation, i.e. which is textually or situationally accessible according to Lambrecht's terminology. Where a new referent is introduced to a discourse and is marked by $k\partial h$, this referent is identified as being important or relevant to the discourse, i.e. it is activated rather than already active, as in sentence (23) above. This can be seen as pragmatic accommodation of a new element that is introduced as if it were identifiable (cf. Lambrecht 1994:65ff).

Haiman (1978) suggests that topic expressions and conditional clauses are comparable (or even identical) in many respects. A clause in Mon containing $k\partial h$ as boundary marker can be interpreted as conditional or sequential, as illustrated in sentence (31). This use of clauses ending in $k\partial h$ is quite frequent in the spoken language, especially in non-final clauses within complex sentences.

(31)kàh ?uə ?a thp?. nèh kp kəpac toə ràn çiə? person give part FINISH MEDL buy 1 go eat DISCARD After he gave me half [of the money] I went to buy something to eat.

Obviously $k\partial h$ has acquired a function in organising information structure, i.e. marking identifiable or accessible chunks of information, nominal, verbal and clausal. Once this function is established, $k\partial h$ can also be used to activate pieces of information as relevant to the discourse by pragmatic accommodation. In this information structural function, $k\partial h$ is directly opposed to the focus marker ra?. It is not, however, incompatible with it. Cooccurrence of $k\partial h$ and ra? in the same clause is not infrequent, especially in adverbial expressions like hot $k\partial h$ ra? 'it is for this reason'. In this context $k\partial h$ clearly has demonstrative value, i.e. it anaphorically refers to information given earlier in the discourse. Therefore it also marks the 'reason', i.e. circumstances leading to some ensuing situation, as known or accessible, and relevant to the following discourse, i.e. $k\partial h$ here also expresses topicality. The focal particle ra? indicates that it is new and important in the discourse that the circumstances mentioned earlier are the reason for what follows.

To conclude this section, we can state that $k \partial h$ is used to mark topical pieces of information in an ongoing discourse. The topicality may be based on the previous discourse, i.e. anaphoric referential accessibility, or on the relevance to the following discourse, or both. The accessibility may be due to shared general, situational, or contextual information.

6. Conclusion

We have seen that among the three demonstratives in Mon, $k\partial h$ has a special position in many respects including semantics, syntax and phonology (stress pattern). Basically (and probably originally) a demonstrative particle indicating medial distance from the centre of interest, $k\partial h$ has acquired a wide range of functions. The exact stages of the extension of function can not be traced from the documented material available in Mon. Already in Old

70 Mathias Jenny

Mon, the particle goh had anaphoric and maybe boundary marking function, but the material is too scarce to draw a final conclusion. In Modern Mon, the polysemy (or multifunctionality) of $k\partial h$ is rather far-reaching, but the different functions are interconnected and overlapping, so that in many cases it is not easy or even impossible to decide which function is prevalent in a given expression. The two main functions of $k\partial h$ are as DEMONSTRATIVE with MEDIAL DEIXIS and marking IDENTIFIABLE INFORMATION RELEVANT to the ongoing discourse. This second function corresponds to a TOPIC MARKER, which can be seen as an extension of anaphoric uses of the demonstrative. Marking a constituent (phrase or clause) as identifiable or topical, $k\partial h$ sets it apart from the new information given in the sentence, i.e. the PREDICATE. This leads to the frequent use of $k\partial h$ as marker of a NON-PREDICATIVE expression.

The question of stress needs further investigation based on more extensive recorded texts. At the present stage of research it seems like $k\partial h$ is fully stressed when it functions as exophoric demonstrative or as resumed topic marker (anaphoric), but unstressed when marking a given topic or antitopic.

Figure 1 summarises the possible development of the different functions of $k \partial h$ in Mon.

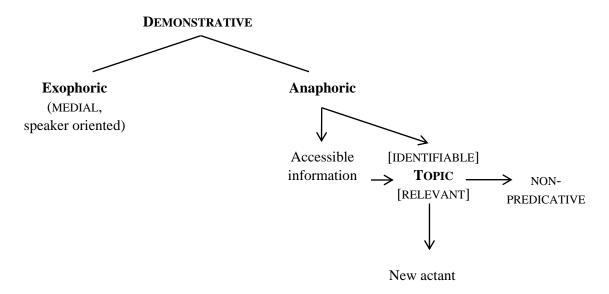


Figure 1: Development of kh

Abbreviations

| ABL | Ablative | MEDL | Medial |
|-------|--------------|------|--------------------------------|
| ADD | Additive | NEG | Negation |
| ATTR | Attributive | NSIT | New situation ('already, now') |
| COM | Comitative | OBL | Oblique |
| DIST | Distal | PFV | Perfective |
| EMPH | Emphatic | p/PL | Plural |
| FOC | Focus | PROX | Proximal |
| INSTR | Instrumental | Q | Question |
| LOC | Locative | REF | Referential |

Mon Particle kòh

References

Bauer, C. 1982. *Morphology and syntax of Spoken Mon.* Doctoral dissertation at SOAS, London.

- Diessel, H. 1999. *Demonstratives. Form, function, and grammaticalization*. Amsterdam/Philadelphia: John Benjamins.
- Dik, S. C. 1989. The theory of functional grammar. Part I: The structure of the clause. Dordrecht/Providence: Foris Publications.
- Duroiselle, C. (ed.) 1962 [1921]. *Epigraphia Birmanica. Vol. II. The Talaing plaques on the Ananda plates.* Rangoon: Government Printing and Stationery.
- Erteschik-Shir, N. 2007. Information structure. The syntax-discourse interface. Oxford: OUP.
- Givón, T. 2001. Syntax. Vol. 2. Amsterdam/Philadelphia: John Benjamins.
- Haiman, J. 1978. Conditionals are topics. Language vol. 54/3, 564-89.
- Halliday, R. 1955 [1922]. A Mon-English dictionary. Rangoon: Ministry of Union Culture.
- Himmelmann, N. P. 1996. Demonstratives in narrative discourse. In Fox, B. 1996. *Studies in anaphora*. Amsterdam/Philadelphia: John Benjamins; 201-254.
- Jenny, M. 2003. New infixes in Spoken Mon. In Mon-Khmer Studies 33, 183-194.
- Jenny, M. 2005. The verb system of Mon. Zurich: ASAS 19.
- Jenny, M. 2006. Mon ra? and non: Assertive particles? Mon-Khmer Studies 36, 21-38.
- Ketumati, Bh.. 1965. *4 languages English, Mon, Burmese and Pali*. Moulmein: Suvannavam sa Printing House.
- Lambrecht, K. 1994. Information structure and sentence form. Topic, focus and the mental representations of discourse referents. Cambridge: CUP.
- Lyons, C. 1999. Definiteness. Cambridge: CUP.
- Okell, J. and A. Allott. 2001. Burmese/Myanmar dictionary of grammatical forms. London: Curzon.
- Shorto, H. L. 1962. A dictionary of Spoken Mon. London: Oxford UP.
- Shorto, H. L. 1971. A dictionary of Mon inscriptions from the VIth to the XVIth centuries. London: Oxford UP.
- Yanson, R. A. 2002. A list of Old Burmese words from 12th century inscriptions. In Beckwith C. I. (ed.) *Medieval Tibeto-Burman languages*. Leiden/Boston/Köln: Brill, 163-167.

TONE NEUTRALIZATION DUE TO CONSONANTS IN MULAO

Seunghun J. Lee

Central Connecticut State University <leeseu@ccsu.edu>

0 Abstract

Mulao requires that syllables with onsets that are aspirates, glottals, voiceless sonorants and fricatives have high tone. These consonants phonologize in a way that input tone neutralizes when such consonants are onsets of these syllables. This paper argues that this requirement is due to the ranking, in which markedness constraints on consonant-tone interaction outrank faithfulness constraints on tone. Crucial to the analysis is the proposal that no faithfulness constraints preserve the relationship between tone and consonant. Finally, the typology of consonant-tone interaction is presented by including data from African languages in which consonants also interact with tone.

1 Introduction

The aim of this paper is to propose that markedness constraints on the tone-segment relationship are responsible for the high tone requirement in syllables with laryngealized consonants, such as aspirates and glottals, voiceless sonorants and fricatives, in Mulao. While various phonetic studies show that there is a physiological relationship between high register and these laryngealized consonants, I argue that phonologization of H tone in Mulao results from constraints that penalize the low register tone on these consonants. ¹

Mulao (or Mulam) is a Tai-Kadai language, and it is spoken by about 20,700 speakers in Luocheng County of Guangxi Zhuang Autonomous Region, China (based on a 2000 census). One of the characteristics of Tai-Kadai languages is that laryngealized consonants often interact with tone. For example, laryngealized consonants such as glottals and aspirates in Zhuang favor high tone (Luo, 2005: 1217-1218). On the other hand, unaspirated stops inhibit high tone in Thai (Ruangjaroon, 2006).

In Mulao, laryngealized onsets such as glottals, voiceless sonorants, and aspirates require a high register tone and never occur with a low register tone as in (1), while other (non-laryngealized) onsets as in (2) do not have such a requirement. In other words, the onsets in (2) can occur with both high and low register tone. Laryngealized consonants that require H register tone are called 'elevator consonants' in this paper.² Tonal representations are based on Optimal Domains Theory (ODT, Cassimjee and Kisseberth, 1998). An output tonal domain is marked with square brackets, and the tonal specification of each domain is marked with a subscript H or L.

High register tone and low register tone will be used interchangeably with high tone and low tone in this paper.

² Hyman and Schuh (1974) refer to these pitch raising segments as 'uplifter consonants'.

74 Seunghun J. Lee

(1) Onsets that require a high tone (*elevator consonants*)³

| | | a. <i>high</i> | | b. <i>low</i> |
|-----------------|---------|--|--------------|--|
| Glottals | [?] | [? uk ⁵⁵] _H | 'to exit' | *[?uk ¹¹] |
| | [?j] | [²j em ⁴²] _H | 'to borrow' | *[?jem ¹¹] |
| Voiceless | [n] | $[\mathbf{n} \mathbf{o}^{53}]_{\mathrm{H}}$ | 'rat, mouse' | $*[no^{24}]$ |
| sonorant | , | • | | , and the second |
| Aspirates | $[k^n]$ | $[\underline{\mathbf{k}^{\mathbf{h}}} \epsilon : \mathbf{k}^{42}]_{H}$ | 'guest' | * $[k^h \varepsilon: k^{13}]$ |
| | $[p^h]$ | $[\mathbf{p^h}^{044}]_{H}$ | 'bed' | $*[p^{h}o^{11}]$ |

(2) Onsets that do not require a high tone

| | | a. <i>high</i> | | b. <i>low</i> | |
|------------------|-----|--|---------------|---|---------------|
| Sonorants | [j] | [j a ⁴²] _H | 'cloth' | [j a ¹¹] _L | 'also, too' |
| | [1] | [] au ⁴⁴] _H | 'snail' | [l au ¹¹] _L | 'to leak' |
| Voiceless | [k] | $[\underline{\mathbf{k}}\mathbf{u}\mathbf{n}^{53}]_{\mathrm{H}}$ | 'stalk, stem' | $[\underline{\mathbf{k}}$ uŋ $^{121}]_{\mathrm{L}}$ | 'grasshopper' |
| | [p] | [p a:k ⁴²] _H | 'mouth' | [p a:k ¹¹] _L | 'white' |
| | [t] | [<u>t</u>aŋ ⁴⁴] _H | 'stool' | $[{\bf t}a{\bf y}^{24}]_{\rm L}$ | 'to stand' |

As summarized in Downing (2009), formal theories have argued that consonant-tone interaction results either from single features belonging to both segments and tone (Halle and Stevens, 1971, Bradshaw, 1999 among others), or from independent segmental and tonal features (Khumalo, 1987, Peng, 1992, Hyman and Mathangwane, 1998, Hansson, 2004 among others). The latter position adopts implicational constraints to account for consonant-tone interaction.

While these two types of theories have contributed to the advance of our understanding of consonant-tone interaction in phonology, there is a lack of explicitness of formal requirements in the grammar that make consonant-tone interaction possible. The proposal in this paper addresses this issue by assuming that there is no faithfulness constraint in the grammar that preserves tone associated with consonants. The lack of faithfulness means there is a lack of tonal contrast in consonants. As a result, surface-oriented markedness constraints play a crucial role in consonant-tone interaction.

The structure of this paper is as follows. In section 2, the proposal on consonant-tone interaction is presented in Optimality Theory (Prince and Smolensky, 1993/2004). The basic phonology and tonology of Mulao appears in section 3. In section 4, an analysis about high-tone-inducing consonants in Mulao is presented with markedness constraints that ban these consonants from being associated with low tone. Such an analysis is supported by a typological study on consonant-tone interaction presented in section 5.

2 Proposal: How consonants interact with tone

Mulao requires high tone in the output on syllables with laryngealized onsets. The interaction between these elevating consonants and tone is based on the assumption that all segments are able to form a dependency relationship with tone (cf. Myers, 1997: 851-853).

All data come from Wang and Zheng (1993). Tone is marked in Chao tone letters, in which number 1 represents the lowest pitch, and number 5 is the highest pitch.

While the claim that consonants are directly associated with a tone is contentious⁴, I argue that consonants and tone form a dependency relationship in the output. This proposal accounts for the co-occurrence restriction in Mulao, and it also explains consonant-tone interaction in general.

Formally, consonant-tone interaction results from the requirement that all segments in the output must belong to a tonal domain, which is ensured by the markedness constraint ROOTNODE \rightarrow T (RT \rightarrow T). This constraint assigns violation marks to segments that do not belong to any tonal domain in the output (cf. Yip, 2002: 83).

It is crucial to this proposal that there are no faithfulness constraints that preserve the dependency relationship between tone and consonants. This lack of faithfulness constraints means that tones are never contrastive on consonants. So, no language will have a contrast between [pa] with low tone on the [p] and [pa] with high tone on the [p].

Consonant-tone interaction is the result of markedness constraints dominating tonal faithfulness constraints. The laryngeally marked elevator consonants affect the tone on their syllable due to the markedness constraint *ELEVATOR-L (*ELV-L). The *ELEVATOR-L constraint gives a violation mark to elevator consonants that form a dependency relationship with a low register tone in the output domain.

The main points of the proposal are summarized in (3).

(3) Proposal

- a. Consonants form a dependency relationship with tone.
- b. There is **no faithfulness constraint** that preserves tone on consonants.
- c. Markedness constraints
 - i. **ROOTNODE**→**T**: All root nodes in the output should be parsed into a tonal domain.
 - ii. *ELEVATOR-L: Elevator consonants should not be parsed into a low register tone domain.

The tonal faithfulness constraints are proposed in order to specifically target tone bearing units (TBUs) that are prosodic nodes such as moras (or syllables).

- (4) Tonal faithfulness constraints (cf. Yip, 2002: 83)
 - a. **DEP-T** The tone <u>associated to a TBU</u> in the output has a corresponding tone in the input.
 - b. **MAX-T** The tone <u>associated to a TBU</u> in the input has a corresponding tone in the output.
 - c. **IDENT-T** The tone <u>associated to a TBU</u> in the output has the same specification with the correspondent tone in the input.

The representation of a syllable with an elevator onset is shown in (5). Association lines in this representation indicate the dependency relationship between tone and a segment (cf. Myers, 1997). Segments associated with the same tone in the output form a

⁴ The dependency relationship between tone and consonants should be distinguished from the relationship between tone and prosodic nodes, which are called Tone Bearing Units (TBUs). For a discussion on TBUs being segments or prosodic nodes, see Bradshaw (1995: 267-269).

76 Seunghun J. Lee

domain. As before, domains are marked with brackets, and the tone of a domain is represented with a subscript at the end of a bracket.

(5) Representation of a syllable with an elevating consonant



3 Mulao

3.1 Mulao consonants

In (6), the consonant inventory of Mulao is presented. The elevating consonants, which require high tone in the output, are in the heavy-lined boxes. Other consonants do not require specific tones in the output.

In Mulao, the possible codas are /p, t, k, m, n, η /, which do not belong to the set of elevating consonants. So, even if codas are present, the consonants should not influence the tone of a syllable, since there are no elevator consonants allowed in coda position.⁵

(6) Consonants⁶ (Wang and Zheng, 1993: 4-5)

| | lab | ial | alve | olar | pal | atal | ve | lar | glottal |
|-----------------|---------------------------|----------------|-----------------|------|-----|------|---------------------------|-----|---------|
| Stops | p | | t | | c | | k | | ? |
| aspirated | $\mathbf{p}^{\mathbf{h}}$ | | th | | ch | | $\mathbf{k}^{\mathbf{h}}$ | | |
| fricatives | f | | S | | ç | | | γ | h |
| affricates | | | ts | | | | | | |
| aspirated | | | ts ^h | | | | | | |
| nasals | m | m | ņ | n | ĵ | ŋ | ŋ̊ | ŋ | |
| liquids | | | ļ | l | | | | | |
| glides | | W | | | | j | | | |
| pre-glottalized | | [?] W | | | | ³j | | γ° | |

3.2 Mulao tone

The tonal inventory of Mulao is shown in (7). While the number of tones in Mulao is traditionally described as having about ten tones as in (8), the phonological tones I assume are simplified based on models proposed in Bao (1990) and adopted in Yip (2002). Mulao has high register tone and low register tone. Each register has a level tone and a contour tone.

In other languages, however, codas may affect tone. In this paper, such languages will not be discussed because Mulao does not show laryngealized consonants in coda position. For codas that contextually vary with respect to mora, see studies on Capanahua by Elías-Ulloa (2006) and on Kashmir by Morén (1999).

⁶ Onsets can be labialized, palatalized or uvualarized in Mulao. See Wang and Zhang (1993) for details.

| (7) | Tonal | inventory ⁷ |
|-------|-----------|------------------------|
| (/) | , i Oliai | III v CIIIOI y |

| register level | | contour (falling) | | | |
|-----------------------|--------------|-------------------|--|--|--|
| High | 55 and 44 | 53 and 42 | | | |
| | level | contour (rising) | | | |
| Low | 11 (and 121) | 12 and 24 | | | |

In the tonal description based on phonetic properties in (8), odd-numbered tones are high and even-numbered tones are low. Tone 7 and tone 8 mark tones in syllables with obstruent codas, which have different surface tones depending on the length of the vowel.

(8) Description of tones in Mulao (Wang and Zheng, 1993: 13)

| odd | 1 | 3 | 5 | 7short | 7long |
|-------------|-----------|------|-------|--------|-------|
| pitch | 42 | 53 | 44 | 55 | 42 |
| description | mid | high | mid | high | mid |
| | fall | fall | level | level | fall |
| even | 2 | 4 | 6 | 8short | 8long |
| pitch | 121 | 24 | 11 | 12 | 11 |
| description | low | mid | low | low | low |
| | rise-fall | rise | level | rise | level |

While the phonetic tonal description in (8) is an important reference, caution is urged when phonological interpretation is solely based on phonetic descriptions (see Maddieson, 1978). For example, in his phonological analysis, Chen (2000: 17-19) acknowledges that phonetics of citation tones of Chinese spoken in Jianyang is a confounding fact because the phonetic data can result from "the final lowering" or "the undershooting of phonological targets". See also Yip (2002: 21-24) for relevant discussions.

4 Tonal Neutralization

In this section, I analyze the H tone neutralization caused by elevating consonants (aspirates, glottals, voiceless sonorants and fricatives) as the markedness requirement on laryngeally marked consonants and tone in the output.

In (9), the H tone requirement in Mulao is illustrated. In the output, syllables with laryngealized onsets must belong to a high tone domain, which is marked with a subscript H. In other words, both an input with a high tone as in (9a) and an input with a low tone as in (9b) neutralize to a H tone in the output. By contrast, non-laryngealized onsets do not have such a requirement. Thus, tone in the input surfaces faithfully in the output. Inputs with high tone have high tone output as in (9c), and inputs with low tone have low tone output as in (9d).

Mulao has a falling contour tone in high register, and a rising contour tone in low register. In his work on historical phonology, Ferlus (2006: 11) reconstructs that low pitch is related to pitch-raising in vowels, and high pitch is related to pitch-lowering in vowels. This reconstruction was achieved by comparing Southeast Asian languages to Old Chinese.

78 Seunghun J. Lee

| (| 9 |) Data to | be | accounted | for | bv | an | analy | vsis |
|----|---|-----------|----|-----------|-----|-----|----|-------|-------|
| ٠, | _ | , Dan io | ~ | accounted | 101 | · , | ~~ | ullul | , 515 |

| | | input | | | output | | |
|----------------------|----|--------------------|------|---------------|--------------------------------|------|---|
| laryngealized onsets | a. | /c ^h ý/ | high | \rightarrow | $[\acute{c}^{h}\acute{v}]_{H}$ | high | $[\underline{\mathbf{p}^{\mathbf{h}}}$ o ⁴⁴] _H 'bed' |
| | b. | /c ^h v/ | low | \rightarrow | $[\acute{c}^{h}\acute{v}]_{H}$ | high | |
| non-laryngealized | c. | /cý/ | high | \rightarrow | $[\acute{c}\acute{v}]_{ m L}$ | high | [j a ⁴²] _H 'cloth' |
| onsets | d. | /cv/ | low | \rightarrow | $[\grave{c}\grave{v}]_{L}$ | low | [j a ¹¹] _L 'also' |

- 'c' stands for consonants, 'v' for vowels, 'h' for aspirates
- acute accent ''' for high tone, grave accent ''' for low tone.

In section 4.1, onsets that allow both high and low tone are analyzed. The output tone remains faithful to the input due to tonal faithfulness constraints that preserve the tone. Second, in section 4.2, the neutralization to high tone in syllables with laryngealized onsets is analyzed as markedness constraints on tone and consonants outranking tonal faithfulness constraints.

4.1 Faithful realization of tone

In syllables with non-laryngealized onsets H tone inputs surface faithfully with a high tone and L tone inputs surface faithfully with a low tone. The faithful realization of input tones in the output is because there are no markedness constraints that dominate any of the faithfulness constraints in the ranking.

An input with a low tone as in (10) surfaces faithfully because the input tone should not be deleted due to MAX-T, and the input tone should not be changed because of IDENT-T (both constraints are defined in (4)). An input with a high tone also emerges faithfully as in (11) due to the constraints MAX-T and IDENT-T (both defined in (4)).⁸

(10) L tone input surfaces faithfully

$\begin{array}{|c|c|c|c|c|} & L & & MAX-T & IDENT-T \\ \hline & /ja/ \ 'also' & & & \\ \hline a. & & [ja]_L & & & \\ \hline b. & [ja] & & W^* & & \\ \hline c. & [já]_H & & W^* & \\ \hline \end{array}$

(11) H tone input surfaces faithfully

| | H /já/ 'cloth' | MAX-T | IDENT-T |
|------|-------------------------|-------|---------|
| a. 🖘 | [já] _H | | |
| b. | [ja] | W* | |
| c. | [jà] _L | | W* |

The faithful realization of input tone in non-laryngealized onsets is due to the fact that no markedness constraints outrank these faithfulness constraints. This is not the case when onsets are laryngeally marked as will be discussed in section 4.2, in which tone neutralizes to high tone in the output.

⁸ In each comparative tableau, asterisks indicate violation profiles. 'W' means that the optimal candidate is the winner under a constraint. 'L' means that the optimal candidate loses against other candidates under a constraint. In each row of the tableau, all L's should be dominated by at least one W. See Prince (2002).

4.2 Syllables with elevating onsets

Laryngeally marked onsets ('elevating onsets') in Mulao must surface in a high tone domain in the output. In other words, inputs with H tone surface faithfully, while L tone inputs neutralize to high tone in the output. I propose that this neutralization to H tone is due the markedness requirement on the elevating consonants, which outranks the tonal faithfulness constraints.

An input with a low tone as in (12) illustrates this point. The optimum in (12a) surfaces with a high tone domain. The faithful candidate in (12b) has a violation of the markedness constraint *ELEVATOR-L, which assigns violation marks to the (laryngeally-marked) elevating consonants within a low tone domain. Forming a separate H tone domain for elevating onsets as in (12c) violates *DOMAINADJACENCY, which favors as few domains as possible in the output. A candidate, in which the elevating onset is not parsed as in (12d), violates ROOTNODE—T (RN—T), which requires all segments to be in some tonal domain. The optimal candidate in (12a), however, violates IDENT-T because the input low tone has changed to high tone in the output.

(12) An input with L tone

| | L p ^h ò/ | RN→T | *ELV-L | *Domain Adjacency | IDENT-T |
|------|---------------------------------|------|--------|----------------------|---------|
| a. 🖙 | [p ⁴ó] _H | | i I | | * |
| b. | [p ^h ò] _L | | W* | | L |
| c. | $[p^h]_H[\delta]_L$ | | | W* | L |
| d. | p ^h [ò] _L | W* | | | L |

(13) *DOMAINADJACENCY

The left edge of a domain should not be adjacent to the right edge of any other domain.

The constraint ROOTNODE→T should dominate DEP-T (defined in (4)) if we consider an input with no tonal specification as in (14). If DEP-T outranks ROOTNODE→T, the candidate in (14b) will be optimal. The optimal output must have a high tone as in (14a). The faithful candidate in (14b) incurs two violations under ROOTNODE→T because all segments in the output do not belong to any tonal domain. This candidate, however, is more harmonic under DEP-T. The output cannot be a low tone domain as in (14c) because it violates the markedness constraint *ELEVATOR-L.

(14) An input with no underlying tone

| | /pho/ | RN→T | *ELV-L | DEP-T |
|------|--|------|--------|-------|
| a. 🖙 | [p ^h ó] _H | | | * |
| b. | p ^h o | W** | | L |
| c. | $[\mathbf{p^h} \grave{\delta}]_{\mathrm{L}}$ | | W* | * |

Let's consider a hypothetical input with L tone on an elevating onset assuming Richness of the Base. The input should not faithfully surface. The faithful candidate in (15b) cannot be the optimum because of the higher ranked constraint *ELEVATOR-L. While

Seunghun J. Lee

there is a change in onset tonal specification from low to high, note that the optimum in (15a) does not violate IDENT-T. The change in the dependency relationship between consonant and tone from L tone to H tone in the output does not violate IDENT-T. This results from the proposal that faithfulness constraints only target tone associated to TBU's, thus excluding tone associated with consonants.

(15) L tone on elevating onsets (a hypothetical input)

| | L H /ṗ ^h ó/ | *ELV-L | IDENT-T |
|------|----------------------------------|--------|---------|
| a. 🖙 | [p ⁴ó] _H | | |
| b. | $[p^h]_L[\delta]_H$ | W* | |

In (16), the ranking argument is presented. The *ELEVATOR-L constraint disfavors any candidate that has a low tone domain with elevating consonants (aspirates, glottals, voiceless sonorants and fricatives). The ROOTNODE→T constraint forces all segments to belong to a tonal domain in the output. Keeping the underlying tone of a moraic TBU (IDENT-T) is not the priority unless the consonant-tone constraint is at stake. In this way, Mulao speakers require syllables with an elevating consonant to belong to a high tone domain.

(16) Ranking by Hasse Diagram

5 Typology: cross-linguistic consonant-tone interaction

In previous sections, the *ELEVATOR-L constraint is proposed to account for the neutralization to high tone induced by laryngealized onsets in Mulao. The reader can easily observe that there is no phonological alternation in Mulao that could independently motivate the constraint *ELEVATOR-L in the set of universal constraints (CON).

An alternative to the proposal would be to describe the requirement of high tone in syllables with laryngealized consonants as a simple phonetic effect. The phonetic explanation would be that these consonants always raise the pitch of the following vowel. Cross-linguistically, voiceless consonants raise the pitch of the onset of a following vowel. If we assume that laryngealized consonants in Mulao relate historically to voiceless consonants, we could then say that the pitch raising effect is a historical remnant linked to the voicelessness (cf. Downing, 2009). ¹⁰

The lack of tone sandhi in Mulao is confirmed by Jerry Edmonson (personal communication). A reviewer pointed out that Kam, a language closely related to Mulao, has tone sandhi, supporting the proposal in this article. In classifier constructions of Kam, consonants show alternations between elevators and non-elevators. When a consonant appears as an elevating variant, the tone becomes the one that is realized with aspirated onsets (cf. Shi, 1983, Yang, 1992).

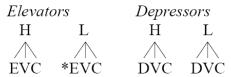
¹⁰ Thanks to an anonymous reviewer who raised this issue.

While acknowledging the possibility of a phonetic analysis of Mulao, this section argues for the current proposal because it not only explains the high tone neutralization in Mulao, but also accounts for the typology of consonant-tone interaction in general. I propose that consonant-tone interaction generally results from the ranking between faithfulness constraints on tone (i.e. IDENT-T) and markedness constraints on tone and consonants (such as *ELEVATOR-L, *DEPRESSOR-H (see below) among others).

An abbreviated typology of consonant-tone interaction is as follows: languages can have elevators (which cause neutralization to H tone), depressors (which cause neutralization to L tone), both elevators and depressors, or neither of them. Each type of language is examined below in (17) - (21).

As shown in the ranking in (17), Mulao is a language in which elevating consonants (marked with "E") require H tone. No restriction on output tone is imposed by depressing consonants (marked with "D") because IDENT-T dominates *DEPRESSOR-H, which bans H tone from being associated to such consonants.

(17) Typology I (Mulao): *ELEVATOR-L >> IDENT-T >> *DEPRESSOR-H



Another type of consonant-tone interaction is found in Nguni or Southern Bantu languages. For example, in Tsonga, spoken in South Africa and Mozambique, there are consonants that block the spreading of high tone. Voiced stops like [g] in Tsonga block the spreading of H tone as in (18). These kinds of consonants are called depressor consonants. For an extensive overview on these types of consonants in African languages, see Bradshaw (1999) and Lee (2008).

(18) Tsonga depressor consonants and H tone blocking

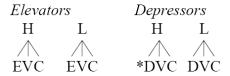
a.
$$H+D\mathbf{L}\mathbf{H} \rightarrow H+D\mathbf{L}\mathbf{H}$$
 'it is a small verandah' [g] gùdá \rightarrow í gùdá b. $H+\mathbf{L}\mathbf{H} \rightarrow H+\widehat{\mathbf{H}}\mathbf{L}\mathbf{H}$ 'it is an ousted impala' $[k]$ $k\mathbf{\hat{o}}mb\acute{e} \rightarrow \mathbf{\hat{n}}$ $k\mathbf{\hat{o}}mb\acute{e}$

The blocking effect in (18) emerges because *Depressor-H bans H tone from being associated to a depressor consonant as in (19). In Tsonga, elevating consonants do not play a role. Therefore, it must be the case that the markedness constraint *Elevator-L is outranked by IDENT-T, which preserves an input tonal contrast in the output with the expense of violating *Elevator-L.

Note that the current version of the proposal assumes a simplified approach to consonant-tone interaction. There are two types of consonants that interact with tone: elevating consonants and depressing consonants.

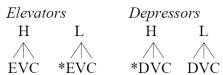
Seunghun J. Lee

(19) Typology II (Tsonga): *Depressor-H >> Ident-T >> *Elevator-L



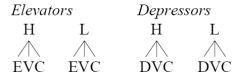
Both elevators and depressors are found in Western Bade, spoken in Nigeria. In Western Bade, depressors block the high tone spreading and elevators block the high tone lowering (Schuh, 2002). I suggest that these two types of blocking demonstrate the presence of depressors and elevators. This is because the proposed markedness constraints on consonant-tone interaction outrank the faithfulness constraint (IDENT-T) as in (20).

(20) Typology III (Bade): *ELEVATOR-L, *DEPRESSOR-H >> IDENT-T



There are also tonal languages with no consonant-tone interaction like Vietnamese. In such languages, consonants do not impose restrictions on tone, or vice versa. In other words, tonal realization and the occurrence of consonants are independent. The ranking of Vietnamese with respect to tone and consonants would be as in (21), ceteris paribus. The faithfulness constraint (IDENT-T) on input tones outranks markedness constraints that restrict types of consonant-tone interaction. ¹²

(21) Typology IV (Vietnamese): IDENT-T >> *ELEVATOR-L, *DEPRESSOR-H



In the proposed typology of consonant-tone interaction, if elevating consonants permit low tone, such consonants must allow high tone as their tonal domain in the output. Similarly, if depressor consonants permit high tone, such consonants must allow low tone as their tonal domain in the output. ¹³

This typology aims to achieve a unified account of consonant-tone interaction in Mulao and in other languages. Although Mulao does not have synchronic evidence such as tonal alternations that interact with consonants, the typology of consonant-tone is enriched by the Mulao case.

¹² Thanks to Thi Thuy Hien Tran who confirmed this property of Vietnamese.

Edmondson and Gregerson (1996) report that Bolyu (also known as Lai), spoken in Guangxi Province, has words with aspirated onsets that only surface in L tone syllables. If aspirated consonants are assumed to be universally elevating consonants, Bolyu is an apparent challenge to the current proposal. However, as other studies reveal, aspirates can behave as depressors as well (e.g. Tsonga). If so, the aspirates in Bolyu are simply (phonological) depressors, and they follow the prediction proposed in this paper.

6 Conclusion

Mulao has laryngealized onsets that cause neutralization to high tone in the output. The laryngealized onsets are aspirates, glottals, voiceless sonorants and fricatives. This high tone neutralization is argued to be a phonological restriction by proposing that consonants can form a dependency relationship with tone, and that a markedness constraint bans low tone on these laryngealized onsets (*ELEVATOR-L).

Crucial to the analysis is that there are no faithfulness constraints that preserve the dependency relationship between consonants and tone. Constraints for consonant-tone interaction are output oriented markedness constraints.

The markedness constraint motivated by the Mulao data is *ELEVATOR-L, which bans elevating consonants from being included in a low tone domain. The prediction borne out by this constraint and other markedness constraints on consonant-tone interaction is presented by examining the typology as in section 5.

Acknowledgement:

I would like to thank two anonymous reviewers for their comments. I also thank Marc Brunelle, James Bruno, Susan M. Burt, Charles Chang, Jerry Edmondson, José Elías-Ulloa, Brian D. McHugh and Jeremy Perkins for their suggestions and comments.

References:

Bao, Zhiming. 1990. On the Nature of Tone, MIT: Ph. D. Dissertation.

- Bradshaw, Mary. 1995. Tone on verbs in Suma. In *Theoretical Approaches to African Linguistics*, ed. Akinbiyi Akinlabi, 255-271. Trenton, NJ: Africa World Press.
- Bradshaw, Mary. 1999. A Crosslinguistic Study of Consonant-Tone Interaction, Department of Linguistics, Ohio State University: Ph. D. Dissertation.
- Cassimjee, Farida, and Kisseberth, Charles W. 1998. Optimal domains theory and Bantu tonology: a case study from Isixhosa and Shingazidja. In *Theoretical aspects of Bantu tone*, eds. L. Hyman and C.W. Kisseberth, 33-132: CSLI Publications.
- Chen, Matthew. 2000. *Tone Sandhi: Patterns across Chinese Dialects* Cambridge Studies in Linguistics: Cambridge University Press
- Downing, Laura. 2009. On pitch lowering not linked to voicing: Nguni and Shona group depressors. *Language Sciences* 31:179-198.
- Edmondson, Jerold, and Gregerson, Kenneth J. 1996. Bolyu tone in Vietic perspective. *Mon-Khmer Studies* 26:117-133.
- Elías-Ulloa, José A. 2006. Theoretical Aspects of Panoan Metrical Phonology: Disyllabic Footing and Contextual Syllable Weight, Department of Linguistics, Rutgers University: Ph. D. Dissertation.
- Ferlus, Michel. 2006. What were the four divisions (deng) of the Middle Chinese. In *The* 39th International Conference on Sino-Tibetan Language and Linguistics, University of Washington, Seattle.
- Halle, Morris, and Stevens, Kenneth. 1971. A note on laryngeal features. *Quarterly Progress Report* 101:198-213.
- Hansson, Gunnar Olafur. 2004. Tone and Voicing Agreement in Yabem. In *WCCFL 23 Proceedings*, eds. B. Schmeiser, A. Kelleher and V. Chand: Cascadilla Press.

84 Seunghun J. Lee

Hyman, Larry, and Schuh, Russell. 1974. Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* 5:81-115.

- Hyman, Larry M., and Mathangwane, Joyce. 1998. Tonal domains and depressor consonants in Ikalanga. In *Theoretical aspects of Bantu tone*, eds. L. M. Hyman and C. Kisseberth. Stanford: CSLI Publications.
- Khumalo, James Steven Mzilikazi 1987. An Autosegmental Account of Zulu Phonology, University of Witwatersrand: Ph. D. Dissertation.
- Lee, Seunghun Julio. 2008. Consonant-Tone Interaction in Optimality Theory, Rutgers University: Ph. D. Dissertation.
- Luo, Liming ed. 2005. Compilation Commission of Zhuang-Chinese-English Dictionary. Beijing: Min zu chu ban she.
- Maddieson, Ian. 1978. The frequency of tones. *UCLA Working Papers in Phonetics* 41:43-52.
- Morén, Bruce. 1999. Distinctiveness, coercion and sonority: a unified theory of weight, Linguistcs Department, University of Maryland at College Park: Ph. D. Dissertation.
- Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15:847-892.
- Peng, Long. 1992. A unified theory of Tone-Voice, Department of Linguistics, The University of Arizona: Ph. D. Dissertation.
- Prince, Alan, and Smolensky, Paul. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Basil Blackwell.
- Prince, Alan. 2002. Arguing optimality. In *Papers in Optimality Theory II*, eds. Andries W. Coetzee, Angela C. Carpenter and Paul de Lacy, 269-304. Amherst, MA: GLSA.
- Ruangjaroon, Sugunya. 2006. Consonant-Tone Interaction in Thai: An OT Analysis. *Taiwan Journal of Linguistics* 4:1-66.
- Schuh, Russell. 2002. Bade /Ngizim Phonology and Morphology. Ms., *UCLA Linguistics* 252, *Winter Quarter* 2002.
- Shi, Lin. 1983. Initial change and tone sandhi in Kam language (in Chinese). *Min Zu Yu Wen* 5:44-47.
- Wang, Jun, and Zheng, Guoqiao. 1993. *An outline grammar of Mulao ['Mulauyü chienchi' translated from Chinese by Luo Yongxian]*. Canberra: National Thai Studies Centre, The Australian National University.
- Yang, Tongyin. 1992. Phonetic Variation and its Mechanism in Kam Classifiers (in Chinese), Central Institute of Nationalities.
- Yip, Moira. 2002. *Tone*. Cambridge, United Kingdom: Cambridge University Press.

EVENTIVITY AND STATIVITY IN THAI VERBS

Natsuki Matsui

Chulalongkorn University <matsuinatsuki@gmail.com>

1. Introduction

As an "isolating" language, Thai lacks morphological markings of inflections and conjugations, and the copula occurs only with nominal complements. Whether Thai has a distinct category of "adjectives", discrete from the category of verbs, has been a highly controversial issue in the grammatical studies of the language. Whereas Prasithrathsint (2000) claims that the so-called adjectives in Thai are subsumed under the category of verbs because their syntactic behavior is not different from that of verbs, Sookgasem (1996) holds that Thai has the syntactic category of adjectives. Post (2008), more recently, also argues that the category of adjectives in Thai should be recognized since a certain cluster of predicates bears a close resemblance to predicates of other languages, both semantically and functionally. If we investigate Thai syntactic categories in terms of the criteria that are standardly applied to European languages, it will be concluded that Thai is a straightforward example of an "adjectival-verb language" (Schachter 1985:18), which does not distinguish between verbs and adjectives. Schachter presents Mandarin Chinese as a typical example of an adjectival-verb language (1985:18). Dixon (2004:13), however, throws doubt on this fairly standard assumption by pointing out that recognizing an adjective class in the "languages in which adjectives show a rather different profile" cannot be standardized according to the study of European languages.

In this paper, I will examine the semantic differences of eventive and stative predicates in Thai, with a view to shedding light on the problem of the categorial distinction. Despite the common assumption that Thai does not distinguish between verbs and adjectives syntactically as well as morphologically (Prasithrathsint 2000), explicit differences are found in the usage of certain predicates. Some researchers treat predicates such as *dii* 'good' and *suăy* 'beautiful' as "adjectives" based on the functional and semantic standards (Sookgasem 1996, Tanaka 2004, Iwasaki & Ingkaphirom 2005, Post 2008). In his analysis of Lao, a language morphosyntactically similar to Thai, Enfield (2004, 2007), on the other hand, classifies "adjectives" as a sub-type of verb in the language, because the words which denote property concepts in Lao have a similar syntactic behavior to verbs but also have distinguishing characteristics with other types of verbs in terms of their function.

By extending Enfield's view to Thai, I will draw the conclusion that Thai "adjectives" should be treated as a particular class, though not as an independent category. Given that the category is determined by the morphosyntactic behaviors or the language, a clear boundary between "adjectives" and "verbs" in Thai cannot be found to distinguish these two varieties as two syntactically distinctive categories. As Post (2008) points out, words denoting property concepts in Thai clearly involve some particular functions which deviate from other types of verbs, however. Thus I will identify Thai "adjectives" as a sub-

type of verb. Importantly, what I propose here does not follow the standard European lexical categories. The number of syntactic categories that one language involves may or may not agree between languages (see Croft 2001). Taking up syntactic categories in Japanese, Uehara (1998) elaborately analyzes the category of "nominal adjective", which holds adjectival properties but morphologically takes on noun-like qualities. This category does not occur in English, for instance, nor does another category "verbal noun" (Shibatani 1990, Uehara 1998). Japanese has an additional category for words denoting property concepts. The additional category is fully motivated to be distinguished in terms of morphosyntactic profiles. Thai, in contrast, does not exhibit sufficient evidence to set a syntactically independent category for words denoting property concepts. My claim is based on the semantic properties of various types of predicates, as represented by Lexical Conceptual Structure (LCS) (Jackendoff 1990, Levin and Rappaport Hovav 1995, Kageyama 1996), which displays "a structured lexical representation of verb meaning" (Levin and Rappaport Hovav to appear) and Argument Structure (Levin and Rappaport Hovav 1995, Pustejovsky 1995), which refers to "the syntactically relevant argumenttaking properties of a verb" (Levin and Rappaport Hovav 1995).

2. Syntactic behaviors of Thai predicates

The question of whether Thai has the category of adjective is rooted in the similarity of the syntactic behavior of predicates. Thai does not require a copula to form the predicative adjectival construction whereas English equivalent expressions do (Prasithrathsint 2000, Iwasaki & Ingkaphirom 2005). The following examples show Thai predicative constructions. The predicate in (1a) is equivalent to the English activity verb *work*, and in (1b), the English stative verb *know*. The predicate in (1c), on the other hand, corresponds to the English verb-derived adjective *drunk*, and (1d), the English adjective *beautiful*.

```
(1)
     a. fâay
                 tham-naan mâak
       Faay
                 work
                            very
       'Faay works a lot.'
     b. fâay
                 rúu mâak
       Faay
                 know very
       'Faay knows a lot.'
     c. fâay
                 maw
                                  mâak
       Faay
                 {be/get} drunk
                                  very
       'Faay is very drunk.'
     d. fâay
                 sŭay
                          mâak
       Faay
                 beautiful very
       'Faay is very beautiful.'
```

It is often observed that Thai "adjectives" and "verbs" behave in the same way in major syntactic environments.

Predicates which belong to this verb sub-type involve the semantic properties with respect to a cross-linguistic perspective introduced by Croft (1991). Adjectives are mono-relational, persistent, and gradable (p.65).

Prasithrathsint (2000) shows two types of relative clauses: one type which includes the relativizer $th\hat{i}i$ ((2a)) and another which does not involve the relativizer ((2b)).

(2) a. nákrian thîi dii
student who good
'good student' (Prasithrathsint 2000:260)
b. nákrian dii
student good
'good student' (Prasithrathsint 2000:260)

Prasithrathsint explains that both types of relative clauses function similarly as they denote quality and modify the head (p.260). But she also points out that the two varieties semantically differ. Post (2008), as well as Kuno & Wongkomthong (1981), also argues that the clause which lacks *thîi* is more functionally restricted than the one with *thîi*. The clause without *thîi* prefers non-dynamic predicates, i.e., states. However, Prasithrathsint (2000) argues that noun phrases like (2b) which appear to involve an "attributive adjective" should be analyzed as noun phrases that involve a verb in which the relativizer *thîi* is omitted. Because there is no feasible syntactic distinction between the two categories, Prasithrathsint thus maintains that Thai is an "adjectival-verb language" in the sense of Schachter (1985).

Predicates like *maw* '{be/get} drunk' and *sŭay* 'beautiful' whose equivalent expressions in English belong to the category of adjectives are also considered to be verbs in the relative clauses in the sentences below.²

(3) a. fâay chôopkhon (thîi) maw Faay like person REL {be/get}drunk 'Faay likes a drunk person.'

> b. fâay chôopkhon (thîi) sǔay Faay like person REL beautiful 'Faay likes a beautiful person.'

Prasithrathsint (2000) also shows that both an active predicate like kin 'eat' and a stative predicate like ruay 'be rich' can cooccur with an aspect marker $l\acute{e}ew$ in Thai.

(4) a. khǎw kin lέεw (s)he eat already

(S)he has eaten.' (Prasithrathsint 2000:262)

b. khắw ruay léεw chiiwít khắw mây mɨan tèε-kòɔn (s)he rich already life (s)he not like past '(S)he has become rich. His/her life is not like in the past.'

(Prasithrathsint 2000:262)

Abbreviation: ACHV=achievement; EMP=emphatic; FINAL PTCL=final particle; IMPR=imperative; NOM=nominal prefix; PFX=prefix; PROG=progressive; RDP=reduplication; REL=relativizer

(5) a. khǎw kamlan kin (s)he now eat '(S)he is eating.' (Prasithrathsint 2000:263) b. khǎw kamlan ruay ləəy chôop khɔ̃ɔŋ síi pheen (s)he rich so like buy thing expensive now '(S)he is rich now. (S)he likes to buy expensive things.'

(Prasithrathsint 2000:263)

Since the categories of words are assumed to be recognized according to their syntactic behaviors, Thai may not need a set "syntactic" category of adjectives. Yet, as is pointed out by Croft (1990/2003, 1991) and Dixon (2004), as long as some group of predicates shows any particular profile which differs from other group, the group can be recognized as a distinct class from another group of predicates.

Considering both semantic/functional and morphological/syntactic aspects, Enfield (2007:242) labels adjectives as a sub-type of stative verbs. He classifies verbs in Lao based on their "lexical aspect" (Rothstein 2004) or "situation aspect" (Smith 1997), which is also called "Aktionsart" and which "covers distinction between properties of event-types denoted by verbal expressions" (Rothstein 2004:1). The classification of verb classes are firstly proposed by Vendler (1957, 1967), who showed four types of verbal aspect: "states", "activities", "achievements", and "accomplishments" (see also Tenny 1994, Smith 1997, Lin 2004, and Rothstein 2004). Each class differs in at least one viewpoint out of three: (i) whether the eventuality described by the verb is eventive (dynamic) or stative (static), (ii) whether the verb conveys a durative event or a punctual event, and (iii) whether the verb involves the event with an natural endpoint (telic) or not (atelic) (Tenny 1994, Smith 1997, Lin 2004, Rothstein 2004).

| Table 1: Vendler | 's | four | types | of verb | classes |
|-------------------------|----|------|-------|---------|---------|
|-------------------------|----|------|-------|---------|---------|

| | State | Activity | Achievement | Accomplishment |
|----------|----------|----------|-------------|----------------|
| Dynamics | stative | eventive | eventive | eventive |
| Duration | durative | durative | punctual | durative |
| Telicity | atelic | atelic | telic | telic |

Enfield (2004, 2007) classifies Lao verbs into majorly two types, "active verbs" and "stative verbs". Active verbs are further subclassified into "achievement verbs" and "activity verbs". The former type is punctual in temporal structure and the latter is durative. Activity verbs include "accomplishment verbs", which involve events "progressing towards a goal" (Tenny 1994:5) and "open activity verbs", which contains no such notion. Stative verbs, on the other hand, are subdivided into "adjective verbs" and "state verbs". The classification is shown in Figure 1.

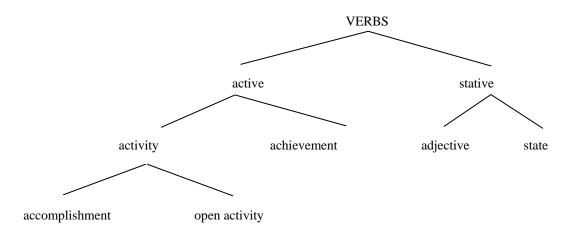


Figure 1: *Taxonomy of verb sub-types* (Enfield 2004:329, 2007:242)

Enfield (2007:242-245) demonstrates subtle differences in semantics and pragmatics between "active verbs" and "stative verbs" by using the cooccurrence test on all types of verbs with aspectual markers. The irrealis markers of $si\theta$ and $ca\theta$ and the negator $b\partial\theta$, for instance, can cooccur with all types of verbs all of which are directly marked, but there is a difference in semantics between active verbs (e.g., $\tilde{n}aangl$ 'walk') and stative verbs (e.g., mii2 'have', suung3 'tall') when they cooccur with the negator $b\partial\theta$.

```
(6)
     a. khòòj5 bòø ñaang1
       1SG.P
                NEG walk
       'I don't/won't walk.'
                                           (Enfield 2007:243)
     b. khòòj5 bòø mii2 pùm4
                NEG have book
       1SG.P
       'I don't have a book.'
                                           (Enfield 2007:243)
     c. khòòj5 bòø suung3
       1SG.P
                NEG tall
       'I am not tall.'
                                           (Enfield 2007:243)
```

Varying from the active verb in (6a), the state verb in (6b) and the adjective verb in (6c) (both of which are subclassed as stative verbs) do not give future reading with the negation marker.

Let us now turn to Thai predicates. In Thai, some predicates also seem to be less preferable than other verbs when they are marked with aspectual markers such as $d\hat{a}y$ (an achievement marker) and kamlan (a progressive marker).

(7) a. fâay {dây /kamlaŋ} tham-ŋaan
Faay ACHV PROG work
'Faay {worked/is working}.'
b. fâay {dây /?kamlaŋ} rúu
Faay ACHV PROG know
'Faay {got to know/knows}.'

```
c. fâay
           {?dây
                      /kamlan}
                                 maw
  Faay
            ACHV
                       PROG
                                  {be/get} drunk
 'Faay {got drunk/is being drunk}.'
d. fâay
           {??dây
                      /?kamlan} sŭay
                                 beautiful
 Faay
            ACHV
                       PROG
 'Faay {was beautiful/is being beautiful}.'
```

All predicates in (7) can cooccur with these aspectual markers either completely or marginally. The differences in acceptability in (7) are not due to syntax but rather to pragmatics. That predicates, similar to Lao predicates, are heavily controlled by a pragmatic preference when they cooccur with aspectual markers. In other words, if the context allows, a less preferable cooccurrence becomes acceptable, as is shown below.

```
(8) a. ?fâay kamlaŋ chalàat
Faay PROG clever
'Faay is being clever.'
```

b. fâay kamlan chalàat phró àan nănsww yá Faay PROG clever because read book many 'Faay is being intelligent since she is reading many books.'

If we hold that the syntactic category of a word should be determined by its syntactic behavior, we can conclude that Thai does not have a "syntactic category" of adjectives. But syntactic behaviors of words per se do not substantiate whether all predicates can be identified as the same type of predicate.

By espousing Enfield's (2004, 2007) perspective of Lao verbs' classification, I claim that "adjectives" should be treated as a particular lexical class in Thai, though not as a syntactically independent category. The data in (7) suggests the verbs in each sentence involve differences and similarities in semantics. Tansiri (2007:54) points out that *kamlaŋ* "profiles a dynamic phase of the situations and construes them as on-going processes". Due to the semantics of the progressive marker and the verb, *kamlaŋ tham-ŋaan* in (7a) describes an on-going process of working, but the rest of the verbs, which can be classified into "stative verbs", imply or emphasize the temporality of the state when they combine with *kamlaŋ*. On the other hand, the stative verbs *rúu* 'know' in (7b) and *maw* '{be/get} drunk' in (7c) display a noticeable difference in semantics when they cooccur with the achievement marker *dây*. In contrast to *sǔay* 'beautiful' in (7d), these predicates have an inchoative reading with *dây*. In the next section, I will examine these stative verbs and classify them into three types according to their semantic differences.

3. Classification of Thai stative verbs

3.1. Eventivity and stativity of the verb

Since Thai words are not marked by morphological change, denoting states by verbs leads to the three-way ambiguity of some Thai change-of-state verbs, depending on context. The possible states are: (i) causative change of state, (ii) change of state (inchoative), and (iii) resultant state, as is pointed out by Thepkanjana (2000:268). Verbs which are associated with this "causative-inchoative-stative alternation" are, for instance, *pît* (*pratuu*) "close

(door)", pàət pratuu "open (door), phaŋ (bâan)"demolish (house)", dàp (thian) "blow out (candle)", khôon (tônmáy) "fell (tree)" (Thepkanjana 2000:267).

The verb $p \partial t$ delivers a causative reading when it is used as a transitive verb as in (9a). In contrast, the interpretation wavers between an inchoative reading and a resultant reading when it is used as an intransitive verb as in (9b). Some Thai causative change-of-state verbs can thus denote the state of the object entity by being used as intransitive verbs. The function of these verbs cannot be identified by their forms alone, but can be identified only by the way they are used in context.

Considering Thepkanjana's (2000) analysis of the interpretational ambiguity of the verbs like $p \partial \sigma t$ 'open', it can be assumed that Thai encompasses an extra type of "stative verb" which indicate either state-change or resultant state (inchoative/stative alternation in Thepkanjana's terminology) other than the stative verb which describe simple states without an additional inchoative meaning.

```
(10) a. fâay maw
Faay {be/get} drunk
'Faay {is/got} drunk.'
b. fâay rúu khwaam-ciŋ
Faay know NOM-true
'Faay knows the truth.' (NOT 'Faay got to know the truth.')
```

The state denoted by this type of verb differs from the state denoted by the verb which does not involve inchoative/stative alternation such as *rúu* 'know' and *sŭay* 'beautiful'. The former denotes a temporary or unsustainable state, while the latter usually indicates a rather constant state which does not involve any "temporal boundary" as is pointed out by Kageyama (1996) with respect to the English stative verb *know*. In other words, the states represented by the verbs without inchoative/stative alternation can be identified with "individual" states, and those by the verbs including the example in (10a) with "stage-level" states in the sense of Carlson (1977).

According to Kageyama (2006:97), stage-level states involve an "event argument" [E-argument], which is linked to "eventivity", in contrast to individual-level states, which lack such a link. He suggests that "only sentences of event description have an E-argument; stative sentences of property description do not have one" (2006:96). I employ the terminology of "eventivity" to signify an element of event dynamics. Following Kageyama's proposal, I claim that Thai stative predicates such as *rúu* 'know' and *mii* 'have', as well as predicates such as *sǔay* 'beautiful' and *dii* 'good', do not have an E-argument, while those like *nùay* '{be/get}tired' and *maw* '{be/get} drunk' in (10) have one. The adverb *bòy* 'often' seems to be able to cooccur with the verbs which involve an

E-argument, since the adverb indicates the frequency of the event denoted by the predicate. Stativity is not related to frequency, although it is related to gradability as is illustrated by the cooccurrence with the adverb *mâak*.

```
(11) a. fâay
                 kin
                       {bòy/ mâak}
                       often/very
        Faay
                 eat
        'Faay often eats./Faay eats a lot.'
      b. fâay
                                   {bòy/ mâak}
                 maw
        Faav
                 {be/get}drunk
                                   often/verv
        'Faay often gets drunk./Faay is very drunk.'
      c. fâav
                      {*bòy/ mâak}
        Faay
                 know often/very
        'Faay knows a lot.'
      d. fâay
                             {*bòy/ mâak}.
                 sŭay
        Faay
                 beautiful
                              often/very
        'Faay is very beautiful.'
```

The frequency adverb $b\partial y$ cannot cooccur with the stative verb $r\hat{u}u$ 'know' in (11c) and $s\check{u}ay$ 'beautiful' in (11d). The activity verb kin 'eat' in (11a) and another type of the stative verb maw in (11b), in contrast, can cooccur with both $b\partial y$ 'often' and $m\hat{a}ak$ 'very'. Activity verbs like kin 'eat', however, do not show any semantic alternation of the verb, varying from verbs like maw '{be/get} drunk'. Notice that the verb maw has an inchoative interpretation when it appears with the frequency adverb $b\partial y$ 'often', but on the other hand, the verb has a stative reading with the gradable adverb $m\hat{a}ak$ as in (11b). The adverb $m\hat{a}ak$ describes quantity but not degree when it appears with "pure" eventive verbs such as kin 'eat' in (11a). The ampleness that the adverb indicates may refer to the amount of the food in (11a). On the other hand, the stative verbs in (11c) and (11d) do not cooccur with the frequency adverb. Furthermore, they do not show an alternation between a stative and inchoative interpretation.

The verbs in (11) are distinguished as to whether they have an E-argument or not. Stative verbs like *maw* '{be/get} drunk,' which show inchoative/stative alternation, encompass an E-argument, as activity verbs like *kin* 'eat'. I will refer to these types of stative verbs as "inchoative stative verbs". The other types of stative verbs, on the other hand, are not associated with an E-argument. Instead, they hold a state-argument (S-argument). A "S(tate)-argument" is proposed by Parsons (1990) for English sentences with adjectives (e.g.,) as well as locatives and state verbs. I will discuss the relation between the E/S-arguments and eventivity in section 4.

3.2. Three types of stative verbs in Thai

A cooccurrence test of $s\check{u}ay$ 'beautiful' with the two types of frequency adverbs $b\partial y$ 'often' and $m\hat{a}ak$ 'very' in (11d) shows the same result as $r\check{u}u$ 'know' in (11c). Although both

In another instance, *mâak* in *fâay tham-ŋaan mâak* 'Faay works a lot.' indicates that the amount of the time of the event of working is ample.

Parsons (1990) argues nouns such as *giraffe* (but not those like *killing*) are not associated with events, but there is no certain evidence that shows those involve states, either.

predicates hold an S-argument, I do not classify these verbs into the same type. The difference between verbs like *rúu* 'know' and *mii* 'have' and those like *sǔay* 'beautiful' and *dii* 'good' can be seen in the cooccurrence with the aspectual marker *phûn* 'just'. Consider the sentences below.

```
(12) a. fâay phûŋ maw
Faay just {be/get} drunk
'Faay just got drunk.'
```

- b. fâay phŵŋ rúu Faay just know 'Faay just got to know.'
- c. ?fâay phûŋ sǔay Faay just beautiful 'Faay just became beautiful.'

Notice that suay 'beautiful' is less preferable than maw '{be/get} drunk' and ruu 'know' when it cooccurs with the aspectual marker *phûn* 'just' without any coercive readings.⁵ The verb rúu 'know', similar to the verb maw '{be/get} drunk', has an inchoative reading when it appears with *phŵŋ* 'just'. I assume that this difference of acceptability between *rúu* 'know' and suay is due to the difference of the semantic structure of these verbs. Although "eventivity" is linked to an E-argument, some verbs encompass eventivity without being associated with an E-argument. Stative verbs like rúu 'know' supposedly include eventivity, but it is inherently suppressed unless elements which tend to be associated with eventivity cooccur (such as phŵn 'just') and coerce the predicate to semantically combine them. In other words, stative verbs like rúu 'know' do not include an E-argument in the Argument Structure when the verb appears without an element which can activate the inherently suppressed dynamics, i.e., eventivity, in Lexical Conceptual Structure (LCS). This semantic property makes the stative verbs like *rúu* 'know' different from the stative verbs like *sŭay* 'beautiful'. I will refer to the former type of stative verb as an "inherently stative verb", in contrast to the latter type, which will be called an "adjectival verb". The knowing-state, for instance, implies that the not knowing-state must exist before the knowing-state is achieved. In other words, the existence of states denoted by inherently stative verbs necessarily implies that the denoted state arises as a result of a state change which is responsible for eventivity. Differing from inherently stative verbs, verbs like sŭay 'beautiful' do not embrace this particular implication of previous state-change. Differences in LCS are illustrated in the next section.

The cooccurrence of *phûŋ* 'just' and *sǔay* 'beautiful' might be acceptable when the two words appear in a sentence like (i).

⁽i) fâay phŵn sǔay toon tèn nâa sèt Faay just beautiful phase decorate face finish

^{&#}x27;Faay just became beautiful as she finished her make up."

Now I will show more examples of the difference in semantics between the three types of stative verbs. Prasithrathsint (2000) uses the example of the imperative with the auxiliary verb *con* in the literary style as evidence to argue that predicates such as *dii* 'good' are verbs.

```
(13) a. con
                phûut
                         mîa
                                 mii khon
                                                  thǎam
                        when
                speak
                                 have person
                                                  ask
        IMPR
        'Speak when somebody asks you.'
                                                             (Prasithrathsint 2000:262)
     b. con
                dii
                      thalòot-pay
                good forever
        IMPR
        'Be good forever.'
                                                             (Prasithrathsint 2000:262)
```

The following colloquial sentences, however, show a semantic difference between the sentences with eventive and stative verbs when they are employed with the particle $s\hat{i}$. The final particle $s\hat{i}$ is used in two ways: one for an imperative sentence, and the other as an emphatic affirmative answer to a question such as 'Do you work?', 'Are you drunk?', 'Do you know?', and 'Is (it) beautiful?'.

```
(14) a. tham-ŋaan sì work FINAL PTCL.{IMPR/EMP} 'Work!'/ '(I) work, indeed.'
b. maw sì {be/get} drunk FINAL PTCL.{IMPR/EMP} 'Get drunk!' / '(I am) drunk, indeed.'
c. rúu sì know FINAL PTCL.{IMPR/EMP} '(I) know, indeed.' (MOSTLY UNACCEPTABLE 'Get to know!')
d. sǔay sì beautiful FINAL PTCL.{IMPR/EMP} 'Beautiful, indeed.' (NOT 'Be beautiful!')
```

As the examples in (14) show, all stative verbs are syntactically able to cooccur with the final particle $s\hat{i}$, although there is a difference in semantics. The difference in semantics between maw '{be/get} drunk', which semantically involves an inchoative/stative alternation, and the inherently stative verbs $r\hat{u}u$ 'know' and adjectival verb $s\check{u}ay$ 'beautiful' is explicit in this case. Note that $s\check{u}ay$ 'beautiful' in (14d) cannot be construed as the imperative but $r\hat{u}u$ 'know' in (14c) may coercively have an imperative reading. Presumably, the final particle $s\hat{i}$ requires eventivity in the LCS of a head verb to produce the imperative construal. If the verb does not embrace eventivity, the outcome of the semantic composition results in an emphatic reading.

Let us observe the next examples which display a clear difference between the verb *rúu* 'know' and the verb *sǔay* 'beautiful'.

⁶ In the case that the LCS of the verb is associated to eventivity, the construal will be pragmatically determined.

```
(15) a. tham-ηaan
                      sà
       work
                      FINAL PTCL. IMPR
       'Work!'
     b. maw
                      sà
      {be/get} drunk FINAL PTCL. IMPR
       'Get drunk!'
     c. rúu
       know
                FINAL PTCL. IMPR
       'Get to know!'
     d. *sŭav
                sà
       beautiful FINAL PTCL. IMPR
       (Intended meaning) 'Be beautiful!'
```

The final particle sa exhibits a more obvious semantic difference between three types of stative verbs, i.e., the inchoative stative verb maw '{be/get} drunk' and the inherent stative verb $r\acute{u}u$ 'know', and the adjectival verb $s\check{u}ay$ 'beautiful'. Notice that inherently stative verbs like $r\acute{u}u$ 'know' have an inchoative interpretation in this case. Adjectival verbs like $s\check{u}ay$ 'beautiful', however, cannot cooccur with the imperative final particle sa. Adjectival verbs display a peculiarity as to their acceptability in (15). I assume the reason that the adjectival verbs like $s\check{u}ay$ cannot cooccur with the imperative final particle sa is that sa is sensitive with respect to eventivity. In other words, the imperative final particle sa requires verbs which involve a dynamic phase. "Pure" stative verbs like $s\check{u}ay$ 'beautiful' do not suppress eventivity in their LCS but inherently lack it, and thus they cannot be combined with the final particle sa which calls for a verb with eventivity. In contrast, inherently stative verbs like $r\acute{u}u$ 'know' de-suppress their eventivity by cooccurring with an eventivity-demanding particle, thus forming an imperative sentence.

The semantic distributions of the imperative markers con, si, and sa can be accounted for as follows: (i) con can be related with either an E-argument or an Sargument of a predicate, assuming that it involves both an E-argument and an S-argument element; (ii) si selects either an imperative or an emphatic marker depending on a predicate. The imperative marker con links only to an E-argument, whereas $s\hat{i}$ can connect with either an E-argument or an S-argument; (iii) sa, which encompasses only an imperative marker, picks the function for an eventivity. The observation of the semantic differences above encourages my claim that Thai stative verbs can be classified into three types: (i) "inchoative stative verbs", (ii) "inherently stative verbs", and (iii) "adjectival stative verbs". Inchoative stative verbs like maw '{be/get} drunk' allow inchoative/stative alternation with or without the assistance of other elements. On the other hand, inherently stative verbs like *rúu* 'know', which do not involve an E-argument but are associated with eventivity, denote only states per se. They can assume an inchoative reading by being modified with other elements such as a final particle or an aspectual marker. Lastly, adjectival stative verbs like sŭay 'beautiful' do not alternate between a stative and inchoative interpretation unless the verbs are coerced to have some definite reading, since they do not lexically include a dynamic phase. Differences in semantic prosperities of Thai stative verbs are summed in Table 2.

| | Stativity | Event argument | Eventivity |
|-------------------------|-------------|----------------|------------|
| Inchoative stative verb | ±stative | S/E | included |
| Inherently stative verb | +(-)stative | S/(E) | suppressed |

S

not included

Table 2: Semantic features of Thai stative verbs

+stative

4. Lexical conceptual structure and argument structure of Thai verbs

4.1 Eventive verbs

Adjectival verb

Enfield (2007) classifies verbs in Lao into major two types, "active" and "stative". Active verbs are subcategorized into two types, "achievement" and "activity". These types, in turn are further classified into two sub-classes, "accomplishment" and "open activity". Stative verbs are classified into "adjective" and "state" (See Figure 1.). Obviously, the active verbs in Lao are the verbs which hold an E-argument. Thai verbs which include an E-argument can be classified into three types, "activity verbs" (e.g. tham-ŋaan 'work', wîŋ 'run'), "achievement verbs" (e.g. phóp 'meet', khâw 'enter'), and "accomplishment verbs" (e.g. tham 'make', sâaŋ 'build'). I will call these verbs "eventive verbs" in contrast to "stative verbs". The LCS representations and the Argument Structure (AS) of the three types of eventive verbs are sketched out in (16). Generally, the AS signifies the valency of verbs. Pustejovsky (1995:63) sees it as "a minimal specification of its lexical semantics". The AS used in this paper is based on the AS which Kageyama (2006:97-98) proposes. What differs from Kageyama's AS is that the AS in this paper involves a state argument, S-argument (St) in addition to an event argument, E-argument (Ev).

```
(16) a. (i) ACTIVITY VERBS (INTRANSITIVE) (e.g. tham-ŋaan 'work', wîŋ 'run')

LCS: [EVENT [ ]x ACT]

AS: (Ev (x<>))

(ii) ACTIVITY VERBS (TRANSITIVE) (e.g. tii 'hit', tê 'kick')

LCS: [EVENT [ ]x ACT ON-[ ]y]

AS: (Ev (x<y>))

b. ACHIEVEMENT VERBS (e.g. phóp 'meet', khâw 'enter')

LCS: [EVENT BECOME [STATE [ ]y BE AT-[ ]z]]

AS: (Ev (<y>))

c. ACCOMPLISHMENT VERBS (e.g. khâa 'kill', sâaŋ 'build')

LCS: [EVENT [EVENT [ ]x (ACT ON-[ ]y/w)] CAUSE [EVENT BECOME [STATE [ ]y BE AT-[ ]z]]]

AS: (Ev (x<y>))
```

Note that the LCS representations above virtually follow Kageyama (1996, 2008) among other types of representations. According to Kageyama's (2006:97-98) theory, the *EVENT* in the LCS representation in (16a-c) is mapped to the AS as the E-argument. Sentences involve an E-argument when the LCS of verbs is related to an "eventivity". The leftmost or "highest" event node in the LCS representation corresponds to the E-argument of the verb. The AS can be divided into two sections: one is the "event section" which is denoted by *Ev*

and the other is the "theta section" which marks whether the verb includes either an "external" argument or an agentive argument (x < y), an "internal" argument (< y >), or both arguments (x < y >) (Kageyama 2006:97-99). []z in the LCS of achievement verbs and accomplishment verbs signifies either a physical space or an abstract state. []w in the LCS of accomplishment verbs, on the other hand, indicates materials or an ingredient of the finished product in the case of verbs such as tham 'make', sâaŋ 'build'. In contrast, []y of ACT ON-[]y/w in (16c) signifies an affected direct object of verbs such as khâa 'kill' and cut 'tat'.

```
(17) a. sùdaa
                pòət pràtuu
        Suda
                open gate
        'Suda opened the gate.'
                                            (Thepkanjana 2000:268)
     b. pràtuu
                pàət
        gate
                open
        'The gate opened.'
                                            (Thepkanjana 2000:268)
     c. pràtuu
                pèət
                open
        gate
        'The gate was open'
                                             (Thepkanjana 2000:268)
```

- (18) ACCOMPLISHMENT/INCHOATIVE STATIVE VERBS (e.g. *pòət* 'open (vt)/open (vi)/be open', *lóm* 'drop down/fall/be down')
 - (i) Transitive usage

 LCS: [EVENT [EVENT []x (ACT ON-[]y)] CAUSE [EVENT BECOME

 [STATE []y BE AT-[]z]]]

 AS: (Ev (x<y>))
 - (ii) Intransitive usage: Inchoative reading

 LCS: [EVENT BECOME [STATE []y AT-[]z]]

 AS: (Ev (<y>))
 - (iii) Intransitive usage: Stative reading LCS: [IMP (EVENT BECOME) [STATE []y AT-[]z]] AS: (St (<y>))

The accomplishment/inchoative stative verbs alternate between an eventive reading, i.e., inchoative reading, and a stative reading, on top of the alternation between the transitive and intransitive form.

First, when the accomplishment transitive verbs alter with their intransitive usages, the superordinate subevent in the transitive usage (EVENT []x (ACT ON-[]y) in the LCS in (18 i)) is not included in the LCS since the external argument of x does not participate in the event, which is shown in the LCS in (18 ii). Second, the mechanism of the alternation of intransitive usage can be accounted for, in that this type of verb may involve a particular operator that switches the interpretation. Note that the eventivity part EVENT BECOME which is enclosed in parentheses (18 iii) in the LCS is suppressed when the verb has a stative reading. Matsui (2007) stipulates that the suppression is induced by the "implication operator (IMP)", which is shown in the LCS representation in (18 iii). The IMP forces an inchoative event into an implication of the word. By virtue of this operation, the causing event is only implicated, but not entailed in the verbs. Since the eventivity is suppressed by the IMP, the verbs do not contain an E-argument, but instead encompass an S-argument. I claim that the IMP fulfills its function when a context or a usage with other elements such as an aspectual marker or a particle triggers the operation. I assume that the basic meaning of the intransitive form of accomplishment/inchoative stative verbs is the inchoative reading, which becomes a stative reading by the rule of "stativalization" (Matsui 2007). Stativalization involves a function where the IMP operator turns a causing event into a background eventivity.

The accomplishment/inchoative stative verbs are not the only class which involves both a transitive/intransitive alternation and an eventive/stative alternation. As is argued by Thepkanjana (2000), Thai includes other types of transitive verbs such as *hak* 'break' whose temporal structure is different from the accomplishment/inchoative stative verbs.

```
(19) a. sùdaa hàk kìŋmáy
Suda break twig

'Suda broke the twig'

(Thepkanjana 2000:269)

b. kìŋmáy phŵŋhàk<sup>7</sup>
twig just break
'The twig has just broken.'

c. kìŋmáy hàk
twig break
'The twig was broken.' [stative reading]

(Thepkanjana 2000:269)
```

Thepkanjana (2000:268) points out that these types of verbs "do not semantically encompass the 'becoming' process or the event in which a non-agentive participant 'acts out' a change leading to the culmination point". That is, the *breaking*-event in a causative usage (19a) and *being broken*-event in an inchoative usage (19b) involve a punctual change, in contrast to the accomplishment/inchoative stative verbs which are associated with a durative change. These types of verbs are referred to as "achievement/ inchoative stative verbs" in this paper. The LCS representations of these types of verbs are as follows.

Sentence (19b) is provided by an anonymous reader who suggests that the verb involves an inchoative reading.

- (20) ACHIEVEMENT/INCHOATIVE STATIVE VERBS (e.g. *hàk* 'break/be broken', *pháp* 'fold/be folded')
 - (i) Transitive usage

LCS: [EVENT [EVENT []x (ACT ON-[]y)] CAUSE [EVENT BECOME [STATE []y BE AT-[]z]]]

AS: (Ev (x < y >))

(ii) Intransitive usage: Inchoative reading

LCS: [EVENT BECOME [STATE []y AT-[]z]]

AS: (Ev (<y>))

(ii) Intransitive usage: Stative reading

LCS: [IMP (EVENT BECOME) [STATE []y AT-[]z]]

AS: (St (<y>))

The LCS representations and AS of these types of verbs are identical to those of accomplishment/inchoative stative verbs. The difference between these two types is whether the transition is related to process.

In sum, eventive verbs can be divided into three types: (i) activity verbs, (ii) achievement verbs, and (iii) accomplishment verbs. Moreover, the achievement verbs and accomplishment verbs are subclassified into two types: (i) verbs which do not show transitive/intransitive alternation, (ii) verbs which involve both a transitive/intransitive alternation and an inchoative/stative alternation.

Table 3: Semantic features of Thai eventive verbs

| | Durativity | Eventivity | AS | Valency |
|------------------------------|------------|------------|----|---------------------|
| Activity verb (intransitive) | durative | including | Е | X |
| Activity verb (transitive) | durative | including | E | <i>x</i> , <i>y</i> |
| Achievement verb | punctual | including | Е | у |
| Achievement/i-s verb* | | | | |
| (transitive) | punctual | including | E | <i>x</i> , <i>y</i> |
| (intransitive/inchoative) | punctual | including | E | у |
| (intransitive/stative) | punctual | suppressed | S | у |
| Accomplishment verb | durative | including | E | <i>x</i> , <i>y</i> |
| Accomplishment/i-s verb | | | | |
| (transitive) | durative | including | E | <i>x</i> , <i>y</i> |
| (intransitive/inchoative) | durative | including | E | у |
| (intransitive/stative) | durative | suppressed | S | у |

... *"i-s" = "intransitive stative"

4.2 Stative verbs

As is discussed above, Thai stative verbs can be classified into three types, (i) inchoative stative verbs (e.g. maw '{be/get} drunk'), (ii) inherently stative verbs (e.g. rúu 'know'), and (iii) adjectival stative verbs (e.g. sǔay 'beautiful'). Now let us observe the LCS representations of these stative verbs. Similar to accomplishment/inchoative stative verbs, the basic semantics of inchoative/stative verbs is assumed to involve an inchoative interpretation. Consider the following LCS representations of inchoative stative verbs.

```
(21) INCHOATIVE STATIVE VERBS (e.g. maw '{be/get} drunk', nway '{be/get} tired')
(i) Inchoative reading
        LCS: [EVENT BECOME [STATE [ ]y AT-[ ]z]]
        AS: (Ev (<y>))
(ii) Stative reading
        LCS: [IMP (EVENT BECOME) [STATE [ ]y AT-[ ]z](])
        AS: (St (<y>))
```

Eventivity is originally encompassed in inchoative stative verbs since they hold the dynamic phase of the state-change, which is designated by the *event BECOME* of the LCS in (21 i). The *IMP* operator, whose function is to shadow the eventivity of the predicate in order to obtain a feature of stativity, can be triggered by pragmatics or cooccurrence with elements such as an aspectual marker or a final particle which can semantically induce the operation (21 ii). Because of this operation, eventivity of the verb is suppressed, and consequently, the verb does not contain an E-argument. Instead, it is associated with an S-argument which is related to the stativity of the subevent and thus, can have a stative reading.

Varying from the inchoative stative verbs, inherently stative verbs such as *rúu* 'know' are not associated with an E-argument. Those verbs, however, are related with "eventivity" in LCS, as we discussed above. The property of holding a dynamic element in the LCS distinguishes these verbs from adjectival verbs, which are "purely" stative. The LCS and AS representations of inherently stative verbs and adjectival verbs are sketched out below.

```
(22) a. INHERENTLY STATIVE VERBS (e.g. rúu 'know', mii 'have')
LCS: [(IMP [EVENT BECOME]) [STATE [ ]x BE AT-STATE-of-[ ]y]]
AS: (St (x<y>))

b. ADJECTIVAL VERBS (e.g. sǔay 'beautiful', dii 'good')
LCS: [STATE [ ]y BE AT-STATE]
AS: (St (<y>))
```

Again, note that the distinctiveness of the inherently stative verbs and adjectival verbs is dependent on whether the verb is associated with "eventivity" or not. I suggest that selecting an aspectual marker and a final particle in Thai involves two aspects in the semantic structure of the verb. One is whether the AS is related with an E-argument or an S-argument; the other is whether the LCS includes eventivity, either evident or suppressed. For this, the final particle sa (an imperative marker) is compatible with the inherent stative verb, but not with the adjectival verb, because having a dynamic feature, i.e., eventivity, is necessary to signify inchoativity. The acceptability of cooccurrence of the verb with these elements hinges on the consistency between the LCS and the AS of the verbs and the semantics of the elements.

A significant difference between inchoative stative verbs and inherently stative verbs is that the former type of verb allows the inchoative/stative alternation with or without the assistance of other elements. The latter type of verb, on the other hand, usually

has a stative reading only, but if it is modified by some other elements, it may also have an inchoative reading. Compare the following sentences and representations.

- (23) a. fâay rúu khwaam-ciŋ Faay know truth 'Faay knows the truth.'
 - b. LCS: [(IMP [EVENT BECOME]) [STATE *fâay* BE AT--*knowing*-of- *khwaam-ciŋ*]] AS: (St (x<y>))
- (24) a. fâay phŵn rúu khwaam-cin Faay just know truth 'Faay just got to know the truth.'
 - b. LCS: [EVENT BECOME [STATE *fâay* BE AT-*knowing*-of- *khwaam-ciŋ*]] AS: (Ev (x<y>))

Again, the *IMP* operator functions to suppress the eventivity in the LCS. Since the aspectual marker *phûŋ* 'just', which requires an inchoativity, inactivates the *IMP* operator in the LCS of the inherently stative verb, the inherently suppressed eventivity emerges in the LCS and sequentially the AS holds an E-argument (24b). Therefore, the event that is described by the sentence in (24a) has an inchoative reading. Intriguingly, Thai adjectival verbs may also be coerced to have an inchoative reading. But this is generally less preferable than other types of stative verbs. I assume the function of an inchoative/stative alternation of adjectival verbs might be a result of the elements which cooccur with the verbs, but not to the semantics of the verb. The Thai aspectual marker is presumably able to coerce a stative reading to the inchoative reading according to the context. However, the final particles such as $s\hat{i}$ and $s\hat{a}$ are more involved with the semantics of the verb, as the differences in interpretation and in acceptability suggest in (14) and (15).

Another test that reveals the difference between inherently stative verbs and adjectival verbs is to check whether they can be used as an adverb. The adjectival verbs can be used as adverbs, but on the other hand, inherently stative verbs cannot. The latter type of verb requires some element such as *mtttan* 'same' to become an adverbial phrase.

- (25) a. *fâay phûut rúu Faay speak know (Intended meaning) 'Faay speaks knowingly.'
 - b. fâay phûut mttian rúu Faay speak same know 'Faay speaks like she knows.'
- (26) fâay tênram sǔay Faay dance beautiful 'Faay dances beautifully.'

In addition, an aspectual marker and a final particle are not the only elements that can trigger an inchoative/stative alternation. The stativalization rule can also take place when the verb appears with provocative elements which can trigger to suppress eventivity in the

LCS. For instance, the inchoative stative verb *tuitun* 'be awake/wake' is forced to have a stative interpretation when it appears in the sentence below.

```
(27) cíap tưừun yùu
Jiap be awake/wake exist
'Jiap is awake.'
```

The grammaticalized verb yuu 'exist' is employed to produce a progressive sentence with the main verb. When this grammaticalized verb appears with an inchoative stative verb, the semantic property of yuu 'exist', which is associated with duration, attracts a stative reading of the main verb. Although the inchoative stative verbs lexically involve such an alternation, the coercion of the reading can be clarified with such elements as an aspectual marker (e.g. phuu), a final particle (e.g. sa), a grammaticalized verb (e.g. yuu), or an adverbial prefix like yau.

```
(28) a. cíap
                phûut
                            yàan maw
        Jiap
                 speak
                            PFX {be/get} drunk
        'Jiap speaks drunkenly.'
     b. cíap
                phûut
                            vàan rúu
                            PFX know
        Jiap
                 speak
        'Jiap speaks knowingly.'
     c. cíap
                phûut
                            yàan suphâap
                            PFX polite
        Jiap
                 speak
        'Jiap speaks politely.'
```

I assume that the inchoative stative verbs undergo the stativalization rule when they form an adverbial phrase with the prefix *yaaŋ*.

5. Conclusion

I have argued that Thai verbs can be divided into two major types, eventive verbs and stative verbs. Eventive verbs include five subtypes: (i) ACTIVITY VERBS, (ii) ACHIEVEMENT VERBS, (iii) ACCOMPLISHMENT VERBS, (iv) ACCOMPLISHMENT/INCHOATIVE STATIVE VERBS, and (v) ACCOMPLISHMENT/STATIVE VERBS. On the other hand, stative verbs can be subclassified into three types: (vi) INCHOATIVE STATIVE VERBS, (vii) INHERENTLY STATIVE VERBS, and (viii) ADJECTIVAL VERBS. Both types of verbs include a subtype which involves an inchoative/stative alternation ((iv) and (vi)). The classification of verbs is based on the property of the verbs associated with "eventivity" (cf. Kageyama 2006). The major distinction lies between eventive and non-eventive verbs. The eventive verbs of group (i)-(iii) are consistently eventive, while group (iv)-(vi) are originally eventive but may be shifted to non-eventive by an operation which is exerted by the implication operator IMP. Group (vii) is seemingly stative but arguably involves backgrounded events in its LCS. This leaves us with group (viii), which I propose to identify as the "adjectival" verb subtype in Thai. This type of verb intrinsically lacks eventivity, and thus holds an Sargument. The categorial distinction of verbs in Thai presented in this paper can account for their syntactic distributions and semantic differences.

Acknowledgements

I would like to express my gratitude to an anonymous reviewer and Prof. Kingkarn Thepkanjana for very helpful comments. I am also grateful to Prof. Taro Kageyama for his valuable suggestions and comments in earlier versions of this paper. An earlier version of this paper was presented at the 18th meeting of the South East Asian Linguistics Society at Universiti Kebangsaan Malaysia, Bangi in 2008. I would like to thank the audience. Prof. Amara Prasithrathsint also gave me valuable comments on the topic of the earlier version. I also owe a special thanks to Kazuya Kudo for his feedback. For providing data and advising me on its interpretation, I would like to thank the graduate school students at the Department of Linguistics, Faculty of Arts, Chulalongkorn University. I am also indebted to Paul Carpenter for stylistic improvement. The responsibility for any mistakes and deficiencies is purely mine.

References

- Carlson, Gregory N. 1977. *Reference to kinds in English*. Ph.D. dissertation. University of Massachusetts, Amherst.
- Croft, William. 1990/2003. *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William. 1991. Syntactic categories and grammatical relations: the cognitive organization of information. Chicago: University of Chicago Press.
- Croft, William. 2001. Radical Construction Grammar: syntactic theory in typological perspective. Oxford: Oxford University Press.
- Dixon, R.M.W. 2004. Adjective classes in typological perspective. *Adjective classes: A cross-linguistic typology*, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald, 1-49. Oxford: Oxford University Press.
- Enfield, Nick J. 2004. Adjectives in Lao. *Adjective classes: A cross-linguistic typology*, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald, 323-347. Oxford: Oxford University Press.
- Enfield, Nick J. 2007. A grammar of Lao. Berlin and New York: Mouton de Gruyter.
- Iwasaki, Shoichi and Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge: Cambridge University Press.
- Jackendoff, Ray. 1990. Semantic structures. Cambridge, MA: MIT Press.
- Kageyama, Taro. 1996. Doosi imiron [Verb semantics]. Tokyo: Kurosio.
- Kageyama, Taro. 2006. Property description as a voice phenomenon. *Voice and grammatical relations*, eds. by Tasaku Tsunoda and Taro Kageyama, 85-114. Amsterdam: John Benjamins.
- Kageyama, Taro. 2008. Goi-gainen-koozoo (LCS) nyuumon [Introduction to the lexical conceptual structure (LCS)]. *Lexicon forum No. 4*, ed. by Taro Kageyama, 239-265. Tokyo: Hituji Shobo
- Kuno, Susumu and Preya Wongkomthong. 1981. Relative clauses in Thai. *Studies in Language* 5: 195-226.
- Levin, Beth and Malka Rappaport Hovav. 1995. *Unaccusativity: at the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.

104 Natsuki Matsui

Levin, Beth and Malka Rappaport Hovav. To appear. Lexical conceptual structure. Semantics: an international handbook of natural language meaning. eds. by Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Berlin: Mouton de Gruyter.

- Lin, Jimmy. 2004. Event structure and the encoding of arguments: the syntax of the Mandarin and English verb phrase. Doctoral dissertation, Massachusetts Institute of Technology.
- Matsui, Natsuki. 2007. The lexical categories of state denoting predicates in Thai. *Jimbun ronkyu [Humanities Review]* 57/2: 73-94. Kwansei Gakuin University.
- Parsons, Terence. 1990. Events in the semantics of English: A study in subatomic semantics. Cambridge, MA: MIT Press.
- Post, Mark. 2008. Adjectives in Thai: implications for a functionalist typology of word classes. *Linguistic Typology* 12: 339-381.
- Prasithrathsint, Amara. 2000. Adjectives as verbs in Thai. Linguistic Typology 4: 251-271.
- Pustejovsky, James. 1995. Generative lexicon. Cambridge, MA: MIT Press.
- Rothstein, Susan. 2004. Structuring events. Oxford: Blackwell.
- Schachter, Paul. 1985. Parts of speech systems. *Language typology and syntactic description, Vol 1, Clause structure*, ed. by Timothy Shopen, 3-61. Cambridge: Cambridge University Press.
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
- Smith, Carlota S. 1997. *The parameter of aspect*. Dordrecht: Kluwer Academic Publishers.
- Sookgasem, Praya. 1996. The predicative-adjective construction in Thai. Pan-Asiatic linguistics: Proceedings of the fourth international symposium on languages and linguistics, Bangkok, July 8-10, 1996: 579-603.
- Tanaka, Hiroshi. 2004. Toogokoozoo o tyuusin tosita nihongo to Taigo no taisyookenkyuu [Comparative study of Japanese and Thai, focusing on syntactic structure]. Tokyo: Kurosio.
- Tansiri, Kachen. 2007. Interactions between viewpoint aspect and situation aspect: A case study of alternating intransitive constructions in Thai. *Manusya: Journal of Humanities* 10/2:54-81. Chulalongkorn University.
- Tenny, Carol. 1994. *Aspectual roles and the syntax-semantics interface*. Dordrecht: Kluwer Academic Publishers.
- Thepkanjana, Kingkarn. 2000. Lexical causatives in Thai. In *Constructions in cognitive linguistics*, ed. by Ad Foolen and Frederike van der Leek, 259-281. Amsterdam: John Benjamins.
- Uehara, Satoshi. 1998. Syntactic categories in Japanese: a cognitive and typological introduction. Tokyo: Kurosio.
- Vendler, Zeno. 1957. Verbs and Times. The Philosophical Review 66: 143-160.
- Vendler, Zeno. 1976. *Linguistics in philosophy*. Ithaca: Cornell University Press.

VIETNAMESE PASSIVE SENTENCES FROM A TYPOLOGICAL PERSPECTIVE

Nguyễn Hồng Cổn Vietnam National University, Hanoi <nghcon@yahoo.com>

0 ABSTRACT

There exist different views on passive sentences in Vietnamese. Some researchers claim that the Vietnamese language does not have a passive voice, other researchers argue that though Vietnamese may not have passive voice as a morphological category, it still has passive sentences as syntactic constructions. Yet, there is no clear consensus among these researchers as far as identification criteria for this kind of constructions is concerned.

Aiming at a more relevant solution to this issue, the present paper will critically review the different approaches to Vietnamese passive sentences and discuss their corresponding syntactic structures from a typological perspective. The paper has three parts: the first presents a review of two different approaches to passive sentences in Vietnamese; the second discusses Vietnamese passive sentences from a typological perspective; the third differentiates passive sentences from other types of sentences in Vietnamese.

1 Two different approaches to passive sentences in Vietnamese

The issue of passive voice/passive sentences in Vietnamese has always been most controversial among Vietnamese linguists. The different approaches to Vietnamese passive sentences can be put into two groups - the morphological approach and the syntactic approach. The former denies and the latter acknowledges the existence of passive sentences in Vietnamese.

1.1 The morphological approach

Some researchers (Trần Trọng Kim 1936, M.B. Emeneau 1951, L. Cadière 1958, etc.) claim that Vietnamese is an isolating language, and its verbs do not have passive voice. Therefore it does not have passive sentences as do inflecting languages (such as Russian, French, etc.), in which the verb has active and passive forms. The verbs in Vietnamese do not change their forms so they do not satisfy this strict morphological criterion of passive voice as a grammatical category. Not intending to contrast active voice and passive voice in Vietnamese, L.C. Thompson (1965: 217) also considers that the constructions with *được*, *bi* are just the translation equivalents of passive constructions in Indo-European languages. He calls these *logical passive expressions* and does not consider them as real passive constructions.

Besides the absence of passive voice as a morphological category, some researchers (e.g. Nguyễn Thị Ảnh 2000, Cao Xuân Hạo 2001) rely on the fact that Vietnamese is a topic-prominent rather than subject-prominent language to deny the existence of passive sentences. They argue that in topic-prominent languages there must not be passive

constructions because passive constructions are typical of subject-prominent languages that have passive voice. This argument can be traced back to Li & Thompson's (1976) typological classification that distinguishes between "topic-prominent languages" and "subject-prominent languages". These authors claim that passive constructions are very common in subject-prominent languages but are usually absent or rarely present in topic-prominent languages. If they do occur in topic-prominent languages, they usually carry a special meaning, like the adversity passive in Japanese.

Those who support the view that there are no passive sentences in Vietnamese also regard *được*, *bi* as transitive verbs so they can not be considered as passive markers. Nguyễn Kim Thản (1977) considers that được, bi are main verbs because in terms of meaning, được means receiving or undergoing something pleasant (for example: được ăn - be able to eat), and by means suffering from something unhappy (for example: by $\hat{o}m$ suffer illness). In terms of grammatical characteristics, bi and được have an object as transitive verbs (được bạc – win the game, bị đòn – got a beat). Based on this argument, Nguyễn Kim Thản concludes that được, bi still keep their full meanings and the grammatical characteristics of main verbs, and does not acknowledge them as passive markers (p.185-191). This view is supported and further elaborated by Nguyễn Minh Thuyết (1986, 1998). Despite the fact that these two scholars do not recognise the existence of passive voice in Vietnamese (as in Indo-European languages), they both consider that Vietnamese has its own ways of expressing passive meanings through syntactic structures (Nguyễn Kim Thản) or lexical means (Nguyễn Minh Thuyết). According to Nguyễn Thị Ảnh (2000), the passive voice as a grammatical category which is expressed by absolutely strict morphological markers is found only in subject-prominent languages. Vietnamese is a topic-prominent language so its passive voice does not have such particular marking devices. To prove that Vietnamese does not have passive voice as a grammatical category, the author has provided many examples showing that được, bị are main verbs and are not passive markers". This view is supported by Cao Xuân Hạo (2002).

1.2 The syntactic approach

In contrast to the first view, some other researchers consider that even though Vietnamese does not have passive voice as a morphological category, it does have passive sentences as syntactic constructions. Nguyễn Phú Phong (1976) acknowledges "the passiveness" as a grammatical category in Vietnamese. He argues that it is possible to identify an alternation of active-passive sentences in Vietnamese which correspond to the translated active-passive sentences in French and points to the formal relations among constituents of each type of sentences in common terms. He also considers dwoc, bi, and do to be passive auxiliaries. Hoàng Trọng Phiến (1980) states that "in Vietnamese the opposition between passive and active voices is not done through purely grammatical ways but through lexico-grammatical ways" (p.167). According to this author, the syntactic structure of a Vietnamese passive sentence is as follows:

- The subject of the passive is the object of the alternative active.
- The predicate of the passive includes an auxiliary dwoc, bi, do + a transitive verb.
- The agentive subject is optional in the passive.

(Hoàng Trọng Phiến, 1980: 166-67)

Vietnamese Passive 107

Lê Xuân Thại (1989) also has a similar view when he claims that even though Vietnamese does not have passive sentences completely similar to those in inflecting languages, it does have sentences which could be characterized passive with the following characteristics:

- The subject denotes patient rather than the agent.
- The predicate is appended by được or bị.
- The predicate may be followed by a clause. For example:
- (1) Em học sinh này được cô giáo khen adr. student this get teacher appraise *This student is appraised by the teacher*.
- (2) Thành phố Vinh bị máy bay giặc tàn phá city Vinh sufferairplane enemy destroy Vinh city is destroyed by enemy planes.

He also acknowledges that duoc, bi could be absent from passive sentences, for example:

- (3) Bữa com dọn ra meal set out The meal (is) set out.
- (4) Ngôi nhà này xây bằng gạch house this build by bricks *This house (is) built of bricks.*

Diệp Quang Ban & Nguyễn Thị Thuận (2000) also support the existence of passive sentences in Vietnamese. They argue that the passive voice in Vietnamese is not marked in the form of verbs but in the form of a special syntactic construction with established grammatical and semantic characteristics. Accordingly, they specify the following characteristics of Vietnamese passive constructions:

- The appropriate grammatical means for expressing passiveness in Vietnamese are *function words (được, bi)* and *word order*.
- Verbs participating in passive constructions are transitive verbs which have semantic relations with entities expressed by noun phrases as subjects before *được*, *bi*.
- Semantically, passive sentences have the following structure: i) The subject of the passive is assigned to the semantic roles of patient, recipient, goal, beneficiary. ii) The types of states of affairs of passive sentences are actions with two semantic characteristics: [+dynamic] and [+control].
- Syntactically, passive sentences have two clauses in their construction: C-V [C-V].

In brief, according to these two authors, the passive voice in Vietnamese is not expressed by the form of the verb, but instead by a syntactic construction with specific grammatical and semantic characteristics.

2 Vietnamese passive sentences from a typological perspective

2.1 Are there passive sentences in Vietnamese?

The review presented so far has indicated that in order to investigate the existence of passive constructions/sentences in Vietnamese it is necessary to clarify 3 points of controversy:

- Vietnamese does not have passive voice as a morphological category; therefore there are no passive constructions/sentences.
- Vietnamese is a topic-prominent language, therefore it lacks a "passive construction" or passiveness is a marginal syntactic phenomenon.
- Được/bị are not function words (verbal auxiliaries), but are modal verbs or lexical verbs, therefore they can not be characterised as passive markers.

With regards to the first point, we agree with the view that the "passive voice" as a morphological category should not be identified with those passive construction. This view has been discussed by many authors from different angles (Nguyễn Kim Thản, 1977; Hoàng Trọng Phiến, 1980; Lê Xuân Thại, 1989; Diệp Quang Ban & Nguyễn Thị Thuận, 2000). Evidences from passive sentences in other languages also show that the morphological forms of verbs are just one of the morpho-syntactic devices used to mark the passive voice (M. Keenan 1985, Shibatani 1994). If we take the strict morphological criteria for "passiveness", even in such languages like English or French the passive voice will not meet this strict requirement, because the passive voice in these languages is not marked only by morphological form of the verb, but also by an auxiliary (be in English, être in French) and word order. So it is possible to conclude that Vietnamese does not express the "passiveness" by morphological markers and therefore does not have passive voice as a morphological category. However, this does not mean it does not have syntactically passive constructions/sentences. We will return to this issue later when we discuss the syntactic characteristics of Vietnamese passive sentences.

With regards to the second point which considers that Vietnamese is a "topic-prominent" language (and not a "subject-prominent" language) and, therefore, does not have passive constructions, we think there is a need to discuss in more details. First of all, we should restate here Li & Thompson's view that the *topic* (what the sentence is about) and the *subject* (denoting the agent of the action expressed by the predicate) do not exclude each other. For instance, it is not the case that in topic-prominent languages there are no subjects, or in subject-prominent languages there are no topics. Even in a typical "topic-prominent" language like Chinese or Vietnamese, these categories do not exclude each other in most sentences. A syntactic description of Chinese sentences offered by Li & Thompson in a subsequent work (1981) also highlights that there are more sentences with subjects (topical or non-topical) than sentences with only topics (without subjects), and sentences with topical subjects prevail. If we apply both the topic and the subject categories in Li & Thompson's understanding to analyze Vietnamese sentences, we see that Vietnamese has a large number of sentences where subjects are identical to topics

Vietnamese Passive 109

(especially in the cases where the predicate is a transitive verb). If the subjects are so prevalent, and the majority of transitive constructions also have subjects (which may be topical or non-topical), there is no reason why the "passiveness" should be considered absent or marginal. Moreover, it should be noted that Ch. N. Li & S. A. Thompson do not absolutely exclude the passive voice from topic-prominent languages, they just do not consider it a typical passive voice, i.e. it is not "purely morphological" passive voice as it is in Indo-European languages.

Because of the typological characteristics of the Vietnamese language as an isolating language, its grammatical categories in general and "passiveness" in particular do not have morphological markings. Dyvik (1984) came to the conclusion that if "subject" is acknowledged as a part of a sentence in Vietnamese, it is not as clear as the subject in Indo-European languages, because grammatical properties of the subject in Vietnamese are more abstract. Just as the subject, "passiveness" could only be identified by more "abstract" criteria. In other words, both "subject" and "passiveness" occur in Vietnamese although not as clearly distinguished as in Indo-European languages (p.7-12).

With regards to the third point which is related to the syntactic functions and meanings of dvoc, bi, we have not considered the fact that these words are grammatically important and to some extent have lexical meanings to exclude their function as passive markers if we look at this issue from the viewpoint of grammaticalization.

- Grammatically, Nguyễn Kim Thản (1977), Nguyễn Minh Thuyết (1976), Nguyễn Thi Ånh (2000) and Cao Xuân Hao (2001) all consider that được, bi are not function words used to mark "passiveness" but modal verbs, or even lexical verbs occupying the central role in predicates. Dyvik (1984), in contrast, tries to prove that *được*, *bi* are gradually losing their roles as main verbs, becoming auxiliaries marking "passiveness". Mainly agreeing with Dyvik, we suggest that even if *được*, *bị* play the central grammatical role in predicates as stated by some researchers, that does not mean these words can not function as passive markers. This is similar to the passive auxiliaries of passive sentences in English (be), in French (être), or in Russian (byt'). An auxiliary like be has almost no semantic role in creating the lexical meaning of a passive state which results from the form of the transitive verb (the past participle), but plays the central grammatical role in the predicate of passive sentences. The evidence is that this role is fulfilled by the auxiliary be, and not the past participle, is the morphological agreement in person and number with the subject of a passive sentence. Thus, grammatically the auxiliary verb be is not different from a main verb in the predicate of active sentences. Yet this does not impede it from being a passive marker.

- Semantically, được and bị indeed still carry the meaning of "enjoy" or "suffer". However, even this semantic feature does not prohibit them from being passive markers if we consider được, bị to be undergoing the process of grammaticalization. Nguyễn Tài Cần (1978) considers that "bị" has shifted from a morpheme to a word, and from a lexical word to grammatical one. Đinh Văn Đức (1986: 118-19) offers more detailed explanations about the grammaticalization of được, bị and the relationships between their grammatical meaning of passiveness and their modal meanings:

"There is a group of Vietnamese verbs such as cần (need), muốn (want), có thể (can), toan (intend), định (intend), dám (dare), bị (suffer), được (get, enjoy), etc., which clearly have no meanings at all. The lexical meanings of these verbs are very insignificant, they have been grammaticalized but they have not yet become true function words. These verbs have very narrow intentions so their extensions should be broad - they are always accompanied by secondary constituents. While expressing the meanings of need ($c\hat{a}n$ - need), possibility ($c\acute{o}$ $th\acute{e}$ - can), intention (toan - inted to, định - intend, dám - dare), desire (mong -wish, muốn-want), passiveness (được-get/enjoy; bi - suffer), etc., these verbs are used according to speakers's attitudes towards and assessments of the realities. These relations reflect the subjective consciousness: When we say "Tôi được khen" (I was appraised) or "Tôi bị phạt" (I am punished), the words được, bị are the grammatical markers of passiveness, but the passive meaning here could be understood depending on the nuances of "good luck" or "bad luck", and the meanings of "good luck" or "bad luck" are completely dependent on the understanding and assessments of the speakers. Consequently được, bi have temporarily become modal words..." (Đinh Văn Đức 1986:139-140).

Agreeing with this explanation, we think the fact that $du\phi c$, bi keep their original lexical meanings while acting as function words (passive markers) is quite normal in grammaticalization.

This analysis has highlighted that it is not surprising if \$\partial uvertice construction of a lexical verb (as a main verb in the predicates) and still keep their original lexical meanings (\$bi\$ means suffering from something unhappy and \$\partial uvertice constructions while playing the roles of an auxiliary marking passiveness. According to Keenan (1985: 257-61), in languages with periphrastic passive, there are at least 4 types of verbs used to mark passive predicates: (i) intensive/relational verbs (like \$be\$ in English, \$byt\$' in Russian, \$\hat{e}tre\$ in French, etc.), (ii) giving-receiving verbs (like the passive constructions with \$get\$ in English), (iii) motion verbs (like \$gayee\$ in Hindi), and (iv) enjoying-suffering verbs (like \$\partial uvertice constructions with \$\p

With the arguments presented above we have come to the following conclusion: Theoretically and practically we have enough evidence to talk about the presence of passive sentences in Vietnamese.

2.2 Typology of Vietnamese passive sentences

As presented above, there are different views on passive sentences in Vietnamese. Even among those who acknowledge passive sentences, there is no consensus as far as their identifications are concerned. We consider that, just like in other languages, Vietnamese passive sentences are syntactically transformed from the alternative active sentences, despite the fact that not all active sentences could be transformed into passive ones. Of course, the syntactic transformations should satisfy certain semantic and pragmatic constraints of passive sentences.

Formally, a prototypical passive sentence in Vietnamese could be identified and differentiated from an active sentence via the following syntactic criteria:

Vietnamese Passive 111

- The subject of the passive is derived from the object of the alternative active. Depending on each sentence, the subject of the passive could be a patient, a recipient, a goal, or an instrument (see also Diệp Quang Ban & Nguyễn Thị Thuân, 2000).

- The predicate of the passive is derived directly from the predicate of the alternative active by appending an auxiliary dwoc or bi before the transitive verb.
- The oblique of the passive is derived from the subject of the alternative active. The oblique is usually optional (in passive sentences without agentive oblique NP). If they are not omitted, they could be replaced before the predicate (in passive sentences with an agentive oblique NP) or after the predicate provided that a preposition *bởi* is added (in passive sentences with an agentive oblique PP).

The transformation of Vietnamese active sentences into alternative passive sentences could be summarized as follows:

```
NP₁ V NP₂ (active sentence)

NP2 được/bị V (non-agentive passive sentence)

NP₂ được/bị NP₁ V (passive sentence with agentive oblique NP)

NP₂ được V bởi NP₁ (passive sentence with agentive PP)
```

(NP₁ – Agentive NP, NP₂ – Patient NP, V- Predicative Verb, Aux – Auxiliary Verbs, Pr - Preposition)

Following are some illustrating examples:

NP₂ được/bởi V

- (5) Nó được khen he get appraise NP_2 Aux V He is appraised.
- (6) Tôi bị mắng
 I suffer scold
 NP₂ Aux V
 I am scolded.

$\underline{NP_2}$ được/bị $\underline{NP_1}$ \underline{V}

- (7) Nó được thầy khen.

 he get teacher appraise

 NP₂ Aux NP₁ V

 He is appraised by his teacher.

NP_2 được/bị V bởi NP_1

- (9) lành nghề Ngôi nhà này được xây dựng bởi những tay thợ House workers skilled this build by get V Pr NP_1 NP_2 Aux This house is built by skilled workers.
- Tiếng con chim sơn ca tiếng còi tàu (10)bi át đi bởi rúc singing lark wristle train suffer drown by hoot NP_2 Aux V Pr NP_1 The lark's singing is drowned by the train's wristle.

Thus, with regards to the structure of passive sentences in Vietnamese, the main means used to mark passiveness are word order plus *được*, *bị* acting as function words (auxiliaries). If we compare Vietnamese passive sentences to passive sentences from languages belonging to different typologies, from highly synthetic ones like Russian to less synthetic ones like English, it becomes clear the fact that Vietnamese, a typical analytic language which uses purely syntactic devices (i.e., word order and function words) to express passiveness, conforms to general laws about the typological differences among languages. These differences are represented in Table 1:

Table 1: *Typological comparison of Passive Constructions*

| Passive | Passive marking means | Language | Examples |
|-----------|---------------------------|------------|--------------------------------------|
| Types | | | _ |
| Synthetic | | | Rabotchie stroili dom |
| | form of verbs | | Workers built the house. |
| | (purely morphological) | | |
| | | Russian | Dom stroilsa rabochimi |
| | | | The house was built by workers. |
| | | | |
| | | | Rabotchie postroili dom |
| Analytic | form of verbs + auxiliary | | Workers have built the house. |
| | + word order | | |
| | (morpho-syntactic) | | Dom byl postroen rabotchimi |
| | | | The house has been built by workers. |
| | | English | Workers built the house |
| | | | The house was built by workers |
| | auxiliary + word order | Vietnamese | Công nhân đã xây xong ngôi nhà |
| | (purely syntactic) | | Workers have built the house. |
| | | | |
| | | | Ngôi nhà đã được công nhân xây xong |
| | | | The house has been built by workers. |

Vietnamese Passive 113

3. Differentiation of passive sentences from other types of sentences

It is necessary to distinguish types of typical passive sentences in Vietnamese (identified according to the above-mentioned criteria) from other types of sentences which are similar in form or meaning but which are not the passive per se.

3.1 Passive sentences NP_2 được/bị V vs. pseudo-passive sentences NP_1 được/bị V

The type of active sentence NP_1 $du\phi c/bi$ V has the subject NP_1 denoting the experiencer or actor of the state of affairs expressed by the predicate V. V could be an intransitive verb $(T\hat{o}i\ bi\ ng\tilde{a}$ - I fell, $N\acute{o}\ du\phi c\ nghi$ - He has a day off) or a transitive verb $(T\hat{o}i\ bi\ nghe\ l\grave{o}i$ $ph\grave{a}n\ n\grave{a}n$ - I have to listen to the complaints; $N\acute{o}\ du\phi c\ xem\ phim$ - He got to watch the movie). $NP_2\ du\phi c/bi\ V$ is a type of passive sentence that has the subject $NP_2\ denoting\ a$ patient (or the subject $NP_3\ denoting\ a$ recipient) and V as a transitive verb $(T\hat{o}i\ bi\ m\acute{a}ng$ - I was scolded; $N\acute{o}\ du\phi c\ khen$ - He was appraised). In this second type of sentence, the agentive oblique $NP_1\ can\ optionally\ appear\ before\ V$. It should be noted that when $NP_1\ is\ absent$, the passive sentences $NP_{2/3}\ du\phi c/bi\ V\ (11,13)$ will have surface structures similar to those of the active sentence type $NP_1\ du\phi c/bi\ V\ (12,14)$:

- (11) Tôi bị mắng I suffer scold NP₂ Aux V I am scolded.
- (13) Tôi được tặng giấy khen
 I get award Certificate of Merit
 NP₃ Aux V NP₂
 I was awarded with a Certificate of Merit.

Some researchers rely on these characteristics to consider sentences of type NP_2 duoc/bi V (like examples 11, 13) as active sentences just as sentences of type NP_1 duoc/bi V (like examples 12, 14), and not as passive sentences. To our understanding, these two sentence types look similar in their surface structures but different in their deep structures: sentences like 12 and 14 have experiencer subjects (i.e., their subjects coincide with the experiencers of process); sentences like 11 and 13 have patient subjects (i.e., their subjects coincide with the patients of action). These two types of constructions can be distinguished from each other by a transformational test - adding an agentive NP_1 before V:

- mắng mắng (11)Tôi bi (11')Tôi bį mę Ι suffer scold Ι suffer mother scold NP₂ Aux NP_2 V Aux NP_1 V I am scolded. I am scolded by my mother.
- Tôi *(12') Tôi (12)bi bi me ngã ngã suffer fall mother fall Ι I suffer NP_1 Aux V NP_1 Aux NP_1 V * I suffer my mother fell. I fell.
- (13)Tôi được tặng giấy khen > (13')Tôi được **trường** tăng giấy khen award Certificate of Merit I get award Certificate of Merit get school NP₂ Aux NP₃ Aux NP₁ V NP_3 NP_2 I was awarded with a Certificate I was awarded with a Certificate of Merit by the school. of Merit.
- Tôi được **trường** xem phim Tôi được xem * (14') (14)phim get watch movie get school watch movie NP₁ Aux V NP₁ Aux NP_1 V NP_2 NP_2 *I got to watch the movie.* * I got the school to watch the movie.

The test shows that sentences 12 and 14 do not accept an agentive NP₁ (me, nhà trurng) before V whereas the addition of an agentive NP₁ before V in sentences 11 and 13 does not change the sentence meaning, but provides clarifying information. From this test, we can conclude that NP_2 dunge/bi V (11, 13) and NP_1 dunge/bi V (12, 14) are two different sentence types: The former represents a typological type of passive sentences (they are passive both in terms of grammar and semantics), while the latter type represents a type of active sentence with the pseudo-passive form (they are passive in terms of grammar but active in terms of semantics), i.e pseudo-passive sentences.

3.2 Passive sentences NP₂ được/bị V vs. active sentences NP₂ do NP₁ V

Some researchers consider a sentence like *Hàng này do xí nghiệp chúng tôi sản xuất (This merchandise is produced by our factory)* a passive one (Nguyễn Phú Phong 1976, Nguyễn Kim Thản, 1977). According to us, this sentence is not a passive for two reasons:

- First, do in this construction does not have the same function as $du\phi c$, bi in passive sentences. The evidence is that do can not independently combine with a transitive verb to form a passive predicate. For example, we cannot say:
- (15) *Hàng này do sản xuất
 Merchandise this by produce
 NP₂ Pr V
 *This merchandise by produced.

In other words, do is always used together with the presence of an agentive subject before a predicative verb, so it's not an auxiliary marking passiveness like duqc, bi.

- Secondly, the transformation of the object in an active sentence to the subject in the sentence type NP_2 do NP_1 V is very limited in scope. Only the patient object NP_2 may

Vietnamese Passive 115

appear in the subject position. Other types of objects (for example, recipient object NP_3) do not have this possibility. For example, we can not perform the following transformation:

(16)Tôi viết thư cho Nam >(16') *Nam do tôi viết thư cho write letter to Nam by I write letter to Nam NP₂ Pr NP₃ Pr NP_1 V NP_2 NP_1 V NP_3 Pr I wrote a letter to Nam. * Nam by I write a letter to.

3.3 Passive sentences NP₂ được/bị V vs. de-transitive sentences NP₂ - V

Sentences with the pattern NP_2 - V have non-agentive NP at the beginning of the sentence (NP is a recipient, pacient, etc.) followed by a predicate (V) which has the original meaning of a transitive verb, usually accompanied by an auxiliary (before V) or an adverb (after V). For example:

- (17) Cửa mở rồi door open already NP₂ V Adv *The door opened*.
- (18) Cầu đang xây
 bridge being build
 NP₂ Aux V
 The bridge is being built.
- (19) Nhà cửa cuốn sạch rồi houses carry away already NP₂ V Adv All the houses have been carried away.

According to some researchers, these NP_2 - V sentences could be considered passive (Nguyễn Kim Thản 1977, Lê Xuân Thại 1994). However, if we take into careful consideration all NP_2 - V sentences as the examples above, it becomes difficult to say whether they are passive or active. At first, they may seem passive because without dwoc, bi they are still understood as having a passive meaning. If we add dwoc, bi after NP_2 , all they become passive sentences:

- rồi (17)Cửa mở rôi (17')Cửa được mở door open already already door get open Adv Adv NP_2 V NP_2 Aux V The door opened. The door was opened.
- (18)Cầu (18')Cầu đang xây đang được xây bridge being build bridge being build get NP_2 NP_2 V Aux V Aux Aux The bridge is being built. The bridge is being built.

sach rồi cuốn sach rồi . > (19') cuốn (19)Nhà cửa Nhà cửa bi houses carry away already houses suffer carry away already NP_2 NP_2 Adv Adv Aux V All the houses were carried away. All the houses were carried away.

But if we add an agentive NP₁ before V, all the sentences lose their passive meanings, or in other words, they have active meanings and become active sentences. For example:

- rồi. rồi (17)Cửa mở (17")Cửa mở me door already door mother open already open NP_2 V Adv NP_2 NP_1 Adv The door is open. The door, mother has opened it.
- Cầu (18") Cầu (18)đang tho xây đang xây bridge being build bridge workers being bulid NP_2 Aux V NP_2 NP_1 Aux V The bridge is being built. The bridge, workers are building it.
- sach rồi > sach rồi (19)Nhà cửa cuốn (19")Nhà cửa lũ cuốn houses carry away already houses flood away already carry NP_2 Adv NP_2 NP_1 V Adv All the houses, the flood carried them away. All the houses were carried away.

In the newly formed sentences, the function of NPs $c\dot{u}a$, $c\dot{a}u$, $nh\dot{a}$ $c\dot{u}a$ could be switched from the subject of sentence to the topic of sentence. These transformational possibilities suggest that the sentences of type NP_2 - V could be a middle type between passive and active sentences, instead of typical passive sentences. We call this a type of *detransitive* sentence (see Nguyễn Hồng Cổn, 2004).

4 Conclusion

The present paper discusses the issue of passive sentences in Vietnamese from a perspective of syntactic typology. Based on a distinction of 3 types of passive sentences which are either purely morphologically, morpho-syntactically, or purely syntactically passive, and based on a variety of evidence, the paper has demonstrated that: 1. Although it has no passive voice as a purely morphological category, Vietnamese still has passive sentences as syntactic constructions, marked by word orders and function words (dwoc, bi); 2. There are 3 types of passive sentences in Vietnamese ($NP_2 dwoc/bi VI, NP_2 dwoc/bi NP_I$), and all could be identified and differentiated from other sentence types by certain syntactic criteria.

References

Cadière, L. 1958. *Syntaxe de la langue vietnamiene*. Paris: Ecole Française d'Extrême-Orient. Publications de l'EFEO, 42.

Emeneau, M.B. 1951. *Studies in Vietnamese (Anamese) grammar*. University of California publications in linguistics (Vol. 8). Berkeley: University of California Press.

Vietnamese Passive 117

Cao Xuân Hạo 1991. *Tiếng Việt: Sơ thảo ngữ pháp chức năng* (Vietnamese Language: Basics of Vietnamese functional grammar, Vol. I. Ho Chi Minh City: Social Sciences Publisher.

- Cao Xuân Hạo 2001. "Hai phép cộng và trừ trong ngôn ngữ" (Addition and subtraction in linguistics), *Tạp chí Ngôn ngữ* (Language Review), No. 10: 1-12. Hanoi: Institute of Linguistics.
- Diệp Quang Ban 1992. *Ngữ pháp tiếng Việt* (Vietnamese grammar), Vol. II. Hanoi: Education Publisher.
- Diệp Quang Ban Nguyễn Thị Thuận 2000. "Lại bàn về vấn đề câu bị động trong tiếng Việt" (On passive sentences in Vietnamese), *Tạp chí Ngôn ngữ* (Language Review), No. 7: 14-21. Hanoi: Institute of Linguistics.
- Dyvik, H.J.J. 1984. Subject or topic in Vietnamese? Norway: University of Bergen.
- Đinh Văn Đức 1986. *Ngữ pháp tiếng Việt: Từ loại* (Vietnamese grammar: Parts of speech). Hanoi: Education Publisher.
- Givón, T.1990. *Syntax: a functional -typological introduction*, Vol. II. Amsterdam/Philadelphia: John Benjamins publishing company.
- Hoàng Trọng Phiến 1980. *Ngữ pháp tiếng Việt: Câu* (Vietnamese grammar: Sentences). Hanoi: Education Publisher.
- Keenan E. L. 1985. "Passive in the world's languages". In *Language typology and language description*, T. Shopen (ed), Vol. I. Cambridge: Cambridge University Press,
- Lê Xuân Thại 1994. *Câu chủ vị tiếng Việt* (Subject-predicate sentences in Vietnamese). Hanoi: Social Science Publisher.
- Li Ch.N. & Thompson S.A.1976. Subject and Topic: a new typology of language, New York –San Francisco –London: Academic Press.
- Li Ch.N. & Thompson S.A. 1981. *Mandarin Chinese: A functional reference Grammar*. Berkeley Los Angeles London: University California Press,
- Nguyễn Hồng Cổn 2004. "Các kiểu cấu trúc phi ngoại động trong tiếng Việt" (Detransitive sentence types in Vietnamese), *Tạp chí Khoa học* (Journal of Sciences), No.2. Hanoi: Vietnam National University.
- Nguyễn Hồng Cổn, Bùi Thị Diên 2004. "Dạng bị động và vấn đề câu bị động trong tiếng Việt" (Passive voice and the issue of passive sentences in Vietnamese), *Tạp chí Ngôn ngữ* (Language Review), No. 7: 1-12, No.8: 8-18. Hanoi: Institute of Linguistics.
- Nguyễn Kim Thản 1964.: *Ngữ pháp tiếng Việt* (Vietnamese grammar), vol. II. Hanoi: Social Science Publisher.
- Nguyễn Kim Thản 1977. Động từ trong tiếng Việt (Verbs in Vietnamese). Hanoi: Social Science Publisher
- Nguyễn Minh Thuyết 1986. "Vai trò của "được", "bị" trong câu bị động tiếng Việt" (The roles of "được", "bị" in Vietnamese passive sentences). In *Những vấn đề các ngôn ngữ phương Đông* (Issues on the Oriental languages). Hanoi: Institute of Linguistics.

- Nguyễn Minh Thuyết Nguyễn Văn Hiệp 1998. *Lý thuyết thành phần câu và thành phần câu tiếng Việt* (Theory of sentence components and sentence components in Vietnamese). Hanoi: VNU Publisher.
- Nguyễn Thị Ảnh 2000. "Tiếng Việt có thái bị động không?" (Is there a passive voice in Vietnamese?), *Tạp chí Ngôn ngữ* (Language Review), No. 5: 36-47. Hanoi: Institute of Linguistics.
- Nguyễn Phú Phong 1976. Le Syntagme Verbal en Vietnamien. The Hague-Paris : Mouton.
- Palmer, F.R. 1994. *Grammatical roles and relations*. Cambridge: Cambridge University Press, 1994.
- Shibatani, M. 1994. *Voice*. In The Encyclopedia of Language and Linguistics/R.E. Asher (ed), Vol. 4. Pergamon Press Ltd.
- Thompson L.C. 1965. *A Vietnamese Grammar*, Seattle & London: University of Washington Press.
- Trần Trọng Kim, Bùi Kỷ, Phạm Duy Khiêm 1936. *Việt Nam văn phạm* (Vietnam Grammar), Hà Nội: Hội Khai trí Tiến đức.

PROTO-SOUTHWESTERN-TAI REVISED: A NEW RECONSTRUCTION*

Pittayawat Pittayaporn

Cornell University

<pp224@cornell.edu>

0 Abstract

Southwestern Tai (SWT) is the best studied among the three main branches of the Tai language family according to Li (1960)'s classification. Because discrepancies among earlier proposals (Brown 1985; Jonsson 1991; Li 1977; and Sarawit 1973) are major obstacles both in subgrouping SWT varieties and in reconstructing PT, this paper makes use of data from 28 varieties incorporating new data from lesser-known languages including the author's own fieldwork, to propose a revised PSWT reconstruction. It first reviews the earlier proposals and then puts forward a new reconstruction of PSWT which makes the following claims: 1) there was a distinct series of uvular consonants in PSWT, 2) there is no evidence for clusters * $p^h r/l$ -, * $k^h l$ -, and *mr- at the PSWT level, 3) mid back unrounded vowel * γ must be reconstructed for PSWT, and 4) vowel length was contrastive. Last but not least, it discusses important features of the proposed PSWT phonology that have implications for the reconstruction of Proto-Tai, including 1) redundancy of length contrast among non-high vowels, 2) gaps within the sub-system initial clusters, and 3) the defective nature of the vowel * γ .

1 Introduction

In the past few decades, the field of Comparative Tai has advanced considerably. Linguists working on Tai now have a good understanding of the phonology of the parent language of Tai languages. Various systems have been proposed (Li 1977; Sarawit 1973). Among the three generally-recognized subgroups of Tai, Southwestern Tai (SWT) is by far the best studied. There is a larger body of literature on SWT languages than on Central Tai (CT) and Northern Tai (NT). This literature includes both language description as well as linguistic analysis of specific problems. However, many questions regarding the history of SWT still await solutions.

In the light of newly available data, I propose in this paper a new reconstruction of Proto-Southwestern Tai (PSWT) phonology, which critically evaluates earlier work as well as take advantage of data that has become available from a wide-range of SWT varieties. Due to space limit, a detailed presentation of the full reconstruction cannot be provided here but will be published as a monograph volume in the near future. This paper aims at highlighting crucial aspects of the new reconstruction of PSWT phonology. I first present an overview of the reconstruction focusing on my central claims about the PSWT

^{*}I would like to thank Paul Sidwell, Laurent Sagart, John Whitman, two anonymous reviewers, and the audience at the 18th Annual Meeting of the Southeast Asian Linguistics Society for their insightful comments. I would also like to thank Mr. Pan Kaewkhamsaeng, my Kapong consultant, for his generosity and enthusiams.

phonology. I then proceed to discuss important features of the proposed sound system of PSWT that have implications for the reconstruction of Proto-Tai (PT).

2 Reconstructing PSWT phonology

Although the reconstruction put forward here agrees largely with those of Li (1977), Sarawit (1973) and Jonsson (1991), it differs in some very crucial aspects. In this section, I first briefly review earlier proposals on the reconstruction of PSWT, then provide some methodological notes, and, lastly, discuss the claims made about the PSWT phonology.

2.1 Review of earlier reconstructions

All earlier PSWT reconstructions agree that PSWT had five places of articulation among consonants: labial, alveolar, palatal, velar, and glottal. Except for the glottal series, these places show four contrastive phonation-types: aspirated voiceless, unaspirated voiceless, glottalized voiced, and plain voiced. Regarding the vowel system, all these proposals posit systems that contrast frontness/backness as well as lip rounding at three vowel heights. As for PSWT tones, they also agree that the three-way tonal contrast of PT was retained at the PSWT level and that tonal splits had not occurred. A comparison between the different systems can be found in Appendix A.

As is the case for PT, Li (1977) is also the standard reconstruction for PSWT. Although his main concern in this monumental work is the reconstruction of PT, he also provides reconstructions for the three subgroups he proposes. In this reconstruction, PSWT can be characterized as PT with a reduced phonemic inventory resulting from mergers of initials and vowels. One important development from PT to PSWT is the introduction of length contrast in high vowels *i/*i:, *uu/uu: and *u/*u: by regular sound changes. Sarawit (1973) is another important attempt at reconstructing the PT vowel system. She modifies the system of initials proposed in an earlier work by Li (1954) and reconstructs the vowel system of each of the three branches. Though different in minor details, her PSWT vocalism is very similar to that of Li (1977)

Two other available reconstructions of PSWT are those of Brown (1965) and Jonsson (1991). Brown's "Ancient Thai" corresponds roughly to PSWT in Li (1960)'s model but only includes varieties spoken in Thailand and adjacent areas. Varieties in Vietnam such as Black Tai and White Tai are not included. This reconstruction is different markedly from the first two with respect to vowel length. Specifically, in this system, contrastive vowel length is not limited to high vowels. The most recent reconstruction is that by Jonsson (1991), which is the only work that deals primarily with PSWT. In contrast to Brown (1965), the only length contrast in this system is that of *a and *a:. As for the consonant system, this reconstruction is almost identical to that of Li (1977). As a point of departure, the PSWT inventory of initials and vowels given in Li (1977) is summarized in Table 1.

The reconstruction put forward in this paper supports many of the basic positions outlined above, but it also makes some new claims about the sound system of PSWT. Specifically, I argue that 1) there was a distinct series of uvular consonants in PSWT, 2) there is no evidence for clusters *p^hr/l-, *k^hl-, ?br- and *mr- at the PSWT level, 3) a mid back unrounded vowel *y must be reconstructed, and 4) vowel length was contrastive in PSWT. As a point of departure, the PSWT inventory of initials and vowels proposed here is summarized in Table 1, Table 2, and Table 3.

 Table 1: PSWT simple consonants

| | labial | alveolar | palatal | velar | uvular | glottal |
|------------|-------------------|-------------------|-------------------|-------------------|--------|---------|
| | *p ^h - | *t ^h - | | $*k^h$ - | | |
| stons | *p- | *t- | *c- | *k- | *q- | |
| stops | * [?] b- | *?d- | | | | *?- |
| | *b- | *d- | * J - | *g- | | |
| c · · | *f- | *s- | | *x- | *χ- | *h- |
| fricatives | *v- | *z- | | *y- | | |
| 1 | *hm- | *hn- | * ^h ɲ- | * ^h ŋ- | | |
| nasals | *m- | *n- | *ɲ- | *ŋ- | | |
| | | *hr- | | | | |
| 1:: do | | *r- | | | | |
| liquids | | * ^h l- | | | | |
| | | *1- | | | | |
| glides | *hW- | | | | | |
| | | | * [?] j- | | | |
| | *w- | | *j- | | | |

 Table 2: PSWT initial clusters

| | labial | alveolar | palatal | velar | uvular | glottal |
|-----|--------|----------|---------|----------------------------|--------|---------|
| | *pl- | | | *kl- | | |
| C+l | *?bl- | | | | | |
| C+1 | *bl- | | | *gl- | | |
| | *ml- | | | | | |
| | | | | *k ^h r | | |
| C+r | | | | | | |
| | *br- | | | *gr- | | |
| | | | | | | |
| | | | | *k ^h w- *kw- | *qw- | |
| C+w | | | | | | |
| C⊤w | | | | *gw- | | |
| | | | | *γw- | | |
| | | | | *gw- *yw- *ŋw- | | |

Table 3: PSWT vowels

| | front | Ba | ick | | |
|------|-----------|-----------|---------|--|--|
| | unrounded | unrounded | rounded | | |
| High | *i, *i: | *w, *w: | *u, *u: | | |
| Mid | *e | *8 | *0 | | |
| Low | *e: | *a, *a: | *o: | | |

| Diphthongs | | | | | | | | |
|------------|------------------------|-----|--|--|--|--|--|--|
| Decreasing | Decreasing *iə *wə *uə | | | | | | | |
| Increasing | | *aw | | | | | | |

2.2 Data

Before proceeding to the discussion of the reconstruction, some methodological notes are in order. As summarized earlier, there are differences among the different systems of PSWT reconstruction. Because such discrepancies are major obstacles both in subgrouping SWT varieties and in reconstructing PT, this paper incorporates new data from lesser-known varieties and proposes a revised PSWT reconstruction. The sound system is reconstructed mainly on the basis of the data from an unpublished comparative Tai wordlist by Gedney (n.d.)¹ but additional data from various sources including wordlists, dictionaries, and my own fieldwork were also incorporated (see Appendix B).

Data from different sources show different notational conventions. Both modern and reconstructed forms from different publications were re-transcribed to conform to the current IPA standard (International Phonetic Association 1999). Notation of tones in modern dialects is the area where the largest variation in transcriptional practice exists. This paper adopts Li's (1977) convention, in which '1' is the assigned code for tones developed from original voiceless stops and '2' for those developed from voiced stops. When there is a need to allow for more fine-grained categorization of the proto-initials, additional notation will be introduced. A survey of different conventions can be found in Edmondson and Solnit (1997). Gedney (1989a) is an excellent reading on the etymological notation of Tai tones.

The data from the wordlist is entered into WordCorr 2.0, which helped organize data, and identify correspondences. Once the data was organized, and the correspondences were identified and tentatively assigned proto-phonemes, the data were converted into a Microsoft Access database for ease of searching and data management. In addition to data from modern SWT varieties (see Appendix B for list of dialects and their abbreviations) loanword, orthography as well as extra-SWT information have also been taken into considerations but only minimally, though they prove to be crucial in some problematic cases.

¹ This wordlist has recently been published as *William J. Gedney's Comparative Tai source book* (Hudak 2008).

2.3 Uncovering the uvular series: *q- and *x

Unlike other simple initials, dorsal consonants present some crucial problem to the reconstruction of the PSWT phonology. Li (1977:192-198, 207-214) reconstructs five phonemes in the velar region for PT: $*k^h$, *k-, *g-, *x-, and *y-. He discusses the fact that PT $*k^h$ and *x- had merged by the time of SWT but the resulting sound may have two different reflexes in White Tai (WT) and some dialects of Lue; k^h or x-. Jonsson (1991: 74), in contrast, proposes that WT x- is reflex of PSWT *x- that contrasted with $*k^h$. Data from Phuan (PH) and Kapong dialect of Phu Thai (KP), which have now become available, indicate that a distinct series of uvular consonants must be reconstructed. Although data unequivocally indicates PSWT *q- and $*\chi$ -, evidence for *- is lacking. The relevant correspondent sets are given in Table 4.

| | BT | WT | TL | TH | KP | PH |
|-------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| *k- | k- | k- | k- | k- | k- | k- |
| *k ^h - | k ^h - |
| *x- | k ^h - | Х- | k ^h - | k ^h - | k ^h - | k ^h - |
| *q- | k ^h - | X- | k ^h - | k ^h - | h- | k ^h - |
| *x- | k ^h - | Х- | k ^h - | k ^h - | h- | h- |

Table 4: Reflexes of PSWT velar and uvular consonants

Although these uvular consonants are not preserved in any modern varieties, the correspondence sets pointing to another series of dorsal initials are very robust. The first newly-reconstructed phoneme in PSWT is the voiceless unaspirated uvular stop *q-. It is reflected as k^h - in most varieties, but as x- in WT and h- in KP. It is then likely that this phonemes became affricated and then fronted to the velar region in all dialects, except in KP where it lenited and merged with PSWT *h-. Similarly, $*\chi$ - merged with $*k^h$ - in most SWT varieties including Black Tai (BT), Tai Long (TL), Thai (TH), and Lao (LA). However, this PSWT phoneme is realized as x- in WT, suggesting that the reflexes in other varieties might have also gone through an affricated stage. KP and PH agree in reflecting this phoneme as *h-. Some etyma reconstructed with uvular consonants are given in Table 5. The forms for 'needle' and 'leg' in KP and PH unexpectedly show k^h - and k- respectively. Note that in PH both the expected $k^h a \cdot k^h$ and the irregular $k a \cdot k^h$ are attested.

Note that many words that go back to PSWT $*k^h$ are words with 'voicing alternation,' i.e. etyma whose SWT and CT reflexes point to earlier voiceless onsets but whose NT reflexes suggest their voiced counterparts. Li (1977:192-198, 207-214) notes that whether the modern WT reflex is a stop or a fricative tends to correlate with the so-called 'voicing alternation'. Diller (1988) adopts this view and claims a strong correlation between WT k^h and forms without voicing alternation on one hand and between x-with etyma that show voicing alternation on the other. This question is however orthogonal to the question of PSWT uvular consonants.

| *q- | | BT | WT | TL | Thai | KP | PH | Saek |
|--------------------------|----------|--------------------|------|--------------------|--------------------|-------------------|--------------------|------|
| 1. to kill | C1 | k ^h a: | xa: | k ^h a: | k ^h a: | har | k ^h a: | ka: |
| 2. knee 3. torn | B1 D1 | $k^{h}aw$ | xaw | k ^h aw | k ^h aw | haw | k ^h aw | kə: |
| 4. needle | A1 | k ^h a:t | xa:t | k ^h a:t | k ^h a:t | ha:t | k ^h a:t | ka:t |
| 5. leg | A1 | k ^h em | xim | k ^h em | k ^h em | k ^h em | hem | kim |
| | | k ^h a: | xa: | k ^h a: | k ^h a: | k ^h a: | ha:~kʰa: | kwa: |
| *χ- | | BT | WT | TL | Thai | KP | PH | Saek |
| 1. horn | C1 | k ^h aw | xaw | k ^h aw | k ^h aw | haw | haw | haw |
| 2. tooth | A1 | k ^h ew | xew | k ^h ew | k ^h iəw | he:w | he:w | - |
| 3. to cross 4. to ascend | C1 C1 | k ^h a:m | xa:m | k ^h a:m | k ^h a:m | ha:m | ha:m | ha:m |
| 4. to ascella | CI | k ^h tun | xuin | k ^h wn | k ^h wn | hwn | hw:n | huin |

Table 5: Some etyma reconstructed with uvular consonants *q- and *χ-

Although [q], [kh], [x] and [h] differ from each other in terms of manners of articulation, places of articulation, and phonation-types, a change from [q] to [kh] or [h] is phonetically very plausible. First of all, the vocal tract shape is very similar in the stops and in the fricatives at dorsal places of articulation (Ladefoged and Maddieson 1996: 166-167), making it likely for the uvular stop [q] to develop a frication component and become an affricate [q χ] or a full-fledge fricative [χ]. If *q- has gone through an intermediate stage when it was realized as [χ], it is not surprising that k^h , x- or h- are the final outcomes in KP and PH respectively. [χ] can easily turn into [x] and [h] as they are all turbulent noises produced lower in the vocal tract, characterized by spectral peaks in the lower region of the frequency range (Jesus and Shadle 2005; Johnson 2003: 121, 129-132). It is also natural for [χ] to be perceived as [kh] because the low-frequency frication noise that makes up the former resembles the low-frequency turbulent noise in the aspiration phase of the latter (Johnson 2003: 121). Therefore, the changes from PSWT *q- to modern k^h -, k^h -,

One may argue that the uvular consonants developed secondarily in certain groups of SWT languages and therefore should not be reconstructed for PSWT. For example, PSWT *x- before back vowels may have been retracted to the uvular point of articulation, i.e. *xaw^A > *\chiaw^A > haw^A 'horn'. However, this cannot be the case because no conditioning environment can be identified. With regards to the vowels, these forms may have either high or non-high, and either front or back vowels. As for tones, forms in these correspondence sets may have PSWT tone *A, *B, *C, or *D. Therefore, the data from KP, e.g. hew^{CI} 'tooth' and haw^{AI} 'horn' as well as PH, e.g. hem^{AI} 'needle' and hat^{AI} 'leg', strongly suggest that a distinct series of uvular consonants consisting of *q- and *\chi\chi\$- should be reconstructed for PSWT. Although no modern dialects have a distinctive set of uvular initials, the systematic correspondence among the varieties included in this paper point to the conclusion.

2.4 Sifting the PSWT cluster inventory

Li (1977) reconstructs a wide range of initial clusters for the labial and velar places of articulation at the PSWT level. For many of these clusters, evidence for their existence at the PSWT level is meagre. In the following section, I show that the aspirated clusters $p^h l/r$, and $p^h l/r$, must be eliminated, and that there is no evidence for $p^h l/r$.

2.4.1 Elimination of *p h l/r- and * k^h l-

Li (1977: 87-89, 226-229) reconstructed * $p^h l/r$ -, * $k^h l$ -, and * $k^h r$ - for PSWT but data from modern dialects support Jonsson (1991: 69-70)'s view that only * $k^h r$ - existed. Li (1977: 87-88) claims that Ahom is the only variety that preserved the * $p^h l/r$ - cluster but "not consistently either in the spelling or in pronunciation." I argue that the Ahom symbol <-r-> should, in fact, be viewed as a redundant graphic marker of aspiration, rather than a symbol representing a separate medial r-. As has been shown in research on Ahom script, this graphic cluster is sometimes absent in words etymologically derived from * $p^h r$ -, i.e. Ahom has $p^h r$ - for PT * $p^h r$ -, 'cliff', * $p^h a$. 'to split', and * $p^h r$ - sky' (Wichasin 1986: 68). On the other hand, the <-r-> is sometimes present in words that come from PT * $p^h r$ -, * $p^h r$ -, * $p^h r$ -, * $p^h r$ - in modern Shan dialects. This suggests that PT cluster * $p^h l/r$ - was a simple $p^h r$ - in Ahom and that the liquid in PT * $p^h l/r$ - had already been lost by the PSWT stage.

However, Diller (1992) and Morey (2005) have recorded sporadic instances of $p^h r$ as well r- in the Aiton speech of Assam, i.e. $p^h r n \cdot 2^n$ 'jungle knife' and $r i^4$ 'long'³. These cases of r- and $p^h r$ - as opposed to the regular h- and $p^h r$ - should be considered cases of archaism or spelling-pronunciation as some of the Aiton r-initial words also have h-initial variants. The explanation for such orthographic interference is supported by the fact that Morey's elicitation of data is based on textual materials (Morey 2005: 99-97) Note that $p^h r a \cdot 2^n$ 'jungle knife' comes from PSWT *b r-, suggesting that many instances of $p^h r$ - are not related to the issue of whether PSWT retained PT * $p^h l l l l$ - at all. Therefore, Aiton $p^h r$ - is not evidence for reconstructing $p^h l l l$ - in PSWT. This stylistic variation resembles the case of r- in TH, in which the older pronunciation r- is kept in formal speech and in reading but the innovative pronunciation l- is expected in natural speech.

As for the velar clusters, Li (1977: 256) assumes without explaination that PT $*k^h I$ - and $*k^h r$ - were kept distinct in PSWT while Jonsson (1991: 77-78) argues that they had merged. Data from modern dialects support the latter position since the two clusters have identical reflexes in all SWT dialects. BT and WT shows s-, and c^h - respectively for both PT $*k^h I$ - and $*k^h r$ - while other varieties agree in showing k^h - or x-. Although there is a lack of internal evidence whether the liquid was *-r-, or *-l-, loanwords from Palaung suggests that the PSWT cluster was $*k^h r$ -, as shown by $*k^h r n r n r$ (Pittayaporn to appear). Therefore, evidence from modern SWT varieties and loanwords shows that PT $*k^h I$ - and $*k^h r$ - merged into PSWT $*k^h r$ -.

2.4.2 Clarifying *ml/r

Another problem regarding labial clusters is whether *ml- and *mr- contrasted. Li (1977: 93, 225) and Jonsson (1991: 64) speculate that a *mr-cluster might have existed in PT and

³ The expected Aiton outcomes for these two etyma are *pha*:² and *hi*:⁴ respectively.

PSWT but merged with *ml- in some modern varieties. Potential evidence for a distinct *mr- at the PSWT level includes the contrast between mr- and ml- in Southern Thai and the alleged split reflex in Thai and Lao. However, a careful examination of the data shows that only *ml- can be reconstructed for PSWT.

Although some Southern Thai dialects distinguish mr- from ml-, the contrast has been shown to be secondary (Diller 1976), that is, it results from processes of syllable reduction, e.g. Thai samrap (Thai) ~ Southern Thai mrap. In addition, while at first glance data from Lao and Thai seem to suggest that two clusters might have existed in PSWT, at a closer look the alleged split in reflex is actually a result of dialect mixing and spelling-pronunciation.

| *ml- | | BT | WT | TL | Yong | Lao | Thai | Saek |
|-----------------|----|------|------|-----|------|------|-------------------|-------------------|
| 1. to destroy | C2 | ma:ŋ | ma:ŋ | - | maːŋ | maŋ | la:ŋ | - |
| 2. to open (the | A2 | mwn | muin | mwn | muin | muin | lw:m ⁴ | mlə:n |
| eyes) | | | | | | | | |
| 3. slippery | B2 | mwn | mwn | mwn | mwn | mwn | lw:n | mlw:l |
| 4. seed | D2 | mit | mit | met | met | met | met | mlek |
| | | | | | | | (let) | |
| 5. insect | A2 | meŋ | meŋ | meŋ | meŋ | me:ŋ | meiŋ | me:ŋ ⁵ |
| | | | | | | | (ໄຍ:ŋ) | |
| 6. body louse | A2 | men | - | men | min | len | len | mlɛl |

Table 6: Etyma reconstructed with *ml-

For Thai, *ml/*mr is reflected as m-, l-, or ml-, superficially suggesting an earlier distinction. However, this hypothesis is unlikely. Cases of Thai ml- are instances of borrowing from archaic or poetic language, i.e. spelling-pronunciation, while cases of Thai m- may be dialect borrowings from either Lao or Yuan (Northern Thai). Thai forms for 'insect' and 'seeds' illustrate this point clearly. The disyllabic form $mal \varepsilon m^{A2}$ written as <mle:n> in modern orthography is used in formal and scientific contexts to refer to sixlegged insects as opposed to $m \varepsilon n^{A2}$, which refers to eight- or ten-legged ones. The latter is also used to refer to any kind of insects in colloquial speech. This is parallel to the case of the formal malet^{DS2} and the informal met^{DS2} for 'seed'. In addition, the Royal Institute's Dictionary 2002 (The Royal Institute 2003) also lists $l\varepsilon\eta^{A2}$ and let^{DS2} 'seed' with the meaning 'insect' and 'seed' respectively but only in very specific uses. This suggests that these 1-initial forms were the original outcomes in Thai but were replaced later by the minitial and the disyllabic forms. Moreover, in many cases only the l-initial form exists e.g. $l\epsilon p^{DL2}$ 'to flash (of lightning)' but not * $m\epsilon p^{DL2}$. In sum, the only direct reflex of the putative *ml/r cluster in Thai is l-, leaving no evidence for the PSWT *ml-/*mrdistinction.

The final -m is likely to be a result of assimilation to the *m- in *ml-, i.e. *mlum > *mlum > luxm.

⁵ This is probably a Lao loan.

In Lao, the regular reflex is m- but forms with l- also exist, e.g. 'body louse'. These may also be cases of dialect mixing since the dialect of Lao used in this paper is spoken in Thailand, where Thai is the national language. If this etymon is found with m-, in other Lao dialects, the dialect mixing hypothesis can be confirmed. Given that unexpected reflexes in Thai and Lao can be explained in terms of dialect mixing, a single *ml- cluster is reconstructed for PSWT.

2.4.3 Filling the gap: the mid vowel *x

In earlier proposals, there is a gap in the middle of the PSWT vowel space. That is, they do not posit the mid back unrounded vowel $*\nu$ although its front counterpart *e and its rounded counterpart *o are reconstructed. Li (1977) and Sarawit (1973) only reconstruct high back vowels *uv and *uv for PSWT, while Jonsson (1991) does not reconstruct $*\nu$ but only reconstructs long *uv. Etyma reconstructed with $*\nu$ here are reconstructed with *uv in earlier proposals.

| Table 7: <i>Re</i> | eflexes of H | PSWT 1 | mid back | unrounded | $*\gamma$ |
|---------------------------|--------------|--------|----------|-----------|-----------|
|---------------------------|--------------|--------|----------|-----------|-----------|

| | BT | WT | TL | Yong | Lao | Thai |
|--------------|----|----|----|------|--------|------|
| *wk | шk | шk | шk | шk | шk | шk |
| *wŋ | шŋ | шŋ | шŋ | шŋ | шŋ | шŋ |
| *xk | γk | γk | шk | шk | γk~wk | шk |
| * ૪ ŋ | γŋ | шŋ | шŋ | un | ຯŋ~ແເŋ | шŋ |

The reflexes of back unrounded vowels before velars are not regular. While TL and Thai agree in having u for every etymon, other varieties have either u or γ . However, languages that show a distinction agree with only a few exceptions. For the high vowels, the length distinction is still retained in KP as illustrated by the contrast between 'one' and 'to steam.' In other varieties, PSWT *-u:u shortened to -u:u across the board. The situation in WT is not very clear as the outcomes of 'late at night' and 'ink' do not agree in vowel height. It is possible that 'ink' was borrowed into Tai languages after PSWT diversified into daughter languages, cf. Ξ Early Middle Chinese m:v

| *wk | | BT | WT | Shan | Yong | Lao | Thai |
|----------------------|----|------|-------|-------|-------|-------------------|--------|
| 1. late at night | D1 | dwk | dγk | dwk | dwk | dwk | dwk |
| 2. ink | D1 | mwk | mwk | mwk | mwk | mwk | mwk |
| | | | | тvk | | | |
| *wŋ | | BT | WT | Shan | Yong | Lao | Thai |
| 3. one | B1 | nwŋ | nwŋ | nwŋ | nwŋ | nw:ŋ ⁶ | nwŋ |
| 4. to steam | B2 | nwŋ | nwŋ | nwŋ | nwŋ | nwŋ | nwŋ |
| *xk | | BT | WT | Shan | Yong | Lao | Thai |
| 5. young male animal | D1 | thyk | thyk | thwk | thwk | thyk | thwk |
| | | | | | | thwk | |
| 6. enemy, war | D1 | syk | syk | swk | swk | sγk | swk |
| | | | | | sγk | | |
| 8. deep | D2 | lyk | lγk | lwk | lwk | lyk | lwk |
| | | lwk | | | | lwk | |
| 9. chest | D2 | ?vk | ?xk | ?xk | ?wk | ?xk | ?ok |
| | | | | | ?ok | ?ok | |
| * ૪ŋ | | BT | WT | Shan | Yong | Lao | Thai |
| 10. long (of time) | A1 | hɣŋ | hwŋ | hwŋ | hwŋ | hviŋ | hwŋ |
| 11. to sift | A2 | syŋ | chuiŋ | khuıŋ | khuıŋ | khɣŋ | khuıŋ |
| 12. to arrive | A1 | thơŋ | thwŋ | thwŋ | thɯŋ | thơŋ | thɯŋ |
| | | | | | | thwŋ | |
| 13. half | B1 | cyŋ | cun | khuıŋ | khuıŋ | khɣŋ | khruiŋ |

Table 8: Etyma reconstructed with *w and *r before velar consonants

As for the mid vowel * γ before velar, BT and Lao retain the vowel intact with while TL and Thai merged *-uuk/*-uuy with *- γk /*- γy to give modern *-uuk/*-uuy. For WT and Yong, γ does not occur before y, suggesting that it has raised to uu before a nasal coda. Note that *e and *o regularly give i and u respectively before nasal finals in WT. Since *- γy is hypothesized to have raised to -uuy, WT - γk is then considered a retention, thus a good diagnostic for PSWT * γ . As for * γ before *- η the best diagnostic is BT and Lao - γy . These two languages still retain the distinction between high and mid back unrounded vowels intact. Note that Yong and Lao have doublets for some etyma possibly because it is going through a change in progress or because of interference from other varieties, such as Thai. However, Lao shows doublets only for etyma reconstructed with * γ .

An interesting case is the etymon 'chest' which shows a rounded vowel o in Thai in contrast with γ in all the other varieties. This is reminiscent of the change *-ut > *-ot in *?utt^D 'to go hungry' and is therefore likely to be a sporadic change on the part of Thai. This interpretation is supported by Kullavanijaya (1992: 653)'s view that the Thai word for $sa'?uuk^{DSI}$ 'hiccups' has the same etymological source for Zhuang sak^{DSI} $?vk^{DSI}$ 'block in the chest'. Having establish that Thai raised *-vk to -uuk across the board, the expected

⁶ That this Phu Thai form has a long vowel may indicate *ur: in PSWT.

outcome for PSWT *?vk strikingly is -?uuk, the second part of the Thai 'hiccups'. Note that the o ~ ut alternation is quite common in spoken Lao, e.g. mot^{DI} ~ mut^{DI} 'all gone'.

2.4.4 In support of contrastive length

The PT vowel system has been the subject of investigation for decades but problems still remain. The PSWT vowels, in contrast, have not been as controversial. Li (1977) and Sarawit (1973) agree in reconstructing vowel length for the high vowels and *a. They also share the view that there is a contrast between short mid vowels and long low vowels, i.e. *e and $*\varepsilon$. However, recently doubts on these interrelated aspects of the vowel system have been expressed (Jonsson 1991; Luo 1997). Here I show that evidence form modern dialects support the position that vowel length was contrastive in PSWT.

While opting not to reconstruct a vowel length contrast in PT, Li (1977) allows for a quantity contrast in PSWT. This choice is in line with Sarawit (1973), whose reconstruction only differs from that of Li in details of the changes involved. In contrast, Jonsson (1991) treats vowel-length as secondary. After excluding secondary lengthening in daughter languages such as Thai and Southern Tai (Diller 1979; Hartmann 1976), the patterns of tone mergers and splits show that all vowels were long in open syllables but vowel-length was contrastive on closed syllables. From the perspective of Li's PT, long vowels in closed syllables in SWT varieties developed from earlier diphthongs or triphthongs.

Although short and long vowels in cognates in modern dialects that have quantity distinction are almost in complementary distribution, the unpredictability of the reflexes favors a reconstruction of PSWT with contrastive vowel length. If vowel length had not been phonemic, conditioning environments should be identifiable. Moreover, languages that have vowel lengths agree in most cases. When they do not, syllable structure or tone can be identified as the conditioning environments. In the Yong cases presented here, the nasal codas are responsible for the neutralization of vowel length, i.e. pre-nasal shortening.

Table 9: Etyma illustrating vowel length contrast among high vowels

| *-i- and *i: | | BT | WT | Shan | Yong | Lao | Thai |
|---------------------------|------------|-------------------|---------|-------------------|--------------------|--------------------|--------------------|
| 1. raw, not ripe | D1 | dip | dip | lip | lip | dip | dip |
| 2. ten | D1 | sip | sip | ship | sip | sip | sip |
| 3. to fly | A 1 | bin | bin | min | bin | bin | bin |
| 4. knife | D1 | mit | mit | mit | mi:t | mi:t | mi:t |
| 5. to pinch | D1 | nip | nip | nip | ni:p | ni:p | niːp |
| 6. foot | A1 | tin | tin | tin | tin | ti:n | ti:n |
| *-u- and *-u:- | | BT | WT | Shan | Yong | Lao | Thai |
| 7. cooked, ripe | D1 | suk | suk | suk | suk | suk | suk |
| 8. to dig | D1 | k^hut | k^hut | k ^h ut | k ^h ut | k ^h ut | k ^h ut |
| 9. mosquito | A2 | ກູຫ | ກູຫ | juŋ | juŋ | ກູຫ | juŋ |
| 10. loom | D1 | hu? | hu:? | huk | hu:k | hu:k | hu:k |
| 11. to stroke | D2 | lup | lup | lup | luːp | lu:p | lu:p |
| 12. tall | A1 | suŋ | suŋ | s ^h uŋ | suŋ | suːŋ | su:ŋ |
| *w and *w: | | BT | WT | Shan | Yong | Lao | Thai |
| 13. to ascend | C1 | k ^h wn | xwn | k ^h wn | k ^h wn | k ^h wn | k ^h wn |
| 14. to go hungry | D1 | ?uit | ?wt | - | ?шр | ?wt | ?ot |
| | | | | | ?ot | | |
| 15. night | A2 | kwn | xwn | k ^h wn | k ^h wn | k ^h w:n | k ^h w:n |
| 16. dark | D2 | mut | mut | mut | mw:t | mw:t | mw:t |
| 17. measure from thumb to | D2 | kwp | xwp | k ^h wp | k ^h w:p | k ^h w:p | k ^h w:p |
| fingertip | | | | | | | |

Another piece of evidence is that in some systems where quantity is not contrastive, other aspects of their phonology indicate that length was distinctive at an earlier stage. In BT and WT, although phonemic length is limited to a, final *-k has become ?- in etyma that have long vowels in Yong, Lao, and Thai, e.g. 'cooked, ripe' and 'loom' above and 'to turn over', 'wing' and 'children' below. In addition, in TL and some other Shan dialects, checked syllables reflected with long vowels in Thai and Lao show different tonal reflexes from those that correspond to short syllables. If vowel length were not contrastive at an earlier stage, there is no explanation for the observed patterns of the tonal splits. The number following each modern form indicates tone within each sound system. In 'tone-box' terms, in D1S forms are reflected with tone 4 while those D1L show tone 2 in TL. This is exemplified by the tonal development of the Shan dialect of Muong Khon shown in Figure 1.

*A*B*C *DS *DL 5 3 1 2 5 2 vcl. aspirated *f-, hm-, ph-,... 1 plain vcl. *p-,... glottalized vd. **?*b-,... 2 plain voiced 4 1 1 *v-, m-, b-,...

Figure 1: Tonal development of the Shan dialect of Muong Khawn (modified from Chamberlain 1975)

Unlike the quantity contrast among high vowels, the issue of length in mid and low vowels is more problematic. Li (1977) and Sarawit (1973) reconstruct long low vowels $*\varepsilon$ and $*\sigma$, contrasting with short mid vowel *e and *o respectively. Jonsson (1991), similarly, reconstructs both mid and low vowels without length contrast. She also argues that the mid vowels might have developed from the short allophone of $*\varepsilon$ and $*\sigma$. However, Luo (1997: 57-63) argues that low vowels in SWT dialects are results of e-lowering. Although he addresses the issue of the PT vowel system, he argues that ε in individual SWT dialects is secondary.

Luo's (1997) analysis is methodologically problematic because different kinds of lowering from various modern dialects are lumped together without working out the development in individual varieties. For example, he shows that several items in Dehong and Tai Lue that show ε have a in Thai and thus assumes that this is evidence for the secondary nature of low vowels. That this particular set of ε in this dialect of Lue is secondary does not provide evidence for or against the reconstruction of PSWT low vowels. Interestingly, he does not provides Lue forms that corresponds to Thai and Lao ε and ε . If all ε and ε were really secondary, it would be likely that PSWT * ε and * ε may have merged with * ε and * ε respectively. He also provides other cases of neutralization and lowering but none of them have been analysed both from the comparative perspective or from that of internal developments within individual languages.

Evidence from modern dialects, though not robust, shows that two PSWT non-high vowels must be reconstructed in each of the front and back vowel series. The first piece of evidence is that the distributional relationship between Li's $*e \sim *\varepsilon$ and $*o \sim *\sigma$ are like that of $*i \sim *i$ and $*u \sim *u$ — almost in complementary distribution but unpredictable. Note the minimal pairs 'duck' vs. 'eight' and 'bird' vs. 'outside'. Secondly, some varieties, such as TL and TY, do not have a quantity contrast in non-high vowels but do have different tonal reflexes for etyma that correspond to Thai and Lao short mid vowels, on one hand, and for those that correspond to Thai and Lao long low vowel on the other. This suggests that there were contrasts between short and long non-high vowels at an earlier stage. Lastly, in languages like WT, Aiton, TY, and Yong where raising from *e and *o to i and u respectively occurred, etyma that correspond to Thai and Lao low vowels did not participate in the change, suggesting that at an earlier stage the two non-high phonemes may have been of different heights.

| *e and *o | | WT | TL | Aiton | TY | Lao | Thai |
|-------------------------------|----------|---------------------|---------|-------------------------|--------------------------|------------------|------------------|
| 1. duck | D1 | pet² | pet² | pit1 | pit ³ | pet1 | pet² |
| 2. to pick up | D1 | kep² | kep² | kip1 | cip³ | kep1 | kep² |
| 3. bird | D2 | nok ⁴ | nok^4 | nuk³ | nok¹ | nok³ | nok ⁴ |
| 4. heel | C1 | sun³ | shon³ | - | - | son ⁴ | son³ |
| | | | | | | | |
| * <i>e</i> : and *o: | | WT | TL | Aiton | Dehong | Lao | Thai |
| * <i>e:</i> and *o: 1. eight | D1 | WT pet ² | TL pet² | Aiton pet ¹⁷ | Dehong pjet ² | Lao pe:t² | Thai pe:t² |
| | D1 D1 | | | _ | | | |
| 1. eight | | pet² | pet² | pet ¹⁷ | | pe:t² | pe:t² |

Table 10: Etyma illustrating vowel length contrast among non-high vowels

Data from some modern dialects show evidence for an earlier quantity distinction, some show evidence for earlier contrast in vowel quality, and some show both. This suggests that at the PSWT stage both quantity and quality were redundantly contrastive among non-high vowels. That is, low vowels *ɛ: and *ɔ: should be reconstructed, contrasting with mid vowels *e and *o. This choice of reconstruction is supported by the fact that varieties that lost quantity distinction still retain quality distinction in non-high vowels.

3 Notable features of PSWT

As argued for above, the new reconstructed PSWT phonology shows characteristics that depart from earlier proposals. For instance, some clusters posited in earlier reconstruction have been eliminated. Another example is the newly reconstructed $*\nu$, which has filled in the gap in the vowel inventory of PSWT. These changes in the reconstruction reveal notable features of PSWT that have implications for the reconstruction of PT. In the following section, I discuss four important features of the PSWT sound systems focusing on their ramifications for deeper-level reconstruction: 1) redundancy of length contrast among non-high vowels, 2) gaps within the sub-system of initial clusters, 3) scarcity of uvular consonants, and 4) defective nature of the vowel $*\nu$.

3.1 Redundancy of length contrast among non-high vowels

As discussed in 2.4.4, both length and height were contrastive in PSWT but mid and low vowels contrast with respect to height as well as for length. Specifically, PSWT mid vowels *e, *r, and *o are always short and cannot occur in open syllables, but low vowels *e; and *r are always long and can occur in open syllables. This peculiar distribution is illustrated in Table 11.

Aiton does not have e/ε and o/o contrasts. Morey (2005) transcribes these phonemes as ε and o respectively because they correspond to contrastive ε and o in Phake. Therefore, Aiton orthography does not distinguish between open and half-open vowels, so it is not helpful here.

Table 11: *Distribution of PSWT non-high vowels*

| | open syllables | | closed syllables | | |
|-----|----------------|-------|------------------|-------|--|
| | long | Short | long | short | |
| mid | | | | *-ok | |
| low | *-0! | | *-ə:k | | |

Aspects of this distributional fact has been noticed by various researchers (Dhananjayananda 1997; Jonsson 1991: 118, 120; Sarawit 1973: 98) but opinions differ on how it relates to PT reconstruction. Li (1977: 259-261) assumes that length was not contrastive in PT and accounted for this PSWT phenomenon by positing low vowels * ε and * σ in PT. In this account, the two low vowels were lengthened regularly in open syllables, leading to the redundant length in PSWT low vowels. He also states that PT mid vowels * ε and * σ were always short and did not undergo lengthening like their low vowel counterparts. In contrast, Sarawit (1973) posits a length contrast through out the vowel system and derived PSWT low vowels ε and * σ by a lowering of PT long * ε and * σ .

The distributional pattern argues against Li's account and lends strong support to Sarawit's proposal. That is, Li cannot account for why PSWT mid vowel *e, and *o did not occur in open syllables. In his scenario, PT *e and *o were long because they were subject to regular open-syllable lengthening. In the same way, PT *e and *o were not lengthened because they did not occurred in open syllables. However, it is not clear why PT mid vowels *e and *o were banned from open syllables in the first place. Sarawit also believes that PT vowels in open syllables are always long but she holds that there was a quantity contrast in PT vocalism. It is precisely this quantity contrast that explains why short *e and *o were not allowed in open syllables. That is, only long vowels were allowed in open syllables. Therefore, the distributional pattern should be view as a gap resulting from a lowering of PT long mid vowel *e: and *o: to PSWT *e: and *o: as proposed by Sarawit.

3.2 Gaps within the sub-system of initial clusters

Another intriguing distributional pattern characteristic of the reconstructed PSWT phonology is gaps within the sub-system of initial clusters. Despite the rich PSWT consonant inventory, only labial and velar stops can be combined with medial *-l- and *-r- as illustrated in Table 12. Unaspirated stops were combined only with medial *-l- and never with *-r-. Aspirated stops only took medial *-r- as the only aspirated cluster in PSWT is *k h-.

| | | labial | velar |
|------|------|--------|--------------------|
| *Ch- | | | |
| *C- | *-1- | *pl- | *kl- |
| *C - | | *bl- | *gl- |
| *Ch- | | | *k ^h r- |
| *C- | *-r- | | |
| *C - | | *br- | *gr- |

Table 12: Clusters consisting of a stop plus a liquid in PSWT

Two opposing interpretations are consistent with SWT data. The first interpretation is that PSWT aspirated clusters developed secondarily from a PT perspective. That is, PT only had unaspirated clusters but some of these clusters acquired aspiration under influence of medial *-r-, i.e. PT *kr- > PSWT k^hr -. Another interpretation is that PT only had aspirated clusters whose aspiration was lost in PSWT due to the presence of the following medial *-l-, i.e. PT * k^hr - > PSWT kl-. According to both interpretations, the gaps in the sub-system of initial clusters are results of changes that impacted the symmetry of PT phonotactics. Although the correct interpretation can only be identified by investigating this issue from a wider Tai perspective, the change from PT *kr- > PSWT k^hr - seems to be more plausible from a typological point of view. Aspiration of stops induced by medial -r- is attested in Tai Yuan (preliminary analysis based on data from Wichiankhiaw 1996) and the colloquial Phnom Penh dialect of Khmer (Wayland and Guion 2005)

3.3 Scarcity of uvular consonants

PSWT consonant inventory can be characterized as having a rich set of distinctive places of articulation and a four-way contrast in phonation-type. Among the six places of articulation, the uvular series have not been reconstructed before as discussed in section 2.3. PSWT etyma reconstructed with uvular initials largely match up with those proposed by Haudricourt (1952) and Ostapirat (2007) who concluded that PT had uvular consonants. This suggests that PSWT uvular consonants were inherited from PT. However, only two uvular consonants have been proposed for PSWT: *q- and * χ -. This observed scarcity of uvular consonants must be directly related to the reconstruction of PT.

There are three possible explanations for the gaps in the uvular series. The first one is to simply view the empty slots as systematic gaps the phonology of PT and PSWT. That is, back places of articulation cross-linguistically tend to have fewer contrasts when compared to more front series. This would make the uvular series parallel to the laryngeal place of articulations, where only the voiceless stop*? and the voiceless fricative *h-existed in PSWT and PT. The second possible explanation is that these gaps resulted from mergers that eliminate certain PT uvular consonants. In other words, more uvular consonants may have existed but some had merged with other sounds before the PSWT stage. For example, aspirated * q^h may have existed in PT but had merged with * k^h or * χ -before the time of PSWT. A seemingly incompatible explanation is that the uvular series is more conservative than more front places of articulation. That is, there were splits in other places of articulation that created contrasts that did not exist in PT. Hypothetically, PT*pr-may have become PSWT (* $p^h r$ ->)* p^h - but PT *qr- became *kr- then * $k^h r$ -. This way, no new phonation-type or manner contrast were created in the uvular series.

Although these three explanations may seem contradictory, they are in fact compatible. One likely possibility is that the PT uvular series started out having fewer contrasts than other series. In later stages, some mergers possibly eliminated certain uvular consonants, e.g. merger of PT $*_{G^-}$ and $*_{g^-}$, while some other splits were splitting consonants in other series, e.g. PT $*_{p}$ became $*_{p}$ before $*_{-r^-}$ but remained $*_{p^-}$ elsewhere. In any case, the correct PT reconstruction will have to account for the scarcity of the uvular consonants in PSWT.

3.4 Defective nature of the vowel *x

Another notable feature of PSWT that has implications for the reconstruction of PT is the defective nature of the mid back unrounded * ν : As pointed out in section 2.4.3, PSWT * ν only occur before velar finals. This curious characteristic of the PSWT * ν contrasts sharply with other vowels, which can be combined with most final consonants. The first interpretation is that PSWT * ν was a reflex of some other vowel in PT. Specifically, there are only three examples of PSWT * ν before velars, all of which are not found in the so-called NT languages: * ν beg 'to strain', * ν ebfore velars, all of which are not found in the so-called NT languages: * ν beg 'to strain', * ν contrasting with newly-introduced * ν -ek and * ν -e ν became PSWT * ν - ν k and * ν - ν 0 contrasting with newly-introduced * ν -ek and * ν -e ν 0 in etyma like 'to strain', 'iron', and 'child'. Another possibility is that * ν 0 was not defective but later sound changes transformed it into some other vowels before non-velar consonants. In this scenario, PSWT * ν - ν 4 and * ν 1 would be directly inherited from PT.

From a PT perspective, the second possibility seems more likely. At the PSWT level, neither Li (1977), Sarawit (1973), or Jonsson (1991) reconstruct this phoneme. At the PT stage, in contrast, Li does posit a mid central vowel *ə but it is reflected in PSWT as short *a. According to this view, it is precisely this lowering that gives rise to the contrast between short *a and long *a: in modern Tai dialects. In contrast, Sarawit reconstructs both long *x: and short *y* for PT. It is this short *y* that relates to the PSWT *y* posited here. She (1973: 433) states that PT *y* became PSWT *uy* before velar consonants, *o* before labial consonants, and *e* before dental consonants. Although Sarawit's account of PT vowel and the second interpretation were arrived at independently, they are in harmony and reinforce each other. Therefore, the defective nature of the PSWT *y* supports Sarawit's view that PT had a distinctive short *y*. Moreover, it argues against Li's use of the mid back vowel slot to account for the contrast between *a and *a: in modern dialects. Having argued in 3.1 that PT had contrast between short and long mid vowels, positing short *y* for PT also raise the question whether long *y* existed in PT.

4 Conclusion

This paper has presented a reconstruction of PSWT that differs markedly from earlier proposals. First of all, a closer look at newly available data shows that clusters $*p^hr/l$ -, $*k^hl$ -, and *mr- reconstructed by Li (1977) and Jonsson (1991) did not exist in PSWT. Moreover, an additional series of dorsal initials also need to be posited, namely the uvular series which consists of *q-, and $*\chi$ -. Regarding the vowels, it has been proposed that a vowel-length distinction be reconstructed at the PSWT stage. Furthermore, it has also been demonstrated that the parent of SWT varieties did have mid back unrounded vowels *x. Most crucially, four notable features of PSWT that have implications for the reconstruction

⁸ *ə and *ə: in Sarawit's original notation.

⁹ **i* in Sarawit's notation.

of PT have been discussed, including the redundancy of length contrast among non-high vowels, the gaps within the sub-system of initial clusters, the scarcity of uvular consonants, and the defective nature of *x.

This reconstruction is a synthesis of earlier works which differ in crucial aspects. It incorporates data from a wide-range of SWT varieties, including newly-available data from Phuan, and Phu Thai dialects. Furthermore, it accounts for previously-unrecognized correspondence sets and evaluates philological, and loanword evidence critically. Last but not least, it considers very carefully development of individual languages. Although this paper has highlighted important aspects of this new PSWT reconstruction, detailed discussions and reconstruction of individual etyma have not been presented. Moreover, many outstanding issues are still left untouched. For example, the phonetic characteristics of the tonal categories *A, *B, *C and *D are still a mystery. The classification of modern SWT dialect is still open to debates. This paper is another step toward a comprehensive historical treatment of SWT languages.

Appendix A
Proposed SWT compared with other reconstructions

| Pittayaporn | Li | Jonsson | Sarawit | Brown |
|-------------------|----------------------|---------------------------|--------------------|--------------------|
| *p- | *p- | *p- | *p- | *p- |
| *p ^h - | $*p^h$ | $*p^{h}$ | *p ^h | *p ^h - |
| | *p ^h l/r- | *p ^h l/r- | *p ^h r- | |
| | | | | *p ^h l- |
| *b- | *b- | *b- | *b- | *b- |
| *?b- | * [?] b- | *?b- | *?b- | * [?] m- |
| *m- | *m- | *m- | *m- | *m- |
| *hm- | *hm- | *hm- | *hm- | *hm- |
| *f- | *f- | *f- | *f- | *f- |
| *v- | *v- | *v- | *v- | *v- |
| *w- | *w- | *w- | *w- | *w- |
| *hW- | *hW- | *hW- | *hW- | *hW- |
| | | | | *pr- |
| *pl- | *pl- | *pl- | *pl- | *pl- |
| *bl- | *bl- | *bl- | *bl- | *bl- |
| *br- | *br- | *br- | *br- | *br- |
| *²bl | * [?] bl/r- | * [?] bl- | * [?] bl- | |
| *ml- | *ml/r- | *ml/r- | *ml- | |
| *t- | *t- | *t- | *t- | *t- |
| *t ^h - | *t ^h - | *t ^h - | *t ^h - | *t ^h - |
| *d- | *d- | *d- | *d- | *d- |
| *?d- | * [?] d- | * [?] d- | * [?] d- | * [?] n- |
| *n- | *n- | *n- | *n- | *n- |
| *hn- | *hn- | *hn- | *hn- | *hn- |
| *hr- | *hr- | *hr- | *hr- | |
| | *xr- | *xr- | *xr- | |
| | | | | * ^h r- |
| *r- | *r- | *r- | *r- | *r- |
| *1- | *1- | *1- | *1- | *1- |
| * ^h l- | * ^h l- | * ^h l- | * ^h l- | * ^h l- |
| *s- | *s- | *s- | *s- | *s- |
| *z- | *z- | *z- | *z- | *z- |
| *c- | *c- | *c- | *c- | *c- |
| | *c ^h - | * f ^h - | *c ^h - | *c ^h - |
| * J - | * J - | * J - | * J - | * J - |

| Pittayaporn | Li | Jonsson | Sarawit | Brown |
|--------------------|-----------------------|--------------------|--------------------|--------------------|
| *ɲ- | *n- | *ŋ- | *ŋ- | *ɲ- |
| *hn- | * ^h ɲ- | * ^h ɲ- | *hŋ- | * ^h ɲ- |
| *j- | *j- | *j- | *j- | *j- |
| *?j- | * [?] j- | * [?] j- | * [?] j- | * [?] ɲ- |
| *k- | *k- | *k- | *k- | *k- |
| *k ^h - | | *k ^h - | *k ^h - | *k ^h - |
| *x- | *k ^h -, x- | | | |
| *q- | · K -, X- | *x- | *x- | *x- |
| *χ- | | | | |
| *g- | *g- | *g- | *g- | *g- |
| *γ- | *γ- | *γ- | *γ- | *γ- |
| *ŋ- | *ŋ- | *ŋ- | *ŋ- | *ŋ- |
| *hŋ- | *hŋ- | *hŋ- | *hŋ- | * ^h ŋ- |
| | | | | *tr- |
| | | | | *dr- |
| *kl- | *kl- | *kl- | *kl- | *kl- |
| | *kr- | *kr- | *kr- | *kr- |
| *k ^h r- | *k ^h l- | *k ^h r- | *k ^h r- | |
| | *k ^h r- | | | |
| | | | | *k ^h l- |
| *gl- | *gl- | *gl- | *gl- | *gl- |
| *gr- | *gr- | *gr- | *gr- | *gr- |
| *kw- | *kw- | *kw- | *kw- | *kw- |
| *gw- | *gw- | *gw- | *gw- | *gw- |
| *γw- | *γw- | *γW- | *γw- | |
| *ŋw- | *ŋw- | *ŋw- | *ŋw- | la |
| *k ^h w- | *k ^h w- | *k ^h w- | *k ^h w- | *k ^h w- |
| *?- | *?- | *?- | *?- | *?- |
| *h- | *h- | *h- | *h- | *h- |
| *i | *i | *i | *i | *i |
| *i: | *i: | 1 | *i: | *i: |
| *u | *u | *u | *u | *u |
| *u: | *u: | u | *u: | *u: |
| *8 | ₩ | | Ψ | * |
| *w | *w | *w | *w | *w |
| *w: | *w: | | *w: | *w: |
| | | | | *8 |

| Pittayaporn | Li | Jonsson | Sarawit | Brown |
|-------------|-----|---------|---------|-------|
| *e | *e | *e | *e | *e |
| *o | *o | *o | *o | *o |
| | | | | *e: |
| | | | | *o! |
| *8! | *&! | *ε | *8! | *ε |
| *3: | *31 | *3 | *3: | *3 |
| *a | *a | *a | *a | *a |
| *a: | *a: | *a: | *a: | *a: |
| *iə | *iə | *iə | *iə | *iə |
| *wə | ew* | ew* | *wə | em* |
| *uə | *uə | *uə | *uə | *uə |
| *a | *a | *a | *a | *a |

Appendix B SWT Languages used in this study

| | Code | Language | Source |
|----|-------|------------------|---|
| 1 | KT | Tai Khamti | Harris (1976) |
| 2 | PK | Tai Phake | Morey (2005) |
| 3 | AT | Tai Aiton | Morey (2005) |
| 4 | LU | Tai Lüa | Harris (1975) |
| 5 | MA | Tai Mao | Harris (1975) |
| 6 | NU | Tai Nüa | Harris (1975) |
| 7 | DH | Dehong Tai | Luo (1999) |
| 8 | TY | Tai Ya | Xing (2000) |
| 9 | LE | Lue | Hudak (1996) |
| 10 | YG | Yong | Gedney (n.d.) and Hudak (1996) |
| 11 | TL | Tai Long | Gedney (n.d.) |
| 12 | BT | Black Tai | Gedney (n.d.) |
| 13 | RT | Red Tai | Gedney (n.d.) |
| 14 | WT | White Tai | Gedney (n.d.) |
| 15 | YU | Tai Yuan | Rungruengsri (1991) |
| 16 | KH | Tai Khün | Petsuk (1978) |
| 17 | LA | Lao | Gedney (n.d.) |
| 18 | PH(1) | Sukhothai Phuan | Trongrat (1998) |
| 19 | PH(2) | Suphanburi Phuan | Trongrat (1998) |
| 20 | PH(3) | Nongkhai Phuan | Trongdi and Thananan (1998) |
| 21 | WG | Phu Thai Wang | Trongdi and Thananan (1998) |
| 22 | KP | Phu Thai Kapong | Pittayaporn (2005) |
| 23 | YO | Yooy | Boonsner (1984) |
| 24 | NY | Nyo | Boonsner (1984) |
| 25 | KL | Kaloeng | Arilak (1985) |
| 26 | TB | Tak Bai Thai | Trongdi and Thananan (1998) |
| 27 | TH | Thai (Siamese) | Gedney (n.d.) |
| 28 | ST | Southern Thai | Trongdi and Thananan (1998) and Umar (2003) |

References

- Boonsner, Thepbangon. 1984. *The Phonological description of Yooy*. M.A. Thesis. Department of Linguistics, Institute of Language and Culture for Rural Development, Mahidol University. Nakhorn Pathom.
- Brown, J. Marvin 1965. From Ancient Thai to modern dialects. In *From Ancient Thai to modern dialects, and other writings on historical Thai linguistics*, pp. 69-254. Bangkok: White Lotus.
- Chamberlain, James R. 1975. A new look at the history and classification of the Tai dialects. In J. G. Harris and J. R. Chamberlain, eds, *Studies in Tai linguistics in honor of William J. Gedney*, pp. 49-60. Bangkok: Central Institute of English Language, Office of State Universities.
- Dhananjayananda, Puttachart. 1997. The emergence of the length distinction in mid-front vowels *e-ee* in Thai. In J. A. Edmondson and D. B. Solnit, eds, *Comparative Kadai: The Tai branch*, pp. 327-336. Arlington, TX: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Diller, Anthony. 1976. A note on Proto-Tai *ml-. In T. Gething, J. G. Harris and P. Kullavanijaya, eds, *Tai linguistics in honor of Fang-kuei Li*, pp. 39-46. Bangkok: Chulalongkorn University Press.
- _____. 1979. Tones, segments and Thai regional society. In T. L. Thongkum, ed, *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie J.A. Henderson*, pp. 60-93. Bangkok: Chulalongkorn University Press, Bangkok.
- ____. 1988. Consonant mergers and Inscription One. *Journal of the Siam Society* 76:46-63.
- _____. 1992. Tai Languages in Assam: Daughters or Ghosts? In C. J. Compton and J. F. Hartmann, eds, *Papers on Tai languages, linguistics and literatures*, pp. DeKalb: Northern Illinois Center for Southeast Asian Studies
- Edmondson, Jerry A. and David B. Solnit. 1997. Introduction. In J. A. Edmondson and D. B. Solnit, eds, *Comparative Kadai: The Tai Branch*, pp. 1-26. Arlington, TX: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Gedney, William J. 1989a. A checklist for determining tones in Tai dialects. In R. J. Bickner, J. Hartmann, T. J. Hudak and P. Peyasantiwong, eds, *Selected papers on Comparative Tai Studies*, pp. 191-206. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.
- _____. 1989b. Future directions in Comparative Tai Linguistics. In R. J. Bickner, J. Hartmann, T. J. Hudak and P. Peyasantiwong, eds, *Selected papers on Comparative Tai Studies*, pp. 7-116. Ann Arbor: Center for Southeast Asian Studies, the University of Michigan.
- _____. n.d. A comparative Tai wordlist. manuscript.
- Harris, Jimmy Ĝ. 1975. A comparative word list of three Tai Nüa dialects. In J. G. Harris and J. R. Chamberlain, eds, *Studies in Tai linguistics in honor of William J. Gedney*, pp. 202-230. Bangkok: Central Institute of the English Language and Office of State Universities.
- _____. 1976. Notes on Khamti Shan. In T. W. Gething, J. G. Harris and P. Kullavanijaya, eds, *Tai linguistics in honor of Fang-Kuei Li*, pp. PAGES. Bangkok: Central Institute of English Language, Office of State Universities.
- Hartmann, John. 1976. The waxing and waning of vowel length in Tai dialects. In T. Gething, J. G. Harris and P. Kullavanijaya, eds, *Tai linguistics in honor of Fangkuei Li*, pp. Bangkok: Chulalongkorn University Press.
- Haudricourt, André-Georges. 1952. Les occlusives velaires en thai. *Belletin de la Société de Linguistique de Paris* 48:86-89.
- Hudak, Thomas J., ed. 1996. William J. Gedney's Lue dialects: glossaries, texts, and translations. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

- ____. ed. 2008. William J. Gedney's Comparative Tai source book. Honolulu: University of Hawaii Press.
- International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Jesus, Luis M. T. and Christine H. Shadle. 2005. Acoustic analysis of European Portuguese uvular [χ, κ] and voiceless tapped alveolar [r] fricatives. *Journal of the International Phonetic Association* 35:27-44.
- Johnson, Keith. 2003. *Acoustic and auditory phonetics*. 2nd. Oxford, UK; Cambridge, MA: Blackwell Publishing.
- Jonsson, Nanna L. 1991. *Proto-Southwestern Tai*. Ph.D. dissertation. Department of Linguistics, State University of New York at Albany. Albany, NY.
- Kullavanijaya, Pranee. 1992. A study of some two-syllabled words in Thai. In *Proceedings* of Third International Symposium on Language and Linguistics, Pan-Asiatic Linguistics, Chulalongkorn University, Bangkok, January 8-10, 1992, pp. 651-656.
- Ladefoged, Peter and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford, UK; Cambridge, MA: Blackwell Publishing.
- Li, Fang-kuei. 1954. Consonant clusters in Thai. Language 30:368-79.
- ____. 1960. A tentative classification of Tai dialects. In S. Diamond, ed, *Culture in history:* Essays in honor of Paul Radin, pp. 951-959. New York: Columbia University Press.
- Li, Fang-Kuei. 1977. *Handbook of Comparative Tai*. Honolulu: University of Hawai'i Press.
- Luo, Yongxian. 1997. *The subgroup structure of the Tai languages: A historical-comparative study*. Berkeley: Project on Linguistic Analysis, University of California, Berkeley.
- _____. 1999. A dictionary of Dehong, Southwest China. Canberra: Pacific Linguistics. Morey, Stephen. 2005. The Tai languages of Assam- A grammar and texts. Canberra: Pacific Linguistics.
- Ostapirat, Weera. 2007. Proto-Tai velars and postvelars. In J. G. Harris, S. Burusphat and J. E. Harris, eds, *Studies in Tai and Southeast Asian Linguistics*, pp. 293-300. Bangkok: Ekphimthai Ltd.
- Petsuk, Rasi. 1978. *General Chracteristics of the Khün language*. M.A. Thesis. Department of Linguistics, Institute of Language and Culture for Rural Development, Mahidol University. Nakhorn Pathom.
- Phothiban, Phuthachart พุทธชาติ โปธิบาล and Thananan ธนานันท์ ตรงดี Trongdi. 1998. Sathana Khong Phasa Tak Bai Nai Phasa Thai Thin สถานะของภาษาตากใบในภาษาไทยถิ่น [The Position of Tak Bai Tai in Tai Dialects (Phase 2)]. Research report. คณะมนุษยศาสตร์ และสังคมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี Faculty of Humanties, Prince of Songkhla University.
- Pittayaporn, Pittayawat. 2005. *Phu Thai Kapong*. fieldnote. Cornell University.
- _____. to appear. Establishing relative chronology of Palaung sound changes using Tai loanwords. *Mon-Khmer Studies* 39:
- Pulleyblank, Edwin G. 1991. Lexicon of reconstructed pronunciation in early Middle Chinese, late Middle Chinese, and early Mandarin. Vancouver: University of British Columbia Press.
- Rungrueangsri, Udom อุคม รุ่งเรื่องศรี. 1991. Photchananukrom Lanna Thai Chabap Mae Fa Luang พจนานุกรมล้านนา-ไทย ฉบับแม่ฟ้าหลวง [Mae Fah Luang's Lanna-Thai Dictionary]. Bangkok: มูลนิธิแม่ฟ้าหลวง [Mae Fah Luang Foundation].
- Sarawit, Mary E. 1973. *The Proto-Tai vowel system*. Ph.D. dissertation. University of Michigan. Ann Arbor.
- The Royal Institute ราชบัณฑิตยสถาน. 2003. Photchananukrom Chabap Ratchabandit Sathan Pho So 2542 พจนานุกรมฉบับราชบัณฑิตสถาน พ.ศ. ๒๕๔๕ [The Royal Institute's Dictionary 2002]. Bangkok: สำนักพิมพ์นานมี [Nanmee Book Publications].

- Thisaphong, Arilak อารีลักษณ์ ทิสาพงศ์. 1985. Rabop Siang Phasa Ka Loeng Thi Ban Dong Mafai Changwat Sakon Nakhon ระบบเสียงภาษากะเลิง ที่บ้านคงมะไฟ จังหวัดสกลนคร [The Phonological description of Kaloeng at Ban Dong Mafai, Sakonnakhorn Province]. M.A. Thesis. สถาบันวิจัยภาษาและวัฒนธรรมเพื่อการพัฒนาชนบท Department of Linguistics, Institute of Language and Culture for Rural Development, มหาวิทยาลัย มหิดล [Mahidol University]. Nakhorn Pathom นครปฐม.
- Trongrat, Phutphong. 1998. A Comparative Phonological Study of Thai Phuan at Suphanburi and Sukhothai Provinces. M.A. Thesis. Department of Linguistics, Institute of Language and Culture for Rural Development, Mahidol. Nakhorn Pathom.
- Umar, Umaiyah Haji. 2003. A Classification of Thai-Kedah dialects using phonological characteristics. Bangkok: Nutcha Printing Limited Partnership.
- Wayland, Ratree and Susan G. Guion. 2005. Sound changes following the loss of /r/ in Khmer: a new tonogenetic mechanism. *Mon-Khmer Studies* 35:55-82.
- Wichasin, Ranoo. 1986. *Tai Ahom paleography*. M.A. Thesis. Department of Oriental Languages, Silpakorn Univerity. Bangkok.
- Wichiankhiaw, Arunrat อรุณรัตน์ วิเชียรเขียว. et al. 1996. The Northern Thai Dictionary of Palm-Leafs Manuscripts พจนานุกรมศัพท์ถ้านนา ฉบับเฉพาะคำที่ปรากฏในคัมภีร์ใบลาน. 2nd. Chiangmai, Thailand: Samnakphim Tratsawin.
- Xing, Gongwan 邢公畹. 2000. Hónghé Shàngyóu Dǎiyǎ yǔ 红河上游傣雅语 [Tai Ya of the Upper Reach of the Red River]. Běijīng 北京: 语文出版社 [Language and Literature Publishing].

COMPLEMENTIZERS AND VERB CLASSIFICATION IN THAI¹

Amara Prasithrathsint²
Chulalongkorn University
<Amara.Pr@chula.ac.th>

ABSTRACT

Despite the important role complementizers play in forming complex sentences, a review of literature shows that there has been no study focusing particularly on complementizers in Thai. Most studies treat them as elements related to such phenomena as parts of speech, embedding, subordinate clauses, etc. Their syntactic behavior is still obscure. Moreover, there is dispute as to what should be analyzed as complementizers and what conditions their occurrence. This study thus aims to identify complementizers in Thai and find out how they are related to verb classification.

Based on an approximately three-million-word corpus of current Standard Thai, the findings show that there are three complementizers in Standard Thai: *thîi*, *wâa*, and *hây*. They are grammaticalized from the noun *thîi* 'place', the verb *wâa* 'to say', and the verb *hây* 'to give', respectively. The occurrence of each of the complementizers depends on the type of verb in the matrix clause. The complementizer *thîi* occurs with **emotive** verbs, such as *dii-cay* 'to be glad', *sĭa-cay* 'to be sad', *palàat-cay* 'to be surprised', whereas *wâa* is found with **communicative** verbs, such as *phûut* 'to say', *athíbaay* 'to explain', *thǎam* 'to ask'. As for *hây*, it occurs with **directive** verbs, such as *sày* 'to order', *môɔp-mǎay* 'to assign', *rîak-ròɔŋ* 'to demand'.

The findings of this study contribute not only to the universal theory of complementation but also to the understanding of the cognitive meanings of complementizers and verb classes in Thai, which can be applied to teaching Thai as a foreign language.

1. Introduction

A complementizer is a function word that marks a complement clause, which is an argument of a matrix verb; for example, that in I think that he will leave tomorrow is a complementizer, and that he will leave tomorrow is a complement clause, which is an argument of the matrix verb think.

The term "complementiser" was apparently used for the first time by Rosenbaum (1967) when he referred to such a construction as *It surprised me that you called last night*, which was claimed to derive from the underlying structure *It you called last night*

The original version of this paper entitled "the Role of Complementizers in Verb Classification in Thai" was presented at the 17th Annual Meeting of the Southeast Asian Linguistics Society (SEALS XVII), August 31-September 2, 2007, University of Maryland, College Park, Maryland, U.S.A. I am grateful to Jerold A. Edmondson, Mathias Jenny, George Bedell, and Kiyoko Takahashi for their valuable comments, which helped me improve this paper.

² Professor Emeritus, Chulalongkorn University, Bangkok, Thailand.

surprised me by extraposition of you called last night and insertion of that complementizer. In case there is no extraposition, we can apply the rules "it" deletion and "that" insertion. The result will be That you called last night surprised me. The clause that you called last night, which functions either as the subject or the complement of the verb surprised, is normally labelled a complement clause or sentential complement.

Studies on complementation and complementizers in several languages reveal that it is typical to have more than one complementizer. In English, for instance, eight complementizers are generally accepted, with *that* is the best-known complementizer accepted by everyone. In addition to *that*, *for-to* and *poss-ing* (gerund) are also regarded as complementizers in Transformational-Generative Grammar (see Akmajian 1977). Five more are also added to the list (see Dirven 1989:113-139, Horie 2000: 12); i.e., *to* (with infinitive), *zero* (bare infinitive), *-ing* (participle), *whether/if*, and *wh-word*. The following are examples of how the eight complementizers in English are used. Note that the + sign means 'occurring with'.

- (1) to (+infinitive): It's easy to park your car here.
- (2) for-to (+NP, infinitive): It's important for us to get there first.
- (3) zero (bare infinitive): I saw him enter the building
- (4) -ing (participle): I saw him entering the building
- (5) *poss-ing* (gerund): I always enjoyed his/him singing that song.
- (6) that (+clause): I saw that he entered the building
- (7) *whether/if* (+clause): I don't know whether he entered the building.
- (8) wh-word (+clause): I don't know who entered the building

Besides English, variation in the use of complementizers is also found in other languages, such as French (Archard 1998, Tasmowski 1989), German (Cinque 1989), Dutch (Besten and Rutten 1989, Koster 1989), Spanish (Delbecque 2000), Portuguese (Quicoli 1982), Japanese (Horie 2000, Suzuki 2000), and Korean (Horie 2000).

Certain studies even arrive at a further step so as to show how complementizers and verb categorization are related. In French, for example, it has been found that the choice of a complementizer depends on the class of verb it occurs with. Archard (1998: 93-94) found that two classes of verbs in French related to infinitival complement clauses; they are **perception verbs**, and **causation verbs**, as in (9) and (10), respectively. Other classes of verbs--verbs of volition, verbs of declaration, and verbs of thoughts and beliefs are compatible with *que* (*that*)-complement clauses, as in (11), (12), (13).

- (9) J'ai *vu* **partir** Paul. (perception) 'I saw Paul leave.'
- (10) Marie *fait* **danser** Paul. (causation) 'Mary makes Paul dance.'
- (11) Je *veux* **que** je sois en mesure de partir dès demain. (volition) 'I want that I be in a position to leave as early as tomorrow.'
- (12) Paul *dit* **que** Marie est belle. (declaration) 'Paul says that Mary is beautiful.'
- (13) Je *pense* **que** J'ai compris le problème. (thought) 'I think I understand the problem.'

Unlike English and other well-known languages mentioned above, Thai has been dealt with very little as far as complementizers are concerned. There is no study that concentrates specifically on complementizers in Thai. They are either ignored or mentioned only as a side-line phenomenon in grammar books. In the recently published Thai grammar textbooks, Higbie & Thinsan (2003) and Iwasaki & Ingkaphirom (2005), three complementizers are recognized: *wâa*, *thîi*, and *thîi-wâa*, as in examples (14)-(16) (from Higbie & Thinsan 2003: 84-85). Note also that Higbie & Thinsan say that the three complementizers are interchangeable.

- (14)man mây thîi chăn rák khăw cin **COMP** him not true I love it 'It is not true that I love him'
- (15)wâa chăn rák khăw man mây ciŋ true **COMP** Ι him it not love 'It is not true that I love him'
- (16)rák khǎw man mây ciŋ thîi- wâa chăn it not true COMP Ι love him 'It is not true that I love him'

Apart from identifying three complementizers as shown in (14)-(16), Iwasaki & Ingkaphirom (2005: 255-268) also classify verbs according to the complementizers.

- Verbs showing **evaluation** and **emotion** take *thîi* complementizer, as *dii-cay* 'to be glad' in (17).
- Verbs of **speech** and **cognition**, such as *phûut* 'to say', *khít* 'to think', take *wâa*, as in (18) and (19).

- (17) phóm **dii-cay** <u>thîi</u> dây phób khun I (m.) be glad COMP get meet you 'I am glad to meet you.'
- (18) khăw **phûut** <u>wâa</u> khun pen khon dii he say COMP you be person good 'He says that you are a good person.'
- (19) khǎw **khít** <u>wâa</u> khun pen khon dii he think COMP you be person good 'He thinks that you are a good person.'

As can be seen above, studies show that several types of complementizers exist in various languages and that they seem to relate to different classes of verbs. In Thai, three complementizers are mentioned, although $h\hat{a}y$ has not been recognized as such. Also, there has been some attempt to sub-categorize verbs in Thai according to the complementizers they occur with, but without in-depth analysis.

It seems that a number of problems still remain with reference to complementizers in Thai, as expressed in the following questions.

- 1) Is $h\hat{a}y$, which is ignored by most linguists, a complementizer in Thai?
- 2) Is it possible to classify verbs according to the complementizers that occur with them?
- 3) How are the verb classes different semantically?
- 4) How are the complementizers different from one another? What are their cognitive meanings?
- 5) Is there any verb that is compatible with more than one complementizer? If so, does it have different meanings when it occurs with different complementizers?

Therefore, this study aims to identify complementizers in Thai and find out the conditions of their occurrence with verbs. The analysis is based on data of cooccurrences of complementizers and verbs taken from a three-million-word corpus of standard Thai. It is hypothesized here that there are three complementizers that mark complement clauses in Thai and that they have different cognitive values, which classify the verbs into three groups with different meanings.

2. Findings

2.1 The three complementizers in Thai

Three basic complementizers are found in Thai: *hây, wâa*, and *thîi*. Their syntactic behavior is in accordance with that of all the complementizers found in other languages; i.e., they mark subordinate clauses which function as complements of the matrix verb.

It must be noted that, due to the limitation of the scope, this study does not analyze the zero complementizer and compound complementizers in Thai, such as *thîi-wâa*, even though they are not less important than the three complementizers dealt with in this paper.

The complementizer $h\hat{a}y$ is grammaticalized from the verb $h\hat{a}y$ 'to give'. Examples (20) and (21) show the occurrence of different $h\hat{a}y$ -words. The clause in square brackets [] is a complement clause. How can we be sure that it is a complement clause? Here I tested it by using this criterion: If the clause can be replaced by aray 'what', then it is clear that it functions as an object or complement of the verb in the main/matrix clause.

- (20) khảw **hây** năŋ-sii chăn he give book I 'he gave me a book.'
- (21) khăw sàŋ [**hây** chăn nâŋ loŋ] he order COMP I sit down 'He ordered me to sit down.'

The complementizer **wâa** is grammaticalized from the verb **wâa** 'to say', as in (22), in contrast with **wâa** in (23), which is a complementizer. Similarly, the complement clause in (23) could be replaced by **aray** 'what'.

- (22) khun **wâa** aray ná you say what PRT 'What did you say?'
- (23) chăn khít [*wâa* khun khuan yùt] I think COMP you should stop 'I think that you should stop.'

The complementizer **thîi** is grammaticalized from the noun **thîi** 'place'. In (24) **thîi** is a noun, but in (25) it is a complementizer. Again, the complement clause in (25) could also be replaced by **aray** 'what'.

- (24) chăn khəy maa *thîi* héŋ níi l w I used to come place CLS This already 'I have been to this place before.'
- (25) chăn sĭa-cay [*thîi* khun mây maa] I be sad COMP you not come 'I was sad that you did not show up.'

2.2 Classification of verbs based on complementizers in Thai

The result of the analysis of the co-occurrences of complementizers and verbs shows that the verbs that need sentential complements in Thai can be divided into three major classes: **directive** verbs, **communicative** verbs, and **emotive** verbs.

2.2.1 Directive verbs (hây-verbs)

Directive verbs are verbs referring to making people do what one wants. They occur with hây and can be divided into two types: 1) verbs of causation, e.g., pen-hèet 'to cause', plian'to cause something to change'; and 2) verbs of pressure, which have some sense of 'pressing' or 'force' at different levels, such as sàŋ 'to order', rîak-ròoŋ 'to demand', môop-mǎay 'to assign'. (See the appendix for a list of hây-verbs.). For example:

- (26) khảw **plìan** <u>hây</u> thúk-yàaŋ dii khîn he change COMP everything good rise 'He changed to make everything better.'
- (27) prachachon **rîak-ròɔŋ** <u>hây</u> nayók laa-òɔk people demand COMP prime minister resign 'The people demand that the Prime Minister resign.'

2.2.3 Communicative verbs (<u>wâa-verbs</u>)

Communicative verbs refer to expressing or communicating ideas, information, and opinions. They occur with the complementizer **wîa**, which signifies **quotation**. The complement clause with **wîa** is, therefore, like something quoted. These class of verbs are divided into six types: 1) **verbs of saying**, such as **phûut** 'to say', **athîbaay** 'to explain', **thăam** 'to ask'³; 2) **verbs of thinking and cognition**, such as **khít** 'think', **rúu** 'to know'; 3) **verbs of feeling**, such as **klua** 'fear', **mîn-cay** 'to be confident'; 4) **verbs of seeing**, e.g. **duu** 'to look at', **càp-taa** 'to watch'; 5) **verbs of hearing**, e.g. **dây-yin** 'to hear'; 6) **verbs of acting**, e.g., **sĕe-sɛ̃eŋ**'to pretend'. (See a list of **wîa-verbs** in the appendix.). For example:

- (28) khǎw **phûut** <u>wâa</u> thúk-yàaŋ dii khên léεw he speak COMP everything good rise already 'He said that everything had been better (improved).'
- (29) khǎw **rúu** <u>wâa</u> thúk-yàaŋ dii khɨn léεw he know COMP everything good rise already 'He knew that everything had been better (improved).'
- (30) khăw **mân-cay** <u>wâa</u> thúk-yàaŋ dii khɨn léεw he was confident COMP everything good rise already 'He was confident that everything had been better (improved).'
- (31) khăw **hěn** <u>wâa</u> thúk-yàaŋ dii khîn léɛw he see COMP everything good rise already 'He saw that everything had been better (improved).'

It should be noted that a <u>wâa</u> complement clause that occurs with the verb *thăam* 'to ask' must end with a question word, e.g. <u>măy</u>, <u>rɨɨ mây</u> 'or not', <u>chây-măy</u> 'is it true?' e.g., <u>khăw **thăam** <u>wâa</u> chăn câ pay măy. 'He asked whether I would go.' In other words, <u>wâa</u>... **question word** equals whether in English. In Thai, there is no word which is an exact equivalent to 'whether' in English.</u>

-

- (32) khǎw **dây-yin** <u>wâa</u> thúk-yàaŋ dii khɨn léεw he hear COMP everything good rise already 'He heard that everything had been better (improved).'
- (33) khăw sěe-sêeŋ <u>wâa</u> thúk-yàaŋ dii khên léew he pretend COMP everything good rise already 'He pretended that everything had been better (improved).'

2.2.3 Emotive verbs (thîi-verbs)

Emotive verbs concern emotions or mental activities. They are divided into four types: 1) **verbs of feeling**, such as *dii-cay* 'to be glad', *sĭa-cay* 'to be sad', *palàat-cay* 'to be surprised', 2) **verbs of intention**, e.g. *tâŋ-cay* 'to intend', *wăŋ* 'to hope', 3) **verbs of mental character**, e.g. *ciŋ-cay* 'to be sincere', *cay-yen* 'to be cool-hearted', 4) **verbs of possibility** (assessed in the mind), e.g. *pen-pay-dâay* 'to be possible', *sadùak* 'to be convenient'. Note that all the verbs with *cay* 'heart' that need a sentential component fall into this category. (See the appendix for a list of *thîi-verbs*.). For example:

- (34) khăw **dii-cay** <u>thîi</u> chăn sòop phàan he be pleased COMP I take exam pass 'He was pleased that I passed the exam.'
- (35)khăw tân-cay thîi cà bàak chăn kòon khon ìin he intend COMP will tell I before people other 'He intended to tell me before other people.'
- (36) khảw **cay-yen** mâak <u>thấi</u> roo chẳn sǒon chuâ-moon he cool-hearted very COMP wait I two hours. 'He was so cool-hearted that he waited for me for two hours.'
- (37) **pen-pay-dâay** <u>thîi</u> chăn cà sòop phàan possible COMP I will take exam pass 'It is possible that I will pass the exam.'

Comparing the findings of the present study to what is stated in Iwasaki & Ingkaphirom (2005: 255-268), I see some agreement, such that verbs showing **emotion** take *thîi* complementizer, as *dii-cay* 'to be glad'. However, the statement in Iwasaki & Ingkaphirom (2005: 255-268) that verbs of **speech** and **cognition**, such as *phûut* 'to say', *khít* 'to think', take *wâa* complementizer does not seem to coincide with what is found in this paper. Here, I label the *wâa*-verbs as "communicative verbs" rather than verbs of speech and cognition because these verbs can occur with all the three complementizers with the denotation of speech and cognition. In other words, they do not exclusively occur with *wâa*. What is interesting is that when they do occur with *wâa* they specifically signify communication. That is why I name them "communicative verbs."

2.3 Overlapping of the three classes of verbs

It is found that certain verbs can occur with more than one complementizer. However, identical verb forms with different complementizers can be shown to have different meanings. It is because each complementizer has its own cognitive meaning, which influences the meaning of the verb in the main clause. For example:

- (38 a) khăw **phûut** <u>hây</u> khun pen khon dii he speak COMP you be person good 'He spoke (persuaded) for you to be a good person.
- (38 b) khăw **phûut** <u>wâa</u> khun pen khon dii he speak COMP you be person good 'He said that you were a good person.'
- (39a) khǎw wăŋ wâa (khǎw) cà tèη-ηaan kàp thəə he hope COMP he will marry with her. 'He to marry her.' hopes
- (39b) khăw **wăŋ** <u>thîi</u> cà tèŋ-ŋaan kàp thəə he hope will go marry with her 'He hopes to marry her.' (with strong intention)

As can be seen in (38a) and (38b), the verb *phûut* 'to speak, to say' can occur with either *hây* or *wâa*, but its meaning varies according to the complementizer it occurs with. With *hây*, *phûut* connotes the meaning of 'force', i.e., 'speak persuasively', whereas *phûut* wâa suggests the meaning of 'his words being quoted to inform others.' The difference between (39a) and (39b) is also due to the choice of a complementizer. When used with wâa the verb wǎŋ 'to hope' suggests the meaning of informing or communicating what the person hopes, but with *thîi*, the verb wǎŋ has the connotation of intention.

2.4 Function as the sentential subject of complement clauses in Thai

Universally, a complement clause can function as subject of a sentence, as in (40). However, the result of the analysis shows that only **thîi**-complement clauses can occur in the position of the subject, as in (41) and (42). The ones marked by **hây** and **wâa** cannot function as subjects, as in the ungrammatical sentences (43) and (44).

- (40) [That she won a scholarship] really satisfied me.
- (41) [<u>thîi</u> chăn sòop phàan] tham-hây khăw dii-cay COMP I take exam pass CAUSATIVE he be pleased 'That I passed the exam pleased him.'
- (42) [<u>thîi</u> khăw khayăn] pen thîi phoo-cay khŏon thúk khon COMP he be diligent be that be satisfied of every person 'That he/she is diligent satisfies everyone.'

- *[wâa thúk-yàan dii (43)khîn léew] pen sìn thîi khăw rúu COMP everything good rise already thing that he know be 'That everything has been better is what he knows.'
- (44) *[<u>hây</u> thúk-khon yiin khîn] pen sìŋ thîi khruu sàŋ COMP everyone stand rise be thing that teacher order 'For everyone to stand up is ordered by the teacher.'

It should be pointed out here that (43) and (44) will be grammatical if we add *thîi* in front of *wâa* and *hây*, as shown in (45) and (46), which are grammatical. The reason why complement clauses with *wâa* and *hây* cannot function as subjects may be related to the fact that these two function words are grammaticalized from verbs, unlike *thîi*, which is derived from a noun. However, further research is needed in order to fully explain such constraints.

- (45) [thî wâa thúk-yàaŋ dii khîn léɛw] pen sìŋ thîik khǎwrúu COMP everything good rise already be thing that he know 'That everything has been better is what he knows.'
- (46) [thîi hây thúk-khon yiin khin] pen sìn thìi khruu sàn COMP everyone stand rise be thing that teacher order 'For everyone to stand up is ordered by the teacher.'

3. Conclusion

This study provides new knowledge of complementizers in Thai. Three complementizers in Thai, hây, wâa, thîi, are found to occur with different verbs. In other words, they classify verbs into three main classes: directive, communicative, and emotive verbs. The verbs are not exclusively divided. Some can actually occur with more than one complementizer. However, they signify different meanings conditioned by the meaning of each complementizer; i.e., hây suggests force or pressing, wâa suggests something quoted, and thîi implies the meaning of mental state. Further studies need to be done on the constraints of the functions and syntactic characteristics of a complement clause with each of the three complementizers.

References

- Akmajian. A. A. 1977. The complement structure of perception verbs in an autonomous syntax framework. In P. W. Culicover, T. Wasow and A. Akmajian. *Formal Syntax*. New York: Academic Press.
- Archard, Michel. 1998. Representation of Cognitive Structures. Syntax and Semantics of French Sentential Complements. Berlin, New York: Mouton de Gruyter.
- Besten, Hans den and Jean Rutten. 1989. On verb raising, extraposition, and free word order in Dutch. In Jaspers, Dany, Wim Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest. Dordrecht, Holland and Providence, U.S.A.: Foris Publications, 21-40.

Cinque, Guglielmo. 1989. On embedded verb second clauses and ergativity in German. In Jaspers, Dany, Wim Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest. Dordrecht, Holland and Providence, U.S.A.: Foris Publications, 77-96.

- Delbecque, Nicole. 2000. Cognitive constraints on complement clause cliticization in Spanish. In Horie, Kaoru (ed.). 2000. *Complementation: Cognitive and Functional Perspectives*. Amsterdam: John Benjamins Publishing Co., 149-198.
- Dirven, René. 1989. A cognitive perspective on complementation. In Jaspers, Dany, Wim
- Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. *Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest*. Dordrecht, Holland and Providence, U.S.A.: Foris Publications, 113-139.
- Higbie, James and Snea Thinsan. 2003. *Thai Reference Grammar*. The Structure of Spoken Thai. Bangkok: Orchid Press.
- Horie, Kaoru (ed.). 2000. *Complementation: Cognitive and Functional Perspectives*. Amsterdam: John Benjamins Publishing Co.
- Horie, Kaoru. 2000. Complementation in Japanese and Korean. In Horie, Kaoru (ed.). *Complementation: Cognitive and Functional Perspectives*. Amsterdam: John Benjamins Publishing Co., 11-32.
- Iwasaki and Preeya Ingkaphirom. 2005. *A Reference Grammar of Thai*. Cambridge: Cambridge University Press.
- Jaspers, Dany, Wim Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest. Dordrecht, Holland and Providence, U.S.A.: Foris Publications.
- Koster, Jan. 1989. Left-right asymmetries in the Dutch complementizer system. In Jaspers, Dany, Wim Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. *Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest.* Dordrecht, Holland and Providence, U.S.A.: Foris Publications, 271-282.
- Quicoli, Antonio Carlos. 1982. *The Structure of Complementation*. Ghent-Antwerp-Brussels: Scientific Publishers.
- Rosenbaum, Peter S. (1967). The grammar of English predicate complement constructions. Cambridge, MA: MIT Press.
- Suzuki, Satoko. 2000. De dicto complementation in Japanese. In Horie, Kaoru (ed.). *Complementation: Cognitive and Functional Perspectives*. Amsterdam: John Benjamins Publishing Co., 33-58.
- Tasmowski, Liliane. 1989. A look at sembler+infinitive from different angles. In Jaspers, Dany, Wim Klooster, Yvan Putseys, and Pieter Seuren (eds.). 1989. *Sentential Complementation and the Lexicon. Studies in Honour of Wim de Geest.* Dordrecht, Holland and Providence, U.S.A.: Foris Publications, 403-422.

APPENDIX: List of verbs, their denotative meanings, and their sub-categories (in terms of their cognitive meanings)

| <i>hây</i> VERB | MEANING | ТҮРЕ | จงใจ | to intend | neoggueo. |
|--------------------|-------------------------|-----------|-------------------|------------------------|-----------|
| กระพื่อ | | causation | ขัง เข จัด | to arrange | pressure |
| ก่อ | to spread out | causation | งัคสรร | to allocate | pressure |
| ทย เกลี่ย | to originate | | | | pressure |
| | to smooth off | causation | 10501 | to talk over | pressure |
| ฆ่า | to kill | causation | ชวน ' | to persuade | pressure |
| คลบันคาล | to cause, inspire, | causation | ช่วย | to help | pressure |
| | motivate | | ช่วยเหลือ | to assist | pressure |
| แต่ง | to decorate, | causation | ชักชวน • | to induce | pressure |
| | compose | | ชำระ ชี้ | to purify | pressure |
| เป็นผล | to result | causation | | to point | pressure |
| เป็นเหตุ | to cause | causation | ชี้แจง | to explain | pressure |
| เป็นโอกาส | to give chance | causation | រ ើ លូ | to invite | pressure |
| เปลี่ยน | to change | causation | เชิญชวน | to persuade and invite | pressure |
| เปิด | to open | causation | โชว์ | to show | pressure |
| เปิดโอกาส | to give an | causation | ดัน | to push | pressure |
| | opportunity | | คำเนินการ | to operate, | pressure |
| ส่งผล | to result | causation | | conduct | pressure |
| สร้าง | to create | causation | ตรวจ | to examine | pressure |
| เสก | to pronounce | causation | ต้องการ | to want | pressure |
| กคคัน | to press | pressure | ตัดสิน | to decide | pressure |
| กระตุ้น | to urge | pressure | ตัดสินใจ | to make up | pressure |
| กำชับ | to insist | pressure | | one's mind | 1 |
| กำหนด | to set, to fix | pressure | ติ | to criticize | pressure |
| เกณฑ์ | to compel | pressure | เติม | to add | pressure |
| เกี่ยง | to let others do | pressure | เตือน | to warn | pressure |
| เกื้อหนุน | to support | pressure | ทน | to be patient | pressure |
| ขอ | to ask for, to | pressure | ทบทวน | to review | pressure |
| | beg | | ทอดทิ้ง | to abandon | pressure |
| ขอพร | to ask for blessings | pressure | ท้า | to do | pressure |
| ขอร้อง | to request | pressure | ท้าทาย | to challenge | pressure |
| ขัด | to obstruct | pressure | ทิ้ง | to throw away | pressure |
| ขับ | to drive away | pressure | แนะนำ | to advice | pressure |
| ขุน | to feed | pressure | บ่น | to complain | pressure |
| า คิด | to think | pressure | บนบาน | to pledge | pressure |
| เคี่ยวเข็ญ | to force | pressure | บริการ | to serve | pressure |
| J | repeatedly | | บังคับ | to force | pressure |
| ง้อ | to conciliate | pressure | บัญญัติ | to prescribe | pressure |
| | | | บันคาล | to cause, create | pressure |

| บำรุง | to support | pressure |
|------------|-------------------|----------|
| โบกมือ | to wave | pressure |
| ปฏิบัติ | to perform | pressure |
| ปรบมือ | to applaud | pressure |
| ประกาศ | to announce | pressure |
| ประทาน | to give | pressure |
| ประยุกต์ | to apply | pressure |
| ประสงค์ | to want | pressure |
| ปรับ | to adjust | pressure |
| ปรากฏ | to appear | pressure |
| ปรารถนา | to desire | pressure |
| ปรารภ | to mention | pressure |
| ปลคปล่อย | to release | pressure |
| ปล่อย | to let go | pressure |
| ปลุก | to arouse | pressure |
| ปลุกระคม | to stir up | pressure |
| ปลูกฝัง | to establish | pressure |
| ผลักดัน | to push | pressure |
| เพยแพร่ | to spread out | pressure |
| ฝึก | to train | pressure |
| พยายาม | to attempt | pressure |
| พัฒนา | to develop | pressure |
| พิจารณา | to consider | pressure |
| พิสูจน์ | to prove | pressure |
| ฟ้อง | to accuse, charge | pressure |
| ภาวนา | to pray | pressure |
| ทอก | to give | pressure |
| มอบหมาย | to assign | pressure |
| มีมติ | to resolve | pressure |
| มุ่งหมาย | to aim for | pressure |
| ยกย่อง | to praise | pressure |
| ถอท | to let, yield to | pressure |
| ย้ำ | to emphasize | pressure |
| ព្ | to incite | pressure |
| รบกวน | to bother, annoy | pressure |
| 50 | to wait | pressure |
| ร้อง | to cry | pressure |
| รักษา | to watch over | pressure |
| รับเกียรติ | to be honored | pressure |
| รายงาน | to report | pressure |
| รายงานให้ | to report, | pressure |
| ทราบ | inform | |
| เร่ง | to accelerate | pressure |
| | | |

| เรียก | to call | pressure |
|-------------|--------------------|----------|
| เรียกร้อง | to demand | pressure |
| เล่า | to tell | pressure |
| วอน | to implore | pressure |
| วาน | to ask | pressure |
| วิงวอน | to ask, request | pressure |
| ส่งเสริม | to promote | pressure |
| สนับสนุน | to encourage | pressure |
| สอน | to teach | pressure |
| สะท้อน | to reflect | pressure |
| สั่ง | to order, command | pressure |
| สาธิต | to illustrate | pressure |
| สำทับ | to stress | pressure |
| เสนอ | to offer | pressure |
| เสริม | to add | pressure |
| เสี้ยมสอน | to teach | pressure |
| | persuasively | |
| ห้าม | to forbid | pressure |
| เห็นควร | to approve | pressure |
| เห็นชอบ | to approve | pressure |
| เหนี่ยวรั้ง | to stop, prevent | pressure |
| อธิบาย | to explain | pressure |
| อธิษฐาน | to wish | pressure |
| อนุญาต | to allow | pressure |
| อนุมัติ | to approve, permit | pressure |
| อนุโลม | to compromise | pressure |
| อยาก | to want, desire | pressure |
| อวยพร | to bless | pressure |
| อ้อนวอน | to plead | pressure |
| | | |

| wâa | | | | | |
|------------------|----------------------|-----------|-------------|--------------------|--------|
| VERB | MEANING | TYPE | ตอบ | to reply | saying |
| เสแสร้ง | to pretend | action | ต่อว่า | to blame | saying |
| กลับกลาย | to become | action | ตะ โกน | to shout | saying |
| กลายเป็น | to become | action | ตั้งคำถาม | to question | saying |
| แข่งขัน | to compete | action | ตั้งชื่อ | to name | saying |
| กลัว | to be afraid | feeling | ตำหนิ | to criticize | saying |
| เกรง | to fear | feeling | ติง | to oppose | saying |
| คำนึง | to think | feeling | ตีความ | to interpret | saying |
| แน่ใจ | to be sure | feeling | เตือน | to warn | saying |
| แปลกใจ | to be surprised | feeling | ใต่สวน | to investigate | saying |
| ภูมิใจ | to be proud | feeling | ถาม | to ask | saying |
| มั่นใจ | to be confident | feeling | ถือ | to regard | saying |
| รู้สึก | to feel | feeling | เถียง | to argue | saying |
| หวัง | to hope | feeling | ทบทวน | to revise | saying |
| หวั่นเกรง | to fear | feeling | ทาย | to guess | saying |
| อับอาย | to be ashamed | feeling | ทำนาย | to foretell | saying |
| ได้ยิน | to hear | hearing | ทึกทัก | to assume | saying |
| ฟัง | to listen | hearing | นับ | to count | saying |
| ได้ความ | to make sense | cognition | แนะ | to advise | saying |
| แปล | to interpret | cognition | บ่งบอก | to indicate | saying |
| หมายความ | to mean | cognition | บ่น | to complain | saying |
| กระซิบ | to whisper | saying | บรรยาย | to describe | saying |
| กล่าว | to say | saying | บอก | to tell | saying |
| กล่าวหา | to accuse | saying | บังคับ | to force | saying |
| กำชับ | to insist firmly | saying | บัญญัติ | to prescribe | saying |
| เก็ง | to guess | saying | ปลอบใจ | to console | saying |
| ขอพร | to ask for | saying | เปรียบเทียบ | to compare | saying |
| νq | blessings | | เปรียบว่า | to compare | saying |
| เข้าใจ | to understand | saying | เปิดประเด็น | to start the point | saying |
| เขียน • • | to write | saying | เปิดเผย | to disclose | saying |
| คัดค้าน | to object to | saying | พยากรณ์ | to predict | saying |
| คาด | to expect | saying | พร่ำพูด | to talk | saying |
| คาดหวัง | to expect | saying | A | continually | |
| คุย | to talk | saying | พิจารณา | to consider | saying |
| จารึก | to write, engrave | saying | พิสูจน์ | to prove | saying |
| แจ้ง | to inform | saying | พูด ม | to talk | saying |
| ง โจมตี | to attack | saying | ฟ้อง | to accuse, charge | saying |
| ชม . งมห | to praise | saying | ย้ำ | to stress | saying |
| ชี้แจง ข้ | to make clear | saying | ยืนยัน | to confirm | saying |
| ซักฟอก | to investigate | saying | รวบรัด | to cut short | saying |
| ซกพอก ตอกกลับ | to respond | saying | ร้อง | to cry | saying |
| аотныв | to respond | 54,1116 | 3 O N | | Suying |

| วะก่ | to specify | saying | อธิษฐาน | to wish | saying, thinking |
|--------------------------|----------------------------|------------------|---------------|-----------------|---------------------|
| ระแวง ราเวง | to suspect to accept | saying saying | จับตา | to watch | seeing |
| รับประกัน | to accept to guarantee | saying | ชัดเจน | to be clear | seeing |
| รับรอง | to guarantee to confirm | saying | ନ୍ | to look at | seeing |
| รายงาน | | saying | จ ดูเหมือน | to seem | seeing |
| ร เอง เน เรียก | to report to call | saying | ุง ตรวจหา | to investigate | seeing |
| เรียน | to can to inform | saying | ปรากฏ | to appear | seeing |
| เวอน ลงความเห็น | to morni to resolve | saying | พบ | to find | seeing |
| ถงท้าย ลงท้าย | to resolve to end | saying | เพ่งเล็ง | to focus | seeing |
| เล่า | to end to tell | saying | ทอง | to look at | seeing |
| ณา วิงวอน | to ask, request | saying | เห็น | to see | seeing |
| สมมติ | _ | | คิด | to think | thinking |
| นรัฦ แททผ | to suppose to conclude | saying | จำได้ | to recognize | thinking |
| สอน | to conclude to teach | saying | ฉุกคิด | to think | thinking |
| ก่อน สั่ง | to teach | saying | 9 | suddenly | 8 |
| ถง สัญญา | to command to promise | saying saying | ชั่งใจ | to weigh in the | thinking |
| แหูหู สันนิษฐาน | to presume | saying | ī | mind | |
| สาบาน | to swear | saying | เชื่อ | to believe | thinking |
| สารภาพ | to swear | saying | เชื่อถือ | to have faith | thinking |
| สำทับ | to stress | saying | เชื่อมั่น | to believe | thinking |
| เสนอ | to sifess to offer | saying | v | firmly | |
| เสริม | to add | saying | ตระหนัก | to realize | thinking |
| เถาม ให้เหตุผล | to add to give a reason | | ทราบ | to know | thinking |
| อธิบาย อธิบาย | • | saying | นับถือ | to respect | thinking |
| อภิปราย | to explain to discuss | saying | นึก | to think | thinking |
| ยมบว เย อวคอ้าง | to discuss to claim | saying | ฟัน | to dream | thinking |
| | | saying | ยอมรับ | to accept | thinking |
| อวยพร | to bless | saying | 9 9 | to know | thinking |
| ออกตัว | to defend | saying | วิเคราะห์ | to analyze | thinking |
| ออกปาก | to say | saying | วิตก | to worry | thinking |
| ออกอุบาย _พ | to trick | saying | สงสัย | to doubt | thinking |
| อ้าง | to claim | saying | สังเกต | to observe | thinking |
| อุทาน I | to exclaim | saying | หยั่งรู้ | to know | thinking |
| อุปมา ' | to compare | saying | | intuitively | |
| เอุ่ถ | to utter a word | saying | | | |
| แอบอ้าง | to claim | saying | | | |

| thîi | | | | | |
|-------------------|-----------------------|-----------|-----------------|----------------------------|--------------|
| VERB | MEANING | TYPE | ใจหาย | to be stunned | feeling |
| จริงใจ | to be sincere | character | ชัง | to hate | feeling |
| ใจกว้าง | to be generous | character | ชื่นใจ | to be delighted | feeling |
| ใจแข็ง | to be hard- | character | เฐา | to be depressed | feeling |
| | hearted | | ดีใจ | to be glad | feeling |
| ใจแคบ | to be narrow- | character | ตกใจ | to be frightened | feeling |
| ใจง่าย | minded to be easy | character | ตัดพ้อ | to express dissatisfaction | feeling |
| ใจคำ | to be merciless | character | ถอดใจ | to be | feeling |
| ใจดี | to be kind- | character | 6101111 | disillusioned | reening |
| | hearted | | ท้อใจ | to be | feeling |
| ใจเย็น | to be cool- | character | | discouraged | |
| a 9/ | hearted | | น้อยใจ | to feel to | feeling |
| ใจร้อน | to be hasty | character | d | belittled | |
| ใจร้าย | to be cruel | character | เบื่อ | to feel tired | feeling |
| ดึงดัน | to be obstinate | character | ประหลาดใจ | to be surprised | feeling |
| คำริ | to contemplate | feeling | ปรารถนา | to desire | feeling |
| ลืม | to forget | feeling | เป็นเกียรติ | to feel honored | feeling |
| กระคาก | to be shy, bashful | feeling | แปลกใจ | to be amazed | feeling |
| กลัว | to be afraid | feeling | ผิดหวัง | to be disappointed | feeling |
| กล้า | to be arraid | feeling | ฝังใจ | to be impressed | feeling |
| กลุ้มใจ | to be depressed | feeling | พร้อม | to be ready | feeling |
| เกรงใจ | to be | feeling | พอใจ | to be satisfied | feeling |
| 811971870 | considerate | recinig | เพียงพอ | to be contented | feeling |
| โกรธ | to be angry | feeling | ภูมิใจ | to be proud | feeling |
| ข้องใจ | to be dubious | feeling | มูนเง มั่นใจ | to be produ | feeling |
| ขอบใจ | to be grateful | feeling | ยินดี | to be pleased | feeling |
| ขัดใจ | to be annoyed | feeling | ลังเล | to hesitate | feeling |
| ขัดหู | to be | feeling | สนใจ | to be interested | feeling |
| • | displeasing to | Ç | สบายใจ | to be happy | feeling |
| | the ear | | สะใจ | to be fulfilled | feeling |
| แข็งใจ | to be | feeling | เสียใจ | to be sad | feeling |
| คร้าน | courageous | faalina | เสียคาย | to be regrettable | feeling |
| หวาน คับข้องใจ | to be lazy to feel | feeling | เสียแรง | to regret | feeling |
| HILIOAIA | distressed | feeling | ใส่ใจ | to be attentive | feeling |
| แคลงใจ | to be suspicious | feeling | หนักใจ | to be heavy- | feeling |
| จนใจ | to be unable to | feeling | | hearted | - |
| | think | | หวั่นเกรง | to be afraid of | feeling |
| จำใจ | to be forced to | feeling | หวั่นใจ | to be afraid | feeling |
| ತ | do | | อยาก | to want | feeling |
| เจ็บใจ | to be hurt | feeling | อ่อนใจ | to be | feeling |
| ใจเสีย | to be disheartened | feeling | | discouraged | |
| | uisiicai telleu | | เจตนา | to intend | intention |

| ตั้งใจ | to intend | intention | จำเป็น | to be necessary | possibility |
|----------|--------------|-----------|-----------|-----------------|-------------|
| ตัดสินใจ | to determine | intention | เป็นไปได้ | to be possible | possibility |
| มีแผน | to plan | intention | ผิด | to be wrong | possibility |
| มุ | to be | intention | พยายาม | to try | possibility |
| | determined | | ยาก | to be difficult | possibility |
| มุ่ง | to aim | intention | สมควร | to be | possibility |
| ยืนยัน | to confirm | intention | | appropriate | |
| รับ | to accept | intention | สะควก | to be | possibility |
| ลื๊อ | to rumor | intention | | convenient | |
| เลือก | to choose | intention | สามารถ | to be capable | possibility |
| หวัง | to hope | intention | เหมาะ | to be suitable | possibility |

VILLAGE NAMES IN CHIANG SAEN: AN ETHNOLINGUISTIC STUDY OF PLACE NAMES IN THAI¹

Sorabud Rungrojsuwan²
Mae Fah Luang University, Chiang Rai
<sorabud@yahoo.com>

Abstract

This paper has two main objectives: 1) to investigate the structures of village names from the use of linguistic—structures and meaning—and non-linguistic criteria—reasons of naming, and 2) to examine the development of communities in Chiang Saen District, Chiang Rai Province, Thailand and local cultural aspect—namely Nothern Thai Dialect—reflected from data about villages and their names. Data of 70 villages in Chiang Saen District were explored; documentary review together with interviewing technique were used for data elicitation. It was found that based on four main reasons of naming, the structure of village names in Chiang Saen mostly found contains two elements: a head and a modifier. The names are mainly distributed to the semantic domains [GEOGRAPHICAL LANDSCAPE] and [PLANT]. For community development, the division of Chiang Saen into three periods according to its political change could show the direction of village settlement and expansion. Moreover, for cultural aspect, it is obvious that the use of the local dialect—Northern Thai Dialect—plays an important role in the naming of Chiang Saen's villages.

1. Rationale

The issue about village names has been widely studied by Thai linguists (Chuaichuwong 1989, Kullavanijaya 1992, Thekhachunhathian and Buranaphatana 1993, Phuangchanhom 1996, Maneein 2000, Deephadung 2004). Past research suggested that, structurally, village names in Thai composed of two major components: head and modifier. Regardless of number of words, there can be more than one lexical item functioning as modifier follows the head as shown in the following examples.

kàat khîilèk **HEAD** MODIFIER.1

'market' 'a kind of plant' (Maneein, 2000)

¹ This study was supported by The National Research Council of Thailand (NRCT) during the fiscal year 2007. The author would like to express a sincere appreciation to Professor Emeritus Dr.Pranee Kullavnijaya—project advisor—for her valuable advises and comments.

² Lecturer, School of Liberal Arts, Mae Fah Luang University, Chiang Rai Province, Thailand, E-mail: sorabud@yahoo.com

Sorabud Rungrojsuwan. 2009. Village Names in Chiang Saen: An Ethnolinguistic Study of Place Names in Thai. *Journal of the Southeast Asian Linguistics Society* 2:161-173. Copyright vested in the author.

| nŏəŋ | klaaŋ | khôok | |
|---------|-----------|------------|---------------------|
| HEAD | MODIFIER1 | MODIFIER.2 | |
| 'creek' | 'central' | 'mound' | (Chuaichuwong 1989) |

Semantically, it was reported that the head element usually belongs to geographical landscape—such as, mountain and river—and surrounded environment—such as plants and animals.

| HEAD dooj | MODIFIER tâaj | |
|--------------|-------------------------|---------------------|
| 'mountain' | 'south' | (Maneein, 2000) |
| hûaj | hêeŋ | |
| 'brook' | 'dry' | (Deephadung 2004) |
| mákhǎam | thâw | |
| 'tamarind' | 'old' | (Deephadung 2004) |
| mên | jàj | |
| 'hedgehog' | 'big' | (Chuaichuwong 1989) |

It should be noted that the structural and semantic analyses of village names in the past research are mostly done upon the working definition which is purely linguistic especially the semantic part. Normally linguists directly translate the meanings of lexical elements of the particular names without considering other factors that might lead to misinterpretation. In other words, in naming a village such as /mêɛ kham/ (แม่คำ), for example, the lexical items /mɛ̂ɛ/ and /kham/ (คำ) are combined. Accordingly, the structure of name consists of /mɛ̂ɛ/ as head and /kham/ as modifier and the meanings of these two words—which are [RIVER] and [GOLD]—are directly analyzed. However, from primary survey, consider from the point of view of the local people who name their village, it was found that in some cases villagers do not pay attention to the basic meanings of all elements in village names but would rather view those names in relation to reasons of naming—for example, the history or the surrounded environment of the villages—which sometimes might not be semantically related to the names. In the case of "Mae Kham Village", for example, the name of the village was from Kham River (แม่น้ำคำ) which is the most important river in the village. Accordingly, in stead of the meaning [GOLD], the word /kham/ here would rather be viewed as a proper name of the river, because the village name is named after the name of the river and no one in the village recognize the relationship between the basic meaning of the word /kham/ which is [GOLD] and the river or the village itself—that is there was no gold in the river. There are some studies by nonlinguists which pay much attention to the reasons of naming (Yodying, J. 1976, Banchuen 1984, Sarnsamuth n.d.). However, those studies did not integrate linguistic analysis with the reasons, but rather be just a kind of nonacademic documentary works. From the above reason, the present study tries to employ the use of reasons of naming—which is nonlinguistic by nature—in structural and semantic analyses of village names.

In addition to linguistic, findings from the studies of village names have been extended to the explanation about the relationship between village names and cultures of the particular research sites (Chuaichuwong 1989, Kullavanijaya 1992, Thekhachunhathian and Buranaphatana 1993, Maneein 2000, Deephadung 2004) as shown in the following examples.

Village Names

Cultures Water

nšonkòk, wanthoon, sábua, hûajsaaj, bòojàj, náamsǎj

nn
n
o
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
o
r
<p

sa 'pool', huaj 'brook', bòo 'well', náam 'water'

(Chuaichuwong 1989)

bùpphǎaraam, phránɔɔn, fáahaam, sǐiphoothaaraam Religion

(names of temples)

(Maneein, 2000)

Although cultural explanation of the particular places will be clarified, other social aspects—such as history and community development—are ignored. This might be because reasons for the selection of research area are not clear such as, it is just because the researchers are working in the particular area and just want to investigate the names of villages and the reflection of those names on local cultures.

In this study, Chiang Saen District of Chiang Rai Province is purposively selected. It is because Chiang Saen is a dynamic area in terms of community settlement. It is said to be the first kingdom of Lanna and was colonized by Burma shortly at the beginning of Rattanakosin Period. People in Chiang Saen were all migrated to other parts of Thailand during the period of King Rama I. After that, during the period of King Rama V, civilians who lives nearby—in Chiang Mai and Lamphun Province—were moved and settled there. History shows that Chiang Saen was once a deserted town. The gradual movement and settlement of people from Chiang Mai and Lamphun during the period of King Rama V until present is an appropriate timing for the study of community development. It should be noted that the establishment of villages—which is a part of administrative decentralization of Thai government—also started from the period of King Rama V.

Therefore, in addition to linguistic, social aspects of village names will also be explored in order to not only help understand the concept of naming but also show how the investigation could provide valuable conclusion to social science.

2. Objectives

The present study aims to investigate:

- a) Structures and meaning of village names from the use of data about reasons of naming
- b) Community development and culture reflected from the information about villages and their names

3. Methodology

3.1 Scope of the study

This study investigate 70 village names in Chiang Sean District, Chiang Rai Province—the province to the northern most of Thailand.

3.2 *Data*

Data were gathered from documents, e-documents, and fieldwork of the author in various resources as follows:

- a) Books, textbooks, theses, and research papers concerning the study of village names in Thai
- b) Websites of six Sub-district Administrative Organizations in Chiang Saen District (Wiang, Yonok, Pasak, Sri Don Moon, Baan Saew, and Mae Ngen)
- c) Interviewing village headmen, assistant village headmen, members of Sub-district Administrative Organizations, and elderly people in the villages

2.3 Data analysis

The gathered data were analyzed into two major parts: language and culture. In relation to language, structural, semantic and reasons for naming were primarily analyzed as follows:

| <สันต้นเปา> | /sǎn | tôn paw/ | |
|-------------|--|-----------------------------------|--|
| Gloss | highland | a kind of tree (Shorea siamensis) | |
| Structure | head | modifier | |
| Semantic | [NATURE] | [NATURE] | |
| Reason | This village is situated on a highland with abandon of | | |
| | "Shorea siamensis" | | |

After the preliminary analysis, all data were put together and discussed in section 3.

In relation to culture, the aspects of community development and the influence of the local languages, such as Northern Thai Dialect were focused. In addition, in order to demonstrate the relationship between culture and names, maps were also used.

3. Structure of village names

In order to make clarification about the semantic which will further lead to the analysis in terms of structure of village names, the analysis about reasons for naming—which is nonlinguistic by nature—is done. Findings from such analysis, then, will be used for semantic analysis—the classification of words into different semantic domains according to the background knowledge of each village which helps make decision about word boundary and its possible meaning—and structural analysis—in what manner that component parts of a particular name are combined—of village names.

Consider the reasons for naming, it was found that village names in Chiang Saen were named upon four different reasons or sources.

A) Geographical landscape and surrounded natural environment of the villages

This type of naming was most found in Chiang Saen. That is 42 out of 70 villages used this criterion for naming as shown

| Names of river | /mɛ̂ɛ má/ | [river + Ma] | 'Ma River' |
|-------------------|--------------|---------------------|--------------------------|
| | /mêe ŋən/ | [river + Ngen] | 'Ngen River' |
| Names of mountain | /dɔj kham/ | [mountain + Kham] | 'Kham Mount.' |
| | /dɔj saŋóo/ | [mountain + Sa Ngo] | 'Sa Ngo Mount.' |
| Names of plants | /daaj/ | [daaj] | 'Celosia argentia Linn.' |
| | /ŋíw thâw/ | [old + Ngiw] | 'Old Bombax ceiba L.' |
| Landscapes | /dɔj ŋaam/ | [mountain + nice] | 'beautiful mountain' |
| | /khóoŋ ŋaam/ | [curve + beautiful] | 'beautiful curve' |

In relation to this finding, semantically, it was found that village names of this type belong to the [NATURE] and [CHARACTER] domains. The former is subclassified into four subcategories [GEOGRAPHICAL LANDSCAPE], [PLANT], [ANIMAL], and [OTHER NATURAL ENTITY], while the latter is divided into three subcategories: [ATTRIBUTE], [ACTION], and [POSITION]. It is especially for the case of words in [NATURE] domain, both common and proper nouns are used.

It addition, it was found that the meanings of some elements in the names cannot be taken directly because there seems to be no relationship between the semantic of those elements and what the villagers knew about. They are, for example, the cases of /mɛɛ ŋən/ and /dɔj kham/. The word /ŋən/ semantically means [money] and the word /kham/ originally means [gold]. However, according to the interview with the village headmen of the two villages, there is no explanation about [money] and [gold] in relation to the naming of the villages, but rather, names /ŋən/ and /kham/ are used in close connection to the river and the mountain located in the villages as their names, respectively. Accordingly, nonlinguistic data about the reasons of naming is used here as the supporting evidence for the determination of whether the particular elements in the names should be translated semantically or they should be taken as just proper names.

B) Important communities or religious places in the past

About ten percent of villages used the names of ancient communities or religious constructions such as, pagodas and temples as their names as shown

| Names of ancient communities | /wɨaŋ nɨa/ /wɨaŋ tâaj/ | 'northern city', 'southern city' |
|----------------------------------|---------------------------|---|
| Names of religious constructions | /cɔɔmkìttì/ /kùu tâw/ | 'Jom Kitti Temple', 'Ku Tao Pagoda', |
| | /nuumnccbiĭa/ | 'Sridonmoon Temple' |

In relation to semantics, lexical items used for this type of naming belong to the [HUMAN AND ARTIFACTS] domain which can be subclassified into three subcategories: [HUMAN AND OCCUPATION], [LIVING PLACES], and [BELIEF]

In this case, nonlinguistic findings show that some names are solely taken from the name of the ancient religious constructions which are well known and are respectful by the people in the communities. Accordingly, the elements in the name—although can be separated in different lexical items—should not be torn apart such as /cɔɔmkíttì/ 'Jom Kitti Temple', /kùu tâw/ 'Ku Tao Pagoda', and /sĭidɔɔnmuun/ 'Sridonmoon Temple'.

C) Exiting communities or villages

It was found that nine villages have been established from the expansion of some particular existing villages. "Sridonmoon Tai Village" (/sĭidɔɔnmun tâaj/), for example, was from "Sridonmoon Village". People in this village had been affiliated to "Sridonmoon Village" for many years before split to "Sridonmoon Tai Village" or 'southern Sridonmoon'. It should be noted that the villagers of the new village did not move to anywhere but it is just the split of a village into two villages politically. It can be observed that this type of naming mostly combine two lexical elements: the name of the existing village as head and the lexical item in [POSITION] subcategory as modifier (It is used to refer to the position of the new village in relation to the existing village.)

Semantically, villages named under this criterion employs two lexical elements: the names of existing villages and lexical items in [POSITION] subcategory. Accordingly, in relation to structural analysis, names of this type will be mostly analyzed as two combined elements.

D) Important event in the past

Nine village names were formed according to some important events of the particular villages in the past. Take "Wang Lao Village" (/wan laaw/) for example. The naming of /wan laaw/ is from an event once when there was a dead body of Lao soldier floating along the river and got stuck in the whirlpool—/wan/—of the village.

The meaning of lexical items used in this type of naming involve in all domains discussed so far: [NATURE], [HUMAN AND ARTIFACTS], and [CHARACTER]

With the understanding about the background of the naming in this category, the analysis in terms of structure can be properly made. Take "Wang Lao Village" for example. The name consists of two elements /wan/ and /laaw/. The word /wan/ can be taken as a homonym with two possible meanings: [palace] or [whirlpool] while the word /laaw/ can possible used to mean [Lao country], [Lao people], [Lao language], or [Lao's]. Without the nonlinguistic knowledge about the village, the possibility of misinterpretation can be easily occurred—in this case /wan/ means [whirlpool] and /laaw/ means [Lao people].

Findings from the above analyses which are both nonlinguistic (reasons for naming) and linguistic (semantic) parts, then, are used as the basis for the analysis of structures of village names as pointed out in A)-D) above.

In relation to structure, it is proposed that names, in this study, consist of "lexical element/s". Lexical elements are structurally defined as components which form village names. A lexical element contains either a lexical item or more than one lexical item according to the results from the analysis of reasons for naming and meanings (semantic) mentioned earlier. From the data, four major types of structure were found according to the

number of lexical elements. It should also be noted that the concepts of two subcategories of nouns namely "common nouns" and "proper nouns" can be helpful for the description in this section as shown.

3.1 Names with one lexical element

Nine names contain only one lexical element. They can be divided into two common names and six proper names as shown.

Common names

/thûn/ 'field'

/râj/ 'farm, vegetable garden'

Proper names

/dâaj/ 'Celosia argentia Linn' /coom kìttì/ 'Jomkitti Temple' 'Sridonmoon Temple' 'Sribunyuen Temple'

/pàa dèɛt/ 'Paa Daet Village in Chiang Mai Province'

/sǔan dòɔk/ 'Suan Dok Village in Lao'

3.2 Names with two lexical elements

Village names with two lexical elements contain the first lexical element functioning as head and the second lexical element functioning as modifier. The head can be common nouns, proper nouns, adjectives, and conjunction, while the modifier can be common nouns, proper nouns, adjectives, and prepositions as in the following examples.

Common noun + Proper noun /mêɛ kham/ 'river + Kham River'
Common noun + Common noun /thaan thoon/ 'water + gold'

Common noun + Adjective /dɔɔj ŋaam/ 'mountain + beautiful'

Common noun + Preposition /wian nia/ 'city + north'

Proper noun + Common noun /mêe kham kasèet/ 'Mae Kham Village + agriculture'

Proper noun + Adjective /chian sĕɛn nɔ́ɔj/ 'Chiang Saen City + small'
Proper noun + Preposition /sĭi dɔɔn muun nɨa/ 'Sridonmoon village + north'

Adjective + Adjective /khóoŋ ŋaam/ 'bendwise + beautiful'
Preposition + Common noun /hŭak wáan/ 'starting point + lake'

It should be noted that names with two lexical elements are most found. It consists of 56 from the total of 70 names.

3.3 Names with three lexical elements

There are only four names which contain three lexical elements. The first element is head while the second and the third elements are modifier. Heads can be either common or proper noun whereas modifiers might be the combination of two elements in followings: a common noun, a proper noun, and an action verb as shown.

Proper noun + Common noun + Proper noun

/mêe kham nɔɔŋ bua/

'Mae Kham Village + pond + lotus'

Proper noun + Preposition + Common noun

/pàa sàk hǎaŋ wɨaŋ/

'Paasak City + edge + city'

Common noun + Common noun + Action verb

/thûn fáa hâam/

'field + sky + to prohibit'

Common noun + Proper noun + Action verb

'pond + lotus + to surface'

3.4 Names with four lexical elements

It was found that only one village name contains four lexical elements: "San Sai Kong Ngam Village" (/săn saaj kɔɔŋ ŋaam/). This name consists of a common noun as head and two common noun together with an adjective as its modifiers as shown.

Common noun + Common noun + Adjective

/săn saaj koon naam/

'ridge + sand + pile + beautiful'

4. Social aspects of village names

Apart from linguistics, related information about the establishment of villages and languages used in the names could also shed light on the social aspect about community development and influence of local languages on the naming. In this section, data about year of establishment of villages will be used for the explanation about community development (This is said to be an extra nonlinguistic finding which could provide much more understanding about the dynamicity of Chiang Saen since the period of governmental reform during King Rama V) while languages used in the names will be examined in order to show the influence of local languages on the formation of official village names.

4.1 Community development

Chiang Saen has been colonized by Siam Kingdom in 1804 during King Rama I period. By that time about 20,000 civilians were evacuated to the cities nearby including Chiang Mai, Lampang, Nan, Lao, Ratchaburi, and Saraburi. The reemergence of communities in Chiang Saen has been started again nearly 100 years ago when King Rama V moved people from Chiang Mai and Lamphun to Chiang Saen (Kinawong 2003, Ongsakul 2003, Chumdee 2006, Chiang Saen District Summary Document 2007). The establishment of villages in Thailand has been done upon decentralization policy—the distribution of administrative power—during this period. From literature review, it was found that the status of Chiang Saen has been politically changed from the period of King Rama V until present for many times and can be divided into three critical periods as shown.

a) Period I (Before 1957)

This period started from the migration of people from Chiang Mai and Lamphun to Chiang Saen during King Rama V period to until 1956 when the central part of the higher northern Thai has been moved from Chiang Saen to Mae Chan District. Chiang Saen has been

changed from district to subdistrict consisting of three administrative areas: Wiang, Paasak, and Baan Saew

b) Period II (1957-1983)

The second period started when Chiang Saen has been reestablished as a district consisting of three subdistricts: Wiang, Paasak, and Baan Saew

c) Period III (1984-present)

The last period started when Chiang Saen split its subdistricts from three to six: Wiang, Paa Sak, Baan Saew, Yonok (split from Wiang), Sridonmoon (split from Paa Sak), and Mae Ngen (split from Yonok)

From the interview of village headmen and other related local people data of year of village establishment were elicited and rearranged chronologically. Together with the three critical period of the district the establishment of each village was plotted. Data from fieldwork study shows that villages established on the early period of community development (Period I) are mostly in the areas of Paa Sak Subdistrict on the western part of the district. The settlement of people and communities moved continuously to the east in Wiang Subdistrict and Baan Saew Subdistrict, respectively as shown in Figure 1.

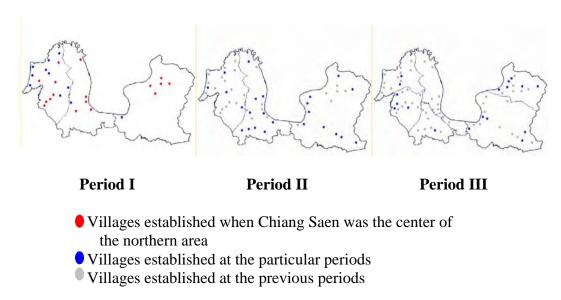


Figure 1: The establishment of villages in the three critical periods

Consider from geographical characteristics of the districts, it was found that the western part of Chiang Saen is mostly plain field which is suitable for agricultural activities while the eastern part is mostly mountainous area. Although there are a number of rivers distributed throughout the districts on the western area, the ease in traveling seems to be an important factor for the settlement and development of communities.

4.2 Cultural aspect

The term "culture" in this study focuses its attention on local languages or dialects used for naming villages. In order to deal with the local languages, findings from reasons of naming

in section 3 can be very useful source of information. Although the names of villages are officially approved by the government, it was found that local languages still significantly influence choices of words used. Results show that—comparing to the Standard Thai which is officially used by Thai government—Northern Thai dialects and some tribal language—Akha—are used in the formation of village names. Such influences can be found in two major categories: a) spelling and pronunciation and b) lexical items.

a) Spelling and Pronunciation

Some names are found to be spelled according to the pronunciation of Northern Thai dialect which is different from that of standard Thai as shown in the following examples.

| Names | Local dialects | Standard Thai |
|----------------|----------------------|---------------|
| ปง <u>ของ</u> | /poŋ khɔ̌ɔŋ / | โขง /khŏoŋ/ |
| ดอย <u>จัน</u> | /dooj can/ | ชัน /chan/ |

In addition, some names deviated from the actual spelling and pronunciation of the local languages as shown.

| Names | | Correct forms according to | | |
|---------------------|------------------------|----------------------------|-------------------------------|--|
| | | the investigation | n about the reasons of naming | |
| ทุ่งฟ้า <u>ฮ่าม</u> | /hâam/ | ห้าม | /hâam/ | |
| หนองบัว <u>สด</u> | /nŏɔŋ bua sòt / | ซด | /sót/ | |
| ดอย <u>สะโง้</u> | /dɔɔj sa ŋóo / | ช้างงู | /cháaŋ ŋuu/ | |

In the case of the deviation from ห้าม to ฮาม, it can be explained in terms of homophones /hâam/ with two different Thai initial letters "ฮ" and "ห". These two words ฮาม and ห้าม are also meaningful in the Northern Thai dialect—mean [bright] and [prohibit], respectively. However, from the interview about the reasons of naming, it was found that the correct form of /hâam/ should be ห้าม not ฮาม—the full meaning of ทุ่งฟ้าฮาม /thûŋ fấa hâm/ [field + sky + prohibit] is 'the land where rain does not heavily fall in the rainy season' (The name is set against the fact that in the old days, this area was flooded almost every year during rainy season).

Secondly, the case of สด /sôt/ in หนองบัวสด /nɔ๊ɔŋ bua sôt/ which actually has to be spelled as ซด /sót/ according to the Northern Thai dialect, the misspelling might due to the higher degree of co-occurrence and semantic compatibility of บัวสด /bua sôt/ 'fresh lotuses' than that of บัวซด /bua sôt/ 'the appeared lotuses'.

For the case of สะโง้ /saŋóo/, it can be explained in terms of deviation in pronunciation of tribal people—which is Akha who has been living in the village which is on Sa Ngo Mountain—who cannot pronounce the original word ช้างงู /cháaŋŋuu/ 'elephant and snake'—the name of the mountain—correctly, instead, they pronounced the word as สะโง้ /saŋóo/.

b) Lexical items

In addition to spelling an pronunciation, it was found that lexical items which belong to Northern Thai dialect are widely used for the naming of village names in Chiang Saen as shown.

| Names | Northern Thai | Standard Thai | Meaning |
|--|---------------|---------------------|----------------------|
| <u>แม่</u> มะ | แม่ | แม่น้ำ | 'river' |
| /mêɛ má/ | /mɛ̂ɛ/ | /mêe náam/ | |
| <u>ดอย</u> จำปี | ดอย | ภูเขา | 'mountain' |
| /dวoj campii/ | /dɔɔj/ | /phuu khǎw/ | |
| <u>สบ</u> รวก | สบ | ปากแม่น้ำ | 'mouth of river' |
| /sòp rûak/ | /sòp/ | /pàak mɛ̂ɛnáam/ | |
| แม่ <u>คำ</u> | คำ | ทองคำ | 'gold' |
| /mêɛ kham/ | /kham/ | /thooŋkham/ | |
| สัน <u>สลี</u> | สลี | Tw | 'Ficus religiosa L.' |
| /sǎn salǐi/ | /salĭi/ | /poo/ | |
| ป่า <u>ก๋อย</u> /pàa kǒj/ Dennst.' | ก๋อย /kŏj/ | กลอย /klooj/ | 'Dioscorea hispida |
| หนอง <u>ปลาสะเด็ด</u> | ปลาสะเด็ด | ปลาหมอ | 'Anabas testudeneus' |
| /nɔၴɔŋ plaa sadèt/ | /plaa sadèt/ | /plaa mɔၴɔ/ | |
| <u>เชียง</u> แสนน้อย | เชียง | เมือง | 'city' |
| /chiaŋ sĕɛn cɔ́ɔj/ | /chiaŋ/ | /m i aŋ/ | |
| <u>ปง</u> ของ | ปง | ท่า | 'dock, pier' |
| /poŋ khŏɔŋ/ | /poŋ/ | /thâa/ | |

5. Summary and discussions

Structural and social aspects of 70 village names in Chiang Saen District, Chiang Rai Province—the province to the northern part of Thailand—were investigated. In addition to documentary information, the interview of village headmen, members of Sub-district Administrative Organizations, and elderly people in the villages was done.

In terms of structure, meaning of village names and findings about the reasons of naming were used as basis of analysis. It was found that there are four major factors/reasons for naming: 1) geographical landscape and surrounded environment, 2) Important communities or religious places in the past, 3) existing communities or villages, and 4) important event in the past. Using these reasons, it was found that structurally, village names can be constructed from one to four lexical elements. Both common nouns and proper nouns are widely used and most villages contain two lexical elements.

From social perspective, Chiang Saen was divided into three critical periods according to its political changes. Data about the establishment of villages show that villagers started their settlement in the western part of the district (Pasak Sub-district) which is mostly plain field suitable for plantation. In later period, the communities

expanded across the district to the mountainous area in the east (Mae Ngen Sub-district). In addition, cultural examination shows that Northern Thai Dialect is widely used for the naming of villages in Chiang Saen as found in spelling, pronunciation, and the selection of lexical items in the names.

It was suggested that the analysis of structures of village names should be done in accompany with reasons of naming because villagers did not consider a particular names as only the combination of many single lexical items but rather, names are formed by strings of words which have once been established as proper names of entities and events. Accordingly, the viewing of names as holistic concepts might be another alternative for the analysis of village names.

For future research, although the study about village names is quite popular in Thailand, most of them did not show significant difference in terms of contents and research methodology. Accordingly, it is suggested that other aspects about village names—such as village names of tribal people, the distribution of villages which have totally the same names, and naming criteria for old and newly established villages—should be explored. The expansion of topics might ultimately lead to much more understanding about naming and might also uncover the nature of naming in relation to ways of living.

References

- Banchuen, P. 1984. *Meaning and History of Naming of Sub-district, Villages, and Important Places in Loey Province*. OS Printing House. (In Thai).
- Chiang Saen District Summary Document. 2007. Chiang Saen District Administration Office. (In Thai)
- Chuaichuwong, W. 1989. *Linguistic Culture of Village names in Maha Sarakham Province*. Master Thesis of Arts. Srinakarinvirote Maha Sarakham University. (In Thai.)
- Chumdee, N. 2006. The Settlement and Development of Ancient Communities in Muang Chiang Saen between 19 and 24 A.D. Master Thesis of History, Graduate School, Chiang Mai University. (In Thai)
- Deephadung, S. 2004. Village Names in Western Region of Thailand: Kanchanaburi, Nakorn Prathom, Pracuapkirikhan, Petchburi, Ratburi, Suphanburi, Samut Songkhram, and Samut Sakorn. Nakorn Prathom. Institute of Language and Cultural Research for Rural Development, Mahidol University. (In Thai)
- Kinawong, B. et al. 2003. *History of Chiang Rai : Chiang Saen*. Chiang Mai : Ming Muang Printing House. (In Thai)
- Kullavanijaya, P. 1992. Village Names in Guangxi Province and the Northeastern Part of Thailand. Linguistics Department, Faculty of Arts, Chulalongkorn University. (In Thai)
- Maneein, N. 2000. *The Adjustment of Village Names in Chiang Mai Province*. Master Thesis. Linguistic Department, Faculty of Arts. Chulalongkorn University. (In Thai)
- Ongsakul, S. 2003. *Phun Muang Chiang Saen* (History of Chiang Saen). Bangkok: Amarin Printing House. (In Thai)
- Sarnsamuth, N. n.d. *Place Names in Chiang Rai Province (Vol.1)*. Chiang Rai: Chiang Rai Cultural Center, Chiang Rai Teacher College. (In Thai)

- Thekhachunhathian, R. and Buranaphatana, M. 1993. Research Report Entitled "Village Names in Tha Uthen District, Nakorn Phanom Province". National Cultural Commission, Ministry of Education. (In Thai)
- Yodying, J. 1976. *History of Sub-district and Village Names in Muang District, Phrae Province*. Master of Education Thesis, Srinakarinvirote University. (In Thai.)

ARRIVAL EXPRESSIONS IN THAI

Kiyoko Takahashi Kanda University of International Studies <kiyoko@kanda.kuis.ac.jp>

0 Abstract

This paper aims to provide an in-depth analysis of the semantic and syntactic structures of Thai expressions for arrival (i.e. an entity arrives at a goal after locomotion). I present a new perspective in which Thai arrival expressions are viewed as a subtype of 'accomplishment¹ construction' consisting of two equipollent verbal components for cause and effect events (Takahashi 2007). The combination of a preceding locomotion event denoted by the first component and a subsequent arrival event denoted by the second component constitutes a macro-event² of accomplishment expressing that the locomotion event gives rise to the arrival event. Thai grammar does not require the formal distinction between finite and non-finite verbs, and therefore more than one verb in a plain form is allowed to co-occur in a clause. This fundamental morphosyntactic property of Thai enables the speakers to produce arrival expressions as well as other types of the accomplishment construction with a coordinate, yet mono-clausal, structure.

-

Kiyoko Takahashi. 2009. Arrival Expressions in Thai. *Journal of the Southeast Asian Linguistics Society* 2:175-193.

Copyright vested in the author.

Originally, the term 'accomplishment' was used by Vendler (1967: 102) to refer to one of the four classes of lexical aspect or 'Aktionsart' (namely, 'state', 'achievement', 'activity' and 'accomplishment'). The accomplishment aspect, which resides in the lexical meaning of, e.g., such English verbs or verb phrases as melt, freeze, learn in one hour, draw a circle, etc., is generally characterized to have the following distinctive features: [- static], [+ telic] (i.e. entailing a clear endpoint), and [- punctual]. However, Takahashi (2007) has applied this term to feature the aspectual nature of Thai constructions consisting of two serial verb phrases that express a cause-effect phenomenon which, the speaker construes, naturally occurs in the given pragmatic, physical, social and cultural context, and whose consequence is of interest to the speaker.

Talmy (2000: 213-288) gives an account of the notion 'macro-event' as follows. A macro-event is a single fused event composed of two simpler events holding some relationship, which is a fundamental and pervasive type of event complex in the underlying conceptual organization of language, and it is amenable to expression by a single clause. Thus, the notion of macro-event is meant to be a cross-linguistically valid one, and accordingly, I utilize this notion to account for the underlying structure of Thai expressions for complex event of arrival (i.e. an entity arrives at a goal after locomotion). However, I do not perfectly agree with him; in particular, I doubt the universal validity of his *a priori* postulation that "a macro-event consists of a pair of close-related Figure-Ground events (*ibid.*: 213)", put differently, consists of "a main event and a subordinate event (*ibid.*: 215)". Having examined Thai expressions for a variety of complex events (cf. Takahashi 2007, 2009), I believe that a macro-event may consist of two coordinate sub-events, which we may call 'complex figure' event (Croft 2001: 327). In this paper I try to show that Thai arrival expressions, which form a major category of Thai construction for a macro-event of accomplishment (cause-and-effect), do involve two coordinate sub-events: a prior locomotion event and a posterior arrival event (see Section 2).

1 Introduction

The purpose of this paper is to examine the semantic and syntactic structures of Thai expressions for the spatio-temporal concept 'arrival'. By the term 'arrival', I refer to an event in which an entity arrives at a goal after locomotion, as exemplified in (1) and (2).

- (1) khoom looy khûn pay thǔn dàat fáa lantern float ascend go arrive roof-deck

 The floating-lantern floated up and arrived at the roof-deck.
- (2) kháw looy khoom khûn pay <u>sùu</u> thóoŋ fáa PRONOUN float lantern ascend go arrive and stay sky

 They sent up a floating-lantern which got to the sky and stayed (there).

Unlike a number of previous studies on Thai motion expressions (e.g. Diller 2006, Kessakul 2005, Kölver 1984, Muansuwan 2002, Thepkanjana 1986, Zlatev 2003), I consider motion expressions like those in (1) and (2) as a single clause that represents a complex event consisting of two sub-events in succession: that is, a prior locomotion event and a posterior arrival event. thửn 'arrive' in (1) and sửu 'arrive and stay' in (2) are often considered as allative preposition indicating a path of motion toward an endpoint, which leads to an interpretation of (1) and (2) as simplex motion expressions. However, the use of these lexical items is not necessary for expressing an allative sense (Takahashi 2005). As illustrated in (3) and (4), a simple concatenation of a locomotion verb phrase and a goal noun phrase is enough to encode a situation in which an entity moves to a goal entity. This reveals that thửn 'arrive' in (1) and sửu 'arrive and stay' in (2), which are missing in (3) and (4) respectively, are not responsible for the allative sense.

- (3) khoom looy khûn pay dàat fáa lantern float ascend go roof-deck The floating-lantern floated up to the roof-deck.
- (4) kháw looy khoom khûn pay thóoŋ fáa PRONOUN float lantern ascend go sky *They sent up a floating-lantern to the sky.*

author and are guaranteed to be well-formed by Thai native speakers.

When there is need to explicitly express the meaning of allative, Thai speakers put the allative preposition *yay* 'to' in front of the goal noun phrase, as in (5) and (6). (For the classification of Thai spatial prepositions, see Section 4.2.)

_

The data used for this study were gathered mainly from published literary works which I randomly selected and partly from a computerized corpus of the Thai language that belongs to the National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Thailand, as well as other electronic texts on the Internet which I freely searched with the Google search engine. Besides examples (37) to (39) which are from Kessakul (2005), all examples in this paper are constructed or adapted from expressions in the abovementioned data by the

- (5) khoom looy khûn pay <u>yan</u> dàat fáa lantern float ascend go to roof-deck *The floating-lantern floated up to the roof-deck.*
- (6) kháw looy khoom khûn pay <u>yan</u> thóon fáa PRONOUN float lantern ascend go to sky *They sent up a floating-lantern to the sky.*

In this paper I will claim that *thtun* 'arrive' in (1) and *suu* 'arrive and stay' in (2) above have the same function as *yut* 'halt, stop and stay' in (7) and *patha?* 'collide' in (8) below have. That is, they have the function of describing a substantial arrival event arising from a preceding locomotion event. I name verbs with this function 'arrival verbs' (see Section 3).

- (7) khoom looy khûn pay <u>yùt</u> thîi lǎŋkhaa lantern float ascend go stop and stay at roof *The floating-lantern floated up and stopped at the roof.*
- (8) kháw looy khoom khûn pay <u>pathá?</u> lǎŋkhaa PRONOUN float lantern ascend go collide roof *They sent up a floating-lantern which collided against the roof.*

Generally, Thai is characterized as a phonologically tonal, morphologically isolating, syntactically verb-serializing, and discourse-pragmatically topic-prominent language. These features of the language are well-known. Unfortunately, however, the following most significant morphosyntactic characteristics of the language are still scarcely recognized: in Thai, verbs have no grammatical division between finite and non-finite forms (Diller 1988, Bisang 1995), and noun phrases adjacent to a verb have no clear distinction between required arguments of the verb (i.e. subject and object noun phrases) and non-arguments (e.g. adjunct, complement, oblique noun phrase) (Minegishi 1988). If we define 'main verb' as 'the finite verb determining the argument structure of the clause including the verb' according to the basic principles of the formal analysis in the "standard" linguistics based on Indo-European grammar, we cannot accurately define the main verb in a Thai serial verb construction which is a mono-clause with multi-verbs. Most of Thai linguists conventionally analyze the first verb *looy* 'float' in (1) and (2) as the main verb while thun 'arrive' in (1) and suu 'arrive and stay' in (2) as preposition, presumably because they adopt the dogma of Indo-European grammar that one clause must contain one finite verb.

There are two extreme opinions about finiteness of verbs in a series comprising a mono-clausal structure (viz. serial verb construction or in short SVC). One is presented by Foley & Olson (1985), and the other is by Givón (1991). Although they have the common view that an SVC should be regarded as a single clause, they sharply contrast with each other in their opinions regarding the finiteness value of verbs in an SVC. Foley & Olson explain that each verb in an SVC has the same status as predicate; namely, they are all finite. By contrast, Givón states that verbs in an SVC are co-lexical stems or grammatical morphemes; namely, they are all non-finite.

I would like to emphasize that finiteness is essentially a morphological concept, not a semantic concept, and therefore the distinction 'finite vs. non-finite' of verbs must be morphologically marked by inflection or other morphological devices. For this reason, I reject the idea that we can classify Thai verbs, which have no inflectional coding of finiteness, as finite or non-finite. In my opinion, we can only semantically divide Thai verbs in use into two groups depending on their discrete status of 'factuality' ('factual' vs. 'non-factual') (Takahashi 2006). We use Thai verbs in a particular discourse to express either a factual event (e.g. the event of eating in the past or the present, as expressed by *kin* 'eat' in (9)) or a non-factual event (e.g. the event of eating in the future, as expressed by *kin* 'eat' in (10)).

- (9) tham ?aahǎan maa <u>kin</u>
 make dishes come eat
 (He) cooked and after coming to the place (he) ate [FACTUAL].
- (10) tham ?aahǎan (wáy / phŵa thîi cà?) <u>kin</u> make dishes (for future reference / in order to) eat (*He*) cooked in order to eat [NON-FACTUAL].

In short, finiteness is not the grammatical category for Thai verbs and Thai speakers do not have to morphosyntactically distinguish between finite and non-finite verbs. Therefore, more than one verb in a plain form can co-exist in a single clause. This fundamental morphosyntactic property of Thai enables the language speakers to produce 'arrival expressions' as well as other types of 'accomplishment constructions' with a coordinate, yet mono-clausal, structure (see Section 2). I argue that arrival expressions are composed of two verbal components for locomotion and arrival, and the two components are in a coordinate relationship. In other words, the two components are equal constituents of a single clause for a macro-event of accomplishment expressing that a locomotion event ends up with an arrival event.

2 Arrival expressions: a subtype of accomplishment construction

In a previous study (Takahashi 2007), I maintained that arrival expressions can be regarded as a kind of 'accomplishment construction' consisting of two verbal components for 'cause' and 'effect' events. The gist of the arrival expressions is that a prior locomotion event [CAUSE] gives rise to a posterior arrival event [EFFECT]. There are other semantic types of accomplishment construction, as illustrated in (11) and (12).

- (11) chon <u>lóm</u>
 bump fall over

 (He) bumped into something [CAUSE] and fell over [EFFECT].
- (12) faŋ <u>rúu rûuaŋ</u> listen understand (He) listened to something [CAUSE] and understood it [EFFECT].

The second verb phrase in these constructions encodes a certain result arising from the preceding event denoted by the first verb phrase. The preceding cause event has a more or

less dynamic nature, and the following effect event is a natural consequence brought about by the preceding cause event. We use these accomplishment constructions to comment on whether or not the effect event is realized as a result of the cause event, as respectively exemplified in (13) and (14).

- (13) pay thửn ráan go arrive shop (He) went [CAUSE] and reached the shop [EFFECT].
- (14) pay mây thừn ráan go NEGATIVE arrive shop (He) went [CAUSE] but did not reach the shop [EFFECT].

The effect event, which is capable of being solely negated, is no more subordinate-like than the preceding cause event. Rather, the two events have the same functional weight in comprising a macro-event of accomplishment. There is one strong piece of linguistic evidence to support this view.

Example (15) below, which includes the pre-verbal progressive marker, is unacceptable because the combination of the cause and the effect events as a whole is within the scope of modification of the progressive marker, and the telic nature (i.e. entailing a clear endpoint) of the effect event is incompatible with the progressive aspect (or atelic aspect, i.e. not entailing a clear endpoint). If (15) is a simplex locomotion expression and *thtun* functions as allative preposition which is subordinate to the preceding motion verb *pay* 'go', then it should be compatible with the progressive aspect.

(15)* kamlaŋ pay thừŋ ráan
PROGRESSIVE go arrive shop
(He) was going to the shop. (intended meaning)

I argue that since there is no main-and-subordinate relationship between a pair of events represented by the two components of accomplishment constructions, the two events cannot be analyzed in terms of 'framing-event' and 'co-event' posited in Talmy's (1991, 2000) theory of 'event integration'. Framing-event and co-event are events that constitute a macro-event represented by a single clause. The framing-event is a main event which determines the overall temporal and spatial framework of the macro-event. The co-event, on the other hand, is a subordinate event of circumstances in relation to the macro-event as a whole. It performs functions of support in relation to the framing event. According to Talmy (2000), there are five macro-event types, as listed in (16), which includes the type of 'motion'.

(16) Talmy's (2000) five macro-event types

a. Motion:

The ball rolled in.

Framing-event denoted by *in*: Path Co-event denoted by *rolled*: rolling

b. Temporal contouring:

They talked on.

Framing-event denoted by *on*: Aspect Co-event denoted by *talked*: talking

c. State change:

The candle blew out.

Framing-event denoted by out: Changed property

Co-event denoted by *blew*: blowing

d. Action correlating:

She sang along.

Framing-event denoted by *along*: Correlation Co-event denoted by *sang*: singing

e. Realization:

The police hunted the fugitive down.

Framing-event denoted by down: Fulfillment or Confirmation

Co-event denoted by *hunted*: hunting

Within the framework of Talmy's (1985, 2000) typology of 'lexicalization patterns' in motion-event encoding, English is categorized as a 'satellite⁴-framed' language, where the framing-event is represented fully by the combination of a satellite element being in construction with the verb (e.g. run *out*) and a preposition being in construction with an object nominal (e.g. *of* the house) and a co-event is represented by a verb (e.g. run) (e.g. run *out of* the house). In contrast, Japanese is categorized as a 'verb-framed' language, where the framing-event is represented by a finite verb and a co-event is represented by a non-finite verb or other lexical elements. Compare English locomotion expression (17) with Japanese locomotion expression (18).

(17) The ball rolled in.

Framing-event denoted by *in* (satellite): Path Co-event denoted by *rolled* (verb): rolling

_

Talmy (2000: 102) defines 'satellite (to the verb)' as "the grammatical category of any constituent other than a noun-phrase or prepositional-phrase complement that is in a sister relation to the verb root".

(18) booru wa korogatte haitta (or haitte itta) ball TOPIC rolling entered (or entering, went)

The ball entered (or went in), rolling.

Framing-event denoted by *haitta* (finite verb) 'entered': Path (or *haitte itta* (non-finite verb, finite verb) 'entering, went': Path) Co-event denoted by *korogatte* (non-finite verb) 'rolling': rolling

In the English expression (17), the satellite (*in*) represents the framing-event of Path and the verb (*rolled*) represents the co-event of rolling (Manner). In the Japanese expression (18), on the other hand, the finite verb (*haitta* 'entered' or *haitte itta* 'entering, went') represents the framing-event of Path and the non-finite verb (*korogatte* 'rolling') represents the co-event of rolling (Manner).

A Thai counterpart of these simplex locomotion expressions is shown in (19).

(19) lûuk boon klîŋ khâw (or khâw pay) ball roll enter (or enter go)

The ball entered (or went in), rolling.

Framing-event denoted by *khâw* (verb) 'enter': Path (or *khâw pay* (verb, verb) 'enter, go': Path)
Co-event denoted by *klîŋ* (verb) 'roll': rolling

Slobin (2003) developed Talmy's typology further, and he claims that verb-serializing languages are neither verb-framed nor satellite-framed languages but are 'equipollently-framed' languages. In equipollently-framed languages, both framing-event and co-event are expressed by equivalent grammatical forms. In the case of Thai locomotion expressions, the framing-event of Path is basically expressed by path or deictic verbs and the co-event of Manner is expressed by manner-of-motion verbs and other lexical items.

However, arrival expressions discussed in this paper, e.g. (20) below, unlike simplex locomotion expressions, e.g. (19) above, are composed of the two components for two serial events of locomotion and arrival.

(20) lûuk boon klîn khâw pay thǔn ball roll enter go arrive *The ball went in, rolling, and arrived.*

Each of the two components in (20) (the locomotion component *klîŋ khâw pay* 'roll + enter + go', the arrival component *thửŋ* 'arrive') has an equipollent status and expresses neither a framing-event nor a co-event. A locomotion event and an arrival event denoted by the two components are coordinate events forming a macro-event of accomplishment. Although the two components express different, albeit serial, events (i.e. locomotion and arrival), they constitute a single clause. A piece of evidence for this claim is that an allative prepositional phrase, if any, occurs at the rearmost position, as shown in (21). This complies with the syntactic principle of Thai grammar that in a clause, a prepositional phrase must follow a verb or a series of verbs expressing a single event.

(21) fəərîi dəən thaan khâam maa thǔn yan thâa rua plaay thaan ferry travel cross come arrive to the final stopping port.

The ferry came across, traveling, and arrived at the final stopping port.

Example (22), where the allative prepositional phrase (*yaŋ thâa rua plaay thaaŋ* 'to the final stopping port') is placed in front of the second component instead, sounds odd.

(22)? fəərîi dəən thaan khâam maa <u>yan</u> <u>thâa rua plaay thaan</u> thǔn ferry travel cross come to the final stopping port arrive *The ferry came across, traveling, to the final stopping port and arrived at some other port.* (possible meaning)

As can be seen in (21), Thai speakers put an allative prepositional phrase introducing a goal entity (*yaŋ thâa rua plaay thaaŋ* 'to the final stopping port') at the end of the two serial components for a locomotion and an arrival (*dəən thaaŋ khâam maa* 'came across traveling' + *thǔŋ* 'arrived'), since they consider the two components as a single clause encoding a macro-event of motion.

In what follows, the semantic and syntactic structures of Thai arrival expressions will be examined in detail. In Section 3, I will outline the semantics of arrival verbs. In Sections 4.1 and 4.2, I will account for the well-composed structures of the first and the second components of the expressions, respectively. And in Section 5, I will give concluding remarks.

3 The semantics of arrival verbs

From the corpus data of Thai arrival expressions which I collected from various published books and electronic texts (see footnote 3 for the details), I found that typical arrival verbs were of two categories: 'stop verbs' and 'change-of-state verbs', as in (23).

- (23) Arrival verbs
- a. Stop verbs:

 thửn 'reach, arrive', khâw 'enter (achievement aspect reading)', hǎa 'seek', chon
 'bump', tôn 'meet', thùuk 'touch', doon 'hit', pathá? 'collide', krathóp 'strike
 against', yùt 'halt, stop and stay', càp 'catch and hold', thâap 'lay flat against', sùu
 'arrive and stay'
- b. Change-of-state verbs: *tèek* 'break', *phaŋ* 'tumble down, fall to the ground'
- (24) and (25) provide samples of arrival expressions containing a stop verb (*pathá?* 'collide') and a change-of-state verb (*tèek* 'break'), respectively.
- (24) lom năaw phûn <u>pathá?</u> bay nâa wind cold dart collide face

 A cold wind darted in and collided against the face.

(25) cùt phlú? khûn pay tèek tua klaan fáa ignite cannon cracker ascend go break body in the center of sky (They) ignited cannon crackers which went up and burst in midair.

I assume that each arrival verb has particular semantic values with respect to the following three aspects: (a) the schematic configuration and other characteristics of the goal entity, such as a point-like or surface-like shape; (b) the type of the effect arising from the mover's arrival, such as a punctual or lasting effect; and, (c) the type of the result of the arrival, such as a certain resultant state. The specific semantic values of typical arrival verbs are summarized in Table 1.

Table 1: Semantic values of typical arrival verbs in Thai

| | (a) The nature of goal | (b) Type of effect | (c) Type of result |
|---------------------------|------------------------|--------------------|--------------------|
| thửn 'arrive' | Point | | |
| khâw 'enter' | Enclosed space | | |
| <i>hǎa</i> 'seek' | Attractive entity | | |
| chon 'bump' | Approaching entity | | |
| <i>tôŋ</i> 'meet' | | Momentary effect | |
| thùuk 'touch' | | Durative effect | |
| doon 'hit' | | Clear effect | |
| pathá? 'collide' | | Small effect | |
| krathóp 'strike against' | | Big effect | |
| yùt 'halt, stop and stay' | Point | | Stasis |
| càp 'catch and hold' | Narrow surface | | Stasis |
| thâap'lay flat against' | Wide surface | | Stasis |
| sùu 'arrive and stay' | Space | | Stasis |
| <i>tèɛk</i> 'break' | | | Being broken |
| phan 'tumble down' | | | Being tumbled down |

4 Construction patterns of arrival expressions

Let's now look more closely at the construction patterns of Thai arrival expressions.

I adopt Croft's (1990, 1998) concept 'causal chain' for the representation of the event structure underlying Thai arrival expressions. The causal chain is the causal and aspectual organization of the 'idealized cognitive model for verbal events (event ICM)'. He assumes that all possible causal-aspectual types of verb-meaning stem from the causal chain among three main 'segments' (constituent phases of event) of the event ICM, namely 'CAUSE + CHANGE + STATE', as graphically shown in Figure 1.

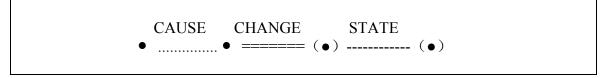


Figure 1: *Idealized Cognitive Model for Verbal Events (event ICM): Causal chain among CAUSE, CHANGE and STATE (Croft 1998: 47)*

As Figure 2 below schematically depicts, the bi-partite structure of Thai arrival expressions (the first CAUSE component + the second EFFECT component) reflects the causal chain linking the preceding locomotion event with the following arrival event.

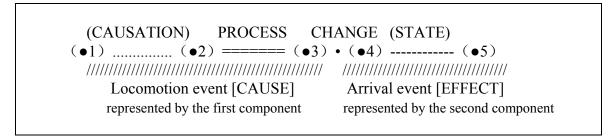


Figure 2: The event structure underlying Thai arrival expressions: Causal chain among CAUSATION, PROCESS, CHANGE and STATE

In order to represent the event structure of Thai arrival expressions as appropriately as possible, I have adapted Croft's simple figure (Figure 1) in a few respects. Firstly, I have changed the label for the first force-dynamic segment 'CAUSE' into 'CAUSATION' to differentiate the label for the first segment in the event structure of the expressions (CAUSATION) from the label for the first syntactic component of the expressions (CAUSE). Secondly, I have analytically divided the second durative segment 'CHANGE' into two segments: the second durative segment 'PROCESS' and the third punctual segment 'CHANGE'. Thirdly, I have parenthesized the first and the last segments 'CAUSATION (which is called CAUSE in Figure 1)' and 'STATE' as well as all the black circles '•' at the beginning or the end of each segment which represent event-participants involved.

In the event structure underlying Thai arrival expressions (Figure 2), the prior locomotion event encompasses two segments, i.e., CAUSATION and PROCESS. Likewise, the posterior arrival event embraces two segments, i.e., CHANGE and STATE. The PROCESS and CHANGE segments in the middle are indispensable to an arrival event (viz. an event in which an entity arrives at a goal after locomotion), while the parenthesized CAUSATION and STATE segments in the periphery are dispensable. Put differently, the latter two peripheral segments may be outside the scope of attention and so be unmentioned. In addition, there are three main participants in an arrival event, i.e., 'Causer', 'Mover' and 'Reference point for determining a path' (such as 'Source' and 'Goal'). A Causer initiates the locomotion of the Mover (CAUSATION); The Mover moves along a path (PROCESS), then it stops at a Goal (CHANGE), and it may stay there for a while (STATE). '•1' at the beginning of the CAUSATION segment and '•2' at the beginning of the PROCESS segment represent the Causer and Mover, respectively. '•3' at

the beginning of the CHANGE segment through '•5' at the end of the STATE segment represent either the Mover or Goal. All of these participants may or may not be named by a noun phrase (and so they are parenthesized), just like the CAUSATION and STATE segments which similarly may or may not be explicitly expressed by a verb.

From Figure 2, we can see that the event structure of Thai arrival expressions entails two sub-events: the CAUSE event of locomotion represented by the first component and the EFFECT event of arrival represented by the second component. Note, however, that the first and the second components are not necessarily always combined. Either of the two components by itself can express a simplex motion event. The first component alone expresses a simplex locomotion event, as in (26) and (27), and the second component alone expresses a simplex arrival event (viz. an instantaneous event of arrival, separated from the preceding locomotion event), as in (28) and (29). Thai syntactic structures are thus quite flexible.

- (26) kháw looy khoom khûn pay
 PRONOUN float lantern ascend go
 They sent a floating-lantern up away. [CAUSATION + PROCESS]
- (27) khoom looy khûn pay lantern float ascend go The floating-lantern floated up away. [PROCESS]
- (28) khoom pathá? lăŋkhaa lantern collide roof

 The floating-lantern collided with the roof. [CHANGE]
- (29) khoom yùt thîi lắŋkhaa lantern stop and stay at roof

 The floating-lantern stopped and stayed at the roof. [CHANGE + STATE]

4.1 Construction patterns of the first component of arrival expressions

The first component of the arrival expressions designates a caused or spontaneous locomotion. Lexical items appearing in the first component can be classified into four main categories: 'causer noun phrase', 'cause-of-motion verb phrase', 'mover noun phrase' and 'locomotion verb phrase'. In turn, the last category subsumes four sub-categories: 'manner-of-motion verb', 'direction verb', 'path verb' and 'deictic verb'. Direction verb and path verb may take a noun phrase indicating a reference point for the path of motion, such as starting point and passing-by point. Table 2 shows the linear order among these main constituents of the first component.

| <1> Causer NP | <2> Cause-of-motion VP | <3> Mover NP | <4> Locomotion VP |
|---------------|------------------------|--------------|--------------------------|
| | [CAUSATION] | | [PROCESS] |
| | | | <4.1> Manner-of-motion V |
| | | | <4.2> Direction V |
| | | | <4.3> Path V |
| | | | <4.4> Deictic V |

Table 2: *Linear order of main constituents of the first component*

Some representative members of the verb categories are given in (30) through (34).

(30) Cause-of-motion verbs:

khwâaŋ 'throw', dìit 'flick', tè? 'kick', paa 'throw', phlàk 'push', yoon 'throw', ltùan 'slide'; khŏn 'load', cuuŋ 'pull, drag', nam 'carry', phaa 'guide someone', yìp 'pick', yók 'lift', lâak 'drag'

(31) Manner-of-motion verbs:

klîŋ 'roll', khlaan 'crawl', khûnup 'creep', dəən 'walk', bin 'fly', ləəy 'float', wîŋ 'run', lǎy 'flow'; kâaw 'step', kraden 'hurtle', tày 'clamber', thalák 'spurt out', phèn 'rush out of', phûŋ 'spout', traween 'wander', bùŋ 'speed', fàa 'break through', hèe 'parade', dândôn 'trudge', lúay 'ramble', trèe 'stroll', yôŋ 'tiptoe'

(32) Direction verbs:

càak 'leave', tòk 'fall', thờy 'retreat', yóɔn 'reverse', rûaŋ 'drop off', lòn 'drop', com 'sink'

(33) Path verbs:

khâw 'enter (durative aspect reading)', ?òɔk 'exit', khûm 'ascend', loŋ 'descend', klàp 'return', khâam 'cross', taam 'follow', phàan 'pass', phón 'pass', lôŋ 'follow along', lát 'cut across', lôʔ 'go along', lîap 'go along', lám 'overstep', ləəy 'exceed', sǔan 'pass each other', ?ôɔm 'take a roundabout way'

(34) Deictic verbs:

pay 'go', maa 'come'

(35) to (37) below exemplify the first component of the arrival expressions. (The number <1> to <4.4> attached to the words in (35) to (39) corresponds to the number indicated in Table 2.) Expressions for caused locomotion such as (35) require the combination of cause-of-motion verb and path/deictic verb, while those for spontaneous locomotion such as (36) and (37) exclude causer noun phrase and cause-of-motion verb which represent the CAUSATION segment.

(35) kháw lûuk boon klîn yóon tè? phàan pratuu khâw pay roll PRONOUN kick ball reverse pass door enter go <2> <3> <4.1> <4.2> <4.3> <4.3> <4.4> He kicked the ball which went back in rolling through the door.

(36)lûuk boon klîŋ yớon phàan pratuu khâw pay ball roll door reverse pass enter go <3> <4.1> <4.2> <4.3> <4.3> <4.4> The ball went back in rolling through the door.

(37)câm kâaw dùm dùm kháw walk quickly **PRONOUN** walk quickly and straight with the step <4.1> <4.1> shoulders bent forward (onomatopoeia) <3> fàa sǎav fǒn ?òon break through falling rain light <4.1>

He hurried and stepped quickly, with his shoulders bent forward, forcing his way through light rain. (Kessakul 2005)

The following are semantic and syntactic constraints on the first component of Thai arrival expressions. Firstly, to express a caused locomotion, one cause-of-motion verb and at least one path or deictic verb must be combined. Secondly, to express a spontaneous locomotion, on the other hand, causer noun phrase and cause-of-motion verb must be absent. Though only one locomotion verb is able to express a locomotion event, normally more than one verb is serialized. It is possible for a manner-of-motion verb, direction verb and path verb to multiply occur in a single clause. This is because the concepts of 'mannerof-motion', 'direction'⁵ and 'path'⁶ can be richly described from more than one perspective. Examples (35) and (36) above, for instance, include two path verbs, i.e., phàan 'pass' and khâw 'enter', which concurrently characterize the motion in question as passing some object and also as going into some enclosed space. Furthermore, example (37) above includes three manner-of-motion verbs, i.e., câm 'walk quickly', kâaw 'step' and fàa 'break through', as well as one onomatopoeia, i.e., dùm dùm 'walk quickly and straight with the shoulders bent forward', all of which together modify the manner of the person's moving. In contrast, the number of cause-of-motion verbs and deictic verbs appearing in a single clause is limited to only one, simply because only one value of 'cause of motion' and 'deictic relation' can be specified for a single motion event.⁷

Interestingly, Kessakul (2005) states that we frequently encounter Thai motion expressions consisting of multi-verb-phrases which express a mover's spontaneous motion along a complex path with intermediate points where one relocation terminates and another relocation starts, as illustrated in (38) and (39) below. This complex path is formed by connecting a number of relocation paths. For example, the complex path in (38) is made up of two paths, i.e. (38a) and (38b); that in (39) four paths, i.e. (39a) to (39d). Kessakul calls

By 'direction' I mean 'relative direction of path being formed with a starting point and/or an endpoint' (Takahashi 1997).

⁶ By 'path' I mean 'relative direction of path arising from interaction with a reference object' (Takahashi 1997) which Talmy (1991, 2000) considers as 'the core schema of motion event'.

However, the combination of the two deictic verbs, viz. *pay maa* 'go + come', may be added to a locomotion verb to indicate a to-and-fro kind of the path of the described locomotion, that is, moving back and forth in a more or less confined space.

this syntactically expanding phenomenon 'structural recurrence' of the locomotion verb phrase. Thai syntactic structures are amazingly elastic indeed.

- (38) a. naawaa kâaw tháaw lon càak fútbàat Nawa step foot descend from footpath <3> <4.1> <4.3>
- b. dəən khâam thanŏn pay yaŋ ráan ?aysakhriim walk cross road go to ice-cream shop

Nawa stepped off the footpath and walked across the road toward an ice-cream shop. (Kessakul 2005)

- (39) a. rút lúk khŵn Rut get up ascend <3> <4.1> <4.3>
- b. dəən thaan təə travel continue
- c. lát ló? pay taam sâak ?aakhaan take a short cut along the side go along the ruin of building <4.3> <4.4>
- d. phàan thanŏn pass road <4.3>

Rut got up and continued his trip, taking a shortcut along the ruin of the building, passing the road. (Kessakul 2005)

It should be noted that normally the first component of arrival expressions examined in this section does not undergo the structural recurrence of the locomotion verb phrase; otherwise, the symmetrical relationship existing between the first and the second components would become ill-balanced.

4.2 Construction patterns of the second component of arrival expressions

The second component of the arrival expressions depicts the final phase of an arrival event, which is the culmination of the prior locomotion event represented by the first component. The second component describes how and where the moving entity has arrived, whereby imposing a telic nature onto the locomotion event. Table 3 shows the linear order of three main categories of lexical items occurring in the second component: 'arrival verb', 'preposition' and 'goal noun phrase'.

| <5> Arrival V | <6> Preposition | <7> Goal NP |
|-------------------------|--|-------------|
| [CHANGE (+ STATE)] | | |
| <5.1> Stop V | <6.1> Endpoint Preposition | |
| <5.2> Change-of-state V | <6.1.1> Allative Preposition | |
| | <6.1.2> Attendant relation Preposition | |
| | <6.2> Global locative Preposition | |
| | <6.3> Local locative Preposition | |

Table 3: Linear order of main constituents of the second component

The semantic and syntactic constraint on the second component of Thai arrival expressions is very simple: to express the final phase of an arrival event, one arrival verb is necessarily used, whereas other constituents are optionally used.

Examples of the second component of the arrival expressions are given in (40) and (41). (The number <5> to <7> attached to the words in (40), (41), (44) to (47) corresponds to the number indicated in Table 3.)

(It relocated and) covered the cabinet.

(It relocated and) arrived in front of the car.

The category of preposition consists of three main sub-categories: 'endpoint preposition', 'global locative preposition' and 'local locative preposition'. There are two endpoint prepositions: the 'allative preposition' (i.e. yan 'to') and the 'attendant relation

In Thai, prepositions for indicating the endpoint of motion (i.e. endpoint prepositions including the allative preposition *yay* 'to' and the attendant relation preposition *kàp* 'with') are prepositions proper; that is, they are no longer content words (verbs or nouns), whereas there are no prepositions proper for indicating the starting point or the passing course of motion. However, in a simplex locomotion expression, the direction verb *càak* 'leave', when occurring after a path verb and/or a deictic verb or before an allative preposition, serves as 'starting-point (ablative) preposition' (i.e. *càak* 'from') (see example (38a)) and the path verb *taam* 'follow', when occurring after a deictic verb or before an allative preposition, serves as 'passing-course preposition' (i.e. *taam* 'along') (see example (39c)).

The two classes of locative nouns functioning as locative prepositions that I name here, 'global locative prepositions' and 'local locative prepositions', roughly correspond to Zlatev's (2003: 322-326) two classes of such nouns, 'class nouns' and 'region nouns', respectively. However, the number of 'class nouns' is smaller than that of 'global locative prepositions' listed in (42), and the lexical item *klaap* 'in the center of, in the middle of, amid; center, middle' is categorized by Zlatev as the latter 'region noun' ('local locative preposition'), but by me as the former 'global locative preposition'.

It is interesting to note that in a simplex locomotion expression, the arrival verbs *thtũn* 'arrive' and *sùu* 'arrive and stay' function as allative preposition when combined with the starting-point (ablative) preposition (i.e. *càak* NP { *thtũn* / *sùu*} NP 'from ... to ...') and *sùu* 'arrive and stay' also functions as allative (or illative) preposition when preceded by the arrival verb *khâw* 'enter (punctual aspect reading)' (i.e. *khâw sùu* NP 'get into ...') (Takahashi 2005: 116-117).

preposition' (i.e. *kàp* 'with'). The allative preposition highlights a path toward a goal entity, while the attendant relation preposition indicates a goal entity with which a mover comes into touch. Unlike these endpoint prepositions, global and local locative prepositions are not prepositions proper, but are nouns that are capable of functioning as locative preposition in a certain context. Various labels have been given to these lexical items, e.g., relational nouns, localizers, locative nominals, locative relator nouns, locative particles, and so on (cf. Bisang 1996: 549). Global locative prepositions specify the global configuration of a locative entity (such as point and side), as in (42).

(42) Global locative prepositions:

thîi 'at; place'
thĕw 'in the region of; row'
rôɔp 'around; surrounding'
thûa 'all over; everywhere'
khâŋ 'on/to the side of; side'
dân 'on/to the side of; surface'
phaay 'in the side of; space'
bûaŋ 'in the direction of; direction'
thaaŋ 'in the direction of; way'
klaaŋ 'in the center of, in the middle of, amid; center, middle'
rawàaŋ 'among'

On the other hand, local locative prepositions specify intrinsic configuration (such as front or back) or relative orientation with respect to a certain viewpoint (such as right and left) or absolute orientation with respect to gravity (such as above and below) or other fixed directions (such as north and south), all of which have relatively fine-grained contrastive values, as in (43).

(43) Local locative prepositions:

bon 'on; upper part'

nay 'in; inner part'

nâa 'in front of; face'

khwăa 'right'

nŭa 'above; north, uphill, upstream'

lâaŋ 'under; lower part'

nôɔk 'out; outer part'

lǎŋ 'behind; back'

sáay 'left'

tây 'below; south, downhill, downstream'

Different types of prepositions may co-occur, but they must take place in the fixed order, as in (44) and (45).

(44)thâap kàp dân bon tûu . . . lay flat against with side upper cabinet <6.1> <6.2> <6.3> (It relocated and) covered the upper side of the cabinet.

(It relocated and) arrived at the front side (or in front) of the car.

Sometimes the combination of two different types of locative proposition, such as dân bon 'upper side' and khân nâa 'front side (or space in front)', is regarded as a compound noun, which may not be followed by a noun phrase, as in (46) and (47).

(It relocated and) arrived at the front side (or at the space in front).

5 Conclusion

In this paper I have characterized Thai arrival expressions as composed of two components expressing a prior locomotion event and a posterior arrival event which are in a coordinate relationship. The two components constitute a single clause for a macro-event of accomplishment. I have shown that Thai arrival expressions have systematic, though very elastic, structures. Thai grammar does not require a single main verb in an SVC. This is crucial for the establishment of accomplishment construction that consists of two equipollent verbal components for cause and effect. My claim is that arrival expressions should be categorized as a major type of this versatile bi-partite construction.

Acknowledgements

I am grateful to Jordan Zlatev and Tasanee Methapisit for discussions that helped me sharpen the ideas presented in this paper. I wish to thank the audience of the SEALS 17 conference (the 17th Annual Meeting of the Southeast Asian Linguistics Society, Collage Park, August 31 - September 2, 2007) for helpful comments and suggestions. I owe my gratitude to anonymous reviewers for their invaluable comments and criticisms on the earlier version of this paper. Thanks are also due to Bruce Horton for stylistic suggestions and helpful comments. Any errors and inadequacies are as usual my sole responsibility.

References

Bisang, Walter. 1995. Verb serialization and converbs—differences and similarities. In Haspelmath, Martin and Ekkehard König (eds.) *Converbs in Cross-Linguistic Perspective: Structure and Meaning of Adverbial Verb Forms—Adverbial Participles, Gerunds—*, 137-188. Berlin, Mouton de Gruyter.

- Bisang, Walter. 1996. Areal typology and grammaticalization: Process of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studieds in Language* 20.3: 519-597.
- Croft, William. 1990. Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information. Chicago, The University of Chicago Press.
- Croft, William. 1998. Event structures in argument linking. In Butt, Miriam and Wilhelm Geuder (eds.) *The Projection of Arguments: Lexical and Compositional Factors*, 21-63. Stanford, CSLI Publications.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford, Oxford University Press.
- Diller, Anthony V. N. 1988. Thai syntax and "national grammar". *Language Sciences* 10.2: 273-312.
- Diller, Anthony V. N. 2006. Thai serial verbs: Cohesion and culture. In Aikhenvald, Alexandra Y. and R. W. Dixon (eds.) *Serial Verb Constructions: A Cross-linguistic Typology*, 160-177. Oxford, Oxford University Press.
- Foley, William A. and Mike Olson. 1985. Clausehood and verb serialization. In Nichols, Johanna and Anthony C. Woodbury (eds.) *Grammar Inside and Outside the Clause: Some Approaches to Theory from the Field*, 17-60. Cambridge, Cambridge University Press.
- Givón, Talmy. 1991. Some substantial issues concerning verb serialization: Grammatical vs. cognitive packaging. In Lefebvre, Claire (ed.) *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, 137-184. Amsterdam, John Benjamins.
- Kessakul, Ruetaivan. 2005. *The Semantic Structure of Motion Expressions in Thai*. Ph.D. dissertation, Tokyo University.
- Kölver, Ulrike. 1984. Local prepositions and serial verb constructions in Thai. In *Arbeiten des Kölner Universalien-Projects*, Nr.56. Cologne, Universität zu Köln, Institut für Sprachwissenschaft.
- Minegishi, Makoto. 1988. Kussetsu mo katsuyoo mo nai gengo: Taigo (A language without inflection: Thai). *Gekkan Gengo* 17.8: 90-95.
- Muansuwan, Nuttanart. 2002. *Verb Complexes in Thai*. Ph.D. dissertation, The State University of New York at Baffalo.
- Slobin, Dan I. 2003. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Strömqvist, Sven and Ludo Verhoeven (eds.) *Relating Events in Narrative, Vol.2: Typological and Contextual Perspectives*, 219-257. Mahwah, N.J., Lawrence Erlbaum Associates.
- Takahashi, Kiyoko. 1997. Verbs for global locomotory body motion in Thai. *Proceedings* of the 16th International Congress of Linguistics.

- Takahashi, Kiyoko. 2005. The allative preposition in Thai. In Sidwell, Paul (ed.) *Papers* from the 15th Annual Meeting of the Southeast Asian Linguistics Society 2005, 111-120. Canberra, Pacific Linguistics.
- Takahashi, Kiyoko. 2006. A two-dimensional classification of complex events represented by basic serial verb constructions. Paper presented at The 4th International Conference on Construction Grammar, Tokyo, September 1-3, 2006.
- Takahashi, Kiyoko. 2007. Accomplishment constructions in Thai: Diverse cause-effect relationships. In Iwasaki, Shoichi, Andrew Simpson, Karen Adams and Paul Sidwell (eds.) *Papers from the 13th Annual Meeting of the Southeast Asian Linguistics Society 2003*, 263-277. Canberra, Pacific Linguistics.
- Takahashi, Kiyoko. 2009. Basic serial verb constructions in Thai. *Journal of the Southeast Asian Linguistics Society*, Vol. 1, 215-229.
- Talmy, Leonard. 1985. Lexicalization patterns. In Shopen, Timothy (ed.) *Language Typology and Syntactic Description, Vol. 3: Grammatical Categories and the Lexicon*, 57-149. Cambridge, Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: A typology of event conflation. *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, 480-519.
- Talmy, Leonard. 2000. Toward a Cognitive Semantics Vol.2: Typology and Process in Concept Structuring. Cambridge, MIT Press.
- Thepkanjana, Kingkarn. 1986. *Serial Verb Constructions in Thai*. Ph.D. dissertation, University of Michigan.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca, Cornell University Press.
- Zlatev, Jordan. 2003. Holistic spatial semantics of Thai. In Casad, Eugene H. and Gary B. Palmer (eds.) *Cognitive Linguistics and Non-Indo-European Languages*, 305-336. Amsterdam, John Benjamins.

A PRELIMINARY DESCRIPTION OF ENDE PHONOLOGY¹

Bradley McDonnell

Arizona State University bradley.mcdonnell@gmail.com

Abstract

In what follows, I describe aspects of the phonology of Ende, a virtually undescribed Austronesian language of Flores in eastern Indonesia.

1. Introduction

Ende (alternatively Endeh, Endenese, or Endehnese) is an Austronesian language primarily spoken on the island of Flores and the smaller island of Ende located off the coast of Flores in the Lesser Sunda Islands of Nusa Tenggara Timur (NTT) in eastern Indonesia. The Ende speaking region makes up the western half of the *Kabupatan Ende*, the Ende Regency, which also includes the larger ethnolinguistic group, Lio, in the eastern half of the regency. The Ende Regency has a reported total population of approximately 254,000 (www.ende.go.id, 2005) of which there is a reported 87,000 speakers of Ende and 130,000 speakers of Lio (Wurm & Hittori, 1981). Ende is classified as a Bima-Sumba language by Esser (1938), in the Central Malayo-Polynesian (CMP) subgroup (Blust, 1993; Grimes 1991b), although classifications of CMP and Bima-Sumba are not considered

Bradley McDonnell. 2009. A Preliminary Description of Ende Phonology. *Journal of the Southeast Asian Linguistics Society* 2:195-226.

Copyright vested in the author.

I would like to thank my Ende informants from the village of Oné Koré (just of the city of Ende) - Fariz Gesiradja and Ignasius Wara - who patiently reviewed the pronunciation of these words many times with me. I want to especially thank Fariz Gesiradja for introducing me to Ende. Furthermore, I would like to acknowledge Rene van den Berg and Marian Klamer for comments on this version of the paper and Dawn Bates for comments on an earlier version of the paper. Finally, I am grateful to Satoshi Nakagawa for allowing me to use the Ende-English dictionary in this paper. Any mistakes are solely my own.

The following abbreviations are used in this paper. 1SG – 'first person singular pronoun', 2SG - 'second person singular pronoun', 3SG - 'third person singular pronoun', 1PL.INCL - 'first person plural inclusive pronoun', 1PL.EXCL - 'first person plural exclusive pronoun', 2PL 'second person plural pronoun', 3PL

pronoun', 1PL.EXCL - 'first person plural exclusive pronoun', 2PL 'second person plural pronoun', 3PL - 'third person plural pronoun', 1SG.POSS - 'first person singular possessive pronoun', 2SG.POSS 'second person singular possessive pronoun', 3SG.POSS - 'third person singular possessive pronoun', 3PL.POSS - 'third person plural possessive pronoun', ASP.PERF - 'perfective aspect', LIG - 'ligature', NUM - 'numeral clitic', CLAS 'classifier'.

² The numbers of speakers from Ende and Lio are taken from the Ethnologue (Gordon, 2006), which exactly matches numbers from Wurm & Hittori (1981). The total population was taken from the Ende Regency website (www.ende.go.id), which appears to be taken from a 2005 census. Combining these two figures, the number of speakers and the number of residents, leaves quite a big discrepancy, resulting in the difference of 37,000. It is unclear if these are speakers of other indigenous languages of Indonesia or incorrect data regarding the speakers or total population.

uncontroversial. According to Wurm & Hittori (1981), Ende and Lio may, in fact, form a single language made up of a dialect chain that can be considered mutually unintelligible at either end of the dialect chain. The extent of their intelligibility, however, is not yet clear. Some have considered Lio to be a dialect of Ende, but have also noted that the cultures see themselves as culturally distinct (see Nakagawa, 1984, 1988b; Aoki, 1988). Even though a clear picture of the dialects of Ende and Lio are unclear, there appears to be three primary dialect groups named after their respective first person pronoun (Wurm & Hittori, 1981; Nakagawa, p.c.). Beginning in the west, the *Nga'o* dialect makes up only about 9,000 speakers (Wurm & Hittori, 1981), but according to Nakagawa (1988b), it may be considered a separate language. Bordering to the east of *Nga'o* is the *Ja'o* dialect, which is spoken in and around the city of Ende in the central region of the Ende Regency. The *Ja'o* dialect is also spoken on the small island of Ende off the south coast of Flores. Lastly, the eastern most dialect is the *Aku* dialect, which borders on Lio speaking regions in the east, which may be mutually intelligible with Lio.

Even though the vast majority of Flores and NTT ascribe to Catholicism, nearly half of the 87,000 in the Ende population ascribe to Islam. Most of the ethnically Ende Muslims reside on the coast and make their living by fishing, while most ethnically Ende Catholics dwell inland and make their living by primarily farming maize and cassava. On the small island of Ende, approximately an hour motorboat ride from the city of Ende, there is 8,000 inhabitants that are almost exclusively Muslims (www.ende.go.id, 2005). Other Ende communities, where the Ende language is still spoken, exist on the north coast of the island of Sumba and the Manggarai region of western Flores (see Nakagawa, 1984; Needham, 1968). Islam apparently came to Ende between the 16th and 17th century from the island of Solor with either direct or indirect influence from the Bugis of Sulawesi (Nakagawa, 1984).

1.1 Previous studies

There have been a handful of anthropological studies on Ende beginning in the midnineteenth century. These studies have focused on kinship terminology and short word lists, mostly for comparative work in linguistics and anthropology. Needham's (1968) study of Ende kinship terminology summarizes many of the earlier studies. He cites the earliest wordlists of Ende as coming from an anonymous Dutch description of the languages of Sumba (Anon., 1855) and Roos (1877). Further, Needham cites van Suchtelen (1921) as presenting Ende wordlists as well as some dialectal information on the Ende speaking region of Flores. Lastly, Needham (1968) states that the other analyses of Ende kinship terminology, including Frazer (1919) and Wilken (1886), are not primary sources, but are based on three aforementioned studies. Needham's own study represents data collected from a small permanent Ende community in Mamboru on the island of Sumba, who had resided on Sumba since at least the mid-nineteenth century (Needham, 1968). In a second attempt to analyze Ende kinship terminology, Needham (1970) cites field notes taken by the late Raymond Kennedy from 1949 to 1950 in the village of Roworeke, closely located to the administrative center of Ende. It is unclear from Needham (1970) how extensive Kennedy's field notes are, but they appear to provide descriptions of material culture and kinship relationships in Ende.

More recent studies on Ende have been carried out by the Japanese anthropologists, Satoshi Nakagawa and Eriko Aoki. They have focused on ritual language of Ende (and Lio) and have compiled a 2,500+ word Ende-English dictionary (Aoki and Nakagawa,

1993). Nakagawa (1988b) focuses on ritual language used during bridewealth negotiations, in which he describes spatial metaphors employed in parallel speech couplets, that are often employed by cultures of eastern Indonesia (see Grimes et. al., 1997; Fox, 1988). Aoki and Nakagawa's (1993) unpublished 2,500+ word Ende-English dictionary is a significant catalog of vocabulary from Ende, primarily from the *Aku* dialect. However, the dictionary also includes dialectal information from all three dialects of Ende. More than any other work, the Ende-English dictionary played a key role in the present study. Furthermore, recent linguistic descriptions have appeared for many of the neighboring Bima-Sumba languages on Flores. These include descriptions of Keo (Baird, 2002), Ngadha (Djawanai, 1977, 1983), and Rongga (Arka et. al., 2007). Other larger languages, such as Manggarai to the west and Sika to the east also have received various descriptions. However, it should be noted that central Flores and CMP languages still remain grossly under-described (see Tryon, 1995; Ross, 1995).

2. Overview of Ende phonology

The following description of Ende phonology, based on the Ja'o dialect, demonstrates various properties that are encountered in the other languages of central Flores, but may be less commonly encountered elsewhere in the languages of eastern Indonesia and Austronesia as outlined by Hajek (2008, forthcoming). First, Ende has a large consonant phoneme inventory with 22 phonemes, which is characteristic of other consonant phoneme inventories on central Flores, including Keo (23; Baird, 2002), Palu'e (21; Donohue, n.d.), Rongga (23; Arka et. al., 2007), and Ngadha (21; Djawanai, 1977, 1983). More particularly, Ende evinces a complex stop system that shows four-way distinction between 1) voiced, 2) voiceless, 3) prenasalized, and 4) implosive stop phonemes that is also common in languages of central Flores (see §2.1.1.1 for discussion of the stop system in Ende). Furthermore, even though Ende still shows phonemic contrast with plain voiced stops, they are much less common. This is notable as another Bima-Sumba language, Kambera, does not show plain voiced stop phonemes, a very marked feature of consonant inventories (Klamer, 1998; cf. Hajek, 2008 forthcoming). Lastly, Ende shares a similar syllable structure with the languages of Flores; all Ende syllables are open, and closed syllables are strictly prohibited (see §2.2 for description of Ende syllable structure; cf. Klamer, 2002:368).

Ende also contains aspects that appear to be less common among the languages of central Flores. The following properties in Ende, although not absent, are more marked. First, Ende has a marked alveolar approximant /I/ phoneme that shows frication during articulation, showing allophonic variation with alveolar fricative [z]. The only other language of Flores that demonstrates a similar speech sound is Rongga (Arka, 2004), although the extent to which these sounds show similar properties is unclear (see §2.1.1.6 for a description of the alveolar approximant in Ende). Secondly, Ende has a phonemic glottal stop in word-initial and word-medial positions. In the word-initial position the lack of a glottal stop affects the vowel quality of such vowels. That is, word-initial vowels show variation in length, duration, and breathiness, although this distinction appears to be a result of the presence or absence of the glottal stop and not a distinction between vowel phonemes as in the case of Kedang (see Samely, 1991 for Kedang; see §§2.1.1.1 and 2.1.2.7 for Ende). The following phonological description of Ende is divided into three

_

³ For an explanation of the phonetic motivation behind this rare phenomena see Hajek (2008 forthcoming).

primary sections: Phonemes in §2.1, Phonotactics in §2.2, and Clitics in §2.3. The description of the Ende phonemes in §2.1 makes up the bulk of this phonological description. §3 summarizes this description of Ende phonology and outlines questions for future research.

2.1 Phonemes

This section offers a description of and establishes the basis for the 24 consonant phonemes in §2.1.1 and the six vowel phonemes in §2.1.2 in Ende. Minimal pairs where possible are given for each of the phonemes in Ende. Where a minimal pair is not available, a near minimal pair is provided. In each minimal or near-minimal pair, both the orthographic and IPA representations of the word are provided. Not all logically possible minimal pairs are provided below, but only those that seem fit to establish the speech sound in question as a phoneme. For instance, minimal pairs are given for the voiced-voiceless distinction in stops for $\frac{1}{2}$ and $\frac{1}{2}$ where $\frac{1}{2}$ and $\frac{1}{2}$ in Ende. Minimal pairs where possible are given for the phonemes in §2.1.2 in Ende. Minimal pairs where possible are given for the phonemes in Ende. Where a minimal pair is not available, a near minimal pair is provided. In each minimal or near-minimal pair, both the orthographic and IPA representations of the word are provided. Not all logically possible minimal pairs are provided below, but only those that seem fit to establish the speech voiceless distinction in stops for $\frac{1}{2}$ and $\frac{1}{2}$ where $\frac{1}{2}$ is a phoneme in §2.1.2 in Ende. Minimal pairs where possible are given for the pair is not available, a near minimal pair is not available, a near mi

2.1.1 Consonants

Ende has 24 consonant phonemes as shown in Table 2.1. The IPA syllables are accompanied by the practical orthography employed in Aoki and Nakagawa's (1993) Ende-English dictionary in the table below.⁴

| | Labial | Alveolar | Palatal | Velar | Glottal |
|-------------------|--|-----------------|-----------|--------------------|-------------|
| Vl. Stop | p | t <t></t> | | k <k></k> | ? <'> |
| Vd. Stop | b | d < d > | | g <g></g> | |
| Affricate | | | ʤ <j></j> | | |
| Implosive Stop | 6 <bh></bh> | d < dh > | | | |
| Prenasalised Stop | $^{\mathrm{m}}b$ $<$ $\mathrm{m}b$ $>$ | $^{n}d < nd >$ | | $^{\eta}g$ < ngg > | |
| Nasal | m <m></m> | n <n></n> | | η < ng > | |
| Vl. Fricative | f < f> | s <s></s> | | | (h) <h></h> |
| Vd. Fricative | | | | $\gamma < gh >$ | |
| Flap/Trill | | r/r < r > | | | |
| Lateral | | (l) <l></l> | | | |
| Approximant | w <w></w> | ı <rh>></rh> | | | |

2.1.1.1 Stop consonants

Ende makes a four-way stop distinction: voiced, voiceless, prenasalised, and implosive. The stops that make up the four-way stop distinction divide into two categories: 1) non-complex stops /b, p, d, t, g, k/ and 2) complex stops /b, d, mb, nd, ng/. These stop are rather symmetric in bilabial, alveolar, and velar places of articulation, except for the absence of a

The conventions in the practical orthography largely follow Aoki and Nakagawa (1993), except in the case of the mid front vowel /e/, which is represented as é, following conventions used for citation forms of Indonesian and other languages of Indonesia in dictionaries. Furthermore, Aoki and Nakagawa (1993) represent long vowels by doubling the vowel, <a> is represented as <aa>. Representing long vowels is not necessary as they only occur in single syllable words that do not already meet the requirement of a bimoriac word as is shown in §2.2.

velar implosive stop. However, only non-complex stops make a voice distinction, baring voiceless prenasalized and implosive stop phonemes. The glottal stop in Ende is both phonemic in word-initial and intervocalic positions, which in the word-initial position affects vowel quality. Ende has only one affricate /ʤ/, which is discussed at the end of this section.

Non-complex stops

Ende makes a voiced-voiceless distinction in non-complex stops /p, t, k/ with /b, d, g/. The voiced-voiceless distinction holds both in word-initial and intervocalic positions.

| /b/ ~ /p/ | ba'i | /ba?i/ | 'bitter' | pati | /pati/ | 'give' |
|-----------|------|---------|----------------|------|---------|----------------|
| | baba | /baba/ | 'father' | 'apa | /?apa/ | 'what' |
| /d/ ~ /t/ | du'u | /du?u/ | 'to stop' | tu'u | /tu?u/ | 'dry' |
| | pota | /pota/ | 'to disappear' | podo | /podo/ | 'tobacco pipe' |
| /g/ ~ /k/ | gajo | /gadzo/ | 'to serve' | kajo | /kacto/ | 'relative' |
| | raga | /raga/ | 'to crawl' | raka | /raka/ | 'to boil' |

The voiced alveolar stop /d/ shows limited occurrences in lexical items in both intervocalic and word-initial positions. However, it still arises as a separate phoneme as it contrasts with other alveolar consonants (see *Implosive stops* and *Prenasalized stops* for minimal pairs).

Further, there is a distinction in the places of articulation in both voiced and voiceless non-complex stops. That is, there is a distinction between labial /b, p/ and alveolar stops /d, t/, respectively.

| /b/ ~ /d/ | baga | /baga/ | 'to step over' | daga | /daga/ | 'goods' |
|-----------|------|--------|----------------|------|--------|----------|
| /p/ ~ /t/ | 'apa | /?apa/ | 'what' | 'ata | /?ata/ | 'person' |

Also, there is a contrast between labial /b, p/ and velar stops /k, g/, respectively.

| /b/ ~ /g/ | ba'i | /baʔi/ | 'bitter' | ga'i | /ga?i/ | 'cheerful' |
|-----------|-------|---------|----------|-------|---------|------------|
| /p/ ~ /k/ | mbopo | /mbopo/ | 'blue' | mboko | /mboko/ | 'seed' |

Finally, there is a contrast between alveolar /d, t/ and velar stops /g, k/, respectively.

| /d/ ~ /g/ | dega | /dəga/ | 'clear' | gera | /gəra/ | 'angry' |
|-----------|-------|---------|---------|-------|---------|---------|
| /t/ ~ /k/ | mbeta | /mbəta/ | 'buy' | mbeka | /mbəka/ | 'flood' |

Implosive stops

Ende has two stop phonemes with ingressive airstream, that is, the voiced labial implosive /6/ and the voiced alveolar implosive /d/. These phonemes contrast in word-initial and intervocalic positions.

| /6/ ~ /d/ | bhéa | /6ea/ | ʻslogan' | dhéo | /deo/ | 'to hold' |
|-----------|-------|--------|-------------|-------|--------|-----------|
| | sodho | /sodo/ | 'to inform' | sobhé | /sobe/ | 'to dive' |

There is a contrast between the voiced bilabial implosive /6/ and voiced and voiceless non-complex stops /b, p/ in word-initial and intervocalic positions.

| /6/ ~ /b/ | boti | /boti/ | 'bottle' | bhoti | /6oti/ | 'to fill (a hole)' |
|-----------|-------|--------|-----------|-------|--------|--------------------|
| | 'ebe | /?əbe/ | 'they' | ébho | /e6o/ | 'long (time)' |
| /6/ ~ /p/ | bhéa | /6ea/ | ʻslogan' | pé'a | /pe?a/ | 'waste' |
| | tebhé | /təbe/ | 'to chug' | tepé | /təpe/ | 'k.o. fish' |

Further, there is a distinction between the voiced alveolar implosive /d/and voiced and voiceless alveolar non-complex stops /d, t/ in word-initial and intervocalic positions.

| /d/ ~ /d/ | du'u | /du?u/ | 'to stop' | dhu | /ɗu:/ | 'until' |
|-----------|------|--------|---------------|-------|--------|-------------|
| | kodo | /kodo/ | 'sago branch' | kodho | /kodo/ | 'to see' |
| /d/ ~ /t/ | tiko | /tiko/ | 'every' | dhéko | /deko/ | 'to follow' |
| | peté | /pəte/ | 'to tie' | pédhé | /pede/ | 'to cook' |

Prenasalized stops

Complex prenasalized stops /mb, nd, ng/ contrast with non-complex stops /b, d, g/ and nasals /m, n, ng/. Prenasalized stops are analyzed here as a single unit in word-initial and intervocalic positions because nasals are only ever followed by voiced non-complex stop consonants of the same place of articulation and show minimal pair contrasts with non-complex stops and nasals.

| $^{\rm m}b/\sim /b/$ | mboti | /mboti/ | 'spinach' | boti | /boti/ | 'bottle' |
|-------------------------|--------|---------|------------|-------|--------|-----------|
| | kombé | /kombe/ | 'night' | 'ebé | /?əbe/ | 'to pick' |
| $^{m}b/\sim/m/$ | mbura | /mbura/ | 'burned' | bura | /bura/ | 'spit' |
| | rhamba | /1amba/ | 'drum' | rhama | /ɹama/ | 'quick' |
| / ^m b/ ~ /6/ | mbana | /mbana/ | ʻgo, walk' | bhara | /6ara/ | 'white' |
| | 'émbo | /?embo/ | 'wave' | ébho | /e6o/ | 'late' |

| $/^{n}d/\sim/d/$ | ndeka | /ndəka/ | 'at, in' | dega | /dəga/ | 'clear' |
|------------------------|--------|----------|--------------|-------|--------|--------------|
| | penda | /pənda/ | 'brag' | peda | /pəda/ | 'k.o. fruit' |
| $/^{n}d/\sim/n/$ | ndeka | /ndəka/ | 'at, in' | neka | /nəka/ | 'wound' |
| | senda | /sənda/ | 'to weave' | sené | /səne/ | ʻlid' |
| $/^{n}d/\sim/d/$ | nduki | /nduki/ | 'persistent' | dhuku | /ɗuku/ | 'to squat' |
| | penda | /pənda/ | 'brag' | pédhé | /pede/ | 'to cook' |
| | | | | | | |
| $/^{\eta}g/\sim/g/$ | ngguru | /ŋguru/ | 'ring' | guru | /guru/ | 'teacher' |
| | tenggo | /təŋgo/ | 'strong' | tegu | /təgu/ | 'to swallow' |
| $/^{\eta}g/\sim/\eta/$ | ngguru | /ŋguru/ | 'ring' | nguru | /ŋuru/ | 'smell' |
| | ronggo | /ro¹¹go/ | 'backside' | rongo | /roŋo/ | 'goat' |

The Glottal stop

The glottal stop /?/ is both phonemic in the word-initial and intervocalic positions. Truly vowel initial words, that is, words lacking a glottal onset, show different phonetic properties in the vowel, such as intensity, length, and breathiness. These properties are assumed here to result from the lack of an initial glottal stop. However, this aspect of Ende phonology needs further investigation. This issue is discussed further in §2.1.2. Minimal pairs for the glottal stop are provided below.

| /?/ ~ /ø/ | 'iwa | /?iwa/ | 'not' | iwa | /iwa/ | 'year' |
|-----------|-------|--------|-------------|--------|---------|--------|
| | nggaé | /¹gae/ | 'to search' | ngga'é | /¹ga?e/ | 'God' |

Furthermore, the glottal stop contrasts with all other stop consonants, complex or otherwise.

| / ? / ~ /b/ | 'apa | /?apa/ | 'what' | baba | /baba/ | 'father' |
|-----------------------------|-------|--------|-------------------|--------|---------|-----------------|
| / ? / ~ / p / | pa'a | /pa?a/ | 'offering' | papa | /papa/ | 'side' |
| /? / ~ /t/ | rha'é | /sa?e/ | 'place' | rhatu | /satu/ | 'to exist' |
| / ? / ~ / d / | 'ari | /?ari/ | 'younger sibling' | dari | /dari/ | 'to stand' |
| / ? / ~ / g / | pa'a | /pa?a/ | 'offering' | baga | /baga/ | 'to step' |
| / ? / ~ / k / | no'o | /no?o/ | 'aunt' | noko | /noko/ | 'thin' |
| $/?/\sim/^mb/$ | rha'é | /ла?е/ | 'space' | rhambu | /ɹambu/ | 'shirt' |
| $/2/\sim/^{n}d/$ | 'até | /?ate/ | 'liver' | ndaté | /ndate/ | 'heavy' |
| /?/ ~ /ŋ/ | wa'u | /wa?u/ | 'go downward' | wangu | /waŋu/ | 'shape of face' |
| /?/ ~ /¹¹g/ | 'aé | /?ae/ | 'water' | nggaé | /ŋgae/ | 'to search' |
| /?/ ~ /6/ | té'é | /te?e/ | 'mat' | tebhé | /tə6e/ | 'to drink' |
| /?/ ~ /d/ | so'o | /so?o/ | 'more' | sodho | /sodo/ | 'to inform' |

2.1.1.2 The Affricate

The only affricate in Ende is the voiced palatal affricate /ʤ/, which occurs in all possible consonant positions. The affricate contrasts with voiced alveolar and velar stop phonemes.

| /ʤ/ ~ /d/ | ja'o | /dza?o/ | 'I' | da'é | /da?e/ | 'arrogant' |
|---------------------|-------|---------|--------------|--------|---------|----------------|
| | ngaju | /ŋaʤu/ | 'to chew' | ngadu | /ŋadu/ | 'to look up' |
| /ʤ/ ~ /g/ | ja | /dza:/ | 'bright' | ga'i | /ga?i/ | 'cheerful' |
| /ʤ/ ~ /ªd/ | joro | /czoro/ | 'to push on' | ndore | /ndore/ | 'to pass' |
| $/dg/\sim/^{n_j}g/$ | ja'o | /d5a?o/ | 'I' | ngga'é | /ŋga?e/ | 'God' |
| /ʤ/ ~ /6/ | jo'é | /d5o?e/ | 'almost' | bhojé | /60dge/ | 'short person' |
| /ʤ/ ~ /ɗ/ | ngaju | /ŋaʤu/ | 'to chew' | ngadho | /ŋaɗo/ | 'to nod' |

2.1.1.3 Fricatives

Fricatives in Ende include a voiceless labiodental fricative /f/, a voiceless alveolar fricative /s/, and a voiced velar fricative / γ /. There is also a glottal fricative /h/ in the Aku dialect with a limited distribution, which is left for discussion in the next section. Besides the fricative phonemes, there are voiced bilabial fricative [β] and voiced alveolar fricative [γ] allophones that occur in free variation with the labiovelar approximant / γ / phonemes, respectively (see §2.1.1.6 for discussion).

The voiceless labiodental, voiceless alveolar, and voiced velar fricatives all show contrast.

| $/s/ \sim /f/$ | sai | /sai/ | 'who' | fai | /fai/ | 'wife' |
|----------------|------|--------|---------|------|-------|----------|
| /f/ ~ /ɣ/ | fu | /fu:/ | 'hair' | ghu | /yu/ | 2sg.poss |
| /s/ ~ /y/ | sa'o | /sa?o/ | 'house' | ghao | /yao/ | 'to mix' |

The voiceless labiodental fricative /f/ shows contrast with all other labial stops. The phoneme /f/ is restricted to word-initial position and does not ever occur in any other position in the word, that is, the intervocalic position.

| /f/ ~ /p/ | fu | /fu:/ | 'hair' | pu'u | /pu?u/ | 'trunk' |
|-------------------------|------|--------|----------|------|---------|----------|
| /f/ ~ /b/ | fai | /fai/ | 'female' | ba'i | /baʔi/ | 'bitter' |
| /f/ ~ /6/ | fi'i | /fi?i/ | 'plate' | bhiu | /6iu/ | 'sore' |
| /f/ ~ / ^m b/ | fu | /fu:/ | 'hair' | mbu | / mbu:/ | 'drunk' |

The voiceless alveolar fricative /s/ contrasts alveolar stop consonants and is not restricted in its distribution in the word. Contrasts for /s/ are provided below for all alveolar stops and the palatal fricatives.

| /s/ ~ /t/ | so'o | /so?o/ | 'more' | to'o | /to?o/ | 'get up' |
|-------------------|-------|--------|--------------|--------|--------|--------------|
| | gheso | /yəso/ | 'to release' | ghetto | /yəto/ | 'to slice' |
| /s/ ~ /d/ | ngasu | /ŋasu/ | 'hundred' | ngadu | /ŋadu/ | 'to look up' |
| /s/ ~ /d/ | koso | /koso/ | 'to rub' | kodho | /kodo/ | 'to watch' |
| $/s/ \sim /^n d/$ | sia | /sia/ | 'clear' | ndia | /ndia/ | 'here' |
| /s/ ~ /ʤ/ | muso | /muso/ | 'to inhale' | mujo | /mu&o/ | 'to rinse' |

The voiced velar fricative $/\sqrt{\gamma}$ contrasts with all other velar stop and palatal affricate consonant phonemes.

| /y/ ~ /g/ | ghawé | /yawe/ | 'to pull' | ga'i | /gaʔi/ | 'cheerful' |
|--------------------------|--------|---------------|-------------|-------|--------|---------------|
| | dhegho | /oyeb/ | 'to huddle' | dhego | /ɗəgo/ | 'old' |
| /y/ ~ /k/ | gheso | $\log \gamma$ | 'to fall' | kesa | /kəsa/ | 'increase' |
| | mogha | /moya/ | 'also' | moké | /moke/ | 'alcohol' |
| $/\gamma/\sim/^{\eta}g/$ | ghawé | /yawe/ | 'to pull' | nggaé | /¹gae/ | 'to look for' |
| /y/ ~ /ʤ/ | ghawé | /yawe/ | 'to pull' | ja | /dza:/ | 'bright' |

Glottal fricative /h/

The Aku dialect of Ende utilizes the glottal fricative /h/, which is analogous to the labiodental fricative /f/ in the Ja'o dialect studied here (Aoki and Nakagawa, 1993). For instance, Aoki and Nakagawa (1993) note hai 'female' in the Aku dialect is analogous to fai 'female' in the Ja'o dialect. Variation also occurs between vowels with glottal-less onsets, described in the §2.1.2, and the glottal fricative /h/. Since these vowels are restricted to word-initial positions, this variation only occurs in the word-initial position. For instance, in comparing the lexicon of the Ja'o dialect and the Aku dialect of Aoki and Nakagawa inga 'ears' and ao 'noisy' in the Ja'o dialect are analogous to hinga 'ears' and hao 'noisy' in the Aku dialect. It appears that the glottal fricative /h/ is not part of the phoneme inventory of the Ja'o dialect, but is included in the Aku dialect. Further investigation is still needed to determine the nature of glottal fricatives and vowels that a glottal onset.

2.1.1.4 Nasals

There are three nasal phonemes in Ende, the bilabial /m/, the alveolar /n/, and the velar /ŋ/. All nasals are voiced and occur in all possible consonant positions. Nasal phonemes most crucially contrast with prenasalized stops to constitute separate phonemes for each, as shown in §2.1.1.1. Furthermore, nasal phonemes contrast in the three aforementioned places of articulation.

| /m/ ~ /n/ | mata | /mata/ | 'eyes' | nata | /nata/ | 'betel nut' |
|-----------|------|--------|--------|-------|--------|--------------|
| /m/ ~ /ŋ/ | mata | /mata/ | 'eyes' | ngata | /ŋata/ | 'to breathe' |
| /n/ ~ /ŋ/ | manu | /manu/ | 'bird' | mangu | /maŋu/ | 'post' |

The bilabial nasal phoneme /m/ contrasts with the voiced bilabial oral stop consonants and the voiceless labiodental fricative /f/.

| /m/ ~ /b/ | mota | /mota/ | 'areca nut' | boti | /boti/ | 'bottle' |
|-----------|-------|--------|-------------|-------|--------|-------------|
| /m/ ~ /6/ | marha | /тала/ | 'to take' | bharo | /baro/ | 'reluctant' |
| /m/ ~ /f/ | mai | /mai/ | 'to come' | fai | /fai/ | 'female' |

The alveolar nasal /n/ shows contrast with voiced alveolar stop consonants, the palatal affricate, and the voiceless alveolar fricative.

| /n/ ~ /d/ | manu | /manu/ | 'bird' | madu | /madu/ | 'honey' |
|------------|------|--------|---------|-------|--------|-------------|
| /n/ ~ /d/ | neka | /nəka/ | 'wound' | dhéko | /deko/ | 'to follow' |
| /n/ ~ /dz/ | na | /na:/ | 'this' | ja | /dga:/ | 'bright' |
| /n/ ~ /s/ | no'o | /no?o/ | 'aunt' | so'o | /so?o/ | 'more' |

The velar nasal $/\eta$ / contrasts with the palatal affricate, voiced velar stop, the glottal stop, and the voiced velar fricative.

| /ŋ/ ~ /g/ | ngambé | /ŋa ^m be/ | 'to sit' | gambé | /gambe/ | 'ingredient in sirih' | |
|-----------|--------|----------------------|-------------|---------|---------|-----------------------|--|
| /ŋ/ ~ /?/ | mangu | /maŋu/ | 'post' | ma'a | /ma?a/ | 'serving' | |
| /ŋ/ ~ /ɣ/ | ngai | /ŋai/ | 'to breathe | ' ghawé | /yawe/ | 'to pull' | |

2.1.1.5 Trill/flap

The alveolar flap [r] and the alveolar trill [r] are allophones of a single phoneme that can occur as a single flap [r] or multiple flaps, producing a trill [r]. This phoneme, represented as the flap /r/ can occur in all possible consonant positions in the word. The flap /r/ contrasts with voiced and voiceless alveolar stops and the palatal affricate.

| /r/ ~ /d/ | ngaru | /ŋaru/ | 'bored' | ngadu | /ŋadu/ | 'to look up' |
|-------------------|--------|---------|----------------|--------|---------|---------------|
| /r/ ~ /?/ | ngara | /ŋara/ | 'name' | 'ata | /?ata/ | 'person' |
| /r/ ~ /d/ | koro | /koro/ | 'chili pepper' | kodho | /kodo/ | 'to watch' |
| $/r/ \sim /^n d/$ | ri'a | /ri?a/ | 'good' | ndia | /ndia/ | 'here' |
| /r/ ~ /ʤ/ | ngguru | /ŋguru/ | 'ring' | ngguju | /ŋguʤu/ | 'mumble' |
| /r/ ~ /n/ | roko | /roko/ | 'to hide' | noko | /noko/ | 'skinny' |
| /r/ ~ /s/ | kera | /kəra/ | 'to shout' | kesa | /kəsa/ | 'to increase' |

2.1.1.6 Approximants

There are two primary approximants in Ende, the labiovelar approximant /w/ and the alveolar approximant /ɪ/, and one more restricted approximant, the lateral approximate /l/. The two primary approximants, the labiovelar /w/ and the alveolar /ɪ/, both produce a high level of turbulent airflow during articulation. Therefore, these approximant phonemes at first pass sound as if they are fricatives. In fact, the turbulent airflow present on each phoneme in some instances gives rise to allophonic variation between the approximant and the respective fricative. That is, for /w/ there is occasionally allophonic variation between [u] and [l] and [l] and [l] and [l] arise when the level of turbulence is increased by a more restricted closure between the passive and active articulators, which in turn causes the approximant properties to become convoluted or lost altogether. The environment for allophonic variation between approximants and fricatives is not predictable.

The voiced labiovelar approximant /w/ is created by the constriction of the velar and lips. During articulation the lips are unrounded. This constriction of the lips creates the increased turbulence in the labiovelar approximant. The increased turbulence creates a marked difference between the labiovelar approximant phoneme and the analogous epenthetic labiovelar approximant that arises between a set of adjacent vowels (see §2.1.2.8). The following contrasts establish the labiovelar approximant /w/ as a separate phoneme. The labiovelar approximant /w/ shows contrast with the approximant alveolar /x/ in all possible consonant positions.

| $/\mathrm{W}/\sim/\mathrm{J}/$ | wé'é | /we?e/ | 'near' | rhé'é | /.ie?e/ | 'sharp' |
|--------------------------------|-------|--------|--------|--------|---------|----------|
| | rhéwa | /iewa/ | 'long' | rhérha | /лела/ | 'to fly' |

The labiovelar approximant /w/ also contrasts with voiced and voiceless bilabial and velar stops. Most crucially, /w/ contrasts with the bilabial stop /b/ in word-initial and intervocalic positions.

| $/w/ \sim /b/$ | waka | /waka/ | 'charismatic' | baga | /baga/ | 'to step' |
|----------------|-------|---------|---------------|--------|---------|---------------------|
| | ndewé | /ndewe/ | 'later' | 'ebé | /?əbe/ | 'they' |
| /w/ ~ /p/ | wé'é | /we?e/ | 'near' | pé'a | /pe?a/ | 'waste' |
| $/w/\sim/^mb/$ | weta | /wəta/ | 'sister' | mbeta | /mbəta/ | 'to buy' |
| /w/ ~ /6/ | warho | /wa.io/ | 'to return' | bharho | /6a.io/ | 'reluctant' |
| /w/ ~ /k/ | rhawo | /Jawo/ | 'sarong' | rhako | /aako/ | 'dog' |
| /w/ ~ /g/ | wawi | /wawi/ | 'pig' | gawi | /gawi/ | 'traditional dance' |

The labiovelar approximant also shows contrast between the labiodental fricative f/ and the velar fricative f/.

| /w/ ~ /f/ | wi | /wi:/ | 'will' | fai | /fai/ | 'female' |
|-----------|--------|----------|---------|-------|--------|--------------|
| /w/ ~ /y/ | weso'o | /wəso?o/ | 'later' | gheso | /osey/ | 'to release' |

The labiovelar approximant contrasts with the bilabial nasal /m/ and the velar nasal /n/.

| /w/ ~ /m/ | rhewa | /awa/ | 'to peel' | rhema | /ɹəma/ | 'tongue' |
|-----------|-------|-------|-----------|-------|--------|-------------|
| /w/ ~ /ŋ/ | wi | /wi:/ | 'will' | ngai | /ŋai/ | 'to breath' |

There is also a contrast between the presence and the absence of an intervocalic labiovelar approximant /w/ in the vowel sequences /ua/ and /ui/, where the vowel sequence /uo/ is unattested in Ende (see §2.1.2.8). In fact, in Aoki and Nakagawa's (1993) Ende-English dictionary the alveolar approximant /w/ is only found only once in a /uwo/ sequence in more than 2,500 lexical items in the lexical item *suwo* 'ear ornament'. The increased turbulence and lack of lip-rounding on the labiovelar approximant /w/ creates a clear contrast between vowel sequences with and without the labiovelar approximants.

 $/w/ \sim /\emptyset/$ nuwa /nuwa/ 'age' nua /nua/ 'village'

The alveolar approximant /ɪ/ is produced when the tongue tip is brought close to the alveolar ridge, creating turbulence in the airflow between the tongue tip and the alveolar ridge. As was noted earlier in this section, the alveolar approximant shows a high level of turbulence, which shows a marked contrast with say the English alveolar approximant /ɪ/. The increased turbulence appears to result from the increased level of constriction between the tongue tip and the alveolar ridge. The following contrasting segments establish the alveolar approximant /ɪ/ as a phoneme. The alveolar approximant /ɪ/ contrasts with all of the alveolar stops /d, t, ⁿd, d/.

| / J / ~ / d / | ngarha | /ŋa.ɪa/ | 'can' | ngadu | /ŋadu/ | 'to look up' |
|-----------------------------|--------|---------|-------------|-------|---------|--------------|
| / J / ~ /t/ | marha | /ma.ra/ | 'to take' | mata | /mata/ | 'eyes' |
| $/I/ \sim /^n d/$ | perha | /pə.a/ | 'to wear' | penda | /pənda/ | 'to shout' |
| / J / ~ / d / | kerho | /orey/ | 'to forget' | kedhu | /kədo/ | 'to open' |

The alveolar approximant /x/ contrasts with the alveolar fricative /s/.

| / J / ~ /S/ | rho | / IO :/ | 'body' | so'o | /so?o/ | 'more' |
|--------------------|-------|----------------|-----------|------|--------|----------|
| | perha | /pə.a/ | 'to wear' | pesa | /pəsa/ | 'to eat' |

The alveolar approximant /ı/ contrasts with the alveolar nasal /n/.

 $/x/ \sim /n/$ rhé'é /xe?e/ 'sharp' né'é /ne?e/ 'aunt'

The alveolar approximant /ı/ also forms a contrast with the palatal affricate /dʒ/.

/x/~/dz/ marha /maxa/ 'to take' ngaju /ŋadzu/ 'to chew'

The alveolar approximant / I contrasts with the flap / f in word-initial and intervocalic positions.

| /1/ ~ /r/ | rhé'é | /se?e/ | 'sharp' | ré'é | /re?e/ | 'bad' |
|-----------|-------|--------|---------|--------|---------|-------|
| | ngara | /nara/ | 'name' | ngarha | /na.ra/ | 'can' |

The lateral /l/ is far more common in Lio than in Ende. Cognates that occur in Lio with the lateral /l/ occur in Ende as the alveolar approximant /J/. For instance, *lawo* 'sarong' in Lio forms a cognate *rhawo* 'sarong' in Ende, and the Lio numeral *telu* 'three' is a cognate with the Ende *terhu* 'three'. In spite of numerous cognates, there is still a small subset of Ende words that contain /l/ that have no /J/ equivalent. Therefore, the lateral approximant /l/ is considered a phoneme, even though it is most likely borrowed from Lio. The lateral /l/ contrasts with the alveolar stop and the palatal affricate phonemes.

| /1/ ~ /d/ | lo | /lo:/ | 'market' | doi | /doi/ | 'money' |
|----------------|-----------|-------|-----------|------|--------|-------------|
| /1/ ~ /d/ | (lala)-lé | /le:/ | 'to joke' | dhéo | /deo/ | 'to hold' |
| $/1/\sim/^nd/$ | lo | /lo:/ | 'market' | ndo | /ndo:/ | 'to dangle' |
| /l/ ~ /ʤ/ | lo | /lo:/ | 'market' | jo | /dzo/ | '1sg. poss' |

The lateral /l/ also contrasts with alveolar fricative /s/.

 $/1/ \sim /s/$ lo'o /lo?o/ 'small' so'o /so?o/ 'more'

Further, the lateral phoneme /l/ contrasts with the alveolar nasal /n/ and flap /r/ phoneme.

| / l / ~ / n / | lo'o | /lo?o/ | 'small' | no'o | /no?o/ | 'aunt' |
|-----------------------------|------|--------|---------|------|--------|----------------|
| /l/ ~ /r/ | gela | /gəla/ | 'glass' | gera | /gəra/ | 'to get angry' |

Most important in establishing the lateral approximant as a phoneme is the contrast between the lateral /I/ and the alveolar approximant /I/ in word-initial and intervocalic positions.

| / J / ~ / l / | rho | \lor\ | 'body' | lo | /lo:/ | 'market' |
|-----------------------------|-----------|-------|-----------|------|--------|----------|
| | (lala)-lé | /le:/ | 'to joke' | arhé | /a.ie/ | 'to fly' |

2.1.2 Vowels

There are six vowels in Ende presented in Table 2.2 below.

Table 2.2: Vowel Phoneme Chart

| | Front | Central | Back |
|------|-----------|-------------|-----------|
| High | i <i></i> | | u <u></u> |
| Mid | e <é> | (a) <e></e> | 0 <0> |
| Back | | a <a> | |

⁵ The Lio cognates are taken from Sawardo et. al. (1987: 50, 35).

The six vowel phonemes in Ende show a distinction between the five basic vowels /i, e, a, o, u/ and the epenthetic mid-central vowel [ə]. There are several reasons for this distinction. First, the mid-central vowel is phonetically shorter than all other vowels and never takes stress or prominence in the root or the word. Secondly, the mid-central vowel is limited in its distribution; it only occurs between two consonants. This means that 1) the mid-central vowel cannot occur in vowel sequences, which is not true of any of the other basic vowels, and 2) the mid-central vowel cannot occur without an initial glottal stop, which again is possible for all other vowels. Lastly, the schwa is limited to the first syllable of bisyllabic words as well as the numeral proclitic sa /sə/ and the prepositional proclitic pa /pə/; that is, it is disallowed in the ultimate syllable. On the other hand, the five basic vowels occur in all possible nucleus positions in syllables.

2.1.2.1 The /i/

The high front unrounded vowel /i/ contrasts with the four other basic vowels and the schwa

| /i/ ~ /u/ | mesi | /məsi/ | 'salty' | mesu | /məsu/ | 'to fall' |
|-----------|-------|--------|-----------------|-------|--------|-------------|
| /i/ ~ /e/ | moki | /moki/ | 'mouth' | moké | /moke/ | 'palm wine' |
| /i/ ~ [ə] | mité | /mite/ | 'black' | meta | /məta/ | 'unripe' |
| /i/ ~ /a/ | weti | /wəti/ | 'to chew sirih' | weta | /wəta/ | 'sister' |
| /i/ ~ /o/ | kérhi | /ke.i/ | 'mountain' | kerho | \orea\ | 'to forget' |

2.1.2.2 The /u/

The high back rounded vowel /u/ shows contrast with the other basic vowels in and the schwa.

| /u/ ~ /e/ | té'u | /te?u/ | 'rat' | té'é | /te?e/ | 'mat' |
|-----------|-------|----------------------|----------|-------|---------|----------|
| /u/ ~ [ə] | muta | /muta/ | 'vomit' | meta | /məta/ | 'unripe' |
| /u/ ~ /a/ | 'əmbu | /?əmbu/ | 'elder' | 'émba | /?emba/ | 'where' |
| /u/ ~ /o/ | rembu | /rə ^m bu/ | 'finish' | rembo | /cəmbo/ | 'fat' |

2.1.2.3 The /e/

The mid front unrounded vowel /e/ contrasts with the basic vowels and the schwa.

| /e/ ~ [ə] | méré | /mere/ | 'big' | meré | /məre/ | 'tight' |
|-----------|------|--------|-------------|------|--------|----------|
| /e/ ~ /a/ | kéré | /kere/ | 'different' | kéra | /kera/ | 'turtle' |
| /e/ ~ /o/ | arhé | /a.ie/ | 'to ask' | arho | /a.io/ | 'orphan' |

2.1.2.4 The /o/

The mid back round vowel /o/ shows contrast with other basic vowels and the schwa.

| /o/ ~ [ə] | pota | /pota/ | 'to disappear' | peté | /pəte/ | 'to tie' |
|-----------|-------|---------|----------------|-------|---------|-----------|
| /o/ ~ /a/ | tembo | /təmbo/ | 'body' | temba | /təmba/ | 'to wash' |

2.1.2.5 The /a/

The back vowel /a/ contrasts with all other basic vowels. Within the root, /a/ contrasts with the schwa. However, in the numeral clitic sa and the prepositional clitic pa 'at', there is allophonic variation between /a/ and /ə/. That is, most commonly the vowel is realized as a schwa in these proclitics, but in careful pronunciation the vowel can be realized as the back vowel /a/.

| /ar e mata / mata/ cycs meta / meta/ umpe | /a/ ~ [ə] | mata | /mata/ | 'eyes' | meta | /məta/ | 'unripe' |
|---|-----------|------|--------|--------|------|--------|----------|
|---|-----------|------|--------|--------|------|--------|----------|

2.1.2.7 Phonetic features of word initial vowels

As discussed in §2.1.1.1., the absence of the glottal onsets affects the phonetic quality of the vowel. More specifically, vowels without a glottal onset begin with a steady increase in intensity, while the vowels with a glottal onset show a much more abrupt beginning and starts at a much higher intensity. Other phonetic qualities of vowels that lack a glottal onset are a longer duration accompanied by lower pitch and breathiness on the vowel. The contrast between these two sets occurs between all of the basic vowels /i, e, a, o, u/. The only vowel that does not show such a contrast is the schwa. Vowel initial words, represented without a glottal stop, contrast with vowels that have a glottal onset, which are represented with a preceding glottal stop in the right-hand column below.

| /i/ ~ /?i/ | iwa | /iwa/ | 'year' | 'iwa | /?iwa/ | 'not' |
|------------|------|--------|-------------|-------|---------|-------------------|
| /e/ ~ /?e/ | éra | /era/ | 'wide' | 'éra | /?era/ | 'children's tool' |
| /a/ ~ /?a/ | arhé | /a.ie/ | 'to fly' | 'arhé | /?a.ie/ | 'to ask' |
| /o/ ~ /?o/ | orho | /oro/ | 'sorghum' | 'orho | /010/ | 'old, used' |
| /u/ ~ /?u/ | uma | /uma/ | 'k.o. tree' | 'uma | /?uma/ | 'dry field' |

2.1.2.8 Vowel sequences

All logically possible vowel combinations occur in Ende except for /uo/. Furthermore, the schwa /ə/ does not combine with any other vowel. The vowel sequences made up the lower vowels /a/ or /o/ and a high vowel /i/ or /u/ are analyzed as diphthongs; that is, /ai/, /au/, /oi/, and /ou/ are analyzed as diphthongs. These diphthongs predominately occur in single syllable words. It is only in the case of the vowel sequence /ai/ that some examples emerge in disyllabic roots, such as *numai* 'yesterday'. All other vowel sequences occur as two syllables. However, when the vowel sequence is made up of the mid front vowel /e/ and a

⁶ The closest example of the vowel sequence /uo/ is *suwo* 'ear ornament', but /w/ in this word is the phonemic /w/ and not the epenthetic glide [w]. See §2.1.1.6 for more discussion of this distinction in Ende.

210 Bradley McDonnell

high vowel /i/ or /u/, the sequence is not analyzed a diphthong, but a two syllable sequence that inserts an epenthetic palatal glide [y]. In other vowel sequences, where a front vowel /i/ or /e/ combines with any other vowel, an epenthetic glide [y] is interested between the two vowels. This includes /ia/, /ie/, /io/, /iu/, /ea/, /ei/, /eo/, and /eu/. Vowel sequences that are made up of a back vowel /u/ or /o/ and any other vowel insert the epenthetic labiovelar glide [w] between the two vowels. This includes the sequences /ui/, /ue/, /ua/, /oe/, and /oa/. It is important to note that the epenthetic glide [w] in vowel sequences differs phonetically from the approximate /w/. The approximant /w/ has much more turbulence present on the phoneme than the epenthetic /w/. See §2.1.1.6 for the phonetic properties of approximants in Ende. Lastly, vowel sequences that consist of a low vowel /a/ and a mid vowel /e/ or /o/ do not insert any transitioning glide. Examples of vowel sequences are provided in Table 2.3.

Table 2.3: *Vowel Sequences*

| Vowel Sequence | Orthography | IPA | Gloss |
|----------------|-------------|---------|------------------|
| oi | moi | [moyi] | 'to store' |
| ai | fai | [fayi] | 'female' |
| ou | bou | [bowu] | 'together' |
| au | kau | [kawu] | 'you' |
| éi | téi | [teyi] | 'to see' |
| iu | niu | [niyu] | 'to call' |
| éu | éu | [eyu] | 'areca palm nut' |
| éo | dhéo | [ɗeyo] | 'to hold' |
| io | nio | [niyo] | 'coconut' |
| ié | (sésé-)sié | [siye] | 'blackbird' |
| éa | bhéa | [6eya] | 'motto' |
| ia | sia | [siya] | 'bright' |
| ui | sui | [suwi] | 'to bake' |
| ua | rua | [ruwa] | 'two' |
| oa | ngoa | [ŋowa] | 'toothless' |
| ué | mbué | [mbuwe] | 'beans' |
| oé | ngoé | [ŋowe] | 'hungry' |
| ao | ngao | [ŋao] | 'whisper' |
| aé | nggaé | [ŋgae] | 'to look for' |
| uo | ** | ** | **** |

2.2 Syllable and stress

The syllable in Ende is minimally V and maximally CVV (where VV represents diphthongs and long vowels); codas are prohibited in native lexical items. If a coda is present on a lexical item, that lexical item is surely a loan word from Indonesian. The only exception to the prohibition on codas is word-medial prenasalized stops /mb, nd, ng/. The nasal consonant forms a coda with initial syllable, while the stop consonant makes up the onset of the following syllable. For instance, *tembo* 'body' is phonetically realized as [təm.ba] and not *[tə.mba]. The possible syllable shapes in Ende are represented in Table 2.5.

Table 2.5: Syllable shapes

| V | arhé | /a.ie/ | 'to fly' | |
|-----|--------------|--------|----------|--|
| CV | ba ba | /baba/ | 'father' | |
| CV: | naa | /na:/ | 'this' | |

Roots in Ende are most always monosyllabic or disyllabic; only some native roots have three syllables, but these are most commonly borrowed words. Single syllable words must minimally be bimoraic. That is, monosyllabic words must have either a long vowel or in some cases show a diphthong. Only clitics can appear as monomoraic, but these are phonologically dependent (see §2.3). The possible word shapes are exemplified in Table 2.6.

Table 2.6: Word shapes

| V: | é | [e:] | 'to think' |
|----------|---------|-----------|------------|
| V.V | éa | [eya] | 'cucumber' |
| CV: | naa | [naː] | 'this' |
| V.CV | arhé | [a.ie] | 'to fly' |
| CV.V | ngaé | [ŋae] | 'to pick' |
| CV.CV | moki | [moki] | 'mouth' |
| CV.CV.CV | tumbé'é | [tumbe?e] | 'really' |

Finally, stress in Ende falls on the penultimate syllable, unless the penultimate syllable contains a schwa. In that case, stress falls on the ultimate syllable. Stress is usually marked by an increase in intensity and vowel length.

2.3 Clitics⁷

Ende, like other Austronesian languages of Flores, is an extremely isolating language that does not show any evidence of prefixes or suffixes. However, there do appear to be a number of clitics, or phonologically dependent words that attach at the phrase or clause level (see Zwicky, 1977). In Ende, all clitics appear to be function words or come from closed word classes. However, this correlation is not possible in the other direction where many function words are not considered clitics. The defining characteristics of clitics in Ende are: 1) clitics are one only syllable, 2) clitics have a phonetically short vowel, and 3) clitics are always unstressed. Clitics in Ende can either be enclitics or proclitics. The following sections outline some of the phonological properties of clitics. The following sections do not provide a complete grammatical account or provide a survey of all the clitics in Ende.

2.3.1 Enclitics

One clear distinction between the phonological word and the clitic in Ende is seen in the proximal demonstrative *na* 'this'. Depending on the placement in the sentence, the demonstrative can be considered a phonological word or a clitic. For example, the demonstrative emerges as a phonological word, i.e., it has a long vowel *naa*, in (1a), when

⁷ In only this section, some long vowels are represented by doubling the vowel, i.e., if <a> is a long vowel it is represented as <aa>. This is only used to make clear the difference between clitics and phonological words.

it acts a constituent on its own in the non-verbal predicate clause. On the other hand, when the demonstrative follows a noun as in (1b), it forms an enclitic on the head noun 'ana 'child'. If the head noun is followed by an adjective or any other modifier, the demonstrative cliticizes to the last element of the noun phrase as in (1c). Apparently, na 'this' can also act as a enclitic of the object of the verb, as in (1d).

- (1) a. **naa** ['ana kami] this child 1PL.EXCL 'This is our child'
 - b. 'ana=**na** mai warho tau 'apa ndé child=this come return make what again 'This child came home to make what again?'
 - c. ['ana lo'o]=**na** mai warho tau 'apa ndé child small =this come return make what again 'This small child came home to make what again?'
 - d. kai bharho kaa=na
 3SG don't.want eat=this
 'She doesn't want to eat this.'

Other enclitics include the perfective aspectual marker, the possessive ligature, and possessive pronouns. Of these three classes of clitics, the perfective aspectual marker and the possessive ligature still make a clear alternation between dependent clitic forms and independent word forms. In the case of the possessive ligature ko'o, the final syllable is lost and the vowel remains short, forming an enclitic with the previous word as in (2). It is important to note that words with the shape $CV_i ?V_i$, in which the two vowels are the same, commonly undergo the same syllable reduction. However, this reduction may be connected cliticization. See §2.3.3.1 for further discussion of syllable reduction in Ende.

- (2) a. ngara **ko'o** kau sai ndé name LIG 2SG who again 'What's your name again?'
 - b. ngara=**ko** kau sai ndé name=LIG 2SG who again 'What's your name again?'

The perfective adverb *peka* /pəka/'already' and the perfective aspectual marker *ka* undergo a similar reduction, but in this case the first syllable containing a schwa is lost as in (3) below.

(3) a. 'embu tegu **peka** aé sa=gela méré grandmother swallow already water NUM=glass big 'Grandmother had drunk a big glass of water.'

b. 'embu tegu=**ka** aé sa=gela méré grandmother swallow=ASP.PERF water NUM=glass big 'Grandmother had drunk a big glass of water.'

Another class of enclitics is the pronominal possessive enclitics. Ende has possessive enclitics for first, second, and third person singular pronouns and the third person plural pronoun. These pronouns alongside the canonical person pronouns are shown in Table 2.6 and examples of them in sentences are shown in (4).

| T. I.I. A C. D | | | | 1.,. |
|--------------------------|--------------|------------|---------|---------|
| Table 2.6: Person | pronouns and | possessive | pronoun | clitics |

| | Person Pronouns | Possessive Person Pronouns |
|----------|-----------------|-----------------------------------|
| 1SG. | ja'o | =jo/d5o/ |
| 2SG. | kau | =ghu /γu/ |
| 3SG. | kai | =ghi /γi/ |
| 1PL.INCL | kita | |
| 1PL.EXCL | kami | |
| 2PL. | miu | |
| 3PL. | 'ebé | =bé /be/ |

- (4) a. kai nggaé ['iné méré]=**ghi**3SG look.for mother big=3SG.POSS
 'He is looking for his aunt'
 - b. kau seka rhima=**ghu** né'é piso 2SG stab hand=2SG.POSS with knife 'You stabbed your hand with a knife'
 - c. ja'o kaa koro mbiraka dhuu tuka=**jo** rhoo 1SG eat spicy.vegetables too.many until stomach=1SG.POSS hurt 'I ate many spicy vegetables until my stomach hurt'
 - d. baba=**bé** rasi bha father=3PL.POSS wash plates 'Their father washes the plates.'

The possessive pronoun enclitics, however, do not make as clear of a phonological alternation between phonological word and clitic as in the previous examples. That is, one can only hypothesize as to phonological alternations between the person pronouns and the possessive pronouns, which most likely represents established diachronic change rather than synchronic alternation as in other cases of clitics. The only exception to this may be the third person plural pronoun 'ebé 'they', which merely drops the first syllable to produce the third person plural possessive pronoun be 'their', as was seen in the case of the aspectual marker. Further, the possessive enclitics show certain grammatical restrictions, as they cannot co-occur with the possessive ligature ko'o as in (5) below. It is important to note that these possessive enclitics cannot occur with any form of ko'o.

- (5) a. kai nggaé ['iné méré] *ko('o)=ghi
 3SG look.for mother big LIG=3SG.POSS
 'He is looking for his aunt' (Lit. He is looks for aunt of him)
 - b. baba *ko('o)=bé rasi bha father LIG=3PL.POSS wash plates 'Their father washes the plates.'

For more discussion of the grammatical behavior these enclitics see McDonnell (2008, forthcoming).

2.3.2 Proclitics

Proclitics in Ende are far less common than enclitics. It appears that both of the proclitics that have been found contain a schwa. This is the numeral proclitic sa /sə/ and the prepositional proclitic pa /pə/ 'at'. These proclitics are shown in (6) below. The extent to which proclitics occur in Ende is still unclear, but the proclitics below are common.

- (6) a. ana ja'o imu rua, **sa=**'imu=ghi 'ata fai child 1SG person two NUM=person=3SG.POSS person female 'I have two children, one (of them) is a girl.'
 - b. **pa**=ndia sa'o=jo at=here house=1SG.POSS 'Here is my house'

2.3.3 Phonological processes

2.3.3.1 Syllable reduction

Final syllables in Ende are commonly reduced in two syllable words with identical vowels separated by a glottal stop. That is, words of the shape: $CV_i?V_i$. For example, ne'e 'with' is commonly reduced to $n\acute{e}$ and the possessive ligature ko'o is commonly shortened to ko. Syllable reduction in these cases appears to be connected to cliticization processes, as the vowels in the aforementioned examples remain short vowels. This means that these examples would qualify as clitics (see §2.3 for the qualifying characteristics). Furthermore, in other examples, vowel reduction is not possible. For example, lo'o 'small' is never reduced to *lo.

2.3.3.2 Nasalization

In Ende, vowels following nasal consonants are always nasalized within the same syllable. That is, *mata* [mãta] 'eyes' shows nasalization on the first vowel, but not the second vowel. While in the word *rongo* [roŋõ] 'goat' the second vowel following the nasal is nasalized, but the first vowel is not nasalized. However, if a nasal precedes a vowel sequence, then both of the vowels in the sequence are nasalized. For example, in the word *nio* [nĩyõ]'coconut' both vowels are nasalized.

2.3.3.3 Lenition and voicing of /k/ in clitics

A common phonological process that occurs across a clitic boundary is that of the voiceless velar stop /k/ changing to the voiced velar fricative [γ]. Inside the phonological word, the voiced velar fricative / γ / contrasts with both voiced and voiceless velar stop phonemes /g/ and /k/ in intervocalic positions. However, across a clitic boundary, the voiced velar fricative / γ / occurs in free variation with the voiceless velar stop /k/. For example, the temporal adverb *peka* /pəka/ 'already' in (7a) is the source for the perfective aspectual marker ka /ka/, which acts as a clitic, as in (7b). The phoneme /k/ in this clitic occurs in free variation voiced velar fricative / γ /, which occurs as *gha* [γ a] in (7c). However, this variation never occurs within the temporal adverb *peka* /pəka/ 'already', a phonological word, as in (7d).

- (7) a. 'embu tegu **peka** aé se=gela méré grandmother swallow already water NUM=glass big 'Grandmother had drunk a big glass of water.'
 - b. 'embu tegu=**ka** aé se=gela méré grandmother swallow=ASP.PERF water NUM=glass big 'Grandmother had drunk a big glass of water.'
 - c. 'embu tegu=**gha** aé se=gela méré grandmother swallow=ASP.PERF water NUM=glass big 'Grandmother had drunk a big glass of water.'
 - d. 'embu tegu ***pegha** aé se=gela méré grandmother swallow already water NUM=glass big 'Grandmother had drunk a big glass of water.'

Lenition across the clitic boundary also occurs with the possessive ligature ko'o. The clitic form may occur as ko [ko] or gho [yo] as in (8b) and (8c), respectively. However, the full form only occurs as ko'o as in (8a), but never occurs as *gho'o as in (8d).

- (8) a. 'embu suka sené **ko'o** kepé grandmother open lid LIG box 'Grandmother opened the lid of the box.'
 - b. 'embu suka sené=**ko** kepé grandmother open lid=LIG box 'Grandmother opened the lid of the box.'
 - c. 'embu suka sené=**gho** kepé grandmother open lid=LIG box 'Grandmother opened the lid of the box.'

d. 'embu suka sené ***gho'o** kepé grandmother open lid LIG box 'Grandmother opened the lid of the box.'

3. Conclusion and future research

This paper has provided a preliminary description of the Ende phonology. First, the consonant and vowel phonemes and allophonic variants have been outlined with important minimal or near-minimal pairs provided for each phoneme. Second, issues related to phonotactics, i.e., the syllable, the phonological word, and stress have also been outlined. Lastly, special attention has been given to the phonological nature of clitics. However, as mentioned in §1.2, there are still many outstanding questions regarding Ende phonology. In particular, questions relating to the phonetics of the 1) word initial (or glottal-less) vowels and 2) the alveolar approximant is needed. Also, the productivity and pervasiveness of clitics in Ende is still largely unknown. For instance, can clitics be stacked? If so, how many are permitted and in what direction? Lastly, this description outlines two phonological alternations, but there is bound to be more that have not yet become clear. For example, Ende appears to devoice final vowel, but the nature of this is not clear enough to include in this paper. I hope that future studies build off of the present study of Ende phonology.

References

- Anon. 1855. Beschrijving van het eiland Soemba of Sandelhout. *Tijdschrift voor Nederlandsch-Indië*, 17: 277-312.
- Aoki, Eriko. 1988. The case of the purloined statues: the power of words among the Lionese. In James J. Fox, ed. *To speak in pairs: essays on the ritual languages of eastern Indonesia*. Cambridge: Cambridge University Press. pp. 2??-2??.
- Aoki, Eriko and Satoshi Nakagawa. 1993. Unpublished Endenese-English Dictionary. Ms.
- Arka, I Wayan, Jeladu Kosmas, and I Nyoman Suparsa. 2007. *Bahasa Rongga: Tatabahasa acuan ringkas*. Jakarta: Penerbit Universitas Atma Jaya.
- Arka, I Wayan. 2004. Palatography in a fieldwork setting: investigating and analysing alveolar continuant [r] and [s] in Rongga. In *Wibawa Bahasa*, edited by I. W. Pastika and I. N. Darma Putra, 40-50. Denpasar: Program Pascasarjana Linguistik, Universitas Udayana.
- Esser, S. J. 1938. Talen. In *Atlas van Tropisch Nederland*. Amsterdam: Koninklijk NederlandschAardrijkskundig Genootschap.
- Frazer, J.G. 1919. Folk-lore in the Old Testament: studies in comparative religion, legend, and law. 3 vols. London: Macmillan.
- Fox, James J. 1988. To speak in pairs: essays on the ritual languages of eastern Indonesia. Cambridge: Cambridge University Press. pp. 228-245.
- Baird, Louise. 2002. A grammar of Kéo: An Austronesian language of East Nusantara. Ph.D. dissertation, Australian National University.
- Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–293.
- Donohue, Mark. n.d. The prosodic structure of Palu'e. ms. http://users.monash.edu.au/~mdonohue/Palu'e/DonohuePalu'estress.pdf

Djawanai, Stephanus. 1977. A description of the basic phonology of Nga'da and the treatment of borrowings. *NUSA: Linguistic studies of Indonesian and other languages of Indonesia*, 5:10-18.

- Djawanai, Stephanus. 1983. Ngadha text tradition: the collective mind of the Ngadha *people, Flores*. Series D-55. Canberra: Pacific Linguistics.
- Esser, S. J. 1938. Talen. In *Atlas van Tropisch Nederland*. Amsterdam: Koninklijk NederlandschAardrijkskundig Genootschap.
- Gordon, Raymond G., Jr. (ed.), 2006. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International.
- Grimes, Charles E. 1991b. Central Malayo–Polynesian: an interim evaluation. Paper presented at the Sixth International Conference on Austronesian Linguistics, Honolulu, Hawaii.
- Grimes, Charles E., Tom Therik, Barbara Dix Grimes, and Max Jacob. 1997. *A guide to the people and languages of Nusa Tenggara*. Kupang: Artha Wancana Press.
- Hajek, John. 2008 (forthcoming). Towards a Phonological Typology of the Vowel and Consonant Systems of East Nusantara. In M. Klamer and M. Ewing, eds. *Typological and Areal Analysis: Contributions from East Nusantara*, Leiden: KITLV Press.
- Klamer, Marian.1998. A grammar of Kambera. Berlin/New York: Mouton de Gruyter.
- Klamer, Marian. 2002. Typical features of Austronesian languages in central/eastern languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41:363-383.
- Kennedy, Raymond. 1955. [Unpublished field notes on Endeh] E-5 (1949-50) OB1, Indonesia,
 - Flores, OF9. Human Relations Area Files, New Haven, Conn.
- McDonnell, Bradley J. 2008. Possessive Structures in Ende. In Lou Hohulin, ed. *Austronesian language papers, Studies in Philippines Language and Cultures*, vol. 18.
- Nakagawa, Satoshi. 1984. The Endenese. In Richard V. Weekes, editor, *Muslim Peoples:* A World Ethnographic Survey.
- Nakagawa, Satoshi. 1988b. The journey of the bridegroom: idioms of marriage among the Endenese. In James J. Fox, ed. *To speak in pairs: essays on the ritual languages of eastern Indonesia*. Cambridge: Cambridge University Press. pp. 228-245.
- Nakagawa, Satoshi. 2003. A bibliographical guide to peoples of Flores. http://bunjin6.hus.osaka-u.ac.jp/~satoshi/flores/guide/bibguide.html
- Needham, Rodney. 1970. Endeh, II: Test and confirmation. *Bijdragen tot de Taal-Landen Volkenkunde*, 126:246-258.
- Needham, Rodney. 1968. Endeh: Terminology, Alliance and Analysis. *Bijdragen tot de Taal-, Land- en Volkenkunde* 124:305-335.
- Roos, S. 1877. Iets over Endeh. Tijdschrift voor Indische Taal-, Land- en Volenkunde 24:481-582.
- Ross, Malcolm. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary*. *An introduction to Austronesian studies*, pt. 1: fasc. 1, ed. by Darrell T. Tryon, 45–120. Berlin: Mouton de Gruyter.

Samely, Ursula. 1991. Kedang (Eastern Indonesia). Some aspects of its grammar. Hamburg: Helmut Buske Verlag.

- Sawardo, P., *et al.* 1987. *Struktur Bahasa Lio*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- van Suchtelen, B.C.C.M.M. 1921. Endeh (Flores). Medeelingen van het Bureau voor de Bestuurszaken der Buitengewesten, bewerkt door het Encyclopaedisch Bureau, Aflevering 26. Weltevredem.
- Tryon, Darrell T. 1995. The Austronesian languages. In *Comparative Austronesian dictionary: An introduction to Austronesian studies*, pt.1, fasc. 1, ed. by Darrell T. Tryon. Berlin: Mouton de Gruyter.
- Wilken, G.A. 1886. Plechtigheden en gebruiken bij verlovingen en huwelijken bij de volken van den Oost-Indischen Archipel. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië*, 35:140-129.
- Wurm, S.A. and Shiro Hattori. 1981. *Language atlas of the Pacific area*, part 1, *New Guinea area*, *Oceania*, *Australia*. Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.
- www.ende.go.id. 2005. Official website of Pemerintah Kabupaten Ende, NTT.
- Zwicky, Arnold. 1977. *On Clitics*. Indiana University Linguistics Club, Bloomington, Indiana.

Appendix I – Ende Swadesh 200+ item wordlist

The following wordlist was taken from Ignasius Waka from the village of Oné Koré just outside the city of Ende. The items in the wordlist are made up of the traditional 200-item Swadesh wordlist with additional items specific to Austronesian and Papuan languages. The wordlist was intended for use in the project entitled *Linguistic Variation in Eastern Indonesia: the Alor and Pantar Project* (http://www.let.leidenuniv.nl/aapp/). I kindly thank Marian Klamer for allowing me to use this wordlist to elicit the following items in Ende. Both the orthographic and IPA representation are presented below. Note that the [w] is used to represent the epenthetic glide between vowels as described in §2.1.2.8, while the [ß] is used to represent the fricated labio-velar approximant. Furthermore, the IPA representations here diverge only in the representation in the palatal glide, which is represented here as [y], following conventions in Austronesian linguistics.

| | English | Ende (Orthographic) | Ende (IPA) |
|----|-----------------|---------------------|----------------------|
| 1 | head | 'urhu | ['ʔuɹu] |
| 2 | hair | fu | [ˈfuː] |
| 3 | arm | rhima | [ˈɹima] |
| 4 | hand | rhima | [ˈɹima] |
| 5 | ear | inga | [ˈiŋa] |
| 6 | eye | mata | [ˈmata] |
| 7 | nose | 'iru | [ˈʔiɾu] |
| 8 | mouth | moki | [ˈmoki] |
| 9 | lips | kumba | [ˈkumba] |
| 10 | tongue | rhema | [so ['] ma] |
| 11 | teeth | ngi'i | [ˈŋiʔi] |
| 12 | neck | tengu | [təˈŋu] |
| 13 | breast | susu | [ˈsusu] |
| 14 | stomach | tuka | [ˈtuka] |
| 15 | navel | pusé | [ˈpuse] |
| 16 | back | 'énggé | [ˈʔeŋge] |
| 17 | leg | 'a'i | [ˈʔaʔi] |
| 18 | foot | 'a'i | [ˈʔaʔi] |
| 19 | knee | mbuku | [¹mbuku] |
| 20 | skin | u'i | [ˈuʔi] |
| 21 | bone | took | [¹toko] |
| 22 | blood | ra | [ˈraː] |
| 23 | heart | pusu | [ˈpusu] |
| 24 | guts | tuka | [ˈtuka] |
| 25 | liver | 'até | ['ate] |
| 26 | father | baba | [ˈbaba] |
| 27 | mother | 'iné | ['ine] |
| 28 | older sibling | ka'é | [ˈkaʔe] |
| 29 | younger sibling | 'ari | [¹ʔaɾi] |
| 30 | husband | aki | [ˈaki] |

74

night

kombé

['kombe]

Bradley McDonnell

| pa] |
|-----|
| |
| |
| |
| |
| ri] |
| |
| .] |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |

_

 $^{^{1}\,\,}$ This is a canoe with two supporting poles on each side, commonly used for ocean fishing.

| 117 | old (people) | mbupu | [¹mbupu] |
|-----|------------------|-------------------------------|-------------------------------------|
| 118 | correct | morho | [oroq _{m1}] |
| 119 | ripe | té'a | [ˈteʔa] |
| 120 | rotten | mbaru | [¹mbaru] |
| 121 | round | monda | [ˈmonda] |
| 122 | flat, smooth | ndéna | [¹ndena] |
| 123 | thick | kapa | [ˈkapa] |
| 124 | all | mbeja | [mbəˈʤa] |
| 125 | a few | sedhiki, selo'o | [səˈdîki], [səˈloʔo] |
| 126 | many | woso | [ˈßoso] |
| 127 | other | pésa | [ˈpesa] |
| 128 | here | pendia, pena | [pəˈndiya], [pəˈnaː] |
| 129 | there | penoré | [pəˈnoɾe] |
| 130 | this | na, ndia | [ˈnaː], [ˈndiya] |
| 131 | that | noré | ['nore] |
| 132 | above | pewawo | [pəˈßaßo] |
| 133 | below | (rarhé) pewena | [pəßəˈna] |
| 134 | left | nggéu | [¹ŋgeyu] |
| 135 | right | nggana | [¹ŋgana] |
| 136 | far | réu | [ˈreyu] |
| 137 | near | wé'é | [ˈßeʔə] |
| 138 | to walk | mbana | [¹mbana] |
| 139 | sit | ngambé | [ˈŋambe] |
| 140 | stand | géré, dari | [ˈgere], [ˈdari] |
| 141 | to lie down | mba'a | [¹mba?a] |
| 142 | to sleep | 'éru, nandé | ['?eru], ['nande] |
| 143 | to die | mata | [ˈmata] |
| 144 | to eat | ka | [ˈkaː] |
| 145 | to drink | minu | [ˈminu] |
| | | pamo (hands), po'u (face), | ['pamo], ['po?u], ['rasi], [təm'ba] |
| 146 | to wash | rasi (objects), temba (cloth) | |
| 147 | to bathe | rio | [ˈriyo] |
| 148 | to bathe a child | rio | [ˈriyo] |
| 149 | to sew | sati | [ˈsati] |
| 150 | to cook | pédhé | [ˈpeɗe] |
| 151 | to laugh | tawa | [ˈtaßa] |
| 152 | to cry | rita | [ˈrita] |
| 153 | to sing | kéa | [ˈkeya] |
| 154 | to dance | toja | [ˈtoʤa] |
| 155 | to cut | poro (chicken), nggeté (wood) | [ˈporo], [ʰgəˈte] |
| 156 | to split | teka | [təˈka] |
| 157 | stab | dhusu | [ˈɗusu] |
| 158 | bite | toki | [ˈtoki] |
| 159 | to dig | koé | [ˈkowe] |

| 160 | to suck | miso | ['miso] |
|-----|----------------|--------------------------|------------------------|
| 161 | to flow | mbéré | [mbere] |
| 162 | fly | rhérha (close), wu (far) | ['ieia], ['ßu] |
| 163 | fall | mesu | [məˈsu] |
| 164 | to freeze | tu'a ² | ['tuʔa] |
| 165 | to give | pati | ['pati] |
| 166 | to hold | dhéo | [ˈdeyo] |
| 167 | to fight | pemata | [pəˈmata] |
| 168 | scared | taku | [ˈtaku] |
| 169 | to hit | pongga | [ˈpoŋga] |
| 170 | to throw | poké | ['poke] |
| 171 | to hunt | téndu | ['tendu] |
| 172 | shoot | pasa | [ˈpasa] |
| 173 | to kill | roré | ['rore] |
| 174 | to tie | riké | [ˈrike] |
| 175 | to know | mbéo | [¹mbeyo] |
| 176 | to think | piki | [ˈpiki] |
| 177 | to say | si'i, ngestéi | [ˈsiʔi], [ŋəsˈteyi] |
| 178 | to speak | mbabho | [¹mba6o] |
| 179 | to hear | zhézhé | [ˌreɪe] |
| 180 | to smell | nguru | [ˈŋuɾu] |
| 181 | to breathe | siro | [ˈsiro] |
| 182 | to spit | bura | [ˈbura] |
| 183 | to vomit | muta | [ˈmuta] |
| 184 | to play | éngé | [ˈeŋe] |
| 185 | to pull | 'ésa | [ˈʔesa] |
| 186 | to push | roka | [ˈroka] |
| 187 | rub | koso | [ˈkoso] |
| 188 | swell | mbowo | [¹mboßo] |
| 189 | swim | nangu | [ˈnaŋu] |
| 190 | to turn | rhé'o | [ˈɹeʔo] |
| 191 | to come | mai | [ˈmayi] |
| 192 | how? | ngéré 'emba | [ˈŋere ˈəmba] |
| 193 | what? | 'apa | [¹ʔapa] |
| 194 | who? | sai | [ˈsayi] |
| 195 | when? | wengi | [βəˈŋi] |
| 196 | how much/many? | sa apa | [sa ¹ ?apa] |
| 197 | why | ngéré 'emba | [ˈŋeɾe ˈəmba] |
| 198 | and | né'é | [ˈneʔe] |
| 199 | because | pu'u | [ˈpuʔu] |
| 200 | if | ki | [ˈkiː] |
| 201 | with | né'é | [ˈneʔe] |

_

 $^{^{2}}$ This can more literally be glossed as 'harden'.

| 202 | no | 'iwa | ['ʔißa] |
|-----|----------------|-------------------|----------------|
| 203 | to count | kira | [ˈkira] |
| 204 | one | se'esa | [səʔəˈsa] |
| 205 | two | ('esa) rua | [ˈruwa] |
| 206 | three | ('esa) terhu | [təˈ.ɪu] |
| 207 | four | ('esa) wutu | [ˈßutu] |
| 208 | five | ('esa) rhima | [ˈɹima] |
| 209 | six | ('esa) rhima 'esa | [ˈɹima ʔəˈsa] |
| 210 | seven | ('esa) rhima rua | [ˈɹima ˈruwa] |
| 211 | eight | ('esa) rua mbutu | [ˈruwa ˈmbutu] |
| 212 | nine | ('esa) tera 'esa | [təˈra ʔəˈsa] |
| 213 | ten | ('esa) semburhu | [səmˈbu.ɪu] |
| 214 | twenty | mburhu rua | [ˈmbuɹu ˈruwa] |
| 215 | one hundred | sengasu | [səˈŋasu] |
| 216 | 1sg | ja'o | [ˈʤaʔo] |
| 217 | 2sg (polite) | kau | [ˈkawu] |
| 218 | 2sg (informal) | kau | [ˈkawu] |
| 219 | 3sg | kai | [ˈkayi] |
| 220 | 1pl excl | kami | [ˈkami] |
| 221 | 1pl incl | kita | [ˈkita] |
| 222 | 2pl | miu | [ˈmiyu] |
| 223 | 3pl | 'ebé | [ˈʔəbe] |
| 224 | road | rarha | [ˈraɹa] |
| 225 | axe | taka | [ˈtaka] |
| 226 | comb | kéké | [ˈkeke] |
| 227 | mat | té'é | [ˈteʔe] |
| 228 | rat | té'u | [ˈteʔu] |
| 229 | bird | naké | [ˈnake] |
| 230 | snake | nipa | [ˈnipa] |
| 231 | crocodile | mori | [ˈmori] |
| 232 | cassava | 'uwi 'ai | [ˈʔußi ˈʔayi] |
| 233 | fingernail | kungu | [ˈkuŋu] |
| 234 | grandparent | 'embu | [?əmˈbu] |
| 235 | grandchild | 'ana 'embu | [ˈana ʔəmˈbu] |
| 236 | short | bhoko | [¹60ko] |
| 237 | thin | noko | [ˈnoko] |
| 238 | blue | meta | [məˈta] |
| 239 | blind | gibé | [ˈgibe] |
| 240 | deaf | dhoka | [ˈɗoka] |
| 241 | to see | kodho | [ˈkoɗo] |
| 242 | to wake up | to'o | [¹to?o] |
| 243 | to wake s.o up | reku | [rəˈku] |
| 244 | run | paru | [ˈparu] |
| 245 | to drop | pati mesu | [ˈpati məˈsu] |

| 246 | to live | muri | [ˈmuɾi] |
|-----|----------------|---------------------|------------------------------------|
| 247 | to work | kema | [ˈkəma] |
| 248 | dry in sun | wari | ['Bari] |
| 249 | search | nggaé | [¹ŋqae] |
| 250 | where? | pe'emba | [pəʔəmˈba] |
| 251 | fog | rubu | [ˈrubu] |
| 252 | dust | 'awu | ['ʔaβu] |
| 253 | at | pa, rheka | ['pa] / ['pə], [ɹəˈka] |
| 254 | hit | pongga | [ˈpoŋga] |
| 255 | to pound | waju | [ˈßaʤu] |
| 256 | cloud | rubu | [ˈrubu] |
| 257 | ash | aro | ['aro] |
| 258 | way, path | parhé | ['pa.ie] |
| 259 | rain | ura | ['?ura] |
| 260 | shoulder | wara | [ˈßara] |
| 261 | chew | pana | [ˈpana] |
| 262 | yawn | mesorho | [oros _e m] |
| 263 | dream | nipi | [ˈnipi] |
| 264 | thatch | 'até | ['?ate] |
| 265 | needle | rharu | [ˈɹaru] |
| 266 | steal | naka | [ˈnaka] |
| 267 | living, alive | muri | [ˈmuɾi] |
| 268 | wood | kaju | [ˈkaʤu] |
| 269 | to plant | murha | [ˈmuɹa] |
| 270 | choose | pirhi | [ˈpiɹi] |
| 271 | grow | tembu | [təmˈbu] |
| 272 | squeeze | porhé | [pore] |
| 273 | buy | mbeta | [^m bə ^l ta] |
| 274 | to open | buka | [ˈbuka] |
| 275 | fowl | naké | [ˈnake] |
| 276 | mosquito | kepa | [kəˈpa] |
| 277 | spider | kawo | [ˈkaßo] |
| 278 | branch | ngga'a | [¹ŋgaʔa] |
| 279 | sand | mboko watu | [¹mboko ˈßatu] |
| 280 | water (fresh) | aé | [ˈae] |
| 281 | sea/salt water | aé mesi | [ˈae məˈsi] |
| 282 | thunder | bhérha | [ˈbe.ɪa] |
| 283 | lightning | kirha | [ˈkiɹa] |
| 284 | sick, painful | ro | ['roː] |
| 285 | shy, ashamed | méa | [ˈmeya] |
| 286 | hide | roko | [ˈroko] |
| 287 | climb | nai | [ˈnayi] |
| 288 | eleven | sembuzhu se'esa | [səmˈbu.u səʔəˈsa] |
| 289 | thirty | sembuzhu 'esa terhu | [səmˈbuɹu ʔəˈsa təˈɹu] |

| 290 | one-hundred thousand | riwu sengasu | [ˈrißu səŋˈasu] |
|-----|----------------------|--------------|-----------------|
| 291 | ice | éé | ['e:] |
| 292 | snow | rubu | [ˈrubu] |
| 293 | spear | tumba | [ˈtumba] |
| 294 | grass | kulu | [ˈkulu] |
| 295 | sago | moké | [ˈmoke] |
| 296 | some | réké | [ˈreke] |
| 297 | wipe | pui | [ˈpuwi] |
| 298 | scratch | nasi | [ˈnasi] |
| 299 | float | mbawa | [¹mbaßa] |

ASPECT MARKING AND MODALITY IN CHILD VIETNAMESE¹

Jennie Tran
University of Hawai'i
<jennietr@hawaii.edu>

Abstract

This paper examines the acquisition of aspect morphology in the naturalistic speech of a Vietnamese child, aged 1;9. It shows that while the omission of aspect markers is the predominant error, errors of commission are somewhat more frequent than expected (~20%). Errors of commission are thought to be exceedingly rare in child speech (<4%) (Sano & Hyams 1994), and thus it appears as if these errors in child Vietnamese are more common than in other languages. They occur exclusively with perfective markers in modal contexts, i.e., perfective markers occur with non-perfective, but modal interpretation. I propose, following Hyams (2002), that these errors are permitted by the child's grammar since perfective features license mood. Additional evidence from the corpus shows that all the perfective-marked verbs in modal context are eventive verbs. I thus further propose that the corollary to root infinitives (RIs) in Vietnamese is perfective verbs in modal contexts.

1. Introduction And Background

There has been considerable debate regarding the acquisition of tense and aspect by young children. Studies on the acquisition of tense-aspect morphology have either attempted to make a general distinction between tense and aspect, or the more specific distinction between grammatical and lexical aspect, or they have centered their analyses around Vendler's (1967) four-way classification of the inherent semantics of verbs: achievement, accomplishment, activity, and state. The majority of the studies have concluded that the use of tense and aspect inflectional morphology is initially restricted to certain semantic classes of verbs (Bronkard & Sinclair 1973, Antinucci & Miller 1976, Bloom, Lifter & Hafitz 1980), and that the inherent lexical aspect of verbs (also known as *Aktionsart*) can have a strong influence on the development of tense-aspect morphology (Shirai & Andersen 1995).

Intertwined with the acquisition of tense-aspect morphology is the phenomenon of root infinitives (RIs) and bare forms which have been of great interest for researchers in first language acquisition in the past decade. Extensive cross-linguistic studies on root infinitives have been done to address the question of what the properties of root infinitives are and what temporal references they have as opposed to bare forms and finite verbs. The following representative examples illustrate the RI-phenomenon very common in languages such German, Dutch, French, Swedish.

¹ Editorial note: The present paper was originally submitted for inclusion in the SEALSXIV proceedings volume. However, the paper was misplaced in the confusion associated with delays in the preparation of that volume. Later Jennie brought the oversight to our attention and we decided to present the paper in JSEALS at the first available opportunity.

228 Jennie Tran

(1) a. Thorsten das haben. (German) Thorsten that have-inf. b. Papa schoenen wassen. (Dutch) Daddy shoes wash-inf. c. Michel dormir. (French) Michel sleep-inf. d. Jag också hoppa där å där. (Swedish) I also hop-inf. there and there

A recently growing number of studies on tense-aspect morphology have specifically addressed the interaction and relationship between tense, aspect, and mood. In this section, a brief review of the relevant literature on aspect and mood relationship is provided.

Hoekstra & Hyams (1998) addressed the importance of the properties of root infinitives and underscored the distinction between true root infinitives and English bare forms. They argue that languages that have root infinitives are subject to the Eventivity Constraint (EC): Only eventive verbs can occur in RI-constructions. In other words, non-finite verbs are typically eventive (non-stative), whereas non-eventive verbs (i.e., stative) are exclusively finite. Concurrently RIs also are subject to the Modal Reference Effect (MRE), which means that RIs receive a modal interpretation or an irrealis meaning rather than a deictic tense interpretation. Hoekstra & Hyams argue that the modal interpretation comes from the infinitival morpheme itself. English bare forms, on the other hand, are not subject to EC and MRE as they do not bear infinitival morphology. Their study shows that English bare forms occur predominantly with non-eventive verb.

Varlokosta et al. (1998) found that very young Greek children produce a 'bare perfective' form, which in adult Greek is ungrammatical. Even if Greek is a language without infinitival morphemes, these 'bare perfective' forms occurring in child Greek do share the central properties of root infinitives. The MRE of root infinitives is thus prevalent in child Greek, i.e. Greek bare perfective verbs have a modal or future interpretation. The following adult and child utterances illustrate the difference between the -i affix in adult speech the -i form in child utterance.

- (2) a. O Spíros exi δiavás**i.** (adult Greek) the Spiros has read-participle 'Spiros has read.'
 - b. O Spíros θa/na δiavási. (adult Greek)
 the Spiros fut/subj read-perf.pres.3sg
 'Spiros is going to/should read.'
- (3) Pio vavási. (child Greek)
 Spiros read-perf.3sg
 'Spiros is going to/wants to read.'

Child Vietnamese 229

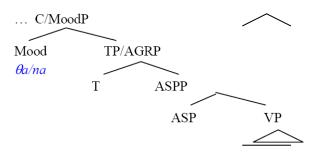
The -i affix in adult Greek reflects two homophonous forms: a perfect participle as in (2a), and the 3rd person singular perfective form as in (2b) occurring in future and subjunctive clauses. The child's verbal utterance in (3), on the other hand, is a perfective verb with an -i affix and is missing either an auxiliary, such as exi in (2a), or the modal particle $\theta a/na$ as in (2b). Due to the absence of an auxiliary or a modal verb, the child's verbal form in (3) is referred to as a "bare perfective", but claimed by Varlokosta et al. to bear the RI-properties. They characterize the -i form in this bare perfective as a participle, on the grounds that the child's bare perfective is a non-finite, non-agreeing form, resembling more the adult version of the participle structure in (2a) rather than the future/subjunctive structure in (2b). They claim that the child -i affix is rather a participle and analogous to the child Italian bare participle, which lacks the support of an auxiliary, as exemplified in (4a) and (4b) below:

- (4a) Presa Checco campana. (child Italian) taken Checco bell
- (4b) Francesco ha pres la campana. (adult Italian) Francesco have-3sg taken the bell 'Francesco has taken the bell.'

Hyams (2002) discusses the development of functional structure in the early grammar of Greek and addresses the resemblance of this bare perfective construction in child Greek to root infinitives found in RI-languages. After a thorough analysis of the structure of the bare perfective, she argues, however, against identifying the Greek bare perfectives as participles, because root infinitives and bare participles are structurally and semantically distinct. She proposes an alternative that captures the semantic property of bare perfectives, namely their very close correspondence to the modal/future structure. Thus the Modal Reference Effect of root infinitives shared by bare perfectives in Greek is a phenomenon that rather results from the match between perfective features and mood, i.e., perfective features check mood features in the syntax.

The following relative ordering of functional heads in Greek is proposed by Roussou (2000) and adopted by Hyams (2002:251).

Adult Greek:

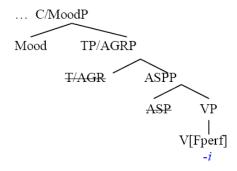


In adult structure, T/AGR and ASP projections are above VP. The modal and future particles $\theta a/na$ are generated in the lower C head specified for Mood or C $_{\rm M(ood)}$, hence license MoodP through Merge.

230 Jennie Tran

Hyams proposes that in child Greek, T/AGR is unspecified because "the bare perfective is necessarily a non-agreeing form, with the unmarked -i affix emerging as the default where AGR is unspecified" (p.258), and ASP is underspecified as the perfective verb does not bear an aspectual meaning. Hence, the irrealis interpretation of the bare perfective is given by the active MoodP, which has to be licensed by syntactic features. Hyams suggests that it is the perfective feature in the bare perfective verb that checks and licenses MoodP because there are no intervening features between Mood and the perfective verb. The structural analysis of bare perfective proposed thus involves an active Mood licensed by the perfective feature in the verb under Attract.

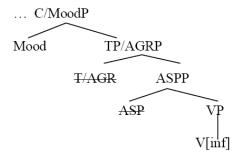
Child Greek bare perfective:



This explains both the modal meaning associated with the bare perfective (from MoodP), which is unexplained in VVR's [Varkolosta et al. 1998] participle analysis, as well as the restriction to perfective verbs...In the absence of modal particles the early grammar avails itself of the feature checking option. The adult system emerges as the modals reach adult-like productivity. We may assume, as proposed in Chomsky (1995), that Merge takes precedence over Move/Attract, such that once the modals are acquired, they push out the feature checking option. (Hyams, 2002:258)

This proposal by Hyams (2002) not only explains the irrealis -i form used by Greek children, but also the vast occurrence of root infinitives in child RI-languages. In child Greek, perfective features license mood in child language. In RI-languages, it is the irrealis feature in the infinitive itself that licenses MoodP through Attract:

Child root infinitive in RI-languages:



Child Vietnamese 231

1.1 Purpose and organization

In this paper, I investigate the following questions:

(1) What aspectual markers occur in child Vietnamese? What errors (if any) are manifested?

(2) Is there a correlation between Aspect and Mood in child Vietnamese?

The present paper is organized as follows. Section 2 provides a brief grammatical sketch on how aspect is used in adult Vietnamese. Section 3 describes the methodology. Section 4 illustrates and analyses the necessary results, from general to specific. Some actual examples from the data are included. Section 5 turns to the gist of the paper, namely the errors of commission and the relationship between aspect and mood with regard to the data. Section 6 concludes the paper.

2. Adult Vietnamese

Vietnamese is an SVO and isolating language, thus has no inflectional morphology. Verbs are not inflected, i.e. they never have a stem change, but there are various markers (separate morphemes) that accompany the verb to express Tense and Aspect.

There are two types of expressions for tense and aspect. One type (a set of three: dang, $d\tilde{a}$, $s\tilde{e}$) comprises preverbal markers. These markers are used in writing and formal conversation, except for the progressive marker dang which is also commonly used in informal conversation. The other type is a postverbal aspect morpheme, $r\hat{o}i$. This morpheme is very common in informal conversational context.

Table 1: Tense and aspect morphemes in declaratives

| Morphemes | Interpretation |
|-----------|---------------------|
| Ø | present stative |
| Ø | present habitual |
| đang | present progressive |
| rồi | perfective |
| đã | past |
| sẽ | future |

Meaning (Ngo, 1999)

dang = to be happening, occurring (progressive)

rồi = finished, completed, already (perfective)

 $d\tilde{a}$ = to have happened, occurred (past)

 $s\tilde{e} = to happen, occur [in the future] (future)$

The next distinction to be made is the positioning of $d\tilde{a}$ and $r \partial i$. All preverbal aspect markers must immediately precede the verb; no lexical item can intervene. On the other hand, the postverbal $r \partial i$ does not have to immediately follow the verb. Other lexical items can come between the verb and this morpheme. The following examples show the contrast:

232 Jennie Tran

rồi. uống uống (5) a. Em bé đã sữa. b. Em bé sữa Baby Pst drink milk. Baby drink milk Perf 'The baby drank milk.' 'The baby drank milk.'

Tense and aspect can also be signaled by the use of a time adverb. When such a time adverb is used, both preverbal and postverbal aspect morphemes are optional.

One idiosyncrasy in the Vietnamese verbal system is the lack of the copula be when used with an adjective. Instead, adjectives can function as verbs. These are referred to as adjectival verbs and they are stative.

(6) Con mèo dể thưởng. cl cat cute 'The cat is cute.'

3. Methodology

The data used for analysis is taken from the naturalistic speech of a monolingual Vietnamese female child, Kim, at the age of 1;9 (MLU = 2.0) The recording was done by the child's father primarily at home, and the last 45 minutes in the car and at the playground. The duration of the recording is approx. 3 hours on audiotape. The total number of utterances in this three-hour-stretch of data including imitations and repetitions is 897. Of these, 201 utterances were eligible for counting. These are spontaneous oneword to multiword utterances excluding imitations, repetitions, and unintelligible utterances. Out of the 201 utterances eligible for counting, 46 were nominal utterances and 155 were verbal utterances. For analysis, only verbal utterances are considered.

Table 2: Types of verbal utterances

| Declaratives | 118 |
|--------------|-----|
| Imperatives | 37 |
| Total | 155 |

The types of verbal utterances used for analysis are only declaratives (118 total) as they contain aspect marking morphemes.

4. Results

4.1 General verbal utterances

At first, a general analysis and counting of the verbal utterances was conducted. This is presented in table 3.

Child Vietnamese 233

| Type of VP / IP | Occurrences | Example |
|------------------------|-------------|----------------------|
| V | 24 | write |
| V + N | 27 | draw dog |
| V + P | 4 | go in |
| V + (N) + Subj. NP | 5 | hug Daddy |
| V + V + N | 1 | draw kick horse |
| V + N + N | 1 | change diapers horse |
| Subj. NP + V | 25 | Daddy eat |
| Subj. NP + V + Obj. NP | 4 | Daddy take milk |
| Subj. $NP + V + V$ | 1 | Daddy come lie |
| Adi. | 26 | dirty too |

Table 3: General breakdown of verbal utterances

4.2 General breakdown into bare verbs and bare verbs with markers

The total of 118 declaratives taken from the corpus consists of bare verbs and verbs with aspect marking.

Make-up of verbs:

/ 73 bare verbs (62%)

118 declaratives

√ 45 verbs with aspect marking (38%)

The next step of analysis involves a more specific breakdown, focusing particularly on correct utterances and errors. Table 4 shows the number of the child's correct uses, as well as errors in declaratives. The errors are categorized into errors of omission and errors of commission. Errors of omission involve errors in which an obligatory inflection, a marker or morpheme, such as agreement or tense/aspect marker, is omitted. Omission errors are the predominant error type in the speech of young children. Errors of commission involve the misuse of an inflection, a marker or morpheme, for example, the misuse of a 3rd person singular marker for 1st person singular. These commission errors are exceedingly rare in child speech (<4%) (Sano & Hyams 1994).

Of 118 declaratives, 36 tokens were excluded because of unclear reference, yielding a total of 82 tokens.

Table 4: Correct use and errors of aspect marking

| aspect marker | | correct | omission | commission | Total |
|---------------|--------------|---------|----------|------------|-------|
| Ø | pr. stative | 16 | 0 | 0 | 16 |
| Ø | pr. habitual | 4 | 0 | 0 | 4 |
| đang | pr.prog. | 0 | 13 | 0 | 13 |
| rồi | perf. | 36 | 4 | 9 | 49 |
| đã | past | 0 | 0 | 0 | 0 |
| | | | | | 82 |

234 Jennie Tran

Table 4 shows a noticeably high number of the perfective morpheme $r \partial i$ (total of 45 out of 82), with 9 tokens used incorrectly. The present progressive dang never occurs in the entire corpus – a total of 13 errors of omission. There were no errors in the present stative and habitual. No utterances with the past aspect marker da and the future marker sa were produced at all. (My interpretation of whether the child's intended meaning in an utterance is stative habitual, progressive, or perfective is based on discourse context. The surrounding utterances, especially the parent's responses, help in determining the appropriate aspect intended by the child.)

Brief summary of the observations taken from table 4:

The child

- always uses a bare verb when a bare verb is required (present stative and habitual);
- produces the postverbal aspect making morpheme $r\hat{o}i$ both with verbs and adjectival verbs;
- omits the copula verb 'be' and has the knowledge of using the adjective as a full verb;
- does not produce any preverbal aspect marker; consequently, she is does not refer to the progressive and the future.

The following actual utterances from the data exemplify the child's typical errors of omission:

(A) Omission of the **progressive** aspect marker:

(7) a. viết adult form: ba đang viết write Dad Prog write

'Daddy is writing.'

b. vẻ cá *adult form:* ba đang vẻ cá draw fish Dad Prog draw fish

'Daddy is drawing a fish.'

(B) Omission of the **perfective** aspect morpheme:

(8) a. má đi tắm adult form: má đi tắm rồi Mom go bathe Mom go bathe Perf

'Mom went bathing / to take a shower.'

b. ba-bi ăn adult form: ba-bi ăn rồi baby eat baby eat

'Baby has eaten.'

The question (1) in section 1.1 can now be answered: The only aspectual marker that occurs in the child's speech is the perfective morpheme $r\partial i$. Errors of omission occurred 8% of the time (4/49) with $r\partial i$, but with the progressive markers dang they occurred all the time (100%). Interestingly, errors of commission are manifested as well: 9 out of 45

Child Vietnamese 235

utterances containing the perfective $r \partial i$, yielding 20%. It is these errors of commission that we now turn to in the remainder of the paper.

5. Discussion

5.1 Errors of Commission

Errors of commission from the data set involve errors occurring exclusively with perfective markers in modal contexts, i.e. perfective morpheme $r\hat{o}i$ occurs with non-perfective, but modal interpretation, as shown in (9):

| (9) | child form: | | | | adult form: | | | |
|-----|-------------|-----------|---------|-------|-------------|------|------|--|
| | vẻ | cá | rồi | Con | muốn | vẻ | cá. | |
| | draw | fish | Perf. | child | want | draw | fish | |
| | 'I wan | t to draw | a fish' | | | | | |

Table 5 below shows the child's use of the perfective marking morpheme $r \hat{o} i$ for meanings other than the perfective. The perfective marking is overextended and overgeneralized to the present (1 token) and the modal meaning (8 tokens). The child overuses the perfective aspect marking morpheme $r \hat{o} i$ to refer to mood.

Table 5: *Temporal and mood reference of perfective verbs*

| Present | Perfective | Modal | Total utterances |
|-------------------|------------|--------------|-----------------------|
| | | | containing <i>rồi</i> |
| 1 | 36 | 8 | 45 |
| (2%) | (80%) | (18%) | |
| • | • | \downarrow | |
| Commission errors | | | |

Figure 1 illustrates the proportion of these commission errors in relation to the correct use of the perfective marker. The rate of errors of commission: 9/45 (20%).

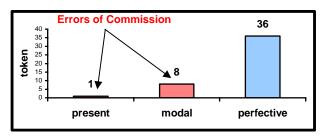


Figure 1: Rates of Commission Errors

Commission Data

The following actual examples from the present data will illustrate this commission phenomenon:

236 Jennie Tran

(A) Present

The only one time the child makes an error in the overuse of the perfective aspect marker $r \hat{o} i$ to mean the present is the following:

(10) Kiến nhỏ quá rồi.

Ant small too Asp

'The ants were too small.'

Intended: 'The ants are too small.'

This utterance was produced when the child was sitting on a bench and ants suddenly started to come out. In this context, the present should have been used (without the aspect marker $r\hat{o}i$), but the child used the perfective. This is evidence for an error of commission.

(B) Modal

All the remaining times, the child's intended meaning is modal, as exemplified in the actual samples from the data. (Due to space limit, only one example can be listed here.)

(11) This utterance takes place at the playground. The child has been playing on the slide for a while and the father spent the last five minutes talking on the phone. The child wanted to sit down on the bench but there were ants that bit her, so she wanted to go home.

```
CHI: đi về — đi về rồi
go home — go home Asp
'Go home — I want to go home.'

DAD: đi về — đi về đi — không đi cầu tuột nữa hả?
go home — go home IMP — not go slide more right?
'Go home, let's go home. You don't go on the slide anymore?'
```

It is clear from the context that Kim wanted to go home. She got bit by the ants and the father's response shows that he agreed that they should go home.

5.2 Eventivity Constraint and Modal Reference Effect

Although the Vietnamese child utterance ve ca ro in (9) is clearly non-infinitival due to the presence of the perfective morpheme, it appears to have a resemblance to the Greek bare perfective verbs and the root infinitives in RI-languages. My proposal is thus the following:

<u>Hypothesis</u>: Vietnamese perfective aspect marker, analogous to the Greek 'bare perfective' (Varlokosta et al. 1998), can also have RI-like properties.

Recall that root infinitives are known to be subject to the Modal Reference Effect, which entails that they typically have an irrealis or modal meaning. Since the Eventivity Constraint is derived from the Modal Reference Effect, the most appropriate way to test my hypothesis is to investigate whether the Eventivity Constraint holds in child Vietnamese verbal utterances.

Child Vietnamese 237

Eventivity Constraint (Hoekstra & Hyams 1998)

- a) Non-finite verbs are typically eventive (non-stative)
- b) Stative verbs are exclusively finite

To determine the eventivity of the child's verbal utterances, I examined all the verbs that occurred with the modal $r \hat{o} i$ and all the verbs that occurred with the perfective $r \hat{o} i$, in terms of their inherent lexical aspect (Shirai & Andersen 1995), along the lines of the four-way inherent-aspectual classification (Vendler 1967):

STATE (non-eventive) - that which has no dynamics, and continues without additional effort or energy being applied, e.g., see, love, hate, want, etc.

ACTIVITY(eventive) - that which has duration, but with an arbitrary endpoint, and is homogeneous in its structure, e.g., *run*, *sing*, *play*, *dance*, etc.

ACCOMPLISHMENT (eventive) - that which has some duration, but has a single clear inherent endpoint, e.g., *run a mile, make a chair, build a house*

ACHIEVEMENT (eventive) - that which takes place instantaneously, and is reducible to a single point in time, e.g., *recognize*, *die*, *reach the summit*, etc.

Each verb was classified according to this schema, and the results are presented in table 7 below.

| VERBS W | VERBS WITH MODAL <i>Rồi</i> | | VERBS WITH PERFECTIVE <i>Rồi</i> | | | | |
|-----------|-----------------------------|------|----------------------------------|-----------|----------------|-------|------------|
| Verb | Gloss | Type | Eventivity | Verb | Gloss | Type | Eventivity |
| vẻ | draw | Act. | V | chật | is tight | Stat. | |
| viết | write | Act. | V | do | is dirty | Stat. | |
| để | put | Acc. | V | nắng | is sunny | Stat. | |
| để dô | put in | Acc. | $\sqrt{}$ | nóng | is hot | Stat. | |
| đánh răng | brush teeth | Acc. | V | thay tả | change diapers | Acc. | V |
| thay tả | change diapers | Acc. | V | ăn (N) | eat (N) | Acc. | V |
| đi về | go home | Ach. | V | đánh răng | brush teeth | Acc. | $\sqrt{}$ |
| | | | | gẩy | break | Ach. | V |
| | | | | về | come | Ach. | V |
| | | | | ngả | fall | Ach. | V |
| | | | | té | fall | Ach. | V |
| | | | | ướt | get wet | Ach. | V |
| | | | | đổ | spill | Ach. | V |
| | | | | đi | leave | Ach. | V |
| | | | | ra | come out | Ach. | $\sqrt{}$ |
| | | | | trời mưa | Start to rain | Ach. | V |

Table 7: Verbs with rồi

The results from table 7 indicate that all the verbs that appear with the modal $r\hat{o}i$ are eventive and all the verbs that appear with the perfective $r\hat{o}i$ are both eventive and non-eventive. There were a total of 8 utterances occurring with modal $r\hat{o}i$ and 36 utterances occurring with perfective roi. In the left part of table 7, one of the verbs occurs twice in

238 Jennie Tran

modal context. Hence 7 verbs are listed but the total number of occurrences is 8. In the right part of table 7, some of the verbs occurred more than once. Hence 16 verbs are listed but the total number of occurrences is 36.

The contrast between modal and perfective use of the perfective morpheme $r\hat{o}i$ is summarized in table 8.

Table 8: Eventivity with $r \hat{o} i$

| | eventive | non-eventive |
|-----------|----------|--------------|
| Modal rồi | 8 | 0 |
| Perf. rồi | 26 | 10 |

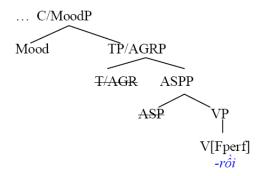
The Eventivity Constraint holds true based on the following pieces of evidence from the data:

All modal $r \partial i$ utterances are eventive. The claim we would like to make is that the Vietnamese child modal $r \partial i$ constructions are equivalent to the RI-construction due to the MRE that obtains in these constructions, i.e. the modal or irrealis meaning obtains in these modal $r \partial i$ utterances. The discussion in the next section can account for this claim.

5.3 Aspect-Mood connection

First, the modal features in the Vietnamese child modal $r \delta i$ construction are licensed in the syntax, hence the incompatibility with non-eventives. Roussou's (2000) syntactic structure proposal and Hyams' (2002) structural analysis of child Greek can be applied here:

Child Vietnamese perfective verb:



In child Vietnamese modal perfective construction, ASP is underspecified as the perfective verb used in modal context does not bear an aspectual meaning. Hence, the modal/irrealis interpretation of the perfective verb, V[Fperf]- $r\hat{o}i$, is provided by its perfective feature under Attract. In other words, analogous to child Greek, perfective features check mood in child Vietnamese, hence there is a modal reference with perfective $r\hat{o}i$.

Second, the properties that Vietnamese modal perfectives share with Greek bare perfectives and root infinitives constitute another convincing piece of evidence for the claim we made above that the Vietnamese corollary to root infinitives corresponds to perfective verbs in modal contexts.

Child Vietnamese 239

Table 9: *Shared properties*

| Root infinitive and Greek bare perfective | Vietnamese modal perfective |
|---|--|
| | |
| DATA: | DATA: |
| (Spiros 1;9) | (Kim 1;9) |
| Ego katiti | Vẻ rồi |
| I sit-PERF-3.sg. | draw PERF. |
| 'I am going to/want to sit' | '(I) am going to/want to draw.' |
| | |
| CHARACTERISTICS: | CHARACTERISTICS: |
| - are overwhelmingly eventive | - all perfective-marked verbs in modal context |
| | are eventive |
| - express modal meaning (MRE) | -express modal meaning (MRE) |

6. Conclusion

Errors of commission are not that rare (up to 20%, 9 out of 45), unlike <4% as reported in Sano & Hyams (1994). The particular errors of commission are not random, but are permitted by the child's grammar because perfective features license Mood. This answers the research question (2): There is a correlation between Aspect and Mood in child Vietnamese. The Vietnamese perfective verb in modal context is the equivalent of the root infinitive in German/Dutch/French/Swedish and the bare perfective in Greek.

References

- Antinucci, Francesco & Ruth Miller.1976. How children talk about what happened. *Journal of Child Language* 3.169-89.
- Bloom, Lois, Karen Lifter & Jeremie Hafitz.1980. Semantics of verbs and the development of verb inflection in child language. *Language* 56: 386-412.
- Bronkard, J.P. & Hermina Sinclair. 1973. Time, tense and aspect. Cognition 2.107-130.
- Hoekstra Teun, and Nina Hyams. 1998. Aspects of root infinitives. Lingua 106.81-112
- Hyams, Nina. 2002. Clausal structure in child Greek: A Reply to Varlokosta, Vainikka and Rohrbacher and a Reanalysis. *The Linguistic Review* 19,3.225-269.
- Ngo, Nhu Binh. 1999. *Elementary Vietnamese*. Boston, Rutland, Vermont, Tokyo: Tuttle Publishing.
- Roussou, Anna. 2000. On the left periphery: Modal particles and complementisers. *Journal of Greek Linguistics* 1.65-94.
- Sano, Tetsuya & Nina Hyams. 1994. Agreement, finiteness, and development of null arguments. In *Proceedings of the NELS 24*. Gonzales, Merce (ed.), volume 2.543-558. Amherst, MG: GSLA.
- Shirai, Yasuhiro and Roger W. Andersen. 1995. The acquisition of tense-aspect morphology: A prototype account. *Language* 71.(4).743-762.
- Varlokosta, Spyridoula, Anne Vainikka, & Bernhard Rohrbacher. 1998. Functional projections, markedness, and 'root infinitives' in early child Greek. *The Linguistic Review* 15.187-207.
- Vendler, Zeno. 1967. Linguistics in philosophy. Ithaca, NY: Cornell University Press.

A NOTE ON INTERFERENCE OF THAI REDUPLICATION ON VIETNAMESE SPOKEN IN UDON THANI PROVINCE OF THAILAND

Songgot Paanchiangwong

University of Social Sciences and Humanities, Hanoi National University <songgot_p@yahoo.com>

1. Introduction

Udon Thani, a province in the northeast of Thailand, has the biggest Vietnamese community in Thailand (Sophana: 2005, 10). These Vietnamese immigrants numbering around 46,700 people came from Northern, Cental and Southern regions of Vietnam in 1945 because of World War II (Khachatphai: 1978, 10). At the present time, there are three generations of Vietnamese immigrants living in Udon Thani Province. They still communicate with each other in Vietnamese, even though some of them are now living outside the community. They live in the Thai language society and they work with Thais, so they need to learn the Thai language as a means of communication. It can be said that the Thai language is involved directly in their daily life, not only when they work outside their community but also in their own family.

Consequently, linguistic interference has occurred in respect of quantitative intensification in Vietnamese, patterning after Thai reduplication. This is documented in this paper.

2. Definition of *Interference*

Various scholars have offered definitions of interference:

Weinreich (1953) defines two languages as being in contact if they are used alternately by the same speakers. Deviation from the language norms as a result of contact is called interference. Weinreich suggests that in speech, interference is like sand carried by a stream; in language, interference is the sedimentary sand deposited on the bottom of a lake.

Barkman (1968) says that whenever a bilingual speaker produces a form in a language which monolingual speakers of that language would not use, either as a result of identification or for other reasons, a case of interference has occurred.

Briere (1968) notes that interference occurs when a speaker of language X attempts to learn language Y. This interference has been attributed to the fact that between any two languages there are similarities and differences on all levels of analysis. The degree of interference would ensue from the partial similarity or difference between the two completing categories, one in the learner's native language and the other in the target language.

Mackey (1972) defines interference as the use of features belonging to one language while speaking or writing another. Mackey regards not only speaking but also writing as having interference. If the text analyzed, however, is not based on the spoken, but on the written form of the language, we will have to look for evidence of another sort

Songgot Paanchianwong. 2009. A Note on Interference of Thai Reduplication on Vietnamese Spoken in Udon Thani Province of Thailand. *Journal of the Southeast Asian Linguistics Society* 2:241-245. Copyright vested in the author.

of interference –graphic interference– the transfers of writing habits from one language to another.

Dulay (1982) points out that *interference* refers to two very different/distinct linguistics phenomena – one that is essentially psychological and the other that is essentially socio-linguistic. The psychological refers to the influence of old habits when new ones are being learned, whereas the socio-linguistic refers to language interactions that occur when two language communities are in contact. Dulay also added that the fewer the bilingual speakers are, the more interference there will be when attempts are mark to communicate with speakers of the target language.

Crystal (1991) identifies interference as a term used in socio-linguistics and foreign-language learning to refer to the errors a speaker introduces into one language as a result of contact with another language, also called negative transfer. The most common source of error is in the process of learning a foreign language, where the native tongue interferes although interference may occur in other contact situations (as in the case of multilingualism).

Lehiste (1998) states that *interference* frequently takes place as deviation from the norm of either language that occurs in the speech of bilinguals as a result of their familiarity with more than one language. A bilingual here is defined as a person who is able to produce grammatical sentences in more than one language. Interference can be found at all levels: phonetics, phonology, morphology, syntax, semantics, and lexis.

3. Reduplication in Vietnamese of Udon Thani compared to Standard Vietnamese and Thai

The following are examples of complete reduplication of Vietnamese that were found in conversations of immigrant Vietnamese in Udon Thani. They are different from complete reduplication of standard Vietnamese in its structure and meaning. Note how in the following examples immigrant Vietnamese in Udon Thani express intensification of verbs such as *mùng* (happy), *đẹp* (beautiful), *yêu* (love), the quantifier *đông*, adjectives *bận* (dirty), *dày* (thick), *đen* (black), *ngon* (delicious) and adverbs *nhanh* (quickly) and *tốt* (well) the forms are reduplicated, with the first iteration given a rising tone (or in the case of *tốt* which already has a rising tone, the first iteration has a marked immediate rising tone compared to the second):

- (1) VnU: Hôm đó chị đi siêu thị, chị đã gặp người Việt, chị *mứng mừng*.
 - Eng: I met a Vietnamese when I went to the supermarkets that day, I was *very happy*.
- (2) VnU: Năm nay hoa hậu Thái Lan đép đẹp.
 - Eng: Miss Thailand this year is *very beautiful*.
- (3) VnU: Lần đầu tiên tôi đến Việt Nam, tôi thấy xe máy trên đường đồng đông.
 - Eng: The first time I went to Vietnam, I saw *a lot of motorbikes* on the road.
- (4) VnU: Tôi không dám đi với nó, nó lái xe máy *nhánh nhanh*.
 - Eng: I do not dare to go with him; he drives his motorbike so fast.

(5) VnU: Anh ta làm việc *tốt tốt*. Eng: He works *very well*.

(6) VnU: Món ăn này *bấn bận* làm sao mà cháu ăn được. Eng: This food is *very dirty*, how can you have it?

(7) VnU: Bác Hồ là người đứng đầu trong cuộc chiến chống Pháp nên người dân Việt Nam *yếu yêu* bác Hồ.

Eng: The president Ho Chi Minh is a leader of fighting against France so Vietnamese people *love* him *so much*.

(8) VnU: Cái cuốn sách dáy dày như thế này tôi đọc cả năm cũng không đọc nổi. Eng: This is a very thick book; I can not complete reading it even thought I spent all year.

(9) VnU: Con chơi kiểu gì, làm sao áo đén đen như vậy.Eng: Son! What did you play? Why your shirt is very black like this?

(10) VnU: Chị Hiền nấu món ăn *ngón ngon* đấy anh ạ. Eng: The foods that sister Hien cooks are *very delicious*.

4. Comparison with Standard Vietnamese and Standard Thai

The writer analyses data into two kinds: structure analysis, meaning analysis as the followings.

The structure of reduplication that the writer found significantly differs from Standard Vietnamese but parallels complete reduplication of Standard Thai. Reduplication in Standard Vietnamese is discussed by Nguyen Tai Can (2004). In his discussion he points out that typically reduplication is complete, with both the base and the copy pronounced similarly. The closest Standard Vietnamese equivalents of the forms discussed above are as

follows:

| mừng mừng | 'quite happy' | bận bận | 'quite dirty' |
|-------------|-------------------|-----------|----------------------|
| đẹp đẹp | 'quite beautiful' | dày dày | 'quite thick' |
| đông đông | 'quite many' | đen đen | 'quite black' |
| nhanh nhanh | 'quite fast' | ngon ngon | 'quite delicious' |
| tốt tốt | 'quite well' | yêu yêu | 'to love quite well' |

However, such use of reduplication in Standard Vietnamese is not common, instead speakers tend to use the adverb *hoi* 'quite'or 'rather' before a verb, adverb, adjective to signal an augmented meaning. For example:

| hơi mừng | 'quite happy' | hơi bận | 'quite dirty' |
|-----------|-------------------|----------|----------------------|
| hơi đẹp | 'quite beautiful' | hơi dày | 'quite thick' |
| hơi đông | 'quite many' | hơi đen | 'quite black' |
| hơi nhanh | 'quite fast'. | hơi ngon | 'quite delicious' |
| hơi tốt | 'quite well' | hơi yêu | 'to love quite well' |

Note, however, that neither the reduplication nor the hoi constructions convey the sense of intensification to the same extent as the VnU reduplication with rising tone on the first iteration. In Standard Vietnamese the senses of very X or X very much are typically handled by constructions with intensifiers *rât* and *lắm*. For example:

Anh ấy sửa chữa đồng hồ của tôi *rất tốt*. (11)SVn:

> Eng: He repairs my watch very well.

Me mua áo *rất đep* cho con. (12)SVn:

> Mother! Buy me a very beautiful shirt. Eng:

Anh ấy sửa chữa đồng hồ của tôi *tốt lắm*. (13)SVn:

> He repairs my watch very well. Eng:

Me mua áo đẹp lắm cho con. (14)SVn:

> Mother! Buy me a very beautiful shirt. Eng:

Yet in Standard Thai senses of very X or X very much are typically handled by complete reduplication with the first iteration given a rising tone regardless of its original tone value (i.e. $oo^{36} + oo^{n}$). For example:

k^haw³²³ t^ham³³ ?5k³² ma³³ **di³⁶ di³³** no³⁵ (15)ST:

> S/he does very well, doesn't s/he? Eng:

 $na\eta^{33} \; saw^{323} \; t^h aj^{33} \; pi^{33} \; ni^{35} \; \textit{suaj}^{36} \; \textit{suaj}^{323} \; no^{35}$ (16)ST:

Miss Thailand this year is very beautiful, isn't she? Eng:

mua³⁵² wan³³ t \int an³²³ hen³²³ luk³⁵ c^haj³³ t^ha³³ k^hap³³ rot³⁵ rew^{36} rew^{36} ST: (17)

Yesterday I saw your son; he was driving very fast. Eng:

(18)ST:

Eng: The president Ho Chi Minh is a leader of fighting against France so Vietnamese people *love* him *very much*.

Thus it is clear that there is structural equivalence in reduplication between Standard Thai and Vietnamese spoken by immigrant Vietnamese in Udon Thani. The result of the study shows that a pattern of complete reduplication of Thai (with the first iteration given a rising tone, regardless of the original tone) is used by the Vietnamese speakers in Udon Thani with effectively the same semantic impact as equivalent Thai constructions, despite having access to Standard Vietnamese strategies of using rât and *låm* constructions...

6. Conclusion

Complete reduplication is widely used in spoken Thai to signal intensification, of the sort translated here as very, very much and so forth. Typically the same kinds of verbal and nominal modifiers can also be reduplicated in a similar - although not exactly the same - manner in Standard Vietnamese, and thus the circumstances exist which can readily induce *syntactic interference*.

The Vietnamese speakers in Udon Thani speak Thai most of the time, so they are fluently acquainted with the style of Thai reduplication, and quite likely to show *syntactic interference*. In terms of the model of Lehiste (1998: 15), we can say that patterns from language A are carried over into language B or that patterns of language B are interpreted in terms of patterns of language A. Specifically in this case, the patterns of reduplication in Thai have been carried over into Vietnamese spoken in Undon Thani.

Abbreviations

SV: Standard Vietnamese

VnU: Vietnamese spoken in Udon Thani, Thailand

ST: Standard Thai

Eng: English

References

- Barkman, L. Bruce. 1968. "Bilingualism and The Teacher of English as a Foreign Language", *Applied Linguistics*, New York: edited by Monica Kehoe, Collier Macmillan.
- Dulay, Heidi. 1982. "The Role of the first Language" *Language two*, New York: edited by Heidi Dulay, Marina Burt, and Stephen Krashen. Oxford University Press.
- Khachatphai Burutphat. 1978. *Immigrated Vietnamese*. Duang Kamol publishing, Bangkok, Thailand.
- Lehiste, Ilse. 1998. *Lecture on Language Contact*, The MIT Press, Cambridge, Massachusetts, London, English.
- Mackey, William, F. 1972. "The Description of Bilingualism", *Reading in the Sociology of Language*. Edited by Joshua A. Fishman, Mounton, the Hague.
- Nguyen Tai Can. 2004. *Vietnamese Grammar*, publishing of Hanoi National University, Hanoi.
- Sophana Srichampa. 2005. *Encyclopedia of ethnology in Thailand*. Institute of Language and Culture for rural development, Mahidol University, Thailand.
- Thomason, Sarah G, 2001. *Language Contact (an introduction)*, George town University Press, Washington D.C.
- Weinreich, Uriel. 1953. *Language in contact*, New York: Publications of the Linguistic Circle of New York No. I.

JSEALS style guide for authors

Papers should be submitted electronically as .DOC or .RTF files, no later than December 31st for inclusion in the following year's issue.

You will be asked to format your paper. Download the template from the homepage at www.jseals.org, and type or paste your text directly into this file, and then apply WORD STYLES according to the model.

Ensure that your pages settings are as follows:

Margins on A4 paper:

| Top | 2.5 cm |
|----------------|--------|
| Bottom | 4.0 cm |
| Left | 2.5 cm |
| Right | 2.5 cm |
| Header | 1.25cm |
| Footer | 3.0 cm |
| Gutter margin: | 0.5 cm |
| | |

Font: all Roman and IPA must be in 12 point Times New Roman or DoulosIPA (or 11 point CharisSIL), for other languages make sure you always use a UNICODE compliant font. Papers which do not comply will be returned.

Line spacing: 14 pt exactly in normal text paragraphs

Set to 'at least' to show larger fonts or graphics

Mirror margins set ON

Image files for Tables or Figures should also be provided separately. Where possible, image files should be *greyscale* or *black & white*, .GIF or .TIF or .JPG, 300 DPI, 10cm wide.

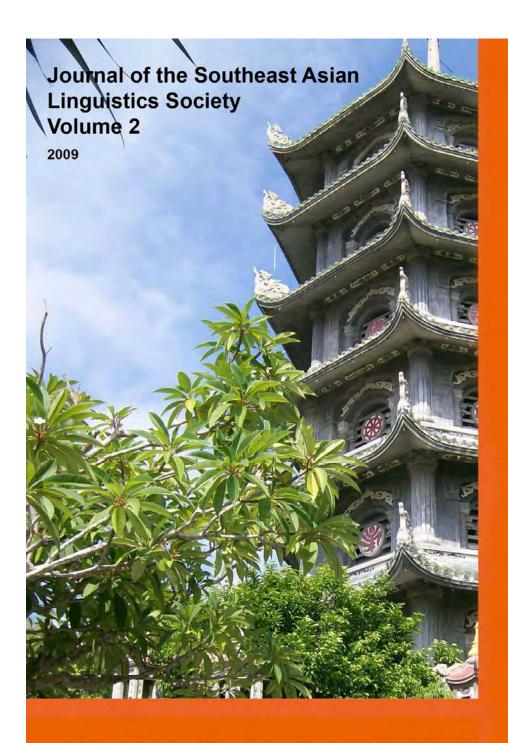
Do not insert page numbers, or other header or footer text.

You do not have to link table or figure numbers in WORD.

Use footnotes, not endnotes.

Your manuscript may be accepted for review if it is not already styled appropriately; in that case you will be asked to apply JSEALS styles to your file at the post-review corrections stage.

If you have any quiries, please consult the website at www.jseals.org or contact Paul Sidwell directly at <paulsidwell@yahoo.com> or <paul.sidwell@anu.edu.au>.



THE AUSTRALIAN NATIONAL UNIVERSITY





PACIFIC LINGUISTICS ISSN 1836-6821

